Jackson *et al. EURASIP Journal on Wireless Communications and Networking* 2015, **2015**:4
http://jwcn.eurasipjournals.com/content/2015/1/4

EURASIP Journal on Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Exploiting and defending trust models in cooperative spectrum sensing

David S Jackson[1*], Wanyu Zang[1], Qijun Gu[2], Wei Cheng[1] and Meng Yu[1]

Abstract

Cognitive radios are currently presented as the solution to the ever-increasing spectrum shortage problem. However, their increased capabilities over traditional radios introduce a new dimension of security threats. Cooperative spectrum sensing (CSS) has been proposed as a means to protect cognitive radio networks from the well-known security threats: primary user emulation (PUE) and spectrum sensing data falsification (SSDF). In this paper, we demonstrate a new threat to CSS protocols that rely on sensor reputations, called the *Rogue Signal Framing* (RSF) intrusion. Rogue signals can be exploited to create the illusion of malicious sensors which leads to the framing of innocent sensors and, consequently, their removal from the shared spectrum sensing. Ultimately, with fewer sensors working together, the spectrum sensing is less robust for making correct spectrum access decisions. The simulation experiments illustrate the impact of RSF intrusions which, in severe cases, shows roughly 40% of sensors removed. To counter the RSF's impact on the cooperative spectrum sensing (CSS), we introduce a new defense based on cluster analysis and community detection from analyzing the network's received signal strength (RSS) diversity. Tests show up to 95% damage mitigation to the integrity of sensor reputations, thus retaining the benefits of trust-based CSS protocols.

1 Introduction

The growing demand for wireless services shows an inevitable overcrowding of the spectrum bands, in large part due to the rapid increase of wireless mobile services in recent years. Conventionally, the Federal Communications Commission (FCC) statically assigned spectrum bands to licensed users for exclusive use on a long-term basis, precluding anyone else from access [1,2]. Yet, analysis of the spectrum bands clearly indicate that current FCC policies have created severely under-utilized channels, causing a bottleneck for new wireless services [1,3,4]. Dynamic spectrum access (DSA) is the proposed solution to alleviate the overcrowding of bands by allowing licensed primary users (PUs) to share unused spectrum with unlicensed secondary users (SUs) in an opportunistic fashion [1,5].

Cognitive radios (CR) utilize the DSA technology that enables autonomous optimization of radio configurations and the scanning of spectrum bands to locate the best available channels on a non-interference basis [6-8]. The cognitive radio network (CRN), consisting of SUs, is given permission to coexist in licensed channels under two preconditions mandated by the FCC: (1) giving spectrum priority to licensed users and (2) minimizing interference to licensed users. The faster the SUs can detect the primary signal and vacate the licensed channels, the smaller the interference. For this reason, the secondary network must achieve accurate spectrum sensing to know exactly when primary users occupy the channel.

The cornerstone of the IEEE 802.22, the first standard for cognitive radio networks, requires the SUs to yield to the PUs immediately after detecting the primary signal within a designated region [9]. The 802.22 WRAN standard is aimed at using DSA technology to allow sharing of geographically unused spectrum allocated for television broadcast services. So in the 802.22 WRAN implementation, the primary network would consist of a TV broadcasting station (primary transmitter) and the corresponding subscribed viewers (primary receivers) [5,9]. Ideally, SUs would occupy unused TV spectrum in geographical locations where the primary network is absent, but may coexist as long as the SUs do not interfere with

*Correspondence: jacksonds3@vcu.edu
[1]Department of Computer Science, Virginia Commonwealth, 401 W. Main St, University, Richmond, VA 23220, USA
Full list of author information is available at the end of the article

the subscribed viewers' reception of the primary signal. However, guaranteeing a minimal level of interference to the primary network is perhaps the biggest obstacle to the commercialization of DSA technology and a very difficult problem to solve [5]. In order to have minimal interference, cognitive radios must be able to reliably detect, in real time, the presence or absence of a primary signal from a given spectrum band. Otherwise, these cognitive radios can unknowingly transmit signals simultaneously with the primary transmitter, causing unacceptable levels of interference to nearby PUs.

Cooperative spectrum sensing (CSS) has been proposed as an effective approach for boosting the detection of primary signals in CR networks [5,10,11]. In centralized CSS, the SUs submit their sensor reports to the fusion center (FC), which is a server for aggregating and cross-examining the network's sensor reports to make a robust analysis of the spectrum availability. The purpose of the FC is to output a global spectrum decision, based on the sensor reports, to notify SUs if they can access a licensed spectrum band, in accordance to the FCC statutes. Research results from [1,12] indicate that shadow fading and multipath fading can be alleviated by requiring multiple SUs to cooperate with each other in determining the spectrum availability.

However, CSS is vulnerable to attacks like the spectrum sensing data falsification (SSDF) where malicious SUs make false reports on the spectrum availability to mislead the FC [13]. To counter SSDF, various trust models have been proposed to protect CCS from malicious SUs [13-17]. These trust-based CSS protocols build reputation profiles for sensors and filter out the sensing reports from those with low reputations. Thus, they can single out attackers and mitigate their influence in the shared spectrum sensing.

Unfortunately, we find that the sensor reputations are exploitable by rogue signals in trust-based CSS protocols. In secondary networks, it is very hard to conclude the root cause of bad sensor reports such as malfunctioning sensors, the hidden node problem, SSDF attacks, and rogue signals. Typically, trust models (from CSS protocols) treat all inaccurate sensors the same way, in a loss of reputation. We consider trust models as overly sensitive intrusion detection systems (IDS) for penalizing sensors without taking into account the root cause of the abnormal sensor reports. As a result, attackers can cause inaccurate sensor reports by transmitting rogue signals in order to destroy the reputation of the targeted sensors. Accordingly, we present a new threat to a variety of trust-based CSS protocols, named the Rogue Signal Framing (RSF) intrusion. To launch this attack, we exploit directional antennas to isolate a radiation pattern to a group of sensors in proximity. The outcome is the emulation of an SSDF attack through sporadic and misleading rogue signals, causing different conclusions of channel availability in the network. The split between local spectrum decisions leads to innocent sensors being treated as malicious and consequently removed from the shared spectrum sensing.

To counteract this new threat, we propose a new defense scheme, named the RSF Clustering Defense (RCD) module, that looks for dense clusters of sensors and examines the proximity and similarity of their reports. Based on the RCD findings, it makes a heuristic decision on whether or not the network was affected by an RSF attack via rogue signals. Thus, the RCD module can distinguish sensors under the RSF intrusion and mitigate the trust damage. In effect, our defense prevents trust models from becoming an overly sensitive IDS by minimizing the false alarms caused by rogue signals but still *relies* on a trust model to stop SSDF attacks. The following are a list of contributions:

- Introduced the Rogue Signal Framing intrusion, an attack on the trust model of CSS protocols
- Developed a solution, the RSF Clustering Defense (RCD), that protects sensor reputations from manipulation in trust models
- Ran simulations that demonstrated the impact of the RSF intrusion and the RCD solution

The rest of the paper is outlined as follows. Section 2 reviews common CRN attacks and trust-based CSS protocols. Then, we present the system model in Section 3, and show the details and analysis of the RSF intrusion in Section 4. We propose the RCD defense and evaluate it in Section 5 and conclude the paper in Section 6.

2 Related works

Our work is mostly related to the following attacks and defenses in CRNs.

PUE and SSDF attacks. Although CRNs are vulnerable to a variety of attacks [6], two attacks received much attention. One is the primary user emulation (PUE) attack [6,18], where an attacker masquerades as the primary transmitter from the vantage point of its neighbors. The other attack is the SSDF [5,13], in which compromised users falsify the local spectrum sensor reports to obscure the existence or create the illusion of a primary signal at the FC [19]. Both of these attacks attempt to deceive the FC on the availability of spectrum resources, causing networks to behave in unintended ways. In contrast, the RSF intrusion disrupts the trust between the FC and sensors, which makes the spectrum sensing less stable.

Tom Clancy et al. [6] lists a host of threats such as sensory manipulation attacks, belief manipulation attacks, and objective function attacks to cognitive radios with embedded learning engines. However, the RSF intrusion focuses on cognitive radio networks with trust schemes

and cooperative spectrum sensing, independent of the learning engine.

Trust-based CSS protocols. To defeat SSDF attacks, several trust-based schemes were developed. Chen et al. [13] presented a sequential probability ratio test (SPRT) that scales the contribution of sensors by their reputation in order to mitigate the impact of SSDF attacks. Their model incorporates sampling votes on the detection or absence of the primary signal and weighing each vote according to the sensor's reputation. For every vote identical to the global decision, the sensor's reputation is incremented, such that their vote carries more weight in future decisions made at the fusion center. Kaligineedi et al. [14] presented a pre-filtering average combination scheme. The scheme's filters are responsible for (1) filtering extreme outlier sensor reports and (2) ignoring sensors that have continuously deviated from the majority over a length of time. Arshad et al. [16] presented a beta reputation system model for hard-decision CSS protocols. Similar to [13], the sensors are rewarded for agreeing with the global spectrum decision, but otherwise penalized. Feng et al. [20] introduced the SensingGuard trust model intended to protect the CSS from rational collusive SSDF attacks, in contrast to sporadic SSDF attacks. Lai et al. [21] introduced a game theory model, based on the Newton-Raphson algorithm, that aims to punish selfish SUs and reward cooperation. In [17,22,23], the authors developed a trust-based CSS protocol that penalized sensors if their reports deviated too far from the expected received signal strength (RSS) values determined by common RSS models. The similarity of these approaches are to build reputation profiles for spectrum sensors in order to filter out sensing reports from untrustworthy sensors. However, our work shows that the reputations can be manipulated and, as a consequence, well-behaved sensors are framed and removed from the shared spectrum sensing.

Received signal strength anomaly detection. Apart from reputation profiles, there are solutions that rely on RSS models and statistical methods to validate the authenticity of sensor reports. Min et al. [24] presented an algorithm that analyzes sensor clusters and their RSS correlation, based on distance and approximated shadow fading, to pinpoint malicious sensors and reduce/remove their input from the fusion center. A big difference in our work and theirs is that they rely (and assume) *a priori* knowledge of the environment's shadow fading to accurately predict the expected RSS value for a cluster of sensors. Secondly, they have no reputation model to go along with anomaly detection, so their solution discards the sensor reports in single intervals instead of penalizing the sensors for an extended duration. In [19,25], the authors developed solutions using RSS estimation models and support vector machines (SVMs), a machine learning technique, to classify sensors as either anomaly or normal. Unlike the various aforementioned solutions, we developed our own defense based on cluster analysis and community detection to safeguard sensor reputations from manipulation, instead of only focusing on the integrity of the CCS.

What makes our solution unique is that our defense protects the integrity of trust models, i.e., sensor reputations, from rogue signal manipulation. Previous literature used trust models to stop malicious SUs (and their sensors) from deceiving the CSS, but did not consider the trust models themselves to be the target of attacks. Trust models were considered reliable solutions against SSDF attacks and malfunctioning sensors, but to our knowledge, none of the papers discussed how to manipulate and disrupt trust models. We realized the vulnerability of trust models due to their coarse threshold of penalizing inaccurate sensor reports, i.e., a sensor is deemed untrustworthy if it does not behave in a predetermined way. However, if an attacker knows how the sensors should behave, then they can leverage rogue signals to disrupt typical sensor behavior and thus destroy their reputations. To protect sensor reputations, we explored techniques from social network analytics, such as cluster analysis and community detection, as opposed to relying on RSS models or shadow fading estimations to predict the correct sensor report.

2.1 Motivation for distinguishing between RSF and SSDF

In an NSF 2009 workshop, the FCC had raised the question, 'What authentication mechanisms are needed to support cooperative cognitive radio networks? Are reputation-based schemes useful supplements to conventional Public Key Infrastructure (PKI) authentication protocols?" [26] Reputation-based schemes in CSS (a.k.a. trust-based CSS protocols) are a popular technique for performing robust and accurate spectrum sensing without any inter-communication with the primary network, but the question remains on how effective they are at satisfying the FCC security requirements. Our work takes a closer look at the robustness of trust-based CSS protocols.

In secondary networks, it is very hard to conclude the root cause of bad sensor reports, which can vary from (1) malfunctioning sensors, (2) the hidden node problem, (3) SSDF attacks (i.e., malicious secondary users), and (4) rogue signals. Yet, the trust-based CSS protocols treat all inaccurate sensors the same way, in that they penalize secondary users and diminish sensor reputation all the same. An important question we wanted to investigate was, 'Should the trust-based CSS protocols treat all inaccurate sensor reports the same way, regardless of the root cause? Or does it cause more harm than good to the system in certain scenarios'.

To test our hypothesis, we simulated multiple directional rogue signals against targeted clusters in a cognitive radio network. The simulation illustrated the impact

of rogue signals negatively affecting sensor reputations which, in severe cases, shows roughly 40% of sensors penalized and eventually ignored in the shared spectrum sensing process. In other words, nearly half of the sensors were removed without any fault of their own, e.g., the sensors were not malfunctioning nor behaving maliciously but were still penalized. That means an outsider has the potential to trick the reputation scheme in order to filter out nearly half of the sensors, thus diminishing the performance of the network's shared spectrum sensing. Trust-based CSS protocols have proven effective against malicious secondary users who report falsified sensing reports, but they did not consider the impact of rogue signals. Hence, based on the outcome of our simulations, we consider trust models as overly sensitive intrusion detection systems (IDS) for penalizing sensors without taking into account the root cause of abnormal sensor reports.

Not being able to determine the origin of inaccurate sensor reports opens the possibility for attackers to use RSF as a stepping stone attack against trust-based CSS protocols. Chen et al. [13] models attacks against CSS protocols as a Byzantine fault tolerance system, in that the CSS protocol can continue functioning as intended as long as there are not too many Byzantine failures, which in this case are generally hidden, malicious, or malfunctioning sensors. In contrast, our work demonstrates that the RSF attack lowers the Byzantine fault tolerance of trust-based CSS protocols, due to having less secondary users participate in the shared spectrum sensing, thus making the system less robust against Byzantine failures.

Clancy et al. [6] warns of a similar threat of rogue signals, but in a different context. They claim that rogue signals can cause faulty statistics, collected from the physical layer (e.g., RSS, channel availability, etc.), and stored in the knowledge base. The cognitive radio's behavior is determined by the learning and reasoning engines which, in turn, depends on the knowledge base of spectrum observations across many channels overtime. Hence, the cognitive radio may not behave as intended, or in fact cause harm, when the knowledge base contains faulty statistics that inhibits good decision-making. Both our work and theirs [6] express the importance of being able to defend against rogue signals. The difference, however, is that our work protects the sensor reputations in trust-based CSS protocols whereas their idea is related towards protecting the integrity of the knowledge base.

3 Attack model

In this section, we define the RSS model and the method of attack for the RSF which employs *directional antennas*. The attacker manipulates sensor reputations by transmitting rogue signals to targeted sensors, thus causing conflicting sensor reports in the network. To ensure that reports do conflict, directional antennas are used to avoid targeting the entire network.

Figure 1 illustrates the system model of trust-based CSS protocols and the different targets of PUE and RSF intrusions. In it, f_0 represents some wireless spectrum frequency, S_i a set of sensors, and R_i the corresponding set of sensor reports. The system model is a stack of dependent layers, starting with the spectrum channel, the network of sensors, the trust model, and finally the FC. The accuracy of the CSS is dependent on the FC receiving reliable input from the above layers. For example, the spectrum channel must be clear enough for communication, the majority of sensors must not be malicious or malfunctioning, and the trust model must filter the malicious sensors to protect the FC from bad input.

Without loss of generality, we use a system as shown in Figure 2 to discuss the proposed security issues. Within the network area, the spectrum sensors are randomly distributed and the attacking antennas are positioned in the middle. The FC collects the sensor reports and cross-examines the local spectrum observations to make a global decision on channel vacancy. Spectrum sensing occurs in scheduled time intervals when all communications from the secondary network stops, called *quiet periods*, in order to listen for the primary signal [5].

3.1 Propagation model

Energy detection. We decided to use energy detection because it is the most widely used spectrum sensing technique for cognitive radio networks [10,24]. Secondly, energy detection is used on three trust-based CSS protocols that we borrow for our simulations, from papers [13,14,16].

When an attacking antenna emits signals, the RSS in decibels per milliwatt (dBm) for any given sensor s_i can be modeled as below according to [27]:

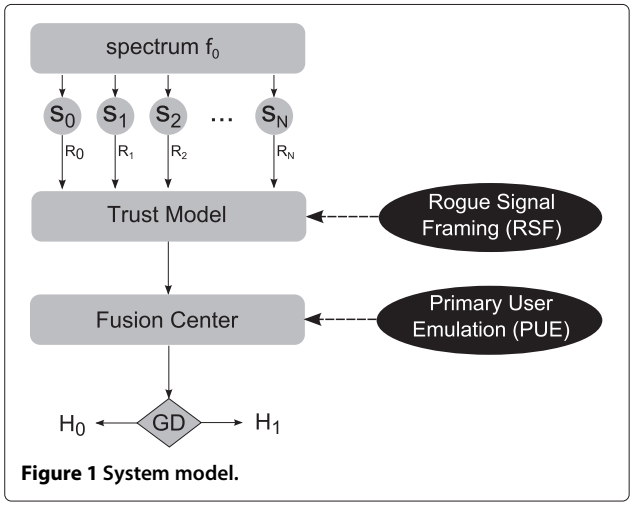

Figure 1 System model.

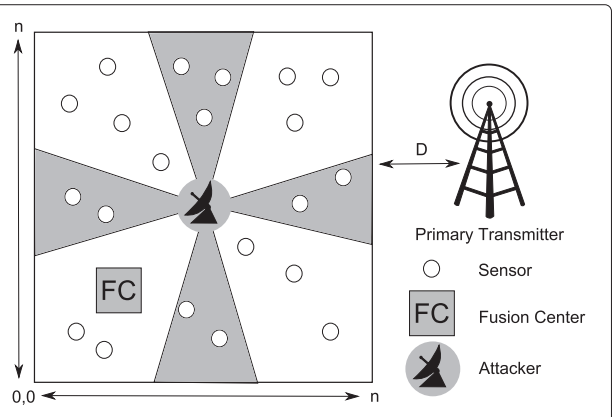

Figure 2 Illustration of attack model.

$$R_i = \begin{cases} \mathcal{N}(\mu_\omega, \sigma_\omega), & H_0 \\ 10\log_{10}(P_{\text{ray}}(d_{ij})) + L_s[x_i, y_i]\,, & H_1 \end{cases} \quad (1)$$

The model gives two possible RSS values. When the antenna is not transmitting (i.e., case H_0), the RSS is actually from the environmental noise, for which μ_ω is the noise power mean and σ_ω is the noise variance. On the other hand, when the antenna is emitting signals (i.e., case H_1), the RSS is determined by the attenuation of signal propagation from the attacker to the sensor plus shadow fading on position $[x_i, y_i]$. In the H_1 case, we use the Rayleigh fading model in milliwatts (mW), expressed as: [28]

$$P_{\text{ray}}(d_{ij}) = \frac{P_t G_t G_r \lambda^2}{\left(4\pi d_{ij}\right)^2}\sqrt{r_1^2 + r_2^2} \quad (2)$$

where d_{ij} is the distance between s_i and the jth attacking antenna, λ denotes the wavelength (meters), P_t is the emission power, G_t and G_r are the antenna gains of the transmitter and receiver, and $r_1, r_2 \sim \mathcal{N}(0,1)$.

The RSS value R_i is measured in decibels per milliwatt (dBm). However, the Rayleigh fading model (from Equation 2) is in milliwatts (mW), so we apply the unit conversion dBm $= 10\log_{10}(mW)$ in Equation 2 under hypothesis H_1. In addition, $L_s[x_i, y_i] \sim \mathcal{N}(0, \sigma_L)$ is the correlated shadow fading gain [29] between s_i's position $[x_i, y_i]$ and the jth antenna's position $[x_j, y_j]$, and σ_L is the shadow fading variance. In the propagation model, we assume that the channel bandwidth is much larger than the coherent bandwidth, so the effect of a multi-path fading is negligible, and thus removed from Equation 1 [9].

3.2 Directional antenna model

Rogue signals are generated by directional antennas to manipulate the sensor reputations. The antenna radiates in a smaller area surface, compressing the radiated energy, and thus raising the signal's strength. Hence, G_t in Equation 2 is substituted by the directional gain according to [30]:

$$G(\theta, \phi) = \left(4\pi r^2\right)\left(\frac{4}{\pi r^2 \sin(\theta)\sin(\phi)}\right) \quad (3)$$

In Equation 3, θ and ϕ are the vertical and horizontal angles of the beam width, respectively. For simplification, we assume $\theta = \phi$. Furthermore, we assume that the rogue signals only affect the sensors inside the beams of the directional antennas. To determine which sensors are attacked, we need to calculate the angle between the attacked sensor and the directional antenna, as illustrated in Figure 3. The angle between position $\vec{p}_i$ of the ith sensor and position $\vec{p}_j$ of the jth antenna is as follows:

$$\theta_{ij} = \arccos\left(\frac{\vec{p}_i \cdot \vec{p}_j}{\|\vec{p}_i\|\,\|\vec{p}_j\|}\right) \quad (4)$$

where $\vec{p}_i, \vec{p}_j \in \mathbb{R}^2$. The ith sensor is affected by the rogue signal if θ_{ij} falls between the lower and upper beam angles θ_l, θ_u of the jth transmitter such that $\theta_l \le \theta_{ij} \le \theta_u$.

4 Rogue signal framing intrusion

In this section, we introduce the RSF intrusion and demonstrate its impact on the network's total trust through simulations.

In the CSS paradigm, the physical layer (i.e., the sensor) provides local signal detection. The FC collects the sensor reports and validates the signal authenticity through cross-examination of the RSS spatial diversity from the network. However, verifying the source of RF waves at the physical layer is incredibly challenging, especially for energy detectors that can only observe the RSS. Since the energy detectors only measure raw RF energy, there is no cryptographic means to identify the source [6].

According to the first CRN standard, the IEEE 802.22, the secondary network must be self-reliant in minimizing interference to the primary network which requires accurate spectrum analysis [18]. In the case of SSDF attacks, trust models have been effective at removing malicious sensors from the shared spectrum sensing [13-17]. However, these trust models cannot distinguish between malicious sensors and accurate sensors misled by rogue signals (as opposed to the legitimate primary signal). In other

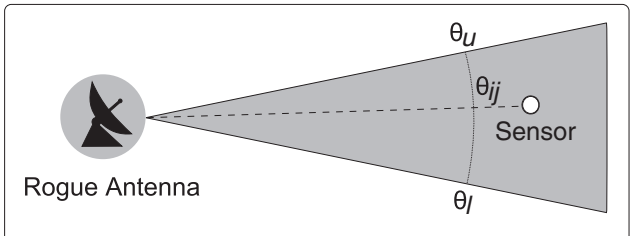

Figure 3 Capturing sensors in the radiation pattern of rogue signals.

words, sensors are labeled untrustworthy when they have a consistent history of abnormal sensor reports, regardless of the cause.

Rogue signals can raise a sensor's RSS well above what is expected, especially in the absence of the primary signal. So a prolonged rogue signal on a group of sensors can cause a sharp contrast in local spectrum observation from the others, thus appearing malicious and no different than SSDF. Consequently, the security protocol brands these sensors as untrustworthy and removes them from the shared spectrum analysis for as long as the stigma remains. As such, launching rogue signals on specific regions of the network over many quiet periods leads to the exploitation of the trust model via the RSF attack. In the context of CSS, we define the term *Rogue Signal Framing* attack as follows:

Definition. *Rogue Signal Framing attack breaks the trust between the fusion center and a group of sensors via rogue signals to create the illusion of malicious sensors.*

To launch this attack, we exploit directional antennas to launch rogue signals on a regional group of sensors and thereby causing them to report abnormally high RSS compared to the rest of the unaffected network. When sensors start reporting differently, the FC interprets the situation as an SSDF attack, when in fact, the sensors reported honestly. In essence, we can use rogue signals to emulate false SSDF attacks to harm innocent sensors and mitigate their cooperation in shared spectrum sensing.

4.1 Motivation for directional antennas

In a CRN with energy detectors, the RSF attacker must limit the rogue antenna's coverage in order to avoid a successful PUE. Directional antennas make it possible to isolate its radiation pattern to a targeted group of sensors (with the rest of the network unaffected), thus convincing the FC that the defecting sensors are malicious. On the other hand, isotropic antennas emit RF waves in all directions and maximize the antenna's coverage. This leaves a massive RF finger print in a network of energy detectors. Chen et al. [18] proposed an RSS-based location verification scheme to detect and pinpoint PUE attacks enforced by a dense network of sensors. However, this scheme was not tested or tailored for pinpointing directional antennas.

Directional antennas are difficult to detect, and even harder to pinpoint, because of their ability to emit rogue signals with narrow and asymmetrical radiation patterns. Any changes made to the beam direction and beamwidth of a directional antenna can drastically change the network's RSS spatial diversity. These observations are supported by work from Bauer et al. [31]. In their experiments, they demonstrated that directional antennas can disrupt localization algorithms on IEEE 802.11 WLANs that resulted in very high errors.

4.2 Trust damage

The main goal of the RSF attack is to compromise the trust between the FC and network sensors. To quantify the trust damage (as a percentage), we use the following equation to measure the network's trust score $T_\Sigma[q]$ on quiet period q with:

$$T_\Sigma[q] = \left(\frac{1}{\sum_{s_i \in S} t_i[0]} \right) \sum_{s_i \in S} t_i[q] \tag{5}$$

where $t_i[q]$ is the trust score of sensor $s_i \in S$. In each trust-based CSS protocol, the trust score is represented differently. In order to compare the trust damage between each protocol, we normalized the trust score t_i such that $t_i[q] \in [0, 1]$ in the equation.

In each quite period, a group of sensors may lose their trust due to the RSF intrusion, so $T_\Sigma[q]$ changes from one quiet period to the next. As the time passes on, sensors exposed to RSF suffer an increasing amount of trust damage, so we expect $T_\Sigma[q]$ will decrease as the number of quiet periods q increases.

4.3 Attack evaluation

To test our proposed framing intrusion, we borrow three different trust-based CSS protocols. The first protocol F_A, by Chen et al. [13], utilizes the sequential probability ratio test (SPRT) and weights the probability by the sensor's reputation to mitigate the impact of SSDF attacks. The second protocol F_B, by Kaligineedi et al. [14], utilizes a pre-filtering average combination scheme. These filters are responsible for (1) filtering extreme outlier sensor reports and (2) ignoring sensors with high-trust penalties. The third protocol F_C, by Arshad et al. [16], utilizes a beta reputation system model for hard-decision CSS protocols. Like F_A, the sensors are rewarded for agreeing with the global spectrum decision, but otherwise penalized.

We make the following assumptions on the simulation's environment according to an IEEE 802.22 WRAN environment that encompasses UHF/VHF TV bands between 54 and 862 MHz [9]. In our simulation, 400 sensors are located inside a $2,000 \times 2,000$ grid. We assume the incumbent broadcasting station operates at the UHF frequency of 615 MHz. Like Figure 2, there are four rogue directional antennas facing the cardinal directions and positioned on the map's center. Protocols F_A, F_B, and F_C are tested on RSF attack scenarios, labeled as RSF-15, RSF-30, and RSF-45 which corresponds to the scenario's antenna beamwidths of 15°, 30°, and 45°, respectively.

Figure 4 shows the network's total trust $T_\Sigma[q]$ over 100 quiet periods for each scenario. Depending on the protocol and different evaluation environment, the

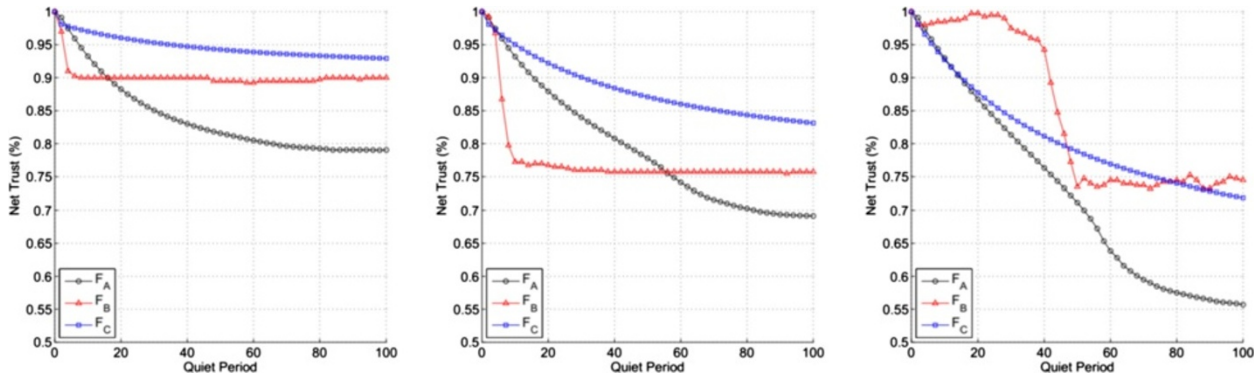

Figure 4 Display of the network's total trust (from Equation 5) over 100 quiet periods for protocols F_A, F_B, and F_C. Like Figure 2, there are four rogue directional antennas facing the cardinal directions and positioned on the map's center. The beamwidth of each rogue antenna is 15°, 30°, and 45° for scenarios RSF-15, RSF-30, and RSF-45, respectively.

RSF intrusion removed nearly 15% to 45% of the network's total trust which correlates to the percentage of sensors removed from the shared spectrum sensing. As expected, $T_\Sigma[q]$ initially decreases and plateaus over time. It plateaus when the misled sensors eventually have no more trust to lose.

In Figure 4, the change in the network's total trust $\Delta T_\Sigma[q]$ per quiet period is different for protocols F_A, F_B, and F_C because a sensor's trust score is adjusted differently for each protocol. Hence, these protocols behave differently against rogue signals, but the overall trend is a net loss of total trust $T_\Sigma[q]$ as the quiet period q increases over time. The protocol differences can be summarized briefly as follows:

- *Protocol* F_A: sensor trust is increased when the local spectrum decision agrees with the FC's global spectrum decision and penalized otherwise, only applies to a random sample of sensors with varying sizes
- *Protocol* F_B: the rate and scope of trust damage depends on the environment's RSS variance, the protocol's penalty threshold scales with the environment's noise variance
- *Protocol* F_C: sensor trust is increased when the local spectrum decision agrees with the FC's global spectrum decision and penalized otherwise, applies to all sensors

From Figure 4, we observe that both protocols F_A and F_C start to plateau because the t_i of misled sensors eventually falls to 0, causing the $\Delta T_\Sigma[q]$ to become stagnant over time. However, protocol F_B differs in that it does not have local spectrum decisions to compare to FC's global spectrum decisions. Instead, it determines if a sensor is malicious when the reported RSS value exceeds a dynamic threshold that correlates with the network's RSS variance. As the attack coverage increases from RSF-15 to RSF-45, so does the RSS variance and the F_B's behavior towards the RSF attack.

The CSS paradigm can be modeled in the context of the Byzantine fault tolerance problem. The authors in [13] describe a Byzantine failure as either a malfunctioning sensor or an SSDF attack. In both cases, the sensors perform unreliable local spectrum sensing that could ultimately mislead the FC to a wrong spectrum decision in the form of a misdetection or false alarm. These decisions are based on the null hypothesis H_0, where the primary signal is presumed absent, and the alternative hypothesis H_1, where the primary signal is presumed present, from Equation 1.

A misdetection is when the FC decides H_0 when in fact the primary signal is present and may result in unacceptable interference to the primary users. Conversely, a false alarm is when the FC decides H_1 when the primary signal is absent and causes a denial of service of spectrum resources for secondary users. The hypothesis tests are represented in Table 1.

The RSF's ability to damage sensor reputations does not directly influence the FC's spectrum decision like in SSDF or PUE attacks. Instead, the RSF lowers the system's *fault tolerance* because the FC has to rely on less sensors to infer the presence of the primary signal. Hence, the RSF weakens the reliability of shared spectrum sensing for trust-based CSS protocols in the aftermath of the intrusion.

Table 1 Hypothesis test

	Primary signal absent (H_0)	**Primary signal present (H_1)**
H_0 is accepted	Correct decision	Misdetection
H_0 is rejected	False Alarm	Correct decision

4.4 Two types of framing

To create an illusion of malicious sensors, there needs to be a separate group of well-behaved sensors to delineate good-from-bad sensor reports. Unfortunately, classifying sensors as either honest or malicious is speculative, as the FCC regulations remove any obligations of the primary network to communicate with the secondary network [6]. Hence, the secondary network is left to assume channel occupancy (i.e., the global spectrum decision) with hypotheses like H_0 and H_1. Therefore, if all sensor reputations are in good standing, such that all sensors equally participate in the shared spectrum sensing, then the global spectrum decision is typically determined by the majority of sensors.

This is especially true for hard-decision combining, which is when the FC makes a global spectrum decision based on a collection of local spectrum decisions, reported by sensors individually, in the form H_0 and H_1. Protocols F_A and F_C use hard-decision combining, with each decision weighted by sensor reputations. Alternatively, the FC can perform soft-decision combining to determine the global spectrum decision based on a collection of non-discrete sensor observations, e.g., energy detectors that report the RSS values instead of a local spectrum decision.

Soft-decision combining not only benefits from using more descriptive data but also becomes more vulnerable to outliers in sensor reports, e.g., extremely high or low RSS values. Generally, CSS protocols are designed to reduce the impact of outliers or remove them entirely, but this still leaves the majority of sensor reports as a strong determinant of the global spectrum decision, just like in hard-decision combining. That is, a majority of sensors will typically decide the global decision, even if that majority is comprised of malicious sensors or affected by a wide-reaching rogue signal, as seen in the case of a PUE attack. In such a case, the FC concludes that the disagreeing minority of sensors, even if well-behaved, are presumed inaccurate.

Hence, we define two outcomes of rogue signals with regard to damaging sensor reputations, called type 1 framing and type 2 framing:

- **Type 1 framing**: the sensors misled by the rogue signal are in the *minority* and lose trust, while the rest of the network gains trust
- **Type 2 framing**: the sensors misled by the rogue signal are in the *majority* and gain trust, while the rest of the network loses trust

For consistency, we will describe sensors affected by a rogue signal as *misled* sensors, and sensors that are not as *unaffected sensors*, like in Table 2.

Prior to this section, type 1 framing has been the designated type of trust manipulation to describe the RSF attack. Type 2 framing, which is also a result of rogue signals, is worthy of discussion for simultaneously accomplishing a PUE attack and harming sensor reputations. Both attacks are manifested through rogue signals but can only be distinguished by the attack's outcome, such as misleading the trust model (via RSF attack) or the FC (via PUE attack). To our knowledge, the fact that a PUE attack may inadvertently affect sensor reputations has not yet been considered in previous literature. We believe that type 2 framing is important in that it highlights the more subtle deficiencies in trust models, like how PUE attacks can also harm sensor reputations as a side effect.

Table 2 Simulation parameters

Parameter	Value	Description
N_s	400	Number of sensors
N_r	4	Number of rogue antennas
γ_θ	-92 dBm	Sensor sensitivity
f	615 MHz	Channel frequency
μ_ω	95.2 dBm	Noise power mean
σ_ω	0.3 dB	Noise power std
d_θ	150 m	distance threshold
σ_L	4.5 dB	Shadow fading variance
$N_x \times N_y$	2,000 m $\times$ 2,000 m	Grid dimensions
$C_{\min}$	5	Minimum cluster size
Z_θ	0.3	Cluster threshold

Figure 5 illustrates two cases of trust damage when the secondary network is bombarded by rogue signals: type 1 framing when the minority of sensors are within the attack coverage and type 2 framing when the minority of sensors are outside the attack coverage. Assuming the network's trust is in a healthy state, the sensors that disagree with the global spectrum decision will be presumed malicious. In type 2 framing, the sensors outside the attack coverage will experience trust penalties.

To show the two types of framing, we tested for the number of misled (attacked) sensors and PUE success rate with respect to antenna beamwidth to identify whether trust damage occurs during a PUE attack, or at least from a rogue signal with a wide attack coverage. We followed the same system parameters from Table 3. The rogue signals are launched for a duration of 100 quiet periods with a transmission power of 10 mW for each integral beamwidth, from 20° to 70°. The recorded trust damage is based on Equation 5 with a fixed quiet period $q = 100$.

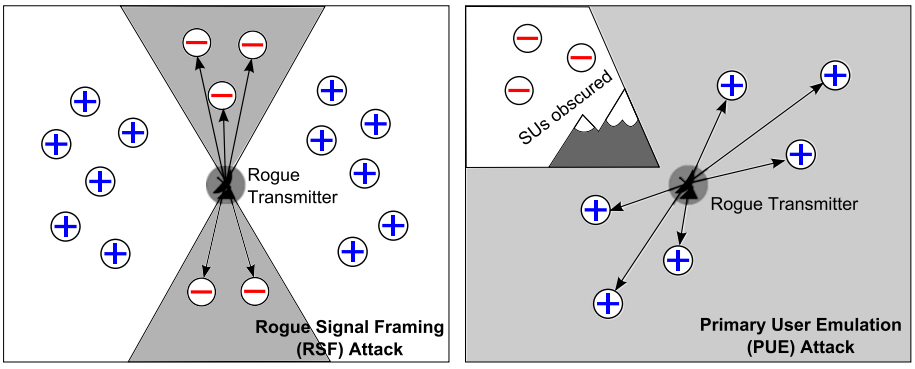

Figure 5 The two outcomes of rogue signals in trust-based CSS protocols. The plus sign indicates an increase of reputation for some sensor, while the minus sign indicates a decrease.

Figure 6 depicts the simulation results of type 2 framing on protocols F_A, F_B, and F_C which shows the trust damage $T_\Sigma[100]$ (on the 100th quite period) and the PUE success rate (%) with respect to antenna beamwidth $\theta°$. Trust damage is evident in all three protocols during successful PUE attacks, i.e., when the PUE success rate is above 0. In cross examining these results, a negative correlation can be observed between the trust damage and the PUE success rate, especially upward of the 60° beamwidth mark. Hence, we use these results to reinforce the notion of type 2 framing as a result of rogue signals from Figure 6.

Table 4 shows the corresponding false alarms (sensors misled by rogue signals) for the beamwidth used on the four attacking directional antennas from Figure 6. The number of false alarms increases sporadically as the beamwidth increases because of the random placement of sensors.

From observing the results in Figure 6 and Table 4 and understanding the trust model algorithms, we see a clear pattern between the relationship of trust damage and false alarms. In the polar cases of 0 or N_s false alarms (where N_s is the number of sensors), the trust damage is virtually 0, since the FC cannot find any disagreements among the sensor reports.

If the trust damage decreases to 0 as the number of false alarms approaches the polar ends (0 or N_s), then it can be surmised that somewhere near the middle should hold the maximum trust damage TD_Ω for a given trust model. In other words, having false alarms equal to roughly $N/2$ produces the maximum trust damage TD_Ω because that is when the sensor network is *most divided* in local spectrum decisions. We will denote FA_Ω as the number of false alarms that produces TD_Ω, as depicted in Figure 7.

Table 3 Attack outcomes on trust models

	RSF	PUE
Misled sensors	Lose trust	Gain trust
Unaffected sensors	Gain trust	Lose trust

The RSF and PUE labels over Figure 7 reflect the likely outcome of an attack from rogue signals. As the false alarms approach N_s due to rogue signals, a successful PUE attack is more likely to occur than the RSF attack. This can be observed in the PUE success rate in Figure 6 as the directional antennas' beamwidth broadens and the number of false alarms increases. It is important to note that regardless of the attack (RSF or PUE), trust damage occurs unless the number of false alarms is either 0 or N_s.

As seen in Table 5, the trust-based CSS protocol F_A can lose over 50% of its sensor trust (essentially removing over half its sensors) because it randomly samples sensors to make decisions, and only the sensors in the current sample are penalized if deemed inaccurate by the FC. Otherwise, protocols F_B and F_C have the same FA_Ω as a result of examining the reports of all sensors instead of sampling. The TD_Ω differs between all three protocols considering that they each use different trust update calculations.

5 Rogue signal framing clustering defense

This section introduces the RSF clustering defense (RCD) module that operates in three steps: 1) analyze the RSS diversity for any clustering behavior, 2) compute the clustering strength in order to conlude the presence of a rogue signal, and if so 3) ignore trust penalties of sensors in the attacked clusters. The defense relies on the fact that directional antennas leave isolated radiation patterns that form dense communities of sensors reporting H_1. Malicious sensors can perform SSDF attacks from the software layer without the need of rogue signals and thus operates outside the physical limitations of signal properties. In contrast, the RSF attack coverage is bound by the rogue signal's radiation pattern. Hence, we look towards a solution involving cluster analysis to exploit the rogue signal's physical characteristics and the finger print it leaves behind in a region of the network.

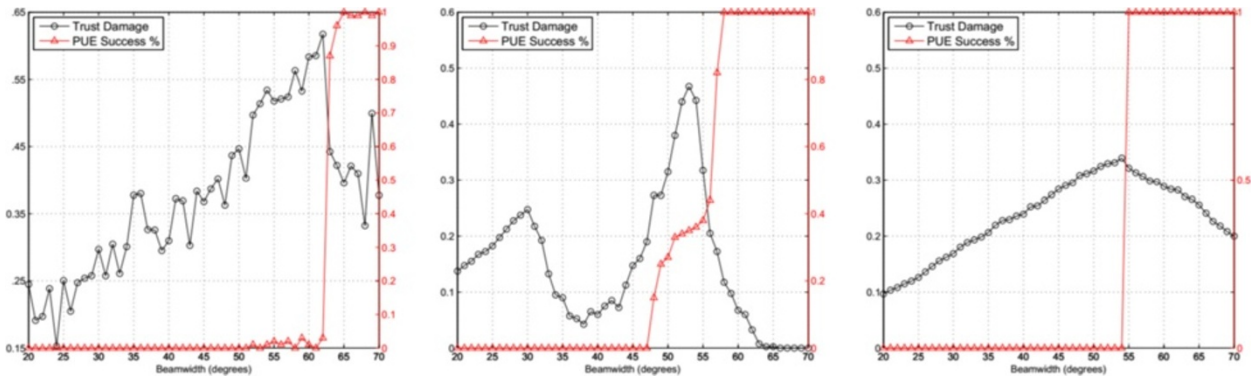

Figure 6 Trust damage over 100 quiet periods with respect to beamwidth and the corresponding PUE success rate. For protocols $\boldsymbol{F_A}$, $\boldsymbol{F_B}$, and $\boldsymbol{F_C}$.

5.1 Network classification and clustering

The beginning of this section briefly examines the necessary network terms and concepts for better understanding the RCD algorithm and its motivation. We use graph partitioning and community detection as the basis for discovering clusters of RSF-attacked sensors. To partition the graph in a meaningful way, we assume that the nodes (e.g., sensors) have discrete characteristics such as a type or class. In our system model, the sensors are classified based on their local spectrum decision such that a given sensor s_i has a corresponding class c_i where $(c_i = -1)$ if s_i reports H_0 and $(c_i = 1)$ if s_i reports H_1. This allows for the measuring of the network's *assortative mixing*, a term defined as the pairing of nodes with the same class [32]. However, the network of sensors also needs meaningful edges for community detection. The RCD module pairs any two sensors s_i, s_j based on their class c_i, c_j and their mutual distance d_{ij} from each other in order to observe spatial clustering.

The goal of the RCD module is to find an isolated and strongly concentrated group of sensors that report H_1. The Kronecker's delta function $\delta(\cdot)$ is a commonly used piecewise constant function in assortative mixing to specify whether or not the two nodes are of the same class [32]:

$$\delta(c_i, c_j) = \begin{cases} 0 \text{ if } c_i \neq c_j \\ 1 \text{ if } c_i = c_j \end{cases} \tag{6}$$

A basic mathematical formula for discretely measuring the assortative mixing in a network can be expressed by [32]:

$$\sum_{\text{edge}(ij)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) \tag{7}$$

where c_i, c_j are the node classes and $\delta(c_i, c_j)$ is the Kronecker's delta function from Equation 6. The left side of the Equation 7 is a summation series that iterates through an edge list and increments for each pair of the same class. The right side of Equation 7 is the matrix formula which iterates through an adjacency matrix and increments the same way. The one-half fraction from the matrix formula is there to remove the double counting of pairs.

Consider Figure 8, a network with two classes of nodes such that one class is designated by black circles and the other by red squares. In such a network, a node can have a degree for each class. Each node n_i keeps track of the number of edges connected to nodes of the same class, denoted as degree k_i^{same}, as well as the number of edges connected to nodes of a different class, denoted as degree k_i^{diff}. The degree k_i^{same} can be computed by Equation 7. Similarly, the degree k_i^{diff} can be computed by the same equation, i.e., Equation 7, with the exception of inverting the sign for the Kronecker's delta function. Figure 8 displays these two types of degrees above each node in the form of $\left(k^{\text{same}}, k^{diff}\right)$ which can be used to measure the strength of the assortative mixing.

Table 4 Number of false alarms for each corresponding beamwidth (degrees) from Figure 6

	Number of false alarms										
Beamwidth	20°	25°	30°	35°	40°	45°	50°	55°	60°	65°	70°
False alarms	56	74	100	123	143	170	190	209	229	249	283

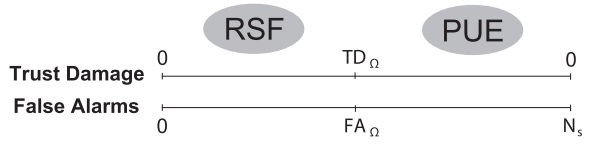

Figure 7 Modeling the trust damage from Figure 6.

Our solution, which involves graph partitioning and community detection, is based on the principle of assortative mixing, but tailored in the context of cognitive radio networks. The RCD has three requirements for operation. First, it needs the local spectrum decision $c_i \in \{H_0, H_1\}$ for all sensors $s_i \in S$. Second, it needs two sets of sensors where $S_{H_0} = \{s_i | c_i = H_0\}$ and $S_{H_1} = \{s_i | c_i = H_1\}$. Lastly, it needs an adjacency matrix A of size $|S| \times |S|$ such that

$$A_{ij} = \begin{cases} 1 \text{ if } d_{ij} \leq d_\theta \\ 0 \text{ if } d_{ij} > d_\theta \end{cases} \tag{8}$$

where d_{ij} is the distance between sensors s_i and s_j and d_θ is the distance threshold.

The RCD module locates k disconnected clusters of sensors C_k such that $s_j \in C_k$, $A_{ij} = 1$, and $c_i = c_j$ for sensors $s_i, s_j \in C_k$. The RCD module's goal is to locate isolated communities C_k that are surrounded by sensors in S_{H_0}. To start, we measure the cluster density of sensors with the same class by counting all connected pairs (s_i, s_j) such that $s_i \in C_k$, $s_j \in S_{H_1}$, and $A_{ij} = 1$. This is computed on all sensors in C_k with:

$$\left\{ d_i^{H_1} \right\}_k = \left\{ \sum_{s_j \in C_k} \left(A_{ij} \delta \left(c_i, c_j \right) \right) - 1 \mid s_i \in C_k \right\} \tag{9}$$

where $\delta(c_i, c_j)$ is a simple Kronecker's delta function from Equation 6 that indicates a difference in a node's class c, i.e., the local spectrum decision. Next, we measure the isolation of sensor $s_i \in C_k$ from $s_j \in S_{H_0}$ by counting all connected pairs (s_i, s_j) such that $A_{ij} = 1$. This is computed on all sensors in C_k by:

$$\left\{ d_i^{\Delta} \right\}_k = D(C_k) = \left\{ \sum_{s_j \in S_{H_0}} A_{ij} \delta'(c_i, c_j) \mid s_i \in C_k \right\} \tag{10}$$

$$\delta'(c_i, c_j) = D'(C_k) = \begin{cases} 0, \text{ if } c_i = c_j \\ 1, \text{ if } c_i \neq c_j \end{cases}$$

Table 5 Trust model comparison

Trust model	FA_Ω	TD_Ω
F_A	235	63%
F_B	201	48%
F_C	201	34%

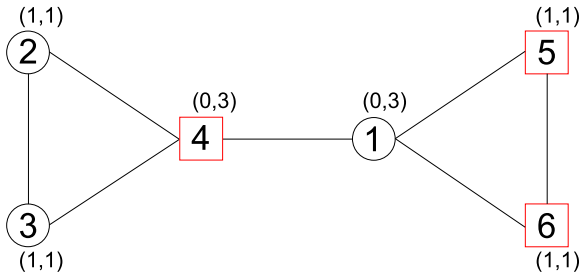

Figure 8 Example of assortative mixing.

Finally, to measure the isolated clustering strength z_k, we use the function:

$$z_k = Z\left(\left\{ d_i^{H_1} \right\}_k, \left\{ d_i^{\Delta} \right\}_k \right) = \frac{\sum_i d_i^{H_1}}{\sum_i \left(d_i^{H_1} + d_i^{\Delta} \right)} \tag{11}$$

In the off chance that a number of malicious sensors from SSDF are positioned near each other, we want to have a level of tolerance Z_θ and a required minimum number of sensors per cluster $C_{\min}$. The restraint $C_{\min}$ prevents a high clustering score Z_k from an insignificant-sized cluster.

Figure 9 shows two scenarios: (1) the RSF-45 where each rogue antenna has a beamwidth of 45° and (2) the SSDF-40 where 40% of the sensors, randomly selected, perform SSDF. The red nodes are sensors reporting H_1, and the blue nodes are sensors reporting H_0. The red edges are formed when $c_i = c_j$ and $d_{ij} < d_\theta$ for sensors s_i and s_j. The blue edges are formed by the same rules except that $c_i \neq c_j$.

The red and blue graph both give valuable information in detecting directional rogue signals by the cluster formations they create. The goal of the red graph is to identify a strong concentration of sensors perceiving a radio signal within a small area. In contrast, the blue graph demonstrates disagreements in spectrum decisions (i.e., H_0 and H_1) between neighboring sensors. As can be seen in the RSF scenario in Figure 9a, the red graphs (created by the rogue signals) is surrounded by the blue graph and lacks any significant overlap between the two graphs. The presence of a red graph, without the intersections of blue edges, outlines a radio's antenna coverage and becomes a clear indication of a rogue signal. However, the SSDF scenario in Figure 9b shows that an overlapping of red and blue graphs reveal a strong likelihood of malicious or malfunctioning sensors, instead of a rogue signal's presence, since there is no apparent pattern of spectrum decisions.

Since we are assuming an environment that conforms to the IEEE 802.22 standard, we assume a network of Customer Premise Equipment (CPE) sensors that infers a static network. This eliminates the option of malicious users moving closer together and forming dense clusters

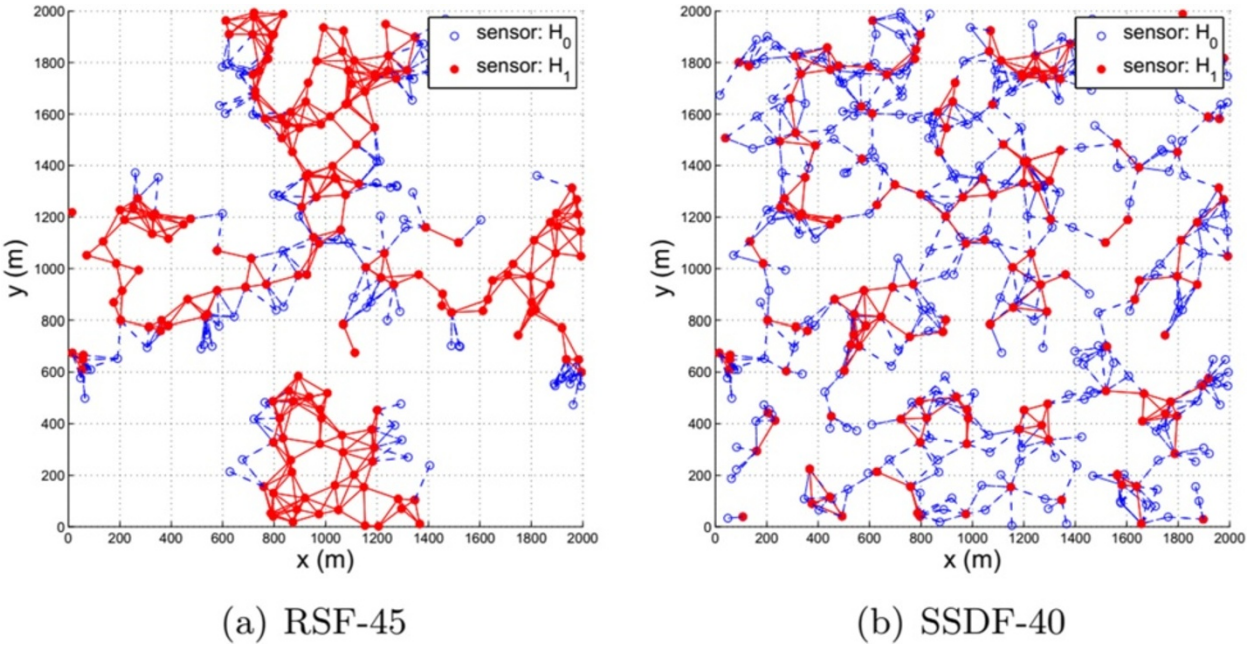

Figure 9 Clustering illustration of our RSF Clustering Defense (RCD) algorithm. The RCD forms two graphs, a red and blue graph, for cluster analysis. The red graph contains edges between sensors reporting H_1. The blue graph contains edges between sensors with opposing local spectrum decisions.

in order to be protected by the RCD module during SSDF attacks. There is a possibility that a group of CPE sensors remain in proximity by coincidence, but the chances can be reduced by adjusting $C_{\min}$, Z_θ, or d_θ accordingly.

Figure 10 illustrates the sequence of operations in a trust-based CSS protocol with our RCD solution added to the system. In the first step, the base station (where the FC resides) collects all sensor reports from the network of sensors S. The second step applies the trust model's filter by removing untrustworthy sensors S_r from S. The third step requires the FC to make a global spectrum decision, denoted as GD, from sensors in $(S - S_r)$ as it normally

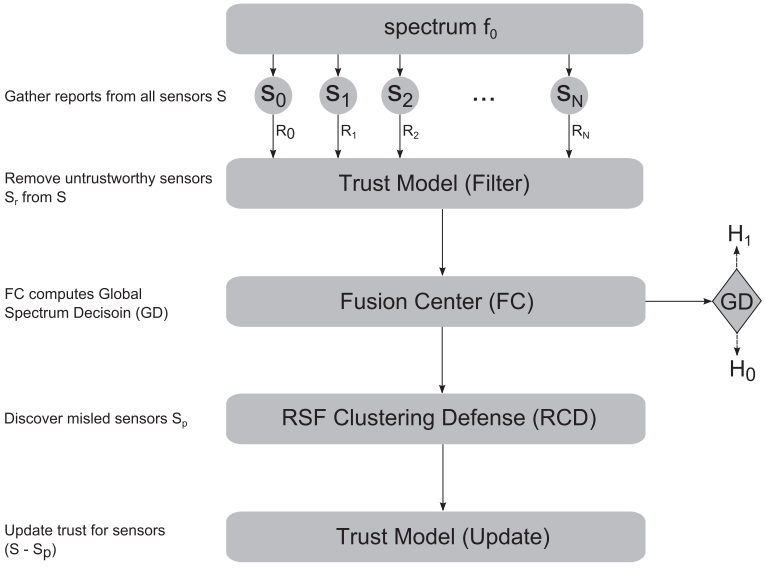

Figure 10 Diagram of the RCD module added to the general framework of trust-based CSS protocols. Sensors protected by the RCD are denoted as S_P.

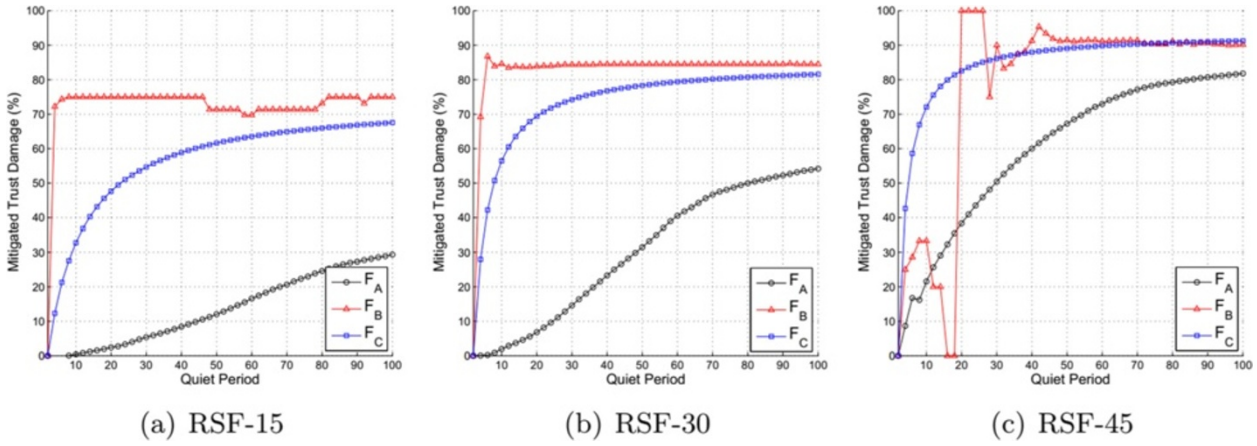

Figure 11 The network's total mitigated trust damage (Equation 12) from the RCD module.

would in trust-based CSS protocols. The fourth step discovers signs of an RSF intrusion and identifies the group of attacked sensors, denoted as S_P. In the final step, the trust model updates the sensor reputations *except* for the set of sensors S_P that are presumed affected by rogue signals. The last step is important because it prevents attackers from exploiting trust models with rogue signals.

5.2 Defense evaluation

In this section, we evaluate the RCD module's performance on its ability to mitigate trust loss from RSF intrusions. Additionally, we compare the RCD module's outcome on RSF and SSDF attacks.

In our simulations, we have two groups of scenarios, the RSF and SSDF. The simulation environment is the same as the one used by the RSF intrusion in Section 4. The beamwidth of each rogue antenna is 15°, 30°, and 45° for scenarios RSF-15, RSF-30, and RSF-45, respectively. The SSDF scenarios simulate malicious sensors by randomly selecting a percentage of the sensors and raising their RSS by 20 dBm from the noise floor. We randomly selected 20%, 30%, and 40% of sensors from the scenarios SSDF-20, SSDF-30, and SSDF-40, respectively.

Figure 11 shows the amount of mitigated trust damage (%) with the RCD module under the same scenarios. The mitigated trust damage is denoted as $T_M[q]$ and calculated by:

$$T_M[q] = \frac{T_\Sigma^R[q] - T_\Sigma[q]}{T_\Sigma[0] - T_\Sigma[q]} \tag{12}$$

where $T_\Sigma^R[q]$ is the network's total trust on quiet period q when using the RCD module, $T_\Sigma[q]$ is the network's total trust without the RCD module (from Figure 4), and $T_\Sigma[0]$ is the initial state of trust scores. We use a minimum cluster size $C_{\min} = 5$, a clustering threshold $Z_\theta = 0.3$, and a distance threshold $d_\theta = 150$ m.

As shown in Figure 11, each protocol benefited from our proposed defense against the RSF intrusion. However, the RCD module offered less protection to protocol F_A due to its sequential random sampling of sensors, instead of cross-examining all sensor reports for a more robust analysis. The spikes from F_B in Figure 11c are due to its protocol design of having a dynamic threshold for deciding malicious sensors. During the spikes, F_B's dynamic threshold is stabilizing as it replaces the old RSS statistics with new data.

Figure 12 shows the rate of false alarms, i.e., the number sensors reporting H_1 when the FC reports H_0, before and after applying the RCD module. In all three RSF scenarios,

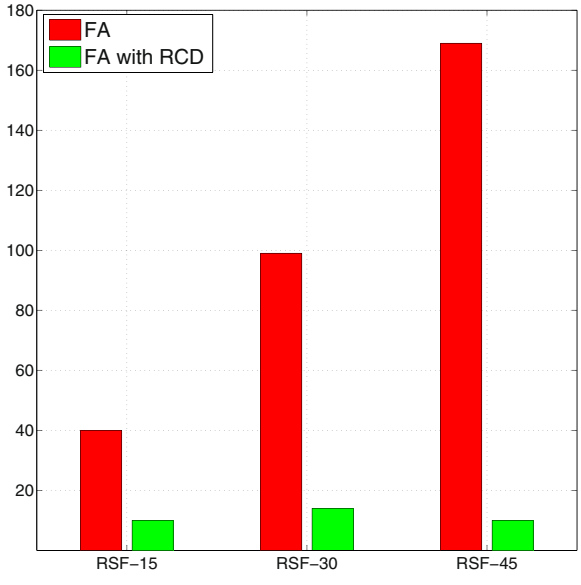

Figure 12 The number of false alarms before and after applying the RCD module.

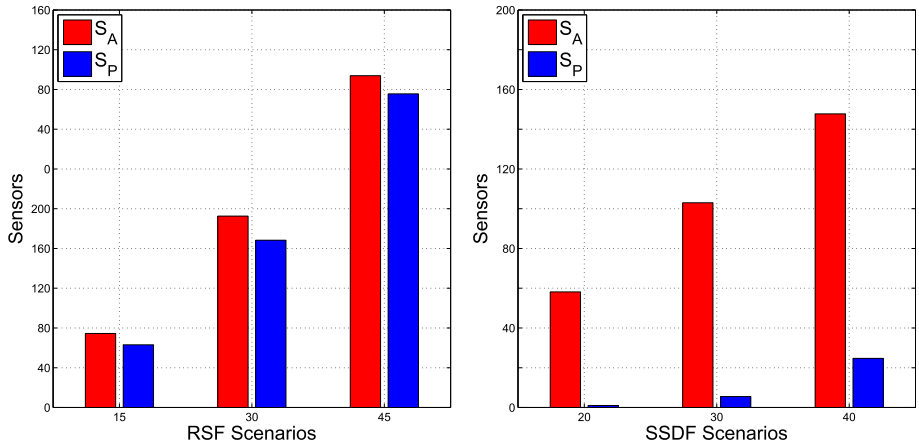

Figure 13 Comparison of the RCD results between RSF and SSDF intrusions. S_A, number of attacked sensors; S_P, number of sensors protected by the RCD.

the RCD module managed to limit the false alarms to a maximum of 3% of total sensors N_s.

Figure 13 compares how the RCD responds to the RSF and SSDF intrusions in terms of the number of sensors attacked S_A and the number of sensors protected S_P by the RCD module. The goal is to maximize S_P for the RSF scenarios and minimize it for the SSDF scenarios so that the reputations of malicious sensors are not protected. In scenario RSF-45, the strongest RSF attack, the RCD module protects 95% of sensors from losing trust due to rogue signals. In contrast, the RCD module erroneously protects 15% of the sensors in scenario SSDF-40. This margin of error is acceptable as 40% of malicious sensors is an unrealistic and profuse amount of attacks in any CR network.

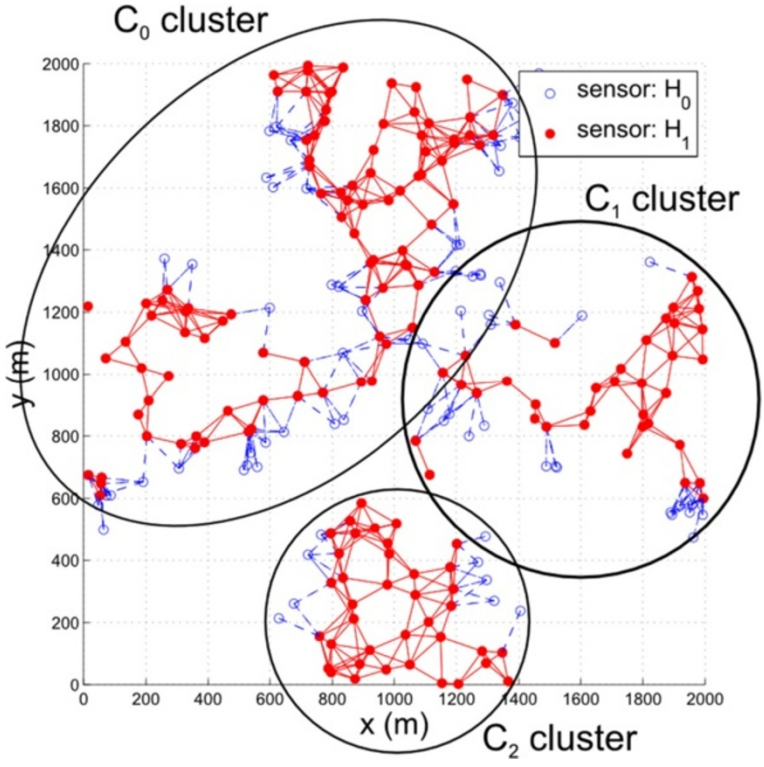

Figure 14 The sensor network is partitioned into a red and blue graph before being analyzed by the RCD module. The red filled nodes are cognitive radios reporting H_1 and are connected to nearby neighbors with similar observations.

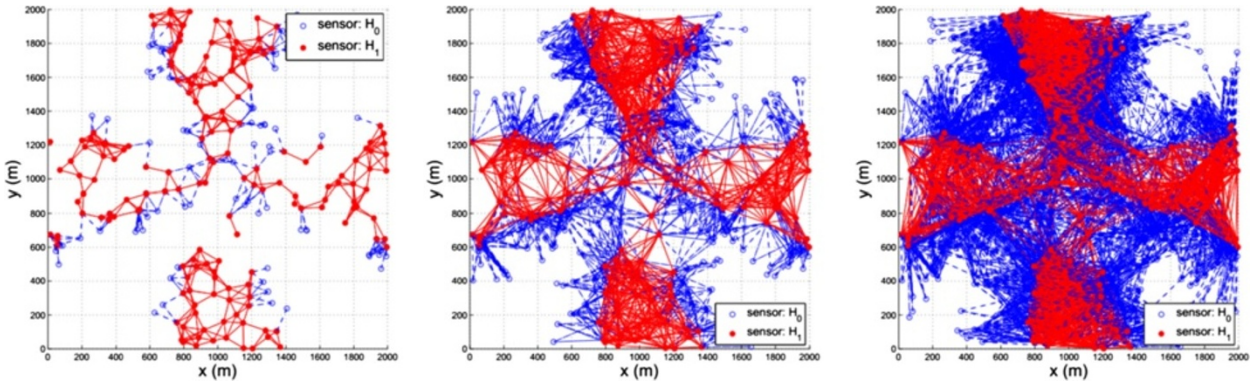

Figure 15 RCD solution applied to a dense network of 400 sensors. From left to right, the RCD's distance threshold is 150 m, 300 m, and 450 m, respectively.

The outcomes of Figure 13 show a high resiliency against the exploitation of SSDF attacks.

5.3 Overhead of defense

To address the time complexity overhead of our defense, we have to examine the algorithm it uses before we can identify the order-of-growth category it belongs to. The proposed RSF Clustering Defense (RCD) algorithm can be separated into three distinct parts: (1) the graph setup, (2) the breadth-first search to identify all the clusters (i.e., subgraphs), and (3) calculating the clustering strength of an identified cluster. Each part can be summarized by the following:

1. Connect all the vertices in the adjacency matrix A_{ij} to its neighbors within a distance threshold d_θ; this step has a time complexity of $O\left(|V|^2\right)$ where $|V|$ is the number of sensors
2. Find all non-overlapping subgraphs (i.e., clusters C_k) using a breadth-first search; this step has a time complexity of $O\left(|V|^2\right)$ since it traverses the adjacency matrix A_{ij} and creates adjacency lists that represent each C_k cluster
3. Calculate the clustering strength of cluster C_k based on the assortative mixing equations (Equations 9 and 10); this step iterates through each C_k adjacency list, thus it has a time complexity of $O(|E| + |V|)$

So the time complexity of the RCD defense is the summation of all three parts: $O\left(|V|^2\right)+O\left(|V|^2\right)+O(|E|+|V|)$. Yet, in a static network, where the cognitive radios do not move, we can ignore the complexity of part 1 since it is only computed once during the program initialization. Hence, the time complexity for each reoccurring quiet period is $O\left(|V|^2\right)+O(|E|+|V|)$. The quiet period is when the cognitive radio network stops transmitting to listen for the primary signal.

The bottleneck of our defense is either in part 2 or part 3, whichever has a worse order of growth between $O\left(|V|^2\right)$ and $O(|E|+|V|)$, depending on the sizes of V and E. The RCD algorihm travereses through K adjacency lists representing each cluster C_k, where $0 \leq k < K$. Figure 14

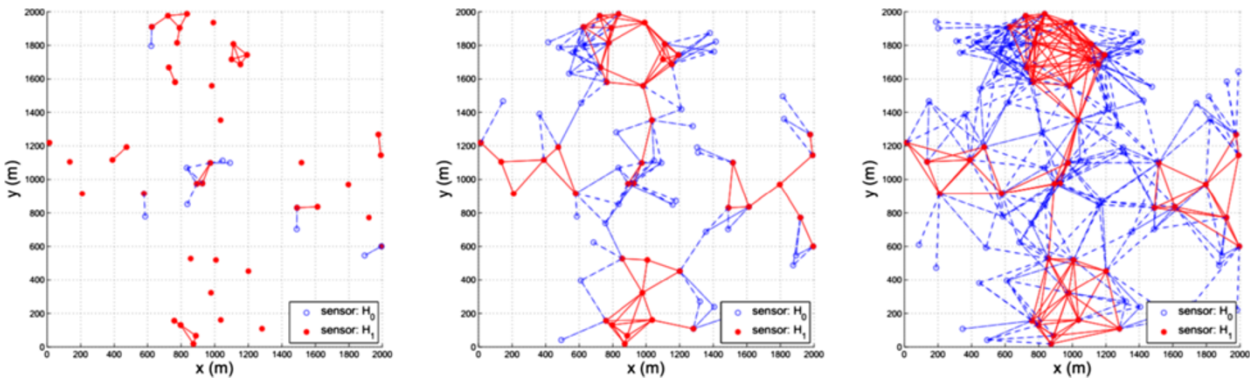

Figure 16 RCD solution applied to a sparse network of 100 sensors. From left to right, the RCD's distance threshold is 150 m, 300 m, and 450 m, respectively.

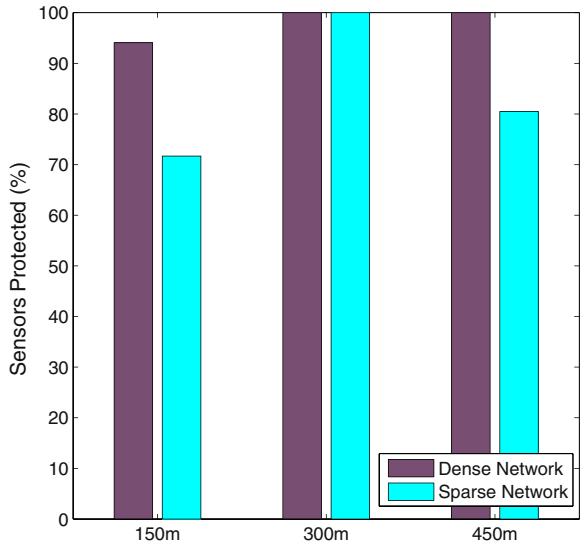

Figure 17 The accuracy of the RCD for dense and sparse networks with $d_\theta = 150, 300$, and 450 m.

shows $K = 3$ clusters present (C_0, C_1, and C_2) in the network where each cluster is roughly $1/4$ to $1/8$ the size of V.

Time complexity can be an issue if an attack is able to impact the network before the defense can adequately prevent or mitigate the damage. However, our algorithm has a descent order of growth, i.e., $O\left(|V|^2\right) + O(|E| + |V|) \approx O\left(|V|^2\right)$, which is smaller than many clustering algorithms such as the Kernighan-Lin algorithm that have an order of growth of $O\left(|V|^3\right)$. Secondly, we are assuming that all intensive processing happens at the base station, with a dedicated server and adequate computing resources performing the analysis, and not on the cognitive radios itself. As such, the time complexity is very feasible for most anticipated network sizes, e.g., no more than several thousand sensors. Furthermore, the calculation of the clustering strength is only applied to small sections of the network, which is usually much smaller than the total number of sensors $|V|$. This occurs in part 2 of our defense where C_k clusters with identical sensor reports are identified using BFS, in similar fashion to the flood fill algorithm.

The need for more intensive processing, like graph algorithms, in radio networks usually raises concerns about the impact it has on a radio's battery life. This is not a concern in our system because the cognitive radios only submit sensor reports every 30 s to a stationary base station that does all the processing on a dedicated server. Hence, the cognitive radios are spared the processing that would otherwise quickly deplete itself of battery life. In a decentralized CSS protocol, each cognitive radio is responsible for computing the shared spectrum algorithms locally, but our system employs a centralized CSS protocol which removes the intensive processing burden on the radio itself.

5.4 Cluster parameters and impact

Naturally, the size and topology of the cognitive radio network has an effect on our RCD solution. A dense network can easily show patterns of rogue signals where as a sparse network gives less information to analyze. To show the difference, we tested our solution on a second network, denoted as the sparse network, consisting of 100 randomly placed sensors. In contrast, the dense network has 400 randomly placed sensors, which is the same network tested and discussed in previous sections. For both dense and sparse networks, we only display the RSF-45 scenario to limit the number of graphs. The RSF-45 scenario emits four rogue signals in the cardinal directions with 45° beamwidth.

The distance threshold d_θ is the condition required to form edges between two sensors. A red graph indicates a strong concentration of sensors perceiving a signal, such that it potentially reveals a rogue signal's antenna coverage. The red graph is formed by sensors that share H_1 reports within the distance threshold, d_θ. Likewise, the blue graph is formed by sensors that simply disagree with their neighbors' spectrum decisions (i.e., H_0 and H_1) within d_θ. The blue graph helps reveal an SSDF attack, especially when the red and blue graph are overlapping, and not clearly segregated. When a rogue signal is present, the red graph should be surrounded by the blue graph, outlining the reach of the rogue signal's antenna coverage.

Figures 15 and 16 show the changing composition in the red and blue graph (created by the RCD) in both dense and sparse networks with different d_θ, where $d_\theta =$

150, 300 and 450 m. For the dense network, the attack coverage of the rogue signals is clearly visible with all three values for d_θ. For the sparse network, the visibility of rogue signals becomes much more difficult to perceive, especially when $d_\theta = 150$ m. Naturally, this occurs from having fewer sensors, randomly placed, over the same area as the dense network. In other words, the sensors are farther away from their neighbors in the sparse network.

At first glance, it might be tempting to just assign an excessive number for d_θ to avoid the sparsity problem, i.e., when clusters are not clearly visible because d_θ is too low. Actually, a very large d_θ can decrease the accuracy of the RCD solution as shown in Figure 17. An infinitely large d_θ will always form complete blue and red graphs across the sensor region, which is not always more informative.

Figure 17 shows the accuracy of the RCD solution for both dense and sparse networks with $d_\theta = 150, 300$, and 450 m. The accuracy is represented by the *number of sensors protected by the RCD solution* divided by the *number of sensors inside the rogue signal's attack coverage*, i.e., S_P/S_A. Notably, the $d_\theta = 300$ m in the sparse network reaches 100% accuracy, but $d_\theta = 450$ m does not, even with more edges to analyze. The reason for this phenomena is due to the blue edges lowering the clustering score Z_k for cluster C_k. This can seen in Equation 11, where the clustering score Z_k decreases because the denominator increases as more blue edges form (from variable d_i^Δ).

There are many variables in our simulations that are worth analyzing at a more comprehensive level. The number of sensors, the number of attackers, the shape and size of the rogue signal, the network's topology, and even the environment's landscape. In future studies, we intend to explore how these variables impact our solution and to establish metrics that fit the parameters according to different scenarios.

6 Conclusions

In this paper, we demonstrated the RSF intrusion, a new threat to trust-based CSS protocols. The attackers can transmit rogue signals onto groups of sensors to emulate SSDF and ruin their reputation with the intent of having them removed from the shared spectrum sensing. Our work cautions the use of trust-based CSS protocols and warrants a line of defense against rogue signals. The RSF simulations were conducted in a realistic environment based on the 802.22 WRAN standard and illustrates the impact of the RSF intrusions on sensor reputation scores. To mitigate the trust damage, we introduced a new defense based on community detection and cluster analysis. The simulation experiments showed that our defense solution, the RCD module, could effectively keep the sensor reputations intact while distinguishing rogue signals from malicious sensors.

Appendix

Rogue Signal Clustering Defense (RCD) Algorithm

The RCD Algorithm is the psuedo code that locates sensors affected by rogue signals in trust-based CSS protocols.

Algorithm 1 The RSF Cluster Detection Module

Function:**RCD**(A, S_{H_0}, S_{H_1})

Initialize cluster index $k \leftarrow 0$
Initialize set of protected sensors S_P
Initialize set of visited nodes V
Initialize breadth-first search queue Q
Initialize set of clusters C_k
Initialize list clustering strength values Z_k
for all $s_i \in S_{H_1}$ **do**
  if $s_i \notin V$ **then**
    $k \leftarrow k + 1$
    add s_i onto C_k, V, and Q
    while Q is not empty **do**
      $s_q \leftarrow$ dequeue(Q)
      for all $s_j \in S_{H_1}$ **do**
        if $s_j \notin V$ **and** $A_{qj} = 1$ **then**
          add s_j onto C_k, V, and Q
        end if
      end for
    end while
    $\left\{d_i^{H_1}\right\}_k \leftarrow D\left(C_k\right)$
    $\left\{d_i^{\Delta}\right\}_k \leftarrow D'\left(C_k, S_{H_0}\right)$
    $z_k \leftarrow Z\left(\left\{d_i^{H_1}\right\}_k, \left\{d_i^{\Delta}\right\}_k\right)$
    add z_k onto Z_k
  end if
end for
for all $z_k \in Z_K$ **do**
  if $|C_k| \geq C_{\min}$ **and** $z_k > Z_\theta$ **then**
    $S_P \leftarrow S_P \cup C_k$
  end if
end for
return S_P

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work was supported by the National Science Foundation under Grant Nos. 0915318, 1048339, and 0916469.

Author details
[1]Department of Computer Science, Virginia Commonwealth, 401 W. Main St, University, Richmond, VA 23220, USA. [2]Department of Computer Science, Texas State University, 601 University Drive, San Marcos, TX 78666, USA.

Received: 1 September 2014 Accepted: 24 November 2014
Published: 8 January 2015

References

1. IF Akyildiz, W-Y Lee, MC Vuran, S Mohanty, Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Compu. Netw. **50**(13), 2127–2159 (2006)
2. M Song, C Xin, Y Zhao, X Cheng, Dynamic spectrum access: from cognitive radio to network radio. IEEE Wireless Commun. **19**(1), 23–29 (2012)
3. S Higginbotham, Spectrum Shortage Will Strike in 2013 Tech News and Analysis. (2013). http://gigaom.com/2010/02/17/analyst-spectrum-shortage-will-strike-in-2013/. Accessed 08 Apr 2013
4. W Li, X Cheng, T Jing, Y Cui, K Xing, W Wang, Spectrum assignment and sharing for delay minimization in multi-hop multi-flow CRNs. IEEE J. Selected Areas Commun. (JSAC) **31**(11), 2483–2493 (2013)
5. R Chen, J-M Park, Y Hou, J Reed, Toward secure distributed spectrum sensing in cognitive radio networks. IEEE Commun. Mag. **46**(4), 50–55 (2008)
6. T Clancy, N Goergen, in *Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on,* Security in cognitive radio networks: threats and mitigation, (2008), pp. 1–8
7. X Xing, T Jing, W Cheng, Y Huo, X Cheng, Spectrum prediction in cognitive radio networks. IEEE Wireless Commun. **20**(2), 90–96 (2013)
8. X Xing, T Jing, Y Huo, H Li, X Cheng, in *IEEE INFOCOM,* April 14–19 2013, Channel quality prediction based on Bayesian inference in cognitive radio networks, pp. 1465–1473
9. SJ Shellhammer, SS N, R Tandra, J Tomcik, in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum, ser.* Performance of power detector sensors of DTV signals in IEEE 802.22 WRANs. TAPAS '06 (ACM, NY, USA, 2006). [Online]. Available: http://doi.acm.org/10.1145/1234388.1234392
10. B Wang, K Liu, Advances in cognitive radio networks: a survey. IEEE J Sel. Topics Signal Process. **5**(1), 5–23 (2011)
11. T Jing, X Chen, Y Huo, X Cheng, in *IEEE INFOCOM,* March 25–30, 2012, Achievable transmission capacity of cognitive mesh networks with different media access control, pp. 1764–1772
12. H Li, X Cheng, K Li, X Xing, T Jing, in *IEEE INFOCOM Mini-Conference,* April 14–19, 2013, Utility-based cooperative spectrum sensing scheduling in cognitive radio networks, pp. 165–169
13. R Chen, J-M Park, K Bian, in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE,* April 2008, Robust distributed spectrum sensing in cognitive radio networks, pp. 1876–1884
14. P Kaligineedi, M Khabbazian, V Bhargava, in *Communications, 2008. ICC '08. IEEE International Conference on,* May 2008, Secure cooperative sensing techniques for cognitive radio systems, pp. 3406–3410
15. F Zhu, S-W Seo, Enhanced robust cooperative spectrum sensing in cognitive radio. J. Commun. Netw. **11**(2), 122–133 (2009)
16. K Arshad, K Moessner, in *Future Network and Mobile Summit 2011 Conference Proceedings,* Robust collaborative spectrum sensing based on beta reputation system, (2011)
17. S Bhattacharjee, S Debroy, M Chatterjee, in *Personal Indoor and Mobile Radio Communications (PIMRC) 2011 IEEE 22nd International Symposium on,* Sept 2011, Trust computation through anomaly monitoring in distributed cognitive radio networks, pp. 593–597
18. R Chen, J-M Park, in *Networking Technologies for Software Defined Radio Networks, 2006. SDR '06.1st IEEE Workshop on,* Sept 2006, Ensuring trustworthy spectrum sensing in cognitive radio networks, pp. 110–119
19. A Min, K Shin, X Hu, Secure cooperative sensing in IEEE 802.22 WRANs using shadow fading correlation. IEEE Trans. Mobile Comput. **10**(10), 1434–1447 (2011)
20. J Feng, Y Zhang, G Lu, L Zhang, in *Trust, Security and Privacy in Computing and Communications (TrustCom) 2013 12th IEEE International Conference on,* July 2013, Defend against collusive SSDF attack using trust in cooperative spectrum sensing environment, pp. 1656–1661
21. J Lai, E Dutkiewicz, RP Liu, R Vesilo, in *Global Communications Conference (GLOBECOM) 2012 IEEE,* Dec 2012, Comparison of cooperative spectrum sensing strategies in distributed cognitive radio networks, pp. 1513–1518
22. M Akbari, A Falahati, in *Telecommunications (IST) 2010 5th International Symposium on,* December 2010, Ssdf protection in cooperative spectrum sensing employing a computational trust evaluation algorithm, pp. 23–28
23. K Zeng, P Paweczak, D Cabric, Reputation-based cooperative spectrum sensing with trusted nodes assistance. IEEE Commun. Lett. **14**(3), 226–228 (2010)
24. A Min, K Shin, X Hu, in *Network Protocols, 2009. ICNP 2009. 17th IEEE International Conference on,* Oct. 2009. Attack-tolerant distributed sensing for dynamic spectrum access networks, pp. 294–303
25. S Liu, Y Chen, W Trappe, LJ Greenstein, in *INFOCOM,* (IEEE, 2009), Aldo: An anomaly detection framework for dynamic spectrum access networks, pp. 675–683
26. GMDR Peter Steenkiste, D Sicker, Future directions in cognitive radio network research. NSF Workshop (2009)
27. N Patwari, P Agrawal, in *IPSN,* Effects of correlated shadowing: connectivity, localization, and RF tomography, (2008), pp. 82–93
28. G Trenkler, *Statistical distributions: M Evans, N. Hastings & B. Peacock*, 2nd edition (John Wiley, New, 1993), p. 170. isbn 0-471-55951, [pound sign] 24.95," *Computational Statistics & Data Analysis*, vol. 19, no. 4 (1995), pp. 483–484, [Online]. Available: http://EconPapers.repec.org/RePEc:eee:csdana:v:19:y:1995:i:4:p:483-484
29. I Forkel, M Schinnenburg, M Ang, in *Proceedings of The 7th International Symposium on Wireless Personal Multimedia Communications, WPMC 2004,* Abano Terme (Padova, Italy, Sep 2004), Generation of two-dimensional correlated shadowing for mobile radio network simulation, p. 5. [Online]. Available: http://www.comnets.rwth-aachen.de
30. AD AIR-4.5, *Electronic Warfare and Radar Systems Engineering Handbook*. (Naval Air Systems Command, Washington, DC 20361, 1999)
31. K Bauer, D McCoy, E Anderson, M Breitenbach, G Grudic, D Grunwald, D Sicker, in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE,* 30 2009-Dec. 4 2009, The directional attack on wireless localization: how to spoof your location with a tin can, pp. 1–6
32. MEJ Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2011)

doi:10.1186/1687-1499-2015-4
Cite this article as: Jackson *et al.*: **Exploiting and defending trust models in cooperative spectrum sensing.** *EURASIP Journal on Wireless Communications and Networking* 2015 **2015**:4.