Trust-based Mechanism Design for Cooperative Spectrum Sensing in Cognitive Radio Networks

Ji Wang[1], Ing-Ray Chen[2], Jeffrey J.P. Tsai[3], and Ding-Chau Wang[4]

[1]Department of Electrical and Computer Engineering, Virginia Tech, Email: traceyw@vt.edu

[2]Department of Computer Science, Virginia Tech, Email: irchen@vt.edu

[3]Dept. of Bioinformatics and Biomedical Engineering, Asia University, Email: jjptsai@gmail.com

[4]Dept. of Information Management, Southern Taiwan University of Science and Technology, Email: dcwang@mail.stust.edu.tw

Abstract

We propose and analyze trust-based cooperative spectrum sensing data fusion schemes against spectrum sensing data falsification attacks in cognitive radio networks. We first consider the case in which a centralized data fusion center is in place for decision making. Then we extend it to the case in which the data fusion center is absent leading to autonomous and distributed decision making. Our trust-based data fusion schemes are based on *mechanism design theory* to motivate users to report authentic sensing data so as to improve the success rate. Further, we decouple erroneous sensing reports due to low sensing capabilities from false reports due to attacks, thus avoiding unnecessary punishments to good users. We conduct a theoretical analysis validated with extensive simulation and identify optimal parameter settings under which our trust-based data fusion schemes outperform existing non-trust based cooperative spectrum sensing data fusion schemes.

Key Words: Spectrum sensing data falsification attacks, Trust management, Consensus-based cooperative spectrum sensing.

I. Introduction

Cognitive radio has aroused a lot of interest as a solution to spectrum scarcity in the next generation of wireless communication. The main idea of cognitive radio is to let the secondary users

(SUs) opportunistically access the channels that are temporarily not occupied by the preassigned primary users (PUs). In a cognitive radio system, the access priorities of PUs have to be guaranteed, i.e., SUs need to learn the PUs' activities to avoid interfering with the PUs on the band. Therefore, SUs need to sense the PU activity on a particular spectrum before transmitting data on that spectrum. Due to limited sensing capabilities of individual SUs, cooperative spectrum sensing is provided as a way to gather SUs' sensing information in order to increase the accuracy of PU occupancy detection. However, cooperative spectrum sensing can be attacked by malicious SUs, who can intentionally report fake sensing results to mislead the final aggregated result. Therefore, how to design a secure data fusion scheme for cooperative spectrum sensing is a big challenge in security management of cognitive radio networks.

This paper proposes and analyzes a trust-based data fusion scheme based on *mechanism design* to aggregate the SUs' reported outcomes in such a way that the correctness of aggregated outcome is stable in the presence of a high percentage of malicious SUs. *Mechanism design* is a sub-field of microeconomics and game theory that considers how to construct and implement a mechanism that provides incentives for the users to communicate and act in such a way so as to further the interest of the designer, despite the fact that the users are strategic and self-interested, and possess private information [1]. We apply mechanism design to implement a scheme providing incentives for all SUs within the system to report their actual sensing capabilities and sensing results, despite the fact that some of the SUs are self-interested with malicious intent to disrupt cooperative spectrum sensing. The basic idea is that the system would like to know the true channel availability but it cannot completely trust sensing reports from the SUs because it is in malicious SU's interest to distort the truth. With mechanism design, the system can design a static game whose rules can influence the SUs to act the way it would like. It is a static game in the sense that all SUs make decisions (or select a strategy) simultaneously, without knowledge of the strategies that are being chosen by the system.

Our trust-based data fusion scheme derived from the static game also employs a reputation/trust system [2]–[5] to identify malicious SUs in the long run. We do not differentiate trust from reputation in this paper, although reputation refers to common belief of the community toward a node, while trust refers to subjective belief of one node toward another node [2]. Henceforth, we will use reputation and trust interchangeably. During data fusion, an SU reports its sensing capability and the sensed channel availability to the system who makes a decision on the channel availability based on majority voting of trusted SUs. The SU sensing capability is taken into consideration in the static game in order to differentiate a fake outcome reported by a malicious SU from an erroneous outcome reported by a good SU with poor sensing capability. Moreover, the system can elect to check the channel availability in order to compare this first-hand evidence with a sensing outcome reported by an SU to detect if the SU lies about the channel availability.

The objectives of our trust-based data aggregation scheme are threefold: (1) under the designed scheme, malicious SUs have no incentive to report fake sensing capabilities; (2) the system's checking probability should be minimized in each time slot since the cost for the system to check the channel availability itself and to detect sensing result discrepancies by SUs is prohibitively high; and (3) the success decision rate should be maximized to match the data fusion outcome with the ground truth channel availability with a high probability.

In order to identify erroneous sensing results due to SUs' poor sensing abilities, our scheme requires SUs to report the sensing capability. A threshold is set to filter out the SUs with low sensing capabilities. On the other hand, to avoid malicious SUs from reporting fake sensing abilities, the system can check the PU activity based on our static game model design. When checking the spectrum, the system punishes the SUs whose report outcomes are different by decreasing their reputation scores based on their reported sensing capabilities. Otherwise, if the system does not sense the spectrum in a particular time slot, it aggregates the reported sensing outcomes from those SUs whose sensing capability is over a predefined threshold, weighed by their reputation scores.

In this paper we consider self-interested malicious attackers with the objective to disrupt cooperative spectrum sensing. In other words, the malicious attackers aim to destroy the functionality of cooperative spectrum sensing, so that the system cannot trust the aggregated sensing results. We consider four types of spectrum sensing data falsification (SSDF) attacks to test the resiliency of the proposed data aggregation scheme: “always yes,” “always no,” “always false,” and “always random.” Under the always yes attack scenario, the malicious SUs always report the presence of PUs ignoring their real sensing results. Under the always no attack scenario, the malicious SUs always report the absence of PUs on the channel ignoring the real detection results. Under the always false attack scenario, the malicious SUs always report the opposite of their sensed outcomes. Under the always random attack scenario, the malicious SUs randomly generate a sensing result to report to the system. We test the resiliency of our trust-based data fusion scheme against these four different attacks. We use simulation to demonstrate that our proposed scheme outperforms a traditional approach using a majority fusion rule under all attacking scenarios despite increasing malicious node population. Also, the malicious nodes can be identified through reputation scores in our scheme.

The research presented in this paper can be situated within the broader class of opportunistic channel selection strategy design in cognitive radio networks. The primary contributions of this paper are as follows:

1) We design a trust-based scheme for cooperative spectrum sensing to enhance the detection accuracy of PU channel occupancy.
2) We develop a static game based on mechanism design to discourage malicious SUs from reporting fake sensing capability.
3) We analyze the impact of SUs’ sensing capability regarding channel occupancy on their ability to dynamically exploit the band.
4) Our design principles apply to both centralized cognitive radio networks in which a Data Fusion Center (DFC) exists and distributed cognitive radio networks in which the DFC does not exist. We identify the best parameter settings under which the performance of our proposed scheme is optimized and outperforms existing schemes in either case. We also discuss the amenability of our design principles as applying to centralized and distributed cognitive radio networks.

This paper has been substantially extended from our previous work [8] [12] as follows: (a) we provide a uniform treatment of applying *mechanism design theory* principles to both centralized cooperative spectrum sensing in which a DFC exists and distributed cooperative spectrum sensing in which a DFC does not exist in cognitive radio systems; (b) we formal prove the correctness of our trust-based data fusion schemes for both centralized and distributed trust-based cooperative spectrum sensing by theorem proving and ascertain the validity by extensive simulation; and (c) we compare centralized and distributed trust-based cooperative spectrum sensing in terms of prediction accuracy and overhead, as well as the amenability of our mechanism design theory principles as applying to centralized and distributed trust-based cooperative spectrum sensing.

The rest of the paper is organized as follows: In Section II, we discuss related work in cooperative spectrum sensing in both centralized and distributed cognitive radio networks. In Section III, we discuss the system model and notation used. In Section IV, we develop, analyze, and validate a trust-based data fusion scheme based on *mechanism design theory* for centralized cooperative spectrum sensing. In Section V we extend our trust-based design methodology and analysis results to distributed cognitive radio networks in which the DFC does not exist so the data fusion decision is made autonomously by every SU in the system. In Section VI we compare centralized and distributed trust-based cooperative spectrum sensing in terms of prediction accuracy and overhead, as well as their amenability to our design principles. We summarize our conclusions and outline directions for future work in Section VII.

II. Related work

Cooperative spectrum sensing is a promising technique to increase PU activity sensing accuracy in cognitive radio networks by aggregating sensing reports from different SUs. In cooperative spectrum sensing, malicious SUs may report false sensing data to degrade the final aggregated sensing outcome.

Cooperative spectrum sensing can be conducted in both centralized and distributed cognitive radio networks. In a centralized system, SUs report sensing outcomes to a DFC and receive instructions from the DFC. In a distributed system, SUs do not rely on a DFC for channel access decision making but autonomously decide the channel availability by aggregating outcomes reported by other SUs. Cooperative spectrum sensing is confronted by SSDF attacks by which malicious SUs intentionally report fake sensing results to mislead decision making. Most existing anti-attack fusion rules in cooperative spectrum sensing are for the centralized infrastructure [6]–[8]. To date, there are only a handful of works on fusion rule design against SSDF attacks in distributed cognitive radio networks [9]–[12]. Our work considers cooperative spectrum sensing in both centralized and distributed cognitive radio networks.

In [13], the sensing information of SUs is weighted to maximize the detection probability of available channels under the constraint of a required false alarm probability. However, the scheme only considers sensing errors from the SUs without considering the malicious behavior of SUs. [14] proposes a modified combinatorial optimization identification (COI) algorithm to defend against malicious attacks. [15] proposes an HMM-based malicious SU detection algorithm to simultaneously estimate two HMMs without requiring separated training sequences. [16] provides an algorithm based on the non-parametric Kruskal-Wallis test to detect malicious users without having a priori knowledge. [17] proposes a decentralized scheme utilizing spatial correlation of received signal strengths and aggregating decisions based on a neighborhood majority voting approach for the secondary users to decide malicious users. However, a common problem related to the works cited above [13]–[17] is that they cannot distinguish a fake sensing outcome reported by a malicious SU from an erroneous outcome reported by a good SU with poor sensing capability.

There is limited work on the use of trust to enhance cooperative spectrum sensing. [18] discusses a simple trust-weighted cooperative spectrum sensing scheme with trust factors of SUs being used as weights for DFC decision making. However, it relies on the DFC to detect result discrepancy between the SU reports and the ground truth to update trust factors of SUs, which incurs a prohibitively high cost. Our static game model minimizes the DFC detection cost by probabilistic periodic check. [6] discusses an innovative idea of decoupling the detection ability of each SU from the reported detection result. According to their model, each SU reports a binary detection result, i.e., whether the targeted frequency is used by PUs or not, together with its detection sensing power to the DFC. The DFC considers the detection ability and trust it has toward each SU, and applies a threshold below which the SU's reported result is filtered out. Therefore, the DFC's final decision is based on trusted SUs' reported outcomes only. However, their scheme may fail when there is a high percentage of malicious SUs in the system. In particular, as each SU reports its own sensing capability, a malicious SU may intentionally report a higher sensing capability to get a higher impact on the final aggregated outcome. Moreover, if malicious SUs collude to report fake sensing capabilities, with a high percentage of malicious SUs within the system, their reported results will finally dominate the DFC's decision making, which will lead to a repeated wrong aggregated decision of the system. Our proposed scheme, on the other hand, can deal with the fake report problem despite the presence of a high percentage of malicious SUs. More specifically, our designed scheme allows the DFC to optionally sense the spectrum, then use the result to assess trustworthiness of SUs, and finally aggregate sensing outcomes from trusted SUs.

III. Assumptions and Notation

A. Threat Model

In this section, we discuss the type of malicious attacks considered in this paper. We assume that attackers are self-interested, i.e., they disrupt cooperative spectrum sensing for own benefits so they can keep the spectrum for their own use. Further, we assume that malicious nodes know each other and can collude to maximize their chance of success. For example, they can all report that they sense the presence of the PUs but in fact the channel is free. They can collude to report fake but high sensing capabilities so that their reported results can dominate decision making. They can also collude to frame a good SU to put it in the blacklist so as to prevent it from accessing spectrum resources. In cooperative spectrum sensing, the main attack is SSDF by which a malicious SU attacks by sending a false sensing report to the DFC. The SSDF attack can be further categorized into four types:

1) "Always yes" attack: malicious SUs always report the PU being active on the channels.
2) "Always no" attack: malicious SUs always report the channels being idle from PUs.

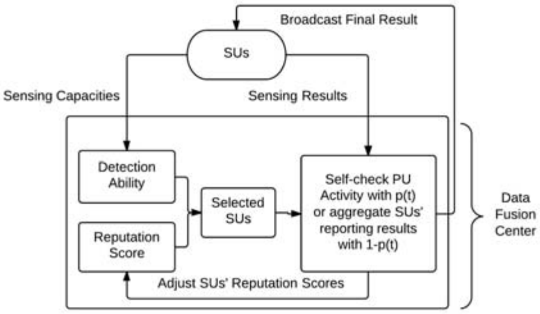


Figure 1: DFC Architecture for Centralized Cooperative Sensing.

3) "Always false" attack: malicious SUs always report the opposite of their sensed channel occupancy.
4) "Always random" attack: malicious SUs report true/false channel occupancy randomly.

B. Cooperative Spectrum Sensing Architecture

In this section, we discuss the system model for centralized cooperative spectrum sensing in cognitive radio networks. Later in Section V we extend it to distributed cooperative spectrum sensing. In a centralized system, SUs report sensing outcomes to the DFC and receive instructions from the DFC. The DFC architecture for centralized cooperative spectrum sensing is shown in Figure 1. The notation used for centralized cooperative spectrum sensing is summarized in Table I. We focus on the design principle and propose a general and flexible utility function design that applies to many scenarios. For example, the cost and utility functions considered in this paper (C, G and L in Table I) can be related to power, money, and/or risk in real scenarios.

We consider a cognitive radio network with N SUs and one DFC adopting the cooperative spectrum sensing technique to learn the PU activities on the channel. Time is slotted in fixed interval length. At the end of each time slot t, SU i reports its sensing result O_i^t (reporting if the PU is using the channel) together with its sensing capability $C_{i,real}^t$ to the DFC. More specifically, O_i^t is a binary value with $O_i^t = 1$ indicating SU i sensed PU existence in time slot t and $O_i^t = 0$ indicating SU i sensed no PU activity in time slot t. During the sensing process, SU i also knows its signal and noise level in time slot t, which can be translated into a continuous value $C_{i,real}^t \in [0, 1]$,

Table I: Notation.

Symbol	Definition
N	the number of SUs.
O_i^t	The sensing result reported by SU i in time slot t. $i \in \{1, 2, ..N\}$ and $O_i^t \in \{0, 1\}$.
O_{DFC}^t	The accumulated outcome of the DFC in time slot t. $O_{DFC}^t \in \{0, 1\}$.
O_{true}^t	The true PU activity outcome sensed by the DFC in time slot t. $O_{true}^t \in \{0, 1\}$.
T_{DFC}^t	The minimum capability threshold used by the DFC in time slot t to filter out sensing reports.
$C_{i,real}^t$	The real sensing capability of SU i in time slot t. $i \in \{1, 2, ..N\}$ and $C_{i,real}^t \in [0, 1]$.
$C_{i,report}^t$	The reported sensing capability of SU i in time slot t. $i \in \{1, 2, ...N\}$, $C_{i,report}^t \in [0, 1]$ and $C_{i,real}^t \leq C_{i,report}^t$
R_i^t	The reputation score of SU i in time slot $t.i \in \{1, 2, ..., N\}$ and $R_i^t \in Z$.
p^t	The probability for the DFC to check the PU activity on the spectrum in time slot t. $p^t \in [0, 1]$.
C	The cost of the DFC for sensing the spectrum in each time slot. $C > 0$.
G	Aggregation function adopted by DFC in each time slot.
L	Punishment function adopted by DFC in each time slot
T_i^t	The minimum capability threshold used by SU i to decide if it can trust its own sensing capability in time slot t.
F_i^t	The minimum trust threshold used by SU i to decide if it can trust another SU's sensing report in time slot t.

reflecting its true sensing capability. This value indicates SU i's certainty for its reported sensing result, i.e., the closer $C_{i,real}^t$ is to 1, the more certain SU i is about its reported sensing outcome in time slot t. Therefore, the closer $C_{i,real}^t$ is to 1, the larger $i's$ reported result should be weighted in the final aggregated outcome. However, a malicious SU i may take advantage of this scheme by intentionally reporting a fake and high sensing capability $C_{i,report}^t$ at time slot t to impact more on the final aggregated outcome. Apparently, a malicious SU does not want to report a low sensing capability because its sensing report will likely be filtered out by the DFC, and would not be able to affect the final accumulated sensing result. Therefore, we assume $C_{i,report}^t \geq C_{i,real}^t$.

After gathering the reported information from SUs, the DFC applies data fusion rules for decision making. The data fusion rules can be categorized into hard decision and soft decision. Under hard decision rules, the DFC applies decision-based rules to combine the results from SUs. Three simple decision-based rules are "or," "and," and "majority" rules. Under soft decision rules, the DFC often makes decisions based on the reported energy from each SU. The soft decision rules usually have a higher communication overhead and require a complicated aggregation algorithm compared to hard decision rules. Therefore, we adopt hard decision rules in our fusion rule design. Let G denote the aggregation function adopted by DFC to generate the final outcome at time t, denoted by $O_{DFC}^t \in \{0, 1\}$.

Besides passively receiving reported results from SUs, the DFC can actively sense the spectrum so as to check if an SU lies about the sensing outcome. More specifically, the DFC checks the PU activity on the channel with probability p^t, with $p^t \in [0, 1]$, in each time slot t. After checking the spectrum availability, the DFC uses its “first-hand” evidence, denoted by O_{true}^t, to punish the SUs whose reported results are different from O_{true}^t based on the punishment function L. Note that malicious nodes will not know whether the DFC will check the PU activity on the channel in a particular slot because the DFC checks the PU activity probabilistically with probability p^t. The punishment to SU i in time slot t is in the form of decreasing SU i’s reputation score R_i^t with $R_i^t \in Z$ for $i \in \{1, 2, ..., N\}$. We assume all SUs have the same initial reputation scores assigned by the DFC, i.e., $R_i^0 = R_j^0$ for $i \neq j$ and $i, j \in \{1, 2, ..., N\}$. Also, the punishment level to i is related to its reported sensing capability $C_{i,report}^t$ because SU $i's$ erroneous sensing outcome is maybe due to its poor sensing capability. The reason that the DFC does not check the spectrum availability in every time slot is that the DFC spectrum sensing is often at a high cost of equipment, technology and energy. Moreover, when the coverage of spectrum is larger than the sensing range of the DFC (multiple channels), the DFC may not be able to check PU activities on all channels at the same time. Therefore, in our scheme, we represent this sensing cost by C which is a fixed value irrelevant of time. If in a particular time slot t the DFC checks the PU existence, it applies the punishment function L to decrease the reputation scores of those SUs with a different sensing result from O_{true}^t;

if the DFC does not check the spectrum availability in t, it applies the aggregation function G to aggregate the opinions from SUs and generate the final outcome O^t_{DFC}. Our trust-based data fusion scheme derived from mechanism design thus has two design objectives:

- Design G and L to force malicious SUs to report the real sensing ability, i.e. $C^t_{i,report} = C^t_{i,real}$ for $i \in \{1, ..., N\}$, $t > 0$.
- Design a scheme to allow the DFC to perform minimum checking with the smallest probability p^t.

IV. Mechanism Design for Centralized Cooperative Sensing and Its Analysis

In this section, we formulate a static game based on mechanism design to model decision making between the DFC and malicious SUs for centralized cooperative spectrum sensing and then present a theoretical analysis to formally prove the correctness.

A. *Trust-Based Data Fusion Rule Design*

We use a static game to model the relationship between the DFC and a malicious SU i in each single time slot. From a malicious SU i's perspective, in time slot t, it has two options on reporting its sensing capability: honestly reporting its sensing capability $C^t_{i,real}$ or intentionally reporting a higher fake sensing capability $C^t_{i,report} > C^t_{i,real}$. On the other hand from the DFC's perspective, in time slot t, it decides to sense the PU activity with probability p^t or not to with probability $1 - p^t$. The payoff matrix for the DFC and a malicious SU i in the game model is shown in Table II. The table entry is in the format of (DFC payoff, malicious SU i payoff). For example, if the DFC checks the PU activity while SU i dishonestly reports a higher fake sensing capability $C^t_{i,report} > C^t_{i,real}$, the payoff to the DFC is $L(C^t_{i,report}, R^t_i) - C$ and the payoff to malicious SU i is $-L(C^t_{i,report}, R^t_i)$.

We explain the payoff matrix in Table II. According to the described static game model of our scheme, in slot t each malicious SU i reports both its sensing result O^t_i and its fake sensing capability $C^t_{i,report}$ to the DFC, who aggregates the reported results to the final outcome based on the aggregation function G. Therefore, the "impact" of i's reported result to the DFC's aggregation

Table II: The Payoff Matrix for the DFC and a Malicious SU.

	SU i reports $C^t_{i,real}$	SU i reports $C^t_{i,report}$
DFC checks	$L(C^t_{i,real}, R^t_i) - C,\ -L(C^t_{i,real}, R^t_i)$	$L(C^t_{i,report}, R^t_i) - C,\ -L(C^t_{i,report}, R^t_i)$
DFC does not check	$-G(C^t_{i,real}, O^t_i, R^t_i), G(C^t_{i,real}, O^t_i, R^t_i)$	$-G(C^t_{i,report}, O^t_i, R^t_i), G(C^t_{i,report}, O^t_i, R^t_i)$

result O^t_{DFC} can be denoted as $G(C^t_{i,report}, O^t_i, R^t_i)$, which can be viewed as the gain to the malicious SU if not caught by DFC. On the other hand in time slot t, the DFC decides to check the spectrum with probability p^t at a fixed cost C. If the DFC decides to check the spectrum, it can detect the true occupancy and then punish the nodes who reported a different outcome from the detected channel occupancy signal. Denoted by $L(C^t_{i,reported}, R^t_i)$ the loss to SU i being punished due to the reported result being different from that by the DFC. A malicious SU i who reports $C^t_{i,report}$ higher than $C^t_{i,real}$ will get a punishment $L(C^t_{i,report}, R^t_i)$. Therefore, the payoff matrix between a malicious SU and the DFC can be defined as in Table II.

In this game, both the malicious SUs and the DFC want to maximize their own utility functions. In particular, the malicious SUs aim at manipulating the DFC's aggregated outcome by reporting higher sensing capabilities. Meanwhile, the DFC aims to minimize the checking probability p^t and leave malicious SUs no motivation to report fake sensing capabilities.

B. Analysis

In this section, we conduct a theoretical analysis to prove the correctness of our data fusion design.

Theorem 1. *To discourage malicious SU i from reporting a higher sensing capability than its actual sensing capability, i.e., $C^t_{i,report} > C^t_{i,real}$, the DFC's checking probability p^t, aggregation function G and punishment function L should satisfy: $p^t[L(C^t_{i,report}) - L(C^t_{i,real})] \geq (1 - p^t)[G(C^t_{i,report}) - G(C^t_{i,real})]$.*

Proof: According to the described static game model and the payoff matrix shown in Table

II, a malicious SU i's payoff of reporting $C^t_{i,real}$, i.e., $u_i(C^t_{i,real}, R^t_i)$, can be expressed as:

$$\begin{aligned} u_i(C^t_{i,real}, R^t_i) = & - p^t L(C^t_{i,real}, R^t_i) \\ & + (1 - p^t)G(C^t_{i,real}, O^t_i, R^t_i) \end{aligned} \tag{1}$$

On the other hand if SU i reports a higher fake sensing capability $C^t_{i,report}$, the payoff to SU, i.e., i $u_i(C^t_{i,report}, R^t_i)$, is:

$$\begin{aligned} u_i(C^t_{i,report}, R^t_i) = & - p^t L(C^t_{i,report}, R^t_i) \\ & + (1 - p^t)G(C^t_{i,report}, O^t_i, R^t_i) \end{aligned} \tag{2}$$

To guarantee that SU i has no incentive to report a higher fake sensing capability, we need $u_i(C^t_{i,real}, R^t_i) \geq u_i(C^t_{i,report}, R^t_i)$. From Equations 1 and 2, we have:

$$\begin{aligned} & p^t(L(C^t_{i,report}, R^t_i) - L(C^t_{i,real}, R^t_i)) \\ & \geq (1 - p^t)(G(C^t_{i,report}, O^t_i, R^t_i) - G(C^t_{i,real}, O^t_i, R^t_i)) \end{aligned} \tag{3}$$

■

Theorem 1 provides a general rule for the design of the aggregation function G and the punishment function L. Let $\triangle G = G(C^t_{i,report}, O^t_i, R^t_i) - G(C^t_{i,real}, O^t_i, R^t_i)$ and $\triangle L = L(C^t_{i,report}, R^t_i) - L(C^t_{i,real}, R^t_i)$ denote the gain and loss of malicious SU i, respectively. Then, we can rewrite Equation 3 as $p^t \triangle L \geq (1 - p^t)\triangle G$. That is, as long as the DFC checks the spectrum with probability no less than $\frac{\triangle G}{\triangle G + \triangle L}$, a malicious SU has no motivation to report a fake sensing capability.

Next, we analyze the checking probability p^t from the DFC's utility perspective. In time slot t, if DFC checks the spectrum availability, it observes the true PU activity result O^t_{true}. Let N_m denote the set of malicious SUs within the system.

The DFC's payoff for checking the spectrum in a particular time slot can be expressed as $\sum\limits_{i \in N_m} L(C^t_{i,report}, R^t_i) - C$; the DFC's payoff for not checking the spectrum is $-\sum\limits_{i \in N_m} G(C^t_{i,report}, O^t_i, R^t_i)$.

Therefore, the DFC's utility function under checking probability p^t is:

$$\begin{aligned} u_{DFC} =& p^t(\sum_{i\in N_m} L(C^t_{i,report}, R^t_i) - C) \\ & - (1-p^t)\sum_{i\in N_m} G(C^t_{i,report}, O^t_i, R^t_i) \end{aligned} \tag{4}$$

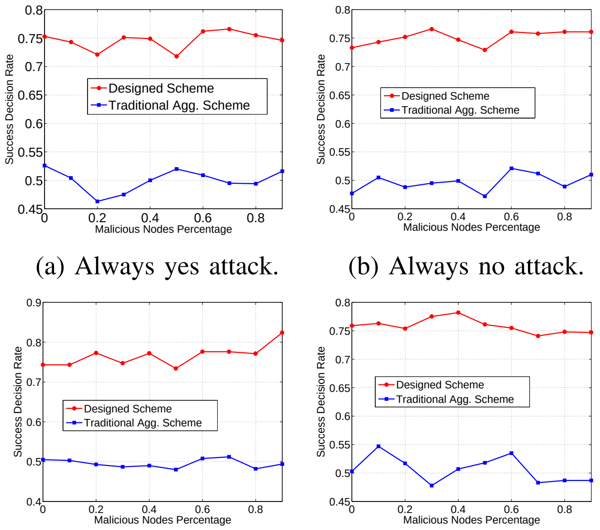
Here we note that the cost to the DFC is equivalent to the negative payoff to the DFC as shown in Equation 4. Consequently, maximizing the DFC's payoff is the same as minimizing the DFC's cost. By taking the derivative of Equation 4 with respect to p^t, we get:

$$\begin{aligned} \frac{\partial u_{DFC}}{\partial p^t} =& \sum_{i\in N_m} L(C^t_{i,report}, R^t_i) \\ & + \sum_{i\in N_m} G(C^t_{i,report}, O^t_i, R^t_i) - C \end{aligned} \tag{5}$$

Equation 5 indicates that the optimized checking probability p^t depends on the evaluation result of $\sum\limits_{i\in N_m} L(C^t_{i,report}, R^t_i) + \sum\limits_{i\in N_m} G(C^t_{i,report}, O^t_i, R^t_i) - C$. In particular, if $\sum\limits_{i\in N_m} L(C^t_{i,report}, R^t_i) + \sum\limits_{i\in N_m} G(C^t_{i,report}, O^t_i, R^t_i) - C > 0$, the optimal p^t value is 1. That is, the DFC should check the spectrum in every time slot to maximize its payoff. On the other side, if $\sum\limits_{i\in N_m} L(C^t_{i,report}, R^t_i) + \sum\limits_{i\in N_m} G(C^t_{i,report}, O^t_i, R^t_i) - C < 0$, the optimal p^t value is 0. Under this scenario, the DFC should not check spectrum to maximize its payoff. However, this analysis needs to be combined with the result generated in Theorem 1, which requires the check probability p^t to be at least $max_i \frac{\triangle G_i}{\triangle G_i + \triangle L_i}$.

C. Simulation Validation

In this section, we validate our trust-based data aggregation design for centralized cooperative spectrum sensing by extensive simulation. The performance of our trust-based data aggregation scheme will be compared with a traditional aggregation scheme where the DFC accumulates the reported results from all SUs and makes the final decision based on majority voting. In contrast, our scheme is designed based on Theorem 1 to discourage malicious nodes from reporting fake sensing capabilities. Also with probability $1 - p^t$, the DFC accumulates the reported results from



(a) Always yes attack.

(b) Always no attack.

(c) Always false attack.

(d) Always random attack.

Figure 2: Comparison of Success Decision Rate between Our Scheme and a Traditional Majority Voting Aggregation Scheme with varying Malicious Node Percentage.

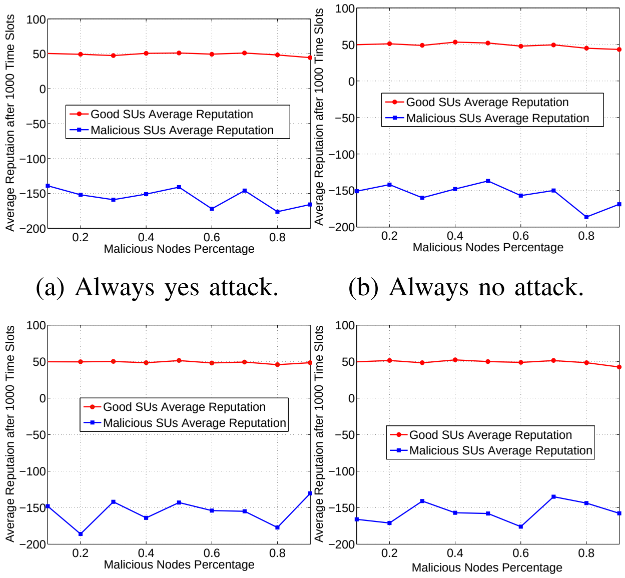
all SUs based on trust-weighted majority voting.

We consider a cognitive radio system consisting 100 SUs and one DFC. We run a simulation experiment with 1000 repeated time slots. The ground truth of the channel occupancy is randomly simulated (either 0 or 1) in these 1000 runs. In each time slot, all SUs report the detected outcome (either 0 or 1) together with their sensing capabilities (within $[0, 1]$) to the DFC. We assume that malicious SUs report the highest sensing capability ($C^t_{i,report}$=1) to maximize its impact, while good SUs report its true sensing capability ($C^t_{i,real}$ following uniform distribution U[0, 1]) to the DFC. Also we assume that malicious SUs report the channel occupancy based on their attack strategies as described in the threat model, while good SUs report the channel occupancy they sense.

The data aggregation function G used by the DFC to aggregate SU sensing reports is based on trust-weighted majority voting. In particular, the DFC first filters out SUs whose reported sensing capability is below the DFC's minimum sensing capability threshold ($T^t_{DFC} = 0.8$). The DFC then categorizes the SUs into two groups S_0 and S_1: S_0 contains the SUs who reported no PU activity on the channel and S_1 contains the SUs who reported PU existence on the channel. Finally, the DFC decides the aggregated outcome as 0 if $\sum_{i\in S_0} R_i^t > \sum_{i\in S_1} R_i^t$, and as 1 otherwise. If the DFC does not check the spectrum, it applies the aggregation function G based on trust-based majority voting discussed above. If the DFC checks the spectrum in a particular time slot t, it senses true channel occupancy O_{true}^t, and uses it to punish the SUs who reported a different outcome. Specifically, the punishment function L on a SU's reputation is $R_i^t = R_i^t - C_{i,reported}^t$ where R_i^t is the reputation of the SU at time t and $C_{i,reported}^t$ is the sensing capability reported by the SU at time t. The initial reputation score for each SU is 0 representing ignorance, i.e., $R_i^0 = 0$ for $i \in \{1, ..., N\}$, with the range of the reputation score being [-200, 200]. We note that with the G and L functions defined above, the DFC will check the spectrum in each time slot with probability $p^t = \frac{1}{2}$ based on Theorem 1 with the design of the aggregation function G and the punishment function L satisfying $\triangle G = \triangle L$. Notice here, we assume the cost for the DFC to check the channel is relatively large, i.e., $\sum\limits_{i\in N_m} L(C_{i,real}^t, R_i^t) + \sum\limits_{i\in N_m} G(C_{i,real}^t, O_i^t, R_i^t) - C < 0$. Therefore, from the DFC's perspective, it is always reluctant to check the channel availability by itself. However, to guarantee the malicious SUs do not have the incentive to fake sensing capabilities, the DFC still needs to sense the PU occupancy with the minimum checking probability given by Theorem 1.

1) Success Decision Rate: We analyze the success rate of the DFC's decision with respect to the percentage of malicious SUs under the four different malicious attacks in our threat model. We vary the percentage of malicious SUs from 0% to 90% and calculate the success decision rate. We also output the success decision rate of the traditional data aggregation scheme as a comparison. The result is shown in Figure 2. Specifically, we analyze the performance under four types of malicious attacks: the "always yes" attack, shown in Figure 2a, where the malicious SUs always report the existence of PU activity; the "always no" attack, shown in Figure 2b, where the malicious SUs always report the absence of PU activity; the "always false" attack, shown in Figure 2c, where the malicious SUs always report the opposite of the sensed PU activity; the "always random" attack, shown in Figure 2d, where the malicious SUs randomly report a binary result as the PU activity.

From Figure 2 we see that our trust-based data aggregation scheme derived from the static game always performs better than the traditional approach. It especially outperforms the traditional



(a) Always yes attack. (b) Always no attack.

(c) Always false attack. (d) Always random attack.

Figure 3: Comparison of Good Node and Malicious Node Reputation Scores.

approach under the "always false" attack behavior because the attackers will be caught whenever the DFC decides to check the channel availability with probability p^t. We conclude two observations. First, the performance of our designed scheme (around 75%) is significantly better than that of the traditional aggregation scheme (around 50%) over all four types of malicious SU attacks. We attribute the superiority of our trust-based majority voting scheme over the traditional majority voting scheme to its ability to accurately track the trust status of nodes in the system so that malicious nodes are filtered out or discounted during trust-based majority voting. On the other hand, the traditional majority voting scheme simply counts the number of 0's and 1's to determine the sensing outcome without any effective mechanism being applied to filter out or discount false sensing outcomes from malicious nodes. Secondly, under all malicious attack scenarios, the performance of our designed scheme is stable over a wide range of malicious node percentage.

2) Reputation Scores: We compare the average reputation scores of the normal and malicious nodes under each attack scenario for our scheme. The results are shown in Figure 3 which demonstrates that our trust scheme can effectively distinguish malicious SUs by reputation scores. In

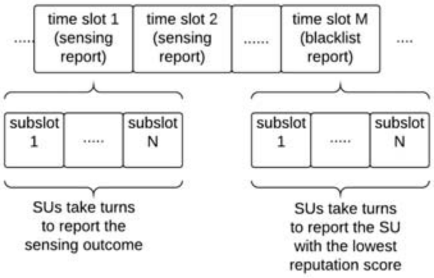

time slot 1 (sensing report)
time slot 2 (sensing report)
......
time slot M (blacklist report)
....
subslot 1
.....
subslot N
subslot 1
.....
subslot N
SUs take turns to report the sensing outcome
SUs take turns to report the SU with the lowest reputation score

Figure 4: Time Schedule on the Common Control Channel.

particular, after 1000 time slots, the average reputation score of the normal nodes is around 50 while that of the malicious nodes is around -150. It means that when our trust-based fusion scheme is in place the malicious node reputation scores will drop dramatically with respect to time. Finally we also note that this reputation gap is stable with respect to the malicious node percentage.

V. Extension to Distributed Cooperative Spectrum Sensing

In this section, we extend our design to the case in which the DFC does not exist. In distributed cooperative spectrum sensing, SUs do not rely on a DFC for channel access decision making but autonomously decide the channel availability by aggregating outcomes reported by other SUs.

We first extend the system model discussed in Section III for distributed cooperative spectrum sensing as follows: We consider a distributed cognitive radio network with N SUs adopting the cooperative spectrum sensing technique to learn the PU's activity on one channel. We assume that all SUs are aware of the existence of each other and are within the communication range. In particular, each SU has a unique identity $i \in \{1, ..., N\}$ which is publicly known by other SUs. To control the overhead communication messages, SUs are not allowed to communicate directly with each other. Instead, the SUs share their sensing outcomes on a common control channel (CCC) in a broadcast manner. Using a CCC for control message exchanges by cognitive radio nodes is indispensable in distributed cognitive radio systems [20]. However, establishing an always-on and reliable CCC in distributed cooperative spectrum sensing is still an open issue. Reserving a portion of the spectrum for establishing a CCC is a possible way but it wastes spectrum resources. Some research efforts [21], [22] attempted not to use a spectrum portion to establish a CCC.

Figure 4 illustrates a time schedule on the common control channel for distributed cooperative spectrum sensing. To avoid communication interference on the CCC, time is slotted and each time slot is further divided into N subslots, one for each SU. An SU with identity i will only broadcast its report in the ith designated subslot while listening to other SUs' reports in other subslots. Since each SU has a unique identity and it can broadcast only in its designated subslot, there is no identity attack possibility. There are altogether M time slots in a reporting cycle. The first $M-1$ slots (each called a sensing report slot) are used for reporting sensing outcomes and sensing capabilities, while the last time slot (called a blacklist report slot) is used for reporting malicious nodes for the purpose of building a blacklist. M is a system parameter and should be sufficiently large to allow each individual SU to assess trust scores of other SUs and report malicious SUs in the blacklist report slot. Here we note that for distributed cooperative spectrum sensing, the processing time is proportional to the number of SUs.

In a sensing report slot, SUs take turns to broadcast their sensing outcomes in their respective subslots. Specifically, SU i broadcasts its outcome O_i^t together with its capability $C_{i,real}^t$ on the CCC. SU i's outcome O_i^t is a binary variable with $O_i^t = 1$ indicating that SU i sensed PU existence and $O_i^t = 0$ indicating that SU i sensed no PU activity. i's sensing capability denoted by $C_{i,real}^t$ is a continuous value $\in [0, 1]$ representing the probability of SU i being able to correctly sense the channel occupancy status. Hence, the closer $C_{i,real}^t$ is to 1, the more confident SU i is about its reported sensing outcome at time t. An SU's sensing capability is characterized by the probabilities of false alarm and missed detection. We follow [9] for estimating the missed detection rate P_{md} and the false alarm rate P_{fa}, as follows:

$$P_{md} = P(p_i^t < \gamma | H_1) \tag{6}$$

$$P_{fa} = P(p_i^t > \gamma | H_0) \tag{7}$$

where H_1 and H_0 denote the hypotheses corresponding to the presence and the absence of PU, respectively, and p_i^t represents the received signal power by SU i at time t which can be estimated by an energy detection sensing method [19]. The sensing capability $C_{i,real}^t$ of SU i can be estimated based on its false alarm rate P_{fa} and missed detection rate P_{md}. Specifically, when SU i senses the existence of PU, its sensing capability can be calculated as $C_{i,real}^t = 1 - P_{md}$. When SU i senses the absence of PU, its sensing capability can be calculated as $C_{i,real}^t = 1 - P_{fa}$.

Let $C_{i,report}^t$ be the sensing capability reported by SU i at time t. A good SU i reports its real sensing capability, i.e., $C_{i,report}^t = C_{i,real}^t$, while a malicious SU i intentionally reports a higher sensing capability $C_{i,report}^t \geq C_{i,real}^t$ to impact other SUs' decision making.

In a blacklist report time slot, each SU reports an SU with the lowest trust score among all SUs it keeps in its database. SU i then updates its blacklist binary vector b_j^i for $j \in \{1, 2, ..., N\}$ based on the blacklist reports gathered in the blacklist report slot. We assume that a malicious node can perform *bad-mouthing* attacks to frame a good node as a bad node.

The goal of our distributed cooperative spectrum sensing design is for SU i to effectively aggregate self and received sensing information, i.e., (O_i^t,$C_{i,report}^t$) for $i \in \{1, ..., N\}$ such that it can achieve high accuracy in sensing PU occupancy and detect malicious SUs in the long run.

A. Trust-Based Data Fusion Rule Design

In this section, we describe our trust-based data fusion rule design consisting of a data fusion process and a blacklist generation process. An SU makes channel availability decisions using sensing reports gathered in a sensing report slot. An SU updates its blacklist in the blacklist generation process using blacklist reports gathered in a blacklist report time slot.

1) Data Fusion Process: In a sensing report slot, SU i makes a channel occupancy decision based on its own and received sensing outcomes from other SUs. Due to a lack of ground truth in decision making, SU i first decides whether to trust its own sensing outcome by comparing its own sensing outcome $C_{i,real}^t$ with a minimum sensing capability threshold T_i^t. There are two cases:

1) If $C_{i,real}^t > T_i^t$, SU i has high confidence about its own sensing outcome and will simply adopt its sensing outcome as the final decision, i.e., $O_{i,final}^t = O_i^t$. Meanwhile, i adjusts other SUs'

trust scores by comparing the received sensing outcomes with its own sensing outcome. SU i updates the trust score of SU j, denoted by $R^t_{i,j}$, only if SU $j's$ reported sensing capability is over SU i's minimum trust threshold, i.e., $C^t_{j,report} > F^t_i$, where F^t_i is SU i's minimum trust threshold in the range of [0, 1]. If j's reported outcome matches i's outcome, SU i increases SU j's trust score by:

$$R^t_{i,j} = R^t_{i,j}(1 + C^t_{j,report}) \tag{8}$$

If j's reported outcome O^t_j does not match i's sensing outcome, SU i decreases SU j's trust score by:

$$R^t_{i,j} = R^t_{i,j}(1 - (C^t_{j,report})^2) \tag{9}$$

2) If $C^t_{i,real} \leq T^t_i$, SU i does not have confidence in its own sensing outcome and will rely on the received sensing information reported by other SUs to decide the final outcome. In this case, given that SU i does not have any basis to judge the trustworthiness of sensing results reported by other SUs, SU i does not update the trust scores of other SUs. SU i first filters out sensing reports from receivers that do not pass SU i's minimum trust threshold, i.e., $C^t_{j,report} < F^t_i$, or on the blacklist, i.e., $b^i_j = 1$ for $j \in \{1, 2, ..., N\}$. After the minimum capability step, the remaining reports are separated into two groups based on if the sensing outcome is 0 or 1. For each group, a group trust score is calculated by a trust sum. SU i then chooses the group with a higher group trust score, and adopts the sensing outcome of the group (either 0 or 1) as the final outcome. If the group scores are of the same value, SU i randomly decides the channel occupancy status for that time slot.

2) Blacklist Generation Process: In a blacklist report slot t, SU i reports the identity of the lowest trust node $B^t_i \in \{1, ..., N\}$. If more than one node are of the same lowest trust score, SU i randomly chooses one to broadcast. Meanwhile, SU i updates a blacklist binary vector $[b^i_1, ..., b^i_N]$ based on the received node identities broadcast by other SUs. Since a malicious SU may perform bad-mouthing attacks and intentionally report a good node as a malicious node, SU i considers B^t_j

(reported by SU j) as malicious only if i trusts j as well as i does not trust B_j^t itself. Specifically, SU i considers that B_j^t should be put on the blacklist if the following two conditions are met:

1) j's trust score is above the average trust score of all nodes maintained by i;
2) B_j^t's trust score is below the average trust score of all nodes maintained by i.

After i decides j as a malicious node, i sets b_j^i to 1 and will exclude j's reports in future decision making.

B. Analysis

In this section, we theoretically analyze our data fusion rule design. We denote by G and L the trust gain and loss, respectively. The theorem below provides the design of G and L to make sure that a good SU will be awarded with trust gain if it reports its true sensing capability and sensing outcome faithfully.

Theorem 2. *For a trust-based data fusion rule design to award SU j who reports its authentic sensing outcome and capability, the trust gain (G) and the trust loss (L) must satisfy:*

$$1 - C_{i,real}^t + C_{j,real}^t - 2C_{i,real}^t C_{j,real}^t \geq \frac{G}{G-L}$$

Proof: The reported sensing capability of node j, $C_{j,real}^t$, is the probability of j being able to sense PU existence status on the channel. According to our designed scheme (described in Section V-A) the conditions to award j's trust by SU i are: i trusts its own sensing outcome, i.e., $C_{i,real}^t > T_i^t$, j's sensing capability is above i's minimum trust threshold, i.e., $C_{j,real}^t > F_i^t$, and j's reported sensing outcome matches that of i, i.e., both i and j sense PU existence the same way either 0 or 1. Therefore, the probability for j being rewarded by SU i with sensing capability $C_{i,real}^t$, denoted by P_{award}, is given by:

$$\begin{aligned} P_{award} =& p(C_{i,real}^t > T_i^t) p(C_{j,real}^t > F_i^t) \\ & (C_{i,real}^t C_{j,real}^t + (1 - C_{i,real}^t)(1 - C_{j,real}^t)) \end{aligned} \tag{10}$$

On the other hand, the conditions to penalize j's trust by SU i are: i trusts its own sensing outcome, i.e., $C^t_{i,real} > T^t_i$, j's sensing capability is above i's minimum trust threshold, i.e., $C^t_{j,real} > F^t_i$, and j's reported sensing outcome conflicts with that of i, which happens if only one of the two SUs correctly senses PU existence. Therefore, the probability of j being punished by i with sensing capability $C^t_{i,real}$, denoted by P_{punish}, is given by:

$$\begin{aligned} P_{punish} = & p(C^t_{i,real} > T^t_i)p(C^t_{j,real} > F^t_i) \\ & (C^t_{i,real}(1 - C^t_{j,real}) + C^t_{j,real}(1 - C^t_{i,real})) \end{aligned} \tag{11}$$

To guarantee j's trust is not penalized when it reports its real sensing outcome and capability, we need:

$$GP_{award} \geq LP_{punish} \tag{12}$$

By plugging in Equation 10 and Equation 11 into Equation 12 we can obtain the expression shown in the theorem. ■

We denote by $\triangle G$ and $\triangle L$ the gap of trust gain and loss between reporting higher sensing capability and real sensing capability. The theorem below provides the design of $\triangle G$ and $\triangle L$ to prevent a malicious SU from gaining trust if it reports high sensing capability and performs SSDF attacks.

Theorem 3. *To prevent a malicious SU j from reporting higher sensing capability to gain trust increase to SU i, $\triangle G$ and $\triangle L$ must satisfy:*

$$C^t_{i,real}\triangle L \geq (1 - C^t_{i,real})\triangle G$$

Proof: For a malicious SU j who reports a fake sensing outcome and a higher capability s.t. $C^t_{j,report} > C^t_{j,real}$, the conditions for j to be caught and punished by SU i are: i trusts its own sensing outcome, i.e., $C^t_{i,real} > T^t_i$, j's sensing capability is above i's minimum trust threshold, i.e., $C^t_{j,report} > F^t_i$, and j's reported sensing outcome disagrees with that of i, which requires i's sensing outcome to be true. SU j uses a higher sensing capability than the minimum trust threshold, i.e., $C^t_{j,report} > F^t_i$, to mislead the data fusion process. Therefore, the probability of j being caught by i with sensing capability $C^t_{i,real}$, denoted by P_{caught}, is given by:

$$P_{caught} = p(C^t_{i,real} > T^t_i)C^t_{i,real} \tag{13}$$

Similarly, the probability of a malicious node j not being punished by SU i, denoted by P_{miss}, is given by:

$$P_{miss} = p(C^t_{i,real} > T^t_i)(1 - C^t_{i,real}) \tag{14}$$

To guarantee j's trust is decreased when it reports a fake sensing outcome and a higher calculated capability, we need:

$$\triangle L P_{caught} \geq \triangle G P_{miss} \tag{15}$$

By plugging in Equation 13 and Equation 14 into Equation 15 we prove the theorem. ■

Corollary 4. *Our trust-based data fusion rule design guarantees that a good SU's trust is increased when it reports authentic sensing outcome and capability, and a malicious SU's trust is decreased when it reports a false sensing outcome and a higher sensing capability.*

Proof: We prove this corollary by showing that our trust-based data fusion rule design satisfies the above two theorems. According to Equations 8 and 9, a good SU (j) who reports true sensing outcome O^t_{real} and capability $C^t_{j,real}$ and will get a trust increase of $G = R^{t-1}_{i,j} \times C^t_{j,real}$ and get a trust loss of $L = R^{t-1}_{i,j} \times (C^t_{j,real})^2$.

On the other hand, a malicious SU (j) who reports a fake sensing outcome and a higher capability $C^t_{j,report} > C^t_{j,real}$ will get an extra trust increase of $\triangle G = R^{t-1}_{i,j} \times (C^t_{j,report} - C^t_{j,real})$ and an extra loss of $\triangle L = R^{t-1}_{i,j} \times ((C^t_{j,report})^2 - (C^t_{j,real})^2)$.

We can easily confirm that G, L, $\triangle G$ and $\triangle L$ satisfy Theorems 2 and 3 and hence we prove the corollary. ■

Therefore, our trust-based data fusion rule design guarantees a trust gain to normal SUs and discourages malicious SUs from reporting false sensing outcomes and capabilities.

C. Simulation Validation

In this section, we conduct a performance analysis of our data fusion rule design for distributed cooperative spectrum sensing using Matlab and compare its performance with three baseline schemes: individual, majority voting, and capability-weighted (CW) majority voting. Under individual data fusion, an SU directly accepts its sensing outcome as the final outcome without considering reported information from other SUs. Therefore, it can be viewed as a non-cooperative scheme. Majority voting counts the number of 0's and 1's and takes the majority as the final outcome. Capability-weighted majority voting is the same as majority voting except that every count is weighted by the SU's reported capability. The performance metric is the individual success rate, or the probability of successfully detecting the actual status of the channel.

The simulation setup is based on $N = 20$ SUs. As in [9], we assume that the SU sensing capability follows the Gaussian distribution. That is, each SU's true sensing capability is modeled by a Gaussian distribution with mean $\mu = 0.6$ and variance $\sigma^2 = 0.2$. The reported sensing capability for a malicious SU is set to a high value at 0.95. The range of the node trust score is [-1, 3], with the initial trust score of 1 representing ignorance. In the experiment, we set T_i^t=0.7 and F_i^t=0.7. The report cycle M is set to a high value at 20 to allow each individual SU to assess trust scores of other SUs and report malicious SUs in the blacklist report slot. So in every 20 time slots, SUs update their individual blacklists based on blacklist reports from other SUs. Each experiment covers 200 time slots. The result is based on 1000 independent repeated experiments.

1) Effect of Malicious Node Population: We first investigate the effect of malicious node percentage on the individual success rate (i.e., the probability of successfully detecting the actual status of the channel). Figure 5 shows the individual success rate of our trust-based data fusion scheme (labeled by trust-based) against the three baseline schemes (labeled by individual, majority, and CW majority, respectively) in the presence of SSDF attacks. It is clear from Figure 5 that our

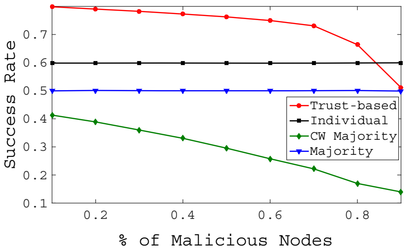


Figure 5: Impact of Malicious Node Population.

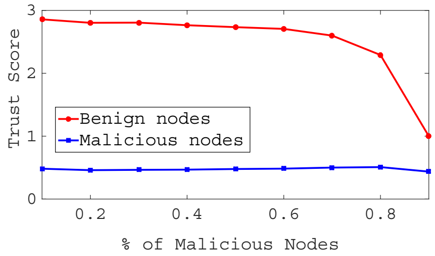


Figure 6: Comparison of Trust Scores.

data fusion rule design outperforms all baseline schemes. The gap between trust-based data fusion and individual data fusion can be viewed as the gain of adopting distributed cooperative spectrum sensing over non-cooperative spectrum sensing. The only exception happens when the percentage of malicious nodes is 90% in which case the number of good nodes is only 2 (10% of N=20) and only one of which has capability higher than the minimum capability threshold, so there is no chance for them to update the trust scores of each other. As a result, the trust score remains at 1 and the success rate remains at 0.5.

We observe that the success rate under individual data fusion stays at 0.6. The reason is that each SU's true sensing capability is modeled by a Gaussian distribution with mean $\mu = 0.6$ and variance $\sigma^2 = 0.2$. We also observe that individual data fusion scheme performs better than majority voting which in turn performs better than capability-weighted majority voting, especially as the percentage of malicious nodes increases. This is because malicious SUs report a higher capability which has an adverse effect on capability-weighted majority voting. Our trust-based data fusion scheme on the other hand takes both trust and capability into consideration and can achieve a much higher accuracy in data fusion.

2) Trust Scores of Benign and Malicious SUs: We compare the average trust scores of good and malicious nodes in our designed scheme. Figure 6 shows the average trust scores of good and malicious SUs at the end of the 50^{th} time slot as recorded by good SUs under SSDF attacks. The results support the claim that our trust-based data fusion scheme can effectively distinguish malicious SUs by their low trust scores. Figure 6 validates the theoretical analysis results that a good SU will be awarded with trust gain if it reports its true sensing capability and sensing outcome faithfully, while a malicious SU's trust will be penalized with trust loss if it falsely reports a high sensing capability and a false sensing outcome. Our trust-based data fusion scheme can efficiently distinguish good nodes from malicious nodes in the long run when the percentage of malicious nodes is below 80%.

3) Impact of Threshold Parameters: We analyze the impact of the minimum capability threshold T_i^t and the minimum trust threshold F_i^t on protocol performance. We consider 4 variants of SSDF attacks: always yes (always saying the channel is free), always no (always saying the channel is not free), always false (always saying the channel is free/not free opposite to what it senses), and always random (always saying the channel is free/not free randomly). Note that the analysis performed so far is for the case of "always false" SSDF attacks, which is the worst case among all.

Figure 7 shows the average individual success rate vs. T_i^t, with F_i^t=0.8 to isolate its effect. The figure is based on 20% malicious nodes. We observe that there exists an optimal T_i^t value under which the success rate is maximized. This is due to our data fusion rule design. Specifically, as the minimum capability threshold T_i^t increases, if SU i's true sensing capability is still above the increasing threshold, then its own sensing outcome is likely to be accurate, so the success decision rate will also increase. However, when T_i^t continues to increase, SU i's true sensing capability will more likely fall below the threshold. In this case, SU i cannot update the trust scores of other SUs effectively and must aggregate sensing outcomes from other SUs with inaccurate trust scores. As a result, the success decision rate decreases. This tradeoff results in the T_i^t optimal point.

Figure 8 shows the average individual success rate vs. F_i^t, with T_i^t=0.8. We observe that there exist an optimal F_i^t value under which the success rate is maximized. This is because as the

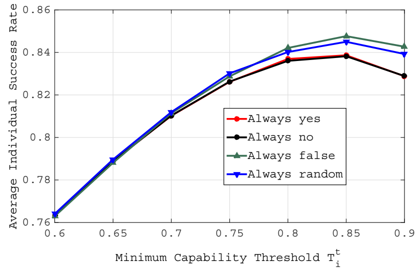


Figure 7: Impact of the Minimum Capability Threshold T_i^t.

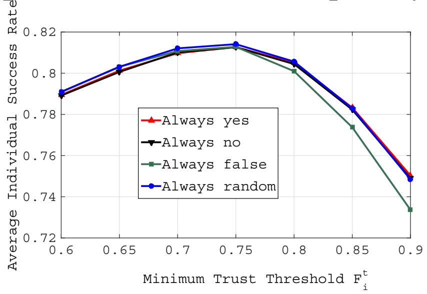


Figure 8: Impact of the Minimum Trust Threshold F_i^t.

minimum trust threshold F_i^t increases, there will be fewer sensing reports passing the threshold but the quality of information is better. This tradeoff results in the F_i^t optimal point.

The optimal T_i^t and F_i^t settings are sensitive to the percentage of malicious nodes (not reported here due to page limit). This result suggests adaptive control based on the percentage of malicious nodes sensed at runtime to maximize protocol performance.

VI. Discussion

In this section, we compare our proposed centralized and distributed cooperative spectrum sensing schemes in terms of prediction accuracy and overhead. We also discuss the amenability of our mechanism design theory methodology as applying to centralized and distributed cooperative spectrum sensing in cognitive radio networks.

A. *Prediction Accuracy*

Figure 9 compares our centralized and distributed cooperative spectrum sensing schemes in terms of prediction accuracy for the “always false” SSDF attack case. The experiment setting is the same as

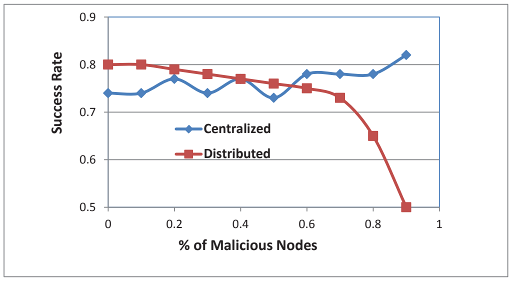


Figure 9: Comparison of Success Rate of Centralized and Distributed Cooperative Spectrum Sensing Schemes.

that in Section V-C. We see that the distributed scheme performs slightly better than the centralized scheme when the malicious node percentage is low (0%-30%). The reason is that in the distributed scheme, good nodes (which are the majority when the percentage of malicious nodes is low) with high capability can update and propagate trust scores through the blacklist reports broadcast on the common control channel, while in the centralized scheme trust scores can only be updated by the DFC with the checking probability p^t. As the malicious node population increases, the centralized scheme outperforms the distributed scheme. The reason is that the centralized scheme can count on the presence of a centralized DFC to check (with the checking probability of p^t) the ground truth PU activity result and can accurately penalize malicious nodes by reducing their trust scores, regardless of the percentage of malicious nodes in the system. When the DFC does not check the ground truth PU activity result (with probability $1 - p^t$), it will accumulate the reported results from all SUs based on trust-weighted majority voting. Since the DFC can track trust status of all SUs accurately, malicious nodes with low trust scores (because the DFC penalizes them) will not adversely affect the PU occupancy outcome. On the other hand, the distributed scheme does not have a centralized DFC that can obtain the ground truth PU activity. As the percentage of malicious SUs increases, an SU with capability lower than the minimum capability threshold can only update the trust scores of other SUs through the blacklist reports broadcast on the common control channel. Because malicious SUs can collude to put good SUs in the blacklist and conversely remove malicious SUs out of the blacklist, an SU is unlikely to keep accurate trust scores of other SUs in the system. The inaccuracy of trust scores kept by an SU in the distributed scheme is most pronounced when the percentage of malicious nodes exceeds a threshold (60%) beyond which the success rate drops rapidly compared with the centralized scheme.

Our analysis indicates that the cutoff point in terms of the minimum percentage of malicious nodes beyond which centralized cooperative spectrum sensing is better than distributed cooperative spectrum sensing is 60%, as suggested in Figure 9. Of course the cutoff point of 60% is not universally true as it depends on the parameter settings. However, one can apply the analysis methodology developed in this paper to derive the cut-off point, when given a set of parameters characterizing the cognitive radio environment.

B. Overhead

The convenience of having a DFC to check the ground truth PU occupancy status in the centralized scheme comes with a cost. In addition to the cost of installing a DFC in the system, it incurs a high overhead every time the DFC checks the ground truth PU activity so that the DFC can properly apply penalty in terms of trust loss to malicious nodes that perform SSDF attacks. The cost is much higher in scale than that of the distributed scheme which only relies on message broadcasting on the common channel for an SU to determine the PU activity as well as trust status of other SUs in the system. The high overhead issue is somewhat mitigated in our work by deriving the smallest DFC checking probability p^t under which malicious SUs do not have the intention to lie about their sensing capability and sensing result. In practice, high overhead is less important than high prediction accuracy. Therefore, the centralized scheme should be the choice especially when there is a high percentage of malicious SUs in the system performing DDSF attacks. Lastly, both schemes are comparable in terms of the decision time delay for an SU to reach a decision, since slotted broadcasting is used in both schemes to avoid communication interference.

C. Mechanism Design as Applying to Centralized and Distributed Cooperative Spectrum Sensing

We argue that centralized cooperative spectrum sensing is more amenable to mechanism design theory principles because of the presence of a DFC that can obtain ground-truth PU occupation status whenever it needs to (with probability p^t in a sensing report slot in our proposed centralized scheme). This ground truth information is used to update trust scores of SUs based on their sensing reports containing PU occupancy and sensing capability information. This allows the DFC to accurately track trust status of SUs and consequently achieve high prediction accuracy. On the other hand in distributed cooperative spectrum sensing, ground-truth PU occupancy status is not available. Each SU can just aggregate self and received sensing information to deduce PU occupancy. This estimated PU occupancy tends to be inaccurate when there is a high percentage of malicious SUs performing attacks. Consequently, Each SU especially the one without a high sensing capability exceeding T_i^t tends to update the trust scores toward other SUs inaccurately, resulting in low prediction accuracy. This is evident from Figure 9 in which distributed cooperative spectrum sensing performs much worse than centralized cooperative spectrum sensing when there is a high percentage of malicious SUs performing attacks.

VII. Conclusion

In this paper we proposed and analyzed trust-based data fusion schemes for cooperative spectrum sensing in cognitive radio networks to cope with data falsification attacks. We designed data fusion rules to distinguish erroneous reports due to low sensing capability from those due to malicious attacks. Our design effectively forces malicious nodes to report true sensing capability and outcome to prevent trust loss, thus allowing a high success rate to be achieved. We also identified optimal trust protocol settings under which the success rate is maximized. The simulation results validated the theoretical analysis and demonstrated that our trust-based data fusion scheme outperforms traditional data fusion rules and can distinguish malicious nodes performing data falsification attacks through their low trust scores in the long run. Finally, we demonstrated that our trust-based design methodology and analysis results can be extended to distributed cooperative spectrum sensing in which the DFC does not exist.

In the future we plan to explore modeling techniques such as Stochastic Petri Nets [23]–[28] to model behaviors of good and malicious SUs in order to study the interaction and exploit the design tradeoffs that exist in the game structure. We also plan to further test the resiliency of our trust-based data fusion scheme against more complicated environmental and operational scenarios such as different received signals at each node because of the geography effect of the SUs, as well as more sophisticated attack behaviors such as opportunistic, collusion, and insidious attacks [29], [30].

ACKNOWLEDGMENT

This work is supported in part by the U. S. Army Research Office under contract number W911NF-12-1-0445 and by the U. S. Air Force Office of Scientific Research under grant number FA2386-17-1-4076.

REFERENCES

[1] L. Canzian, Y. Xiao, W. Zame, M. Zorzi, and M. van der Schaar, “Intervention with private information, imperfect monitoring and costly communication,” *IEEE Transactions on Communications*, vol. 61, no. 8, 2013, pp. 3192–3205.

[2] A. Josang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision,” *Decision Support Systems*, vol. 43, no. 2, 2007, pp. 618–644.

[3] J. Guo, I.R. Chen, and J.J.P. Tsai, “A Survey of Trust Computation Models for Internet of Things Systems,” *Computer Communications*, vol. 97, 2017, pp. 1-14.

[4] I.R. Chen, J. Guo, and F. Bao, “Trust management for SOA-based IoT and its application to service composition,” *IEEE Transactions on Services Computing*, vol. 9, no. 3, 2016, pp. 482–495.

[5] I.R. Chen, F. Bao, and J. Guo, “Trust-based Service Management for Social Internet of Things Systems,” *IEEE Transactions on Dependable and Secure Computing*, vol. 16, no. 6, 2016, pp. 684-696.

[6] Y. Cai, L. Cui, K. Pelechrinis, P. Krishnamurthy, M. B. Weiss, and Y. Mo, “Decoupling trust and wireless channel induced effects on collaborative sensing attacks,” in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, 2014.

[7] W. Wang, H. Li, Y. Sun, and Z. Han, “Attack-proof collaborative spectrum sensing in cognitive radio networks,” in *43rd Annual Conference on Information Sciences and Systems*, 2009.

[8] J. Wang and I.R. Chen, “Trust-based data fusion mechanism design in cognitive radio networks,” in *IEEE CNS Workshop on Cognitive Radio and Electromagnetic Spectrum Security*, Oct 2014, pp. 1–6.

[9] Q. Yan, M. Li, T. Jiang, W. Lou, and T. Hou, “Vulnerability and protection for distributed consensus-based spectrum sensing in cognitive radio networks,” in *IEEE INFOCOM*, 2012, pp. 900–908.

[10] C. Chen, M. Song, and C. Xin, “A density based scheme to countermeasure spectrum sensing data falsification attacks in cognitive radio networks,” in *IEEE GLOBECOM*, 2013, pp. 623–628.

[11] H. Tang, F. Yu, M. Huang, and Z. Li, “Distributed consensus-based security mechanisms in cognitive radio mobile ad hoc networks,” *IET Communications*, vol. 6, no. 8, 2012, pp. 974–983.

[12] J. Wang, I.R. Chen, J.J.P. Tsai, and D.C. Wang, “Trust-based cooperative spectrum sensing against ssdf attacks in distributed cognitive radio networks,” in *IEEE International Workshop on Communications Quality and Reliability*, May 2016, pp. 1–6.

[13] C. Chen, H. Cheng, and Y.-D. Yao, “Cooperative spectrum sensing in cognitive radio networks in the presence of the primary user emulation attack,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, 2011, pp. 2135–2141.

[14] Z. Qin, Q. Li, and G. Hsieh, “Defending against cooperative attacks in cooperative spectrum sensing,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, 2013, pp. 2680–2687.

[15] X. He, H. Dai, and P. Ning, “Hmm-based malicious user detection for robust collaborative spectrum sensing,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, 2013, pp. 2196–2208.

[16] F. Adelantado and C. Verikoukis, “A non-parametric statistical approach for malicious users detection in cognitive wireless ad-hoc networks,” in *Communications (ICC), 2011 IEEE International Conference on*, June 2011, pp. 1–5.

[17] C. Chen, M. Song, C. Xin, and M. Alam, “A robust malicious user detection scheme in cooperative spectrum sensing,” in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 4856–4861.

[18] D. Shu, J. Wang, F. Liu, and X. Song, “A trust-based method for cooperative spectrum sensing in cognitive radio networks,” in *3rd International Conference on Consumer Electronics, Communications and Networks*, 2013, pp. 68–71.

[19] A. Goldsmith, “Wireless communications,” in *Cambridge University Press*, 1996.

[20] B.F. Lo, “A survey of common control channel design in cognitive radio networks,” *Physical Communication*, vol. 4, no. 1, March 2011, pp. 26-39.

[21] A.M. Masri, C.F. Chiasserini, C. Casetti, and A. Perotti, “Common control channel allocation in cognitive radio networks through UWB communication,” *Journal of Communications and Networks*, vol. 14, no. 6, Dec. 2012, pp. 710-718.

[22] K. Bian, J.M. Park, and R. Chen, “Control Channel Establishment in Cognitive Radio Networks using Channel Hopping,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, 2011, pp. 689-703.

[23] I.-R. Chen and D.-C. Wang, “Analysis of replicated data with repair dependency,” *The Computer Journal*, vol. 39, no. 9, 1996, pp. 767–779.

[24] I.R. Chen, T.M. Chen, and C. Lee, “Performance evaluation of forwarding strategies for location management in mobile networks,” *The Computer Journal*, vol. 41, no. 4, 1998, pp. 243–253.

[25] S.T. Cheng, C.M. Chen, and I.R. Chen, “Dynamic quota-based admission control with sub-rating in multimedia servers,” *Multimedia Systems*, vol. 8, no. 2, 2000, pp. 83-91.

[26] I.R. Chen, B. Gu, S.E. George, and S.T. Cheng, “On failure recoverability of client-server applications in mobile wireless environments,” *IEEE Transactions on Reliability*, vol. 54, no. 1, 2005, pp. 115-122.

[27] B. Gu and I.R. Chen, “Performance analysis of location-aware mobile service proxies for reducing network cost in personal communication systems,” *Mobile Networks and Applications*, vol. 10, no. 4, 2005, pp. 453-463.

[28] I.R. Chen and N. Verma, “Simulation study of a class of autonomous host-centric mobility prediction algorithms for wireless cellular and ad hoc networks,” *36th annual symposium on Simulation*, 2003, pp. 65-72.

[29] R. Mitchell and I. R. Chen, “Modeling and analysis of attacks and counter defense mechanisms for cyber physical systems,” *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 350–358, March 2016.

[30] I.R. Chen, R. Mitchell, and J.H. Cho, “On modeling of adversary behavior and defense for survivability of military manet applications,” in *34th IEEE MILCOM*, 2015, pp. 629–634.