A Sub-Optimal Low Complex Multiple Malicious Users' Detection Mechanism in a Cooperative Spectrum Sensing System

Shivanshu Shrivastava

Abstract— In cognitive radio based cooperative sensing systems, some secondary users are likely to act malicious. Such secondary users are called malicious users. An optimal application of Dixon's outlier detection scheme for detecting malicious users in a cognitive radio cooperative spectrum sensing system has recently been termed as the Sliding Window Dixon's test. The Sliding window Dixon's test applies the Dixon's test on the smaller data sets with maximum overlaps. However, it is found that this scheme is highly complex. In this paper, we propose a novel scheme to exploit the robustness of the Dixon's test without compromising on its complexity. Simulations demonstrate that the proposed methods are efficient in suppressing multiple malicious users.

Keywords— Sub-optimal low complex scheme, Cognitive Radio, Cooperative Spectrum Sensing, Secondary User, Malicious User, Dixon's test, Sliding Window Dixon's test

I. INTRODUCTION

The requirement for the improvement in the scant electromagnetic spectrum utilization has brought the need of cognitive radio (CR) as a conceivable solution [1], [2]. It proposes the use of the authorized spectrum by the unlicensed users, otherwise called the secondary users (SUs) or the CR users, while it is not being utilized by the licensed or the primary users (PUs). Clearly, it can be noticed that the basic requirement for an SU is to productively detect the spectrum status of the PU. The multipath fading and shadowing endured by obstacles corrupt the PU spectrum detecting precision by an individual SU. Examinations have demonstrated that detecting the spectrum collaboratively by various SUs enhances the detecting ability in profound fading conditions [3],[4], [5]. SU collaboration may be performed by centralized or decentralized mechanisms. The centralized mechanism uses a fusion center (FC) to perform cooperative spectrum sensing (CSS). In this mechanism, the information or the individual verdicts of the coordinating SUs on the spectrum inhabitance are added at a FC. An ultimate verdict in regards to the spectrum inhabitance is made with the assistance of the added information or the added choice measurement at the FC. The participation of SUs in CSS is susceptible to various security dangers. A broadly explored security hazard is the spectrum detecting information falsification (SSDF) attack. An SSDF attack is the aftereffect of the self centered interests of a coordinating SU. These self centered interests cause an SU to act maliciously, making it a malicious user (MU). An MU purposefully sends false data to the FC, increasing the likelihood of the FC to settle on a wrong choice about the spectrum inhabitance.

The SSDF attacks can essentially influence the CSS system operation. The CSS plot must be sufficiently robust to the influences of such kind of false information. Thus, an MU recognition plan is required. Various works have been proposed for the recognition of MUs launching SSDF attacks. In [6], a weighted expansion of the SU energy values has been done to suppress the MU values.

These weights depend on the historical backdrop of the collaborating SUs. In [7], suspicious levels have been ascertained for each of the SUs. The suspicious levels are computed from the past detecting reports of the SUs. These suspicious levels are utilized to dispose-off the MUs. In [8], weights relegated to all the collaborating SU values for weighted expansion, depend on a raised
cosine function, which is acquired from the deviations of the individual SU's choices from the final choice made by the FC. The issue of appointing introductory reputations to the coordinating SUs has been examined in [9]. The work is highly dependent on allocating the introductory reputations by adding the choices of the collaborating SUs at the FC. To accomplish this, the authors chose some trusted hubs. Trusted hubs are characterized as the users which are ensured to report a right choice to the FC. Further, like [8], the weights are appointed by contrasting the individual choices with the joined outcome. The shortcoming of these works is their reliance on the history of the cooperating SUs over a span of time. Reputation based tests especially need to designate introductory reputations to the SUs participating in CSS. Recent investigations on similar subjects have been carried out in [20-22].

In [10], Kaligineedi et.al. proposed a measurable way to distinguish the MUs statistically. This mechanism has been carried out by some statistical strategies which appoint outlier factors or anomaly figures to all the collaborating SUs. The isolation of an SU from the rest of the SUs depends on the deviation of the anomaly figure from a predefined value. The shortcoming of the technique proposed in [10] is that the mean and the variance of the information data needs to be evaluated. The values from the MUs severely influence the mean and variance of the informational collection. Henceforth, the statistical parameters have to be precisely assessed which is difficult in hostile situations. Especially when the fraction of the MUs goes high, the evaluated estimations of the mean and the variance are probably going to give wrong choices.

The statistical mechanisms gain further escalation when a new perspective to the security issues comes into picture. The data from the MUs behaves as outliers to the rest of the data. The authors in [11] introduce Dixon's statistical test for removing outliers in a general data set. The Dixon's test hence can be used to remove the data reported by the MUs which was been proposed in [12].

The advantage of Dixon's test is that it does not depend on the statistical parameters of the values under consideration. However, the limitation of the application of Dixon's test in [12] was its inefficiency in detecting multiple MUs in a CSS scheme. Further, in [13], 'emph{Sliding Window Dixon's test (SW Dixon's test)} was proposed to overcome the limitations of [12]. The SW Dixon's test was found to be highly efficient in detecting multiple MUs. However, its limitation was its mathematical complexity.

In this paper, a sub-optimal defense mechanism to counter the SSDF attacks by multiple MUs has been proposed. The complexity of proposed mechanism is almost one-third to that of the SW Dixon's test. Considering the overall system assessment which includes its performance and its complexity of operation, the proposed scheme in this paper outperforms the SW Dixon's test.

Similar to the SW Dixon's test, the proposed scheme applies the test on smaller gatherings of the cooperating SUs' data. Treating multiple gatherings separately enables the Dixon's test to detect multiple MUs. However, we do not allow maximum overlaps of the smaller groups as done in SW Dixon's test. This improves the complexity of the systems significantly.

This paper contributes the ongoing researches in detecting the MUs in a CSS system by giving a scheme that is able to efficiently reject multiple MUs in one single iteration. Following are the contributions of the proposed schemes:
- Independence from the knowledge of the probability distribution of the data.
- Efficiency in discarding multiple MUs in one single iteration.
- Separate schemes for average and large number of MUs are proposed.
- The complexity is one-third compared to the SW Dixon's test in [13].
The organization of the rest of the paper is as follows: Section 2 discusses the system model and the types of malicious users (MUs) considered. In Section 3, the effect of MUs on a CR-CSS system is explained. In Section 4, the Dixon’s outlier test is explained. Section 5 discusses the schemes proposed for the detection of single as well as multiple MUs. Section 6 presents the simulation results and section 7 concludes the paper.

II. SYSTEM MODEL

Consider K SUs that are spread randomly in an area around the PU. The K SUs agreeably sense the PU spectrum in cooperation. This is accomplished with the signal received from the PU. The choice about the spectrum inhabitance can be made by choosing one of the accompanying two speculations, termed as hypotheses [14],

- Hypothesis $H_1$: The PU is available.
- Hypothesis $H_0$: The PU is unavailable.

The $k$th SU receives signal from the PU at the $n$th moment. This signal is represented by $y_k(n)$, and can be communicated utilizing the previously mentioned speculations as,

$$y_k(n) = \begin{cases} h_k(n)x(n) + u_k(n) : H_1 \\ u_k(n) : H_0, \end{cases}$$

where $x(n)$ is the PU signal, $h_k(n)$ indicates multipath channel

Furthermore, $u_k(n)$ is the additive white Gaussian noise (AWGN).

Since energy detection is assumed to be performed at the SU nodes, (1) must be recast with the choice measurement as the energy of the received signal. The energy $Y_k(n)$ of the signal received by the $k$th SU for the $n$th detecting interim, separated into $L$ samples, is given by,

$$Y_k(n) = \sum_{i=0}^{L-1} |y_k(nT + iT_s)|^2$$

where $T$ is the detection time (interim), $T_s$ is the sensing time and $LT_s=T$. Since, $u_k(n)$ is considered as AWGN, $Y_k(n)$ in (2) will take after a central chi-square circulation under $H_0$ also, under $H_1$, it will take after a non-central chi square appropriation with $L$ degrees of freedom [14]. Each SU sends its figured energy to the FC through an error free channel. The energies of the K coordinating SUs are combined at the FC as [12],

$$Y(n) = \frac{1}{K} \sum_{k=1}^{K} Y_k(n).$$

At the FC, a comparison of the figured $Y(n)$ is made with a pre-decided threshold $\lambda$ [15], and the administer for settling on the choice about the spectrum inhabitance is given by,

$$Y(n) \begin{cases} H_1 \\ H_0 \end{cases} \begin{cases} \lambda \end{cases}$$

Obviously, $Y(n)$ in (3) follows a chi-square distribution as,

$$Y(n) \sim \begin{cases} \chi^2_{2L} (\gamma) : H_1 \\ \chi^2_{2L} : H_0 \end{cases}$$

where $\chi^2_{2L}$ represents the chi square distribution with 2$L$ degrees of freedom and $\gamma$ is the non-centrality parameter.

The productivity of a spectrum sensing framework is evaluated on the basis of two parameters, the probability of detection $P_d$ and the probability of false alarm $P_{fa}$ [15]. $P_d$ is defined as

$$P_d = \text{Pr}(Y(n) > \lambda \mid H_1)$$
and $P_{fa}$ is defined as

$$P_{fa} = \Pr (Y(n) > \lambda | H_0)$$

--------------(7)

III. THE ATTACK STRATEGY AND ITS EFFECTS

Let there be a spectrum band authorized to a PU. Suppose it is vacant at a given time moment. Apparently, when an SU senses that band, the signal energy from that band ascertained by it will be low. The participating SUs detecting the band will send their computed energies to the FC. The SUs are assumed to have the essential assets for settling on their individual choices with the approaching signal energy [9]. Subsequent to recognizing the band to be empty, the MUs purposefully send higher values to the FC. Thus, $Y(n)$ in (3) will have a higher measure than its actual value at the FC. The likelihood of the FC settling on a wrong choice on the absence of the PU in (1) builds up, consequently incrementing $P_{fa}$ in (7), which thus diminishes the throughput of the SUs. A comparative assault is made by the MUs by sending low values when the spectrum is possessed. This will come about into the decrement of the $P_a$ estimation of the spectrum detecting framework, accordingly expanding the obstruction brought about to the PU by the SUs.

IV. THE BASIC IDEA: THE DIXON’S OUTLIER TEST

The Dixon's test was presented in [11] and [16]. These works proposed the technique which depends on anomaly detection in a data sample. The term anomaly is given to the measurements in an information set which demonstrate extensive deviations from the mean estimation of the information set. In [16], the authors recommended the use of the test on little information sizes of three for non-Gaussian information. We propose to apply the test on gatherings of size three so as to identify numerous MUs.

Assume, at a specific detecting moment $n$, the energy values sent by the coordinating SUs to the FC are indicated by $Y_1(n), Y_2(n),\ldots,Y_K(n)$. Since, the estimations have been considered at the moment $n$, we compose the energy informational index as $Y_1, Y_2,\ldots,Y_K$. The expressions for the Dixon's test outlier factor for informational indexes of various sizes can be found in [17]. Our examination requires the estimation of the anomaly factor for a informational collection of length 3. Consider a case of an informational collection of 3 vitality values $Y_1, Y_2, Y_3$. In the wake of masterminding this informational index in an increasing arrangement, let the new information set be $Y_1, Y_2, Y_3$. The exploratory outlier factor for the considered informational collection is given as [17],

$$O^{H}_{exp} = \frac{Y_3 - Y_2}{Y_3 - Y_1}$$

--------------(8)

And

$$O^{L}_{exp} = \frac{Y_2 - Y_1}{Y_3 - Y_1}$$

--------------(9)

$O^{H}_{exp}$ and $O^{L}_{exp}$ are the Dixon's test outlier factors for the MUs purposefully transmitting high and low values. $O^{H}_{exp}$ and $O^{L}_{exp}$ are contrasted with the critical value $O_{crit}$. The critical values for various estimations of $K$ are given in the table for Dixon's test [18], for various confidence levels. The confidence level directly affects viewing an incentive as an anomaly. For a given confidence level, if the ascertained $O_{exp}$ esteem is lesser than $O_{crit}$, then the last rank information esteem under assessment is expected to have a place with a similar typical populace. Else, it is regarded as an exception.
V. PROPOSED METHOD

The technique for the calculation of the Dixon's test anomaly variable or the outlier factor can be found for the informational collections of various sizes in [17]. In [12], the test is specifically connected on all the approaching vitality information at a specific detecting interim for identifying a solitary MU. Such a use of the Dixon's test may demonstrate less proficient for the accompanying reasons:

1) In usual conditions, it is likely that the quantity of coordinating users’ data ought to be more than 10. Be that as it may, the Dixon's test has been ended up being more effective for information tests of little sizes (10 or less) [16].
2) Owing to the ability constraint of the Dixon's test to sift through a solitary exception in an informational collection, its application all in all approaching vitality informational collection makes it valuable for smothering just a single MU in a CSS framework.
3) The test is appropriate for distinguishing anomalies in typically circulated information tests [16].

The $Y_k(n)$ samples are non-Gaussian as they are certain positive energy values. In [19], the strength of the Dixon's test to any accepted fundamental dissemination was considered. It is demonstrated that the test is powerful to the deviation from normality when K=3. This result inspired us to apply the Dixon's test in gatherings of 3 SUs. Likewise, applying the test in gatherings of SUs clearly distinguishes one MU in each gathering. Let there be P honest SU energy values, ordered as $Y_1(n), Y_2(n), ..., Y_P(n)$, stay after the testing is finished. The energy values will be combined at the fusion center as,

$$Y'(n) = \frac{1}{P} \sum_{p=1}^{P} Y'_p(n).$$

The decision rule will be modified as

$$Y'_n \geq \hat{\lambda}$$

where $\hat{\lambda}$ is as declared in (4). The estimation of $\hat{\lambda}$ is dictated by settling a value of $P_{fa}$ as shown in [15].

Once the MU values have been found out, (6) and (7) can be re-written as,

$$P_d = \Pr(Y'(n) > \hat{\lambda} \mid H_1)$$

and

$$P_{fa} = \Pr(Y'(n) > \hat{\lambda} \mid H_0).$$

For the application of the result proved in [19], the incoming energy data values at the FC are first divided into groups of 3 each. The values in each group are then arranged in ascending order and the corresponding outlier factors are calculated. Then Dixon's test is applied on each group.

This results into detection of one possible outlier in each group, which enhances the possibility of multiple outlier detection in the overall energy data sample. If the number of incoming energy values is not a multiple of 3, overlapping of the energy values at one end of the data set is performed. It is evident that multiple outlier rejection is done in one single time iteration. Algorithm 1 shows the implementation of this scheme.
Algorithm 1: Implementation of the Three point Dixon’s test

N: length of the incoming energy data set
\( e \) : incoming energy data set
\( \lambda \): Detection threshold
I: Rank of the values of \( e \)
Cv: critical value

for \( i \leftarrow 2:3; [3 \times \text{floor}(N/3) - 1] \)
\[
Y \equiv \left[ e(i - 1), e(i), e(i + 1) \right] \quad \{\text{Grouping the values}\}
\]
\[
\left[ Y_s, I_s \right] \equiv \text{sort} (Y) \quad \{\text{Sorting the values with indices stored in } I_s\}
\]
\[
o_{[3]} = \frac{Y_s(I_s(3)) - Y_s(I_s(2))}{Y_s(I_s(3)) - Y_s(I_s(1))} \quad \{\text{For suspected high values}\}
\]
and
\[
o_{[1]} = \frac{Y_s(I_s(2)) - Y_s(I_s(1))}{Y_s(I_s(3)) - Y_s(I_s(1))} \quad \{\text{For suspected low values}\}
\]
if \( o_{[3]} > C_v \)
\[
e\left( I \left( Y_s(3) \right) \right) = 0
\]
end if
if \( o_{[1]} > C_v \)
\[
e\left( I \left( Y_s(1) \right) \right) = 0
\]
end if
end for

if \( \text{mod}(N/3) \neq 0 \)
\[
Y_r \equiv \left[ e(N - 2), e(N - 1), e(N) \right]
\]
\[
\left[ Y_{rs}, I_{rs} \right] \equiv \text{sort} (Y_r)
\]
\[
o_{[3]} = \frac{Y_{rs}(I_{rs}(3)) - Y_{rs}(I_{rs}(2))}{Y_{rs}(I_{rs}(3)) - Y_{rs}(I_{rs}(1))}
\]
\[
o_{[1]} = \frac{Y_{rs}(I_{rs}(2)) - Y_{rs}(I_{rs}(1))}{Y_{rs}(I_{rs}(3)) - Y_{rs}(I_{rs}(1))}
\]
if \( o_{[3]} > C_v \)
\[
e\left( I \left( Y_{rs}(3) \right) \right) = 0
\]
end if
if \( o_{[1]} > C_v \)
\[
e\left( I \left( Y_{rs}(1) \right) \right) = 0
\]
end if
end if

\( P = \text{count}(e(i) > 0, i = 1, 2, \ldots, N) \quad \{\text{Find the non zero values}\} \)

\[
E_p(m) = \frac{1}{P} \sum_{i=1}^{N} e(i)
\]
if \( E_p(m) > \lambda \)
Decide \( H_1 \)
else
Decide $H_0$
end if

VI. COMPLEXITY COMPARISON WITH SW DIXON’S TEST

As it could be seen that the K-point Dixon’s test essentially works on discovering three far-end data points in a data set. To make the complexity analysis, the worst case scenario needs to be investigated. For selecting three far-end data points, the complexity of the worst case scenario is of the order of $O(K)$. The SW Dixon’s test includes sorting of three data-points and the test is rehashed at K-test focuses. It is rehashed at K-test points because maximum overlap is allowed. With the maximum overlap, the test has to be done at the points 1, 2, ..., K. Hence, the sorting is carried out K times which brings its complexity as $O(K)$. On the other hand, the sorting performed in K point Dixon's test will make its complexity $O(K)$. The order of complexities for finding the extreme values is ($O(K)$) in both the K-point Dixon’s test and the SW Dixon's tests connected at K test focuses. The Dixon's remainder $O_{\exp}^H$ (or $O_{\exp}^L$) is to be figured once in the K-point Dixon's test while it is to be processed K times in the proposed 3-point Sliding Dixon's test connected at K test focuses.

On the other hand, in the scheme proposed in this paper, the Dixon's test is applied over non-overlapping smaller gatherings of the data set. The size of the gathering is three. In this manner, the complexity of the proposed mechanism becomes one-third of the SW Dixon's test as well as the K point Dixon's test. The sorting has to be performed $K/3$ times which makes the complexity of the proposed scheme as $O(K/3)$. In this way the proposed technique is outperforms the SW Dixon's test and the K-point Dixon's test in terms of complexity.

VI. SIMULATION RESULTS AND PERFORMANCE ASSESSMENT

Similar to [13], first, we evaluate the proposed test for detecting a single MU. The number of SUs considered for detecting a single MU is 20, so that we can compare the results with the scheme in [12]. Next, we evaluate the proposed test for multiple MUs. We assume that there is no line-of-sight (LOS) component present in the environment. When an LOS component is absent, the environment follows a Rayleigh fading [4]. The time-bandwidth product, which gives the sensing samples for sensing the system is considered to be 50. We consider the SNR for the non-MU channels as 2 dB. When the spectrum is unoccupied, the MUs are assumed to be reporting uniform values corresponding to an SNR 6 dB higher than that of the non-malicious users. Similarly when the spectrum is occupied, the MUs are assumed to be reporting values corresponding to 5dB lower SNR. A significance level of 0.1 is considered for applying the Dixon's test. Performance has been evaluated in terms $P_d$ and $P_{fa}$.

VI. a. PERFORMANCE ASSESSMENT IN SUPPRESSING A SINGLE MU

The performance evaluation for suppressing a single MU is done by observing the improvement brought in taking the ROC curve of the affected system near to the ROC curve of an unaffected system. The unaffected system is referred as an honest system.

We make a slight modification in the original ROC curves in this paper. Instead of plotting $P_{d}$ with $P_{fa}$ directly, we plot each of them separately with lambda. The reason for this change is that we want to study the individual affects on $P_d$ and $P_{fa}$ caused by the MUs, which will not be possible with standard ROC curves. These curves are compared with the performance of an honest cooperating system. We observe the capability of a scheme to bring the affected system close to the honest system. The results are shown in Fig. 1, Fig. 2, Fig. 3 and Fig. 4. The plots show that the proposed scheme outperform the K-point Dixon's test. The improvement brought by the scheme proposed here is comparable to that of [13]. However, it gives a much better performance in terms of
complexity. Hence, the overall performance of this scheme can be regarded as better than the SW Dixon's test in [13].

**Fig. 1** Single MU transmitting higher measurements with 20 SUs while the PU spectrum is unoccupied: $P_{fa}$ vs $\lambda$ plot

**Fig. 2** Single MU transmitting higher measurements with 20 SUs while the PU spectrum is unoccupied: $P_d$ vs $\lambda$ plot

**Fig. 3** Single MU transmitting lower measurements with 20 SUs while the PU spectrum is occupied: $P_{fa}$ vs $\lambda$ plot
VI. b. PERFORMANCE ASSESSMENT IN SUPPRESSING A MULTIPLE MUS

We consider individual non-collusion attacks launched by the multiple MUs. Similar to the case in Section VI. a, we assess the outcome of the proposed scheme from its tendency to bring $P_d$ and $P_{fa}$ curves of an affected system near to those of an honest system. A varying percentage of MUs are considered to signify their multiple numbers. These MUs are present among the cooperating SUs. Fig. 5, Fig. 6, Fig. 7 and Fig. 8 show these assessments. When the fraction of MUs is 10%, the proposed scheme shows a noteworthy performance. The scheme shows a satisfactory outcome till the percentage of MUs is increased up to 20%. Considering the low complexity of the scheme proposed here as compared to the scheme [13], its overall performance is better.
Fig. 6: Multiple MUs transmitting lower measurements with 40 SUs while the PU spectrum is occupied: $P_d$ vs $\lambda$ plot

Fig. 7: Multiple MUs transmitting higher measurements with 40 SUs while the PU spectrum is unoccupied: $P_{fa}$ vs $\lambda$ plot

Fig. 8: Multiple MUs transmitting higher measurements with 40 SUs while the PU spectrum is unoccupied: $P_d$ vs $\lambda$ plot
REFERENCES