Multinomial Trust in Presence of Uncertainty and Adversaries in DSA networks

Shameek Bhattacharjee, Mainak Chatterjee, Kevin Kwiat and Charles Kamhoua

***Abstract*—Dynamic spectrum access (DSA) networks allow opportunistic spectrum access to license exempt secondary nodes. Usually secondary nodes employ a cooperative sensing mechanism to correctly infer spectrum occupancy. However, the possibility of falsification of locally sensed occupancy report, also known as secondary spectrum data falsification (SSDF) can cripple the operation of secondary networks.**

In this paper, we propose a multivariate Bayesian trust model for secondary nodes in a distributed DSA network. The proposed model accurately incorporates anomalous behavior as well as monitoring uncertainty that might arise from an anomaly detection scheme. We also propose possible extensions to the SSDF attack techniques. Subsequently, we use a machine learning approach to learn the thresholds for classifying nodes as honest or malicious based on their trust values. The threshold based classification is shown to perform well under different path loss environments and with varying degrees of attacks by the malicious nodes. We also show the trust based fusion model can be used by nodes to disregard a node's information while fusing the individual reports. Using the fusion scheme, we report the improvements of cooperative spectrum decisions for various multi-channel SSDF techniques.[1]

I. Introduction

Dynamic Spectrum Access (DSA) allows non-licensed users (secondary users) to opportunistically sense-and-use fallow licensed bands when the licensed users (primary users) are not using their bands. In order for the secondary users not to interfere with the primary users, they must continuously sense the bands to evaluate if the bands are vacant. Such sensing done through stand-alone secondary users are subject to the typical wireless channel induced noise; hence they might not correctly infer the true occupancy of a band. Hence a network of secondary users participate in *cooperative spectrum sensing* where each share his local sensing results with others. The final inference on the occupancy of various bands is obtained using some fusion rule that fuses multiple local sensing results as shared by various secondary users. However, such dependence on multiple sensing results from various sources has opened avenues of attacks and made channel access decisions in DSA networks susceptible to various unforeseen vulnerabilities [1].

One of the most common forms of attack is the Secondary Spectrum Data Falsification (SSDF) attack, where some malicious users share false local sensing opinions that might compromise the accuracy of fused spectrum occupancy inference [1], [4]. In SSDF attacks, the malicious user may advertise 'empty' as 'occupied' resulting in denial of spectrum or advertise 'occupied' as 'empty' to induce a punishable breach of regulations. Either way, this would severely cripple the operation of secondary nodes. There has been some research that focused on defense against SSDF attacks via isolation of malicious nodes [4], [8], [9] or disregarding information from nodes whose reputation is lesser than acceptable [2], [3], [4]. Also, most works classify users as either trustworthy or not. Due to unavailability of complete information, such strict *binary* classification of evidence may be impractical. Some of the earlier works [4], [8] analyzed defense strategies for single channel networks, although multi-channel spectrum data is usually shared during cooperative sensing.

In this paper, we propose a framework to cope with smart malicious users in a distributed multi-channel DSA network who employ various extensions of SSDF attack techniques. We introduce a multinomial Bayesian trust model that considers any uncertainty that might arise from an imperfect anomaly monitoring mechanism. First, we identify the strategies that a malicious secondary user might adopt to inflict the maximum damage while remaining undetected as long as possible. In that regard, we propose three variations of multi-channel SSDF attacks and provide the defense mechanisms by evaluating the trustworthiness of the users that share spectrum usage information. Instead of the commonly used binomial models (Beta distribution), we use a multinomial Dirichlet distribution [6] and subjective logic [7] (Josang's Belief model) for computing the trust of the users addressing the shortcomings of binary classification and uncertainties that arise due to lack of complete evidence. We argue that anomaly detection based on signal strength prediction is highly dependent on path loss and hence incorporate it in the proposed trust model. We also validate our model for different path loss environments and different SSDF strategies.

II. System Model and Assumptions

We assume an ad-hoc secondary network with n secondary nodes of which some fraction are malicious. Secondary node i continuously scans N channels to determine whether a channel is occupied or not and constructs its observed local binary occupancy vector as: 1 for 'occupied' and 0 for 'unoccupied'. This decision is arrived upon by comparing the energy sensed on a channel with a common normalization threshold. Once a binary vector is created, secondary node i broadcasts it to its neighbors. Any node j within a sharing radius R from

[1]Bhattacharjee, and Chatterjee are with EECS department at the University of Central Florida, Orlando. Emails: {shameek, mainak}@eecs.ucf.edu. Kwiat and Kamhoua are with Information Directorate at the Air Force Research Laboratory, Rome, NY. Email: {kevin.kwiat,charles.kamhoua.1}@us.af.mil. Approved for Public Release; Distribution Unlimited:88ABW-2015-1368, Dated 24 Mar 15. This research was partially funded by the Florida Center for Cybersecurity under Subagreement: 2108-1072-00-A.

978-1-5090-0073-9/15/$31.00 ©2015 IEEE

node i is considered to be i's neighbor and would hear the broadcast messages (binary occupancy vectors) from its neighbors. An honest node would broadcast the same vector as it observed while malicious node would broadcast a modified one i.e., falsify the occupancy of some channels. Based on the received vectors, all nodes will employ some majority voting fusion technique to obtain a better estimate about the spectrum occupancy of all N channels [4], [5]. We assume the presence of a trust enforcement entity to which all nodes periodically report the calculated trust values of its neighbors. This entity maintains long term history or average reputation of each node as observed/experienced by their neighbors. Given that, we do not consider bad-mouthing attacks where nodes lie about the trust of other nodes to the enforcement entity. Such a feature has been already discussed in [10] and can be incorporated in our model without loss of generality.

A. Anomaly Detection and Evidence Gathering

In this paper, instead of proposing an anomaly detection scheme, we assume that the anomaly detection mechanism as discussed in [2] is already in place. For each channel, the anomaly detection mechanism provides one of the three possible observations: (i) *matches* indicating positive behavior (denoted by φ), (ii) *mismatches* indicating negative/anomalous behavior (denoted by β), and (iii) *undecided* (denoted by μ) indicating inability to classify a channel as either positive or negative. For each neighbor j, node i constructs an evidence vector of length N, which corresponds to the predicted inference on each channel as φ, β or μ. The total number of matches, mismatches and undecided accumulated against node j by node i at any time slot t given its advertised vector is η^t_φ, η^t_β and η^t_μ respectively. Of course $\eta^t_\varphi + \eta^t_\beta + \eta^t_\mu = N$. Each channel wise observation is independent of each other. The advantage of the anomaly detection scheme in [2] is that it works even for densities greater than 50% of malicious nodes. Other schemes that produce binary evidence based on majority voting are not robust to high local densities of malicious nodes in distributed networks.

B. SSDF Attack Models

Traditionally, SSDF attacks have been viewed as 'denial' (changing 0's to 1's) or 'induced' (changing 0's to 1's). Hence given a channel, an SSDF attacker would have two strategies: 'always denial' or 'always induced'. However, we argue that falsifying every bit of its occupancy vector all the time may facilitate easy and quick detection of malicious users. Also, changing every bit at all times can increase the cost of resource limited attackers. Hence a smart malicious attacker would refrain from such a simplistic strategy and employ some probabilistic or randomized attack strategies where some channels are attacked and some are reported correctly. We propose and analyzed three extensions to SSDF attacks.

1. Deterministic magnitude SSDF attack: A malicious node falsifies on a *fixed* number of channels every time slot. The fraction of channels falsified on every time slot is denoted as D_{attack}. $D_{attack} = 0.50$ means half of the total number of channels are falsified on each time slot. However, the channels falsified on are *randomized every time slot*.

2. Probabilistic magnitude SSDF attack: A malicious node falsifies opinions on a random number of channels every time slot, and the channels falsified are also random. However, such nodes follow a long term average, P_{attack}, which represents how aggressive a malicious node is. $P_{attack} = 0.60$ means that the probability of a channel's report being changed is 60%. Thus, any number of channels (from 0 to N) can be changed in a time slot.

3. Collaborative deterministic magnitude SSDF attack: The malicious nodes collaborate to agree upon the channels they falsify. If the channel set remains same, it is called static collaborative attack (channel preference); otherwise it is called dynamic collaborative attack. We denote this attack by D^{col}_{attack}. Collaborative attacks are harder to defend against as the probability of blinding the majority voting fusion rule increases. Here, the channels attacked over time may be random.

The values of D_{attack}, P_{attack}, D^{col}_{attack} may be dictated by the attack budget a malicious user can afford or how conservative (to avoid easy detection) or aggressive a malicious user wants to be. It may be noted that very high values facilitate easy detection while very low values cannot significantly affect the majority voting rule.

III. Motivation for a Multinomial Trust Model

Existing binary trust models like [3], [4], [8] do not account for evidence with three or more possible outcomes. In such anomaly detection models, the central fusing entity gathers all opinions and performs a majority voting to find the fused result for every channel. If advertised opinion of a participating node does not agree with the fused result, it is considered as negative rating; else the node's behavior is considered positive. This only works well in centralized architectures where the number of malicious nodes compared to the total nodes is less than 50% [9]. As a remedy, a signal bound prediction based anomaly detection scheme has been proposed in [2], which produces multinomial evidence namely match, mismatch and undecided as discussed before. While calculating trust, the undecided ones are accounted for by splitting them into either positive or negative ratings. Given η_φ η_β and η_μ as defined in Section II-A, trust computed by node i for neighbor j is given as:

$$E^{j,i} = \frac{\eta_{\varphi^j} + \frac{\eta_{\mu^j}\eta_{\varphi^j}}{\eta_{\varphi^j}+\eta_{\beta^j}}}{\eta_{\varphi^j} + \eta_{\beta^j} + \eta_{\mu^j}} \tag{1}$$

where $0 \leq E^{j,i} \leq 1$.

Partial splitting of the undecided ones in the ratio of observed matches and mismatches is justified only when the channels attacked are uniformly random. However, when the channels are attacked with some pseudo-random preference on certain channels, such splitting is not justified. An intelligent adversary might employ a variety of statistical techniques and have some preference on a subset of channels to be attacked violating the assumption that η_μ can split in the ratio of matches and mismatches.

TABLE I
TRUST-OPINION TUPLE; N=40

Scenario	φ	β	μ	$E^{j,i}$
1	14	13	13	0.51
2	19	18	3	0.51
3	22	0	18	1.00
4	10	0	30	1.00
5	31	0	9	1.00
6	22	14	4	0.61

Binary models only work in environments where number of uncertain ratings are not high. However, if the number of uncertain ratings is very high, such models cannot distinguish between large and small proportions of uncertain ratings. Consider the 6 scenarios as shown in Table I. Scenarios 1, 2, and 6 have sufficient number of mismatches to given a lower values than scenarios 3, 4, and 5 with no mismatches. Note, though there are a varying number of undecided in scenarios 3, 4, and 5, they have the same trust value. Consider scenarios 4 and 5, where both produce a high trust value 1.00 although the number of positive ratings in scenario 4 is three times less than scenario 5, while the undecided in scenario 4 is much more than compared to scenario 5. This trust model assumes that having no mismatches is an index of being an honest node although in reality many opinions may be undecided. In data mining, this is popularly known as not being robust to null invariance where evidence supports conclusions that are neither `True` nor `False`. Hence even if statistical preference is not employed, there is a small non-zero probability that all channels which were attacked were inferred as undecided. Such a possibility is high when there are a large number of undecided opinions in the trust evidence. Hence a better trust model is one which would give more trust value to scenario 5 and a less trust value to scenario 4. This calls for the use of multinomial trust modeling such as the Dirichlet distribution, which is the multivariate generalization of the corresponding binomial models (Beta distribution).

IV. DIRICHLET EXPECTATION BASED TRUST MODEL

Multinomial distribution is the multi-variate generalization of the binomial distribution with $k > 2$ possible outcomes where each trial results in exactly one out of k possible outcomes in a total of N trials. Given this, number of occurrences d_i of each outcome i;$(1 \leq i \leq k)$, the observations data can be denoted with observation vector $\boldsymbol{D} = \{d_i | 1 \leq i \leq k\}$. Similarly let x_i be the unknown probability of occurrence of each outcome denoted by probability vector $\boldsymbol{X} = \{x_i | 1 \leq i \leq k\}$. Given this $\boldsymbol{D}$ is said to have a multinomial distribution with parameters N and $\boldsymbol{X}$. However initially $\boldsymbol{X}$ is unknown and hence the problem is Bayes estimation of parameters in $\boldsymbol{X}$ given data $\boldsymbol{D}$.

In our problem we can model the observation counts match, mismatch and undecided as the possible outcomes on the inference over each channel; hence $k = 3$ and the total number of channels being N. Thus observation counts from the trust evidence fit very well with the concept of multinomial distribution. Hence the problem of trustworthiness can be answered once we know posterior probabilities of occurrences of seeing a match, mismatch or undecided $\vec{\boldsymbol{X}}$ from a node j based the evidence $\boldsymbol{D}$. In Bayesian systems $\vec{\boldsymbol{X}}$ is updated over time with incrementally acquired evidence $\boldsymbol{D}$.

Dirichlet distribution is often used as a conjugate prior for a multinomial distribution because both prior and posterior retain the same form. Hence the unknown degree of belief associated with the three outcomes can be calculated assuming conjugate priors. In Bayesian systems, a prior probability distribution $p(x)$ is said to be conjugate to the class of distributions $p(D|x)$ if the resulting posterior $p(x|D)$ is in the same family as $p(x)$. In such a case, the resultant posterior $p(x|D)$ can be used as prior (if required) for further belief updates as incrementally new evidence D over time is received. The observation counts constitiute evidence parameter $\boldsymbol{D}$; $\boldsymbol{X}$ forms the probability parameter. In short, $\boldsymbol{X}$ is said to have a Dirichlet distribution with parameter $\boldsymbol{D}$ and is denoted as $\boldsymbol{X} \sim Dir(\boldsymbol{D})$. Such degrees of beliefs can be used to model trust and reputation [6].

In terms of Bayesian systems, if $x_1, \cdots, x_k$ are the unknown probabilities associated with k events, and the evidence is d_i, then the posterior degree of belief on each x_i having accounted for evidence parameter d_i is given by Eqn .2. The evidence parameter d_i, is defined as $d_i = r_i + Ca_i$, where r_i represents the most recent count for event i and a_i represents a prior base rate and C represents an a-priori constant which dictates whether an informative or non informative prior is initially assumed [6]. The posterior $p(x_i|d_i)$ can be calculated using the posterior Dirichlet multinomial distribution function given by:

$$f(\vec{x}|\vec{d}) = \frac{\Gamma(\sum_{i=1}^{k} d_i)}{\prod_{i=1}^{k} \Gamma(d_i)} \prod_{i=1}^{k} x_i^{d_i - 1}, \tag{2}$$

where $x_1, x_2, \cdots, x_k \geq 0$, $\sum_{i=1}^{k} x_i = 1$, $d_1, \cdots, d_n > 0$, with restriction that $x_i \neq 0$. The relation between observation parameter d_i and actually observed outcomes r_i is that $r_i + C.a_i = d_i$, where $\sum_{i=1}^{k} a_i = 1$ and $C > 0, a_i > 0$ such that zero occurrence of an outcome preserves the condition that $d_i > 0$.

Since trust is an expectation of a node's behavior, the node's trust is given by the mean vector for Eqn. (2) and is given as

$$E(x_i|\vec{d}) = \frac{d_i}{\sum_{i=1}^{k} d_i} \tag{3}$$

where d_i is a known as the total evidence mass for event i. The degrees of belief associated with the outcomes are expressed as the mean of each outcome.

A. Applying Dirichlet model to trust evidence

The most recent observation vector is the multinomial trust evidence $\boldsymbol{r} = \{\eta_\varphi, \eta_\beta, \eta_\mu\}$. Thus,

$d_1 = \eta_\varphi + Ca_i$; $d_2 = \eta_\beta + Ca_i$ and $d_3 = \eta_\mu + Ca_i$. Since there is no reason to believe a node has a particular pre-disposition to behave in a positive, negative or uncertain way, we assume a uniformly distributed non-informative prior. Since there are 3 outcomes, the prior initial base rate $a_i = \frac{1}{3}$ and $C = 3$. Given this $d_1 = \eta_\varphi + 1$; $d_2 = \eta_\beta + 1$; $d_3 = \eta_\mu + 1$.

Now that we have the parameters of the Dirichlet distribution we can express the expected degrees of belief associated with the events of match, mismatch and undecided in terms of the observed trust evidence using Eqn. (3) as

$$E_{\varphi} = \frac{\eta_{\varphi} + 1}{\eta_{\varphi} + 1 + \eta_{\beta} + 1 + \eta_{\mu} + 1} \tag{4}$$

Similarly, $E_{\beta} = \frac{\eta_{\beta}+1}{\eta_{\varphi}+\eta_{\beta}+\eta_{\mu}+3}$ and $E_{\mu} = \frac{\eta_{\mu}+1}{\eta_{\varphi}+\eta_{\beta}+\eta_{\mu}+3}$.

Hence for each node j, we have $E_{\varphi} = E^b_{ji}$ representing degree of belief, $E_{\beta} = E^d_{ji}$ representing degree of disbelief and $E_{\mu} = E^u_{ji}$ reflecting degree of uncertainty of node j based on gathered trust evidence of node i from the anomaly monitoring phase.

B. Interpreting belief as subjective logic for trust modeling

The proposition that a node will cooperate is either true or false and hence is a binary proposition. However, due to inherent uncertainty and imperfect knowledge caused by lack of evidence it is not possible to infer with certainty that the proposition is true or false. Hence we only have an *opinion* about this proposition and trust is often reported as the *expected opinion* [7]. This translates the problem into degrees of belief, disbelief and uncertainty represented by $E^b_{ji}, E^d_{ji}, E^u_{ji}$ where $E^b_{ji} + E^d_{ji} + E^u_{ji} = 1$. Josang's belief model is typically used to deal with such uncertainty in a proposition of binary state space (honest or not), but the beliefs are multinomial [7] where one of the features express uncertainty. Josang's definition of opinion $\boldsymbol{\omega} = \{b, d, u, a\}$ is a quadruple where the components respectively correspond to the belief, disbelief, uncertainty, and relative atomicity such that $a, b, d, u \in [0, 1]$ and $b + d + u = 1$. The expected opinion pertinent to the positive interaction or belief is given as

$$E(\omega) = b + au \tag{5}$$

where a is known as the relative atomicity which determines how uncertainty contributes to the final expected opinion. Since the proposition that a node will cooperate or not is binary, we treat a as 0.5 which is the value of relative atomicity in our model and $E^b_{bi} = b, E^d_{bi} = d, E^u_{bi} = u$. Hence the expected opinion on the proposition that the node is cooperative or not is given by

$$E^{\omega}_{ji} = E^b_{bi} + (a)E^u_{bi} \tag{6}$$

C. A Conservative Weighted Trust Metric

$E^{\boldsymbol{\omega}}_{ji}$ is a number between 0 and 1 making the separation between non trustworthy and the trustworthiness nodes difficult to depict or visualize. Hence we use a Shapley log value with Eqn. 7 to transform $E^{\boldsymbol{\omega}}_{ji}$ to a generic value on the real line where non-trustworthy nodes have a monotonically decreasing values and trustworthy nodes have monotonically increasing values. Subsequently, we report the normalized weight w_{ji} by giving a value between $[-1, 1]$ using Eqn. 8. The Shapley log value is given as

$$r_{E^{\boldsymbol{\omega}}_{ji}} = log_2 \left(\frac{E^{\boldsymbol{\omega}}_{ji}}{1 - E^{\boldsymbol{\omega}}_{ji}} \right) \tag{7}$$

The normalized conservative trust weight is given by

$$w_{ji} = \begin{cases} 1 - e^{-|r_{E^{\boldsymbol{\omega}}_{ji}}|} & \text{if } r_{E_{ji}{}^{\omega}} > 0; \\ -(1 - e^{-|r_{E^{\omega}_{ji}}|}) & \text{if } r_{E_{ji}{}^{\omega}} < 0; \\ 0 & \text{if } r_{E_{ji}{}^{\omega}} = 0 \end{cases} \tag{8}$$

where $w_{ji} \in [-1, 1]$. For performance and results, we mostly use the absolute value of node j's normalized trust weight which is average of all trust ratings calculated by node j's neighbors so that there is no bias for one particular pair and also because we cannot plot all node pairs. Hence average trust rating of a particular node j can be represented as w_j.

D. An illustrative example of Dirichlet Trust Computation

The scenarios in Table II represent trust evidence on a particular time slot for 40 channels for different nodes. Scenarios 1, 2, and 6 have both occurrences of mismatches while 3, 4 and 5 do not. Intuitively, we would expect 3, 4 and 5 to have higher trust than 1, 2 and 7. However, scenario 4 has high number of uncertain ratings as opposed to 5. The binary models cannot capture relative uncertainty with just one value. Hence $E^{j,i}$ gives the same answer for scenarios 4 and 5. *This ambiguity is resolved by the Dirichlet expectation model.*

If we observed the corresponding values for the same in our model given by $E(\boldsymbol{\omega})$, we observe that our model captures the presence of a high number of uncertain ratings by generating a trust value of 0.61 for scenario 4, whereas giving a higher value of 0.86 to scenario 5, differentiates between scenarios 4 and 5. We can also see that scenario 3 which has less uncertain ratings than 4 but more uncertain ratings than 5, has a trust value intermediate to the scenarios 4 and 5. This demonstrates that given no evidence of mismatch, lower uncertainty should be awarded with higher trust.

Among the scenarios with non-zero evidence of mismatches, scenario 6 has least number of undecided and most number of matches. Hence scenario 6 achieves higher trust value than scenarios 1, 2, and 7 but lower than scenarios 3, 4, and 5.

TABLE II
TRUST-OPINION TUPLE; N=40

Scenario	φ	β	μ	$E^{j,i}$	E^{ω}_{ji}	w_{ji}
1	14	13	13	0.518	0.5116	0.044
2	19	18	3	0.513	0.5166	0.064
3	22	0	18	1.00	0.755	0.664
4	10	0	30	1.00	0.616	0.361
5	31	0	9	1.00	0.860	0.837
6	22	15	3	0.594	0.581	0.281
7	12	22	6	0.353	0.383	−0.476

V. SIMULATION MODEL AND RESULTS

To validate the trust model, we conduct extensive simulation experiments where we consider a 100×100 grid with 30 randomly scattered nodes. Each node scans $N = 40$ channels and has a sharing radius of 30 units. The nodes 1, 3, 4, 5, 6, 9, 20, 27, and 29 are programmed as malicious while the rest are honest. We consider both cases when the malicious node attack independently and collaboratively. For training set learning, we consider an environment with path loss of 4. For testing

sets, we show that proposed model correctly classifies nodes as malicious or honest for networks with different pathloss factors and magnitudes of attack. We vary the path loss factor from 3 to 5. Both D_{attack} and P_{attack} are varied from 0.30 to 0.90.

A. Instantaneous and Average Trust: Node pair perspective

Figure 1 shows the comparison of both instantaneous and average trust values between honest node 16 and malicious node 20, as calculated by one of their respective neighbors 3 and 28. Regardless of the temporal uncertainty of evidence of certain time slots, we observe that honest node 16 clearly has a higher trust value than 20. This is true for both the instantaneous values and the long term moving average. The honest node's trust value is more stable as it does not attack at all while malicious node employs $P_{attack} = 0.50$ over time. Hence, there is a large difference between a malicious and honest node.

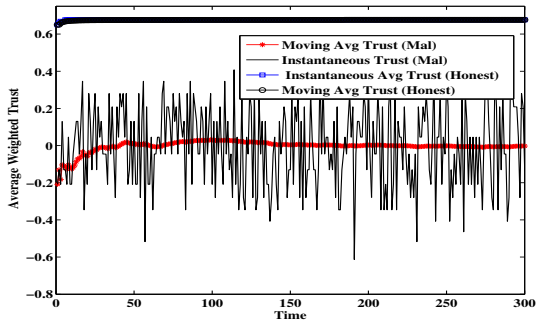


Fig. 1. Comparison of trust of malicious and honest nodes

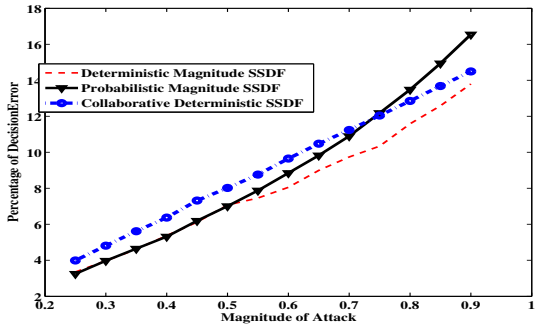


Fig. 2. Errors for different and varying SSDF attack magnitudes

B. Comparison between SSDF attack strategies

Fig. 2 shows how the different SSDF strategies affect the average utility of cooperative sensing with increasing magnitude of attack in the *absence of a defense mechanism*. In general, we see the obvious trend that the fraction of decision errors increases as the malicious nodes increase their magnitude of attack. We observe that for higher magnitudes of attack (above 0.50), P_{attack} produces more mismatches than D_{attack}. This is because, for $P_{attack} = m$, the fraction of channels attacked could be more than m. This, coupled with the fact that different channels are being attacked, increases the chances of circumventing the majority voting rule. We also observe that a collaborative attack is able to do more damage than most values of P_{attack} and D_{attack}, except when $P_{attack} > 0.80$. This validates intuition because collaboratively modifying opinions increases the probability of compromising the fused decision. However, when independent nodes attack with a high magnitude (> 0.80), they automatically will have common channels between them. Hence explicit deterministic collaboration does not yield similar benefits for excessively high attack magnitudes as it loses to additional diversity provided by P_{attack}.

C. Classification Threshold

To identify malicious nodes, there must be an observation or learning phase to make sure that the model is generic enough to correctly classify under a wide range of scenarios. In that regard, we propose a machine learning based approach that considers relevant network, radio, and topological parameters. We generate training data sets for different path loss environments and varying P_{attack} seeking to find an optimal threshold $w_{classify}$ to decide whether a node is malicious or not. We have identified the following:

Effects of pathloss: Given the anomaly detection model, we have found that the variation of pathloss effects the degree of uncertainty in evidence as shown in Table III. In general, we found if the path loss factors are too high or too low the average uncertainty is less evident; whereas an intermediate value increases uncertainty. A network with pathloss factor 4 induces maximum uncertainty while pathloss lower or higher reduce uncertainty. Hence we choose pathloss 4 as a parameter of our training since it induces maximum uncertainty.

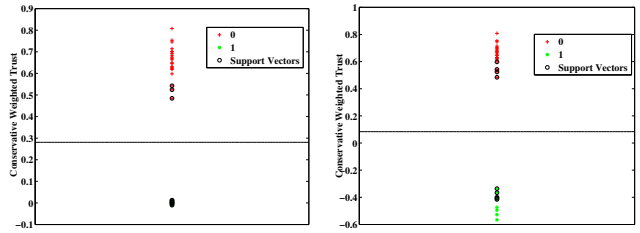
Effects of magnitude of attack: In general, if we can detect for lower magnitudes of attack, then we can detect for higher magnitudes of attack as well. Hence, we use training data sets mostly considering lower magnitudes of attack. However, too low of a magnitude defeats the purpose of an attack when a majority voting is used for decision making. Hence, to strike a balance between attacking and avoiding detection we choose an intermediate attack magnitude of 0.5 for training set.

TABLE III
EFFECT OF PATHLOSS ON UNCERTAINTY; WORST CASE $P_{attack} = 0.50$

Pathloss	Average E_μ
3	0.166065
4	0.426087
5	0.305750
5.5	0.200618

1) Training data sets: We run a support vector machine (SVM) over training examples to map trust values into support vectors and find the optimal hyper- plane (which in our case is a single line due to the linear nature of the data with only one feature i.e., the trust value). This optimal hyperplane that divides the feature space into two regions (upper and lower), can act as a robust classification threshold to distinguish between malicious or honest nodes. The lower and upper regions of SVM contains labels corresponding to malicious nodes and honest nodes respectively. In Figure. 3(a), '+' represents the labels corresponding honest nodes and '*' represents labels corresponding to the malicious node, along

with the optimal hyper-plane. The choice of the threshold is a tunable parameter. Our objective is to mimic the worst case scenario for classification; thus we place emphasis on the lower probabilities of attack when classification is hard. From Figs. 3(a) and 3(b), we choose 0.29 as it is the value of the optimal hyper-plane for the training set $T_1 = T(4, 0.5)$ with the worst case parameters.



(a) (b)

Fig. 3. (a) Pathloss=4, $P_{attack} = 0.50$ (b) Pathloss=3, $P_{attack} = 0.80$

D. Performance of Trust Model

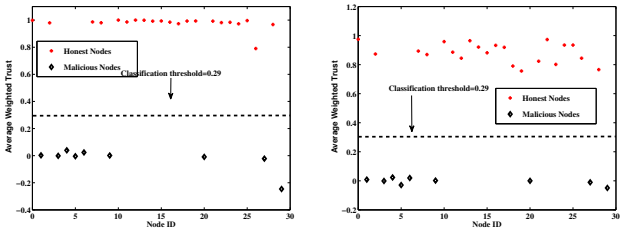
The performance of our trust model has two aspects. First, malicious node detection using steady state values of each node as calculated by its neighbors and then robust fusion using instantaneous trust values.

1) Malicious Node Identification: We use $w_{classify} = 0.29$ as the threshold below which we treat them as malicious. We see from Figs. 4(a), 4(b), 5(a), and 5(b), that for a variety of pathloss environments and for both higher and lower attack magnitudes of 0.8 and 0.5, we are able to distinguish between malicious and honest nodes with a high degree of certainty. Even among honest nodes, trustworthiness varies with the amount of uncertainty in their evidence.

2) Robust Fusion Spectrum: In Fig. 6(a), we see the performance benefit under P_{attack} from blind fusion, and observe much lower percentage of mismatches from ideal. Similarly, Fig. 6(b) also shows similar results for collaborative SSDF although it does not perform very effectively very low magnitudes, (because our training examples did not train for very low values as impractical). However, for most values, the performance benefit derived from a correct decision for all nodes in the network is high.

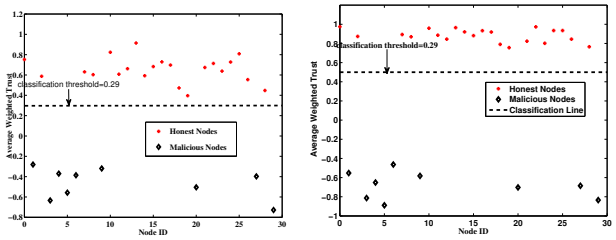
VI. Conclusions

In this paper, we argue that binary trust models are not appropriate when SSDF attacks are launched in a distributed DSA network. In that regard, we propose a multinomial Bayesian trust framework based on a Dirichlet distribution and subjective logic for assigning a trust value that better models the evidence gathered from an anomaly monitoring scheme. We present a machine learning based approach for calculating a classification threshold to decide whether nodes are honest or malicious. We also provide an instantaneous Dirichlet trust based fusion model, whereby we disregard information sent by potentially misbehaving nodes which increases the accuracy of cooperative sensing.



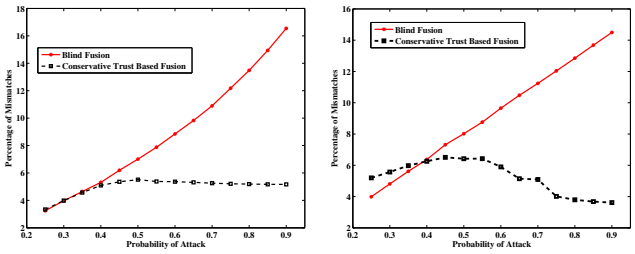
(a) (b)

Fig. 4. (a) Pathloss=3, $P_{attack} = 0.50$ (b) Pathloss=5, $P_{attack} = 0.50$



(a) (b)

Fig. 5. (a) Pathloss=3, $P_{attack} = 0.80$ (b) Pathloss=3, $P_{attack} = 0.80$



(a) (b)

Fig. 6. Performance of Trust based Fusion with (a) P_{attack} (b) D^{col}_{attack}

References

[1] S. Bhattacharjee, S. Sengupta and M.Chatterjee, "Vulnerabilities in Cognitive Radio Networks: A survey", *Elsevier Journal of Computer Communications*, Vol. 36, pp. 1387-1398, 2013.

[2] S. Bhattacharjee, S. Debroy, M. Chatterjee, "Trust Computation through Anomaly Monitoring in Distributed Cognitive Radio Networks", *IEEE PIMRC*, pp. 593-597, 2011.

[3] Y. Cai, L. Cui, K. Pelechrinis, P. Krishnamurthy, M. Weiss, Y. Mo, "Decoupling trust and wireless channel induced effects on collaborative sensing attacks", *IEEE DYSPAN*, pp.224-235, 2014.

[4] R. Chen, J.M. Park, K. Bian, "Robust Distributed Spectrum Sensing in Cognitive Radio Networks", *IEEE INFOCOM*, pp. 1876-1884, 2008.

[5] A. Ghasemi and E. S. Sousa, Collaborative spectrum sensing for opportunistic access in fading environments, *IEEE DySPAN*, Nov. 2005.

[6] A. Josang and J. Haller, "Dirichlet Reputation Systems", *The Second International Conference on Availability, Reliability and Security (AREAS)*, pp. 112-119, Apr. 2007.

[7] A. Josang, "A logic for uncertain probabilities", *Intl. Journal of Logic Fuzziness and Knowledge based Systems*, Vol. 9(3), pp. 279–311, 2001.

[8] A. Rawat, P. Anand, C. Hao, P. Varshney, "Countering byzantine attacks in cognitive radio networks," *IEEE ICASSP*, pp. 3098-3101, 2010.

[9] A. Rawat, P. Anand, C. Hao, P.K. Varshney, "Collaborative Spectrum Sensing in the Presence of Byzantine Attacks in Cognitive Radio Networks," *IEEE Trans. on Signal Proc.*, vol.59(2), pp. 774-786, 2011.

[10] Y. Sun, Z. Han, W. Yu, K. Liu, "A Trust Evaluation Framework in Distributed Networks: Vulnerability Analysis and Defense against Attacks",*IEEE INFOCOM*, pp. 230-236, 2006.

[11] M. Vu, N. Devroye, M. Sharif, V. Tarokh, "Scaling Laws of Cognitive Networks," Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), pp. 2-8, Aug. 2007.