Secure Collaborative Sensing for Crowdsourcing Spectrum Data in White Space Networks

Omid Fatemieh, University of Illinois Urbana-Champaign
Ranveer Chandra, Microsoft Research, Redmond, WA
Carl A. Gunter, University of Illinois Urbana-Champaign

***Abstract*—*Collaborative Sensing* is an important enabling technique for realizing opportunistic spectrum access in white space (cognitive radio) networks. We consider the security ramifications of *crowdsourcing* of spectrum sensing in presence of malicious users that report false measurements. We propose viewing the area of interest as a grid of square cells and using it to identify and disregard false measurements. The proposed mechanism is based on identifying outlier measurements inside each cell, as well as *corroboration* among neighboring cells in a hierarchical structure to identify cells with significant number of malicious nodes. We provide a framework for taking into consideration inherent uncertainties, such as loss due to distance and shadowing, to reduce the likelihood of inaccurate classification of legitimate measurements as outliers. We use simulations to evaluate the effectiveness of the proposed approach against attackers with varying degrees of sophistication. The results show that depending on the attacker-type and location parameters, in the worst case we can nullify the effect of up to 41% of attacker nodes in a particular region. This figure is as high as 100% for a large subset of scenarios.**[1]

I. Introduction

The use of wireless spectrum has been mainly regulated in the form of fixed and long-term license assignments. This form of assignment has proven to create significant inefficiencies in spectrum usage [1], [5]. The emerging paradigm for addressing this issue in the research and regulatory community is Dynamic Spectrum Allocation (DSA). DSA allows for opportunistic access to the licensed bands by unlicensed users on a non-interference basis. As an important regulatory step, FCC has recently adopted rules to allow unlicensed radio operation in the unused portions of the UHF spectrum, commonly referred to as *white spaces*. A key enabling technology for realizing DSA is Cognitive Radio (CR). A CR is a radio that can change its transmitter parameters based on interaction with the environment in which it operates [19], [33]. A CR is typically equipped with the ability to perform spectrum sensing, that is to sense the spectrum to identify frequencies that are unused by the licensed (also known as primary or incumbent) users. This allows the CR (also known as secondary) to operate in the unused frequencies, while avoiding frequencies that are in use by the primaries.

A number of proposed architectures for realizing large-scale DSA require that information about the locations and frequencies of primary signal presence be gathered by a central base station [3], [7]. This entails collecting and combining sensing reports from geographically separated CRs[2]. The recent FCC ruling on unlicensed radio operation in the broadcast TV spectrum requires that spectrum availability information be stored in geo-location databases [2]. In addition, it is shown that collaboration among the CRs yields significant benefits in terms of reducing uncertainties and relaxing individual sensing requirements [16], [39]. In *collaborative sensing* each CR performs spectrum sensing and the results, which we refer to as *spectrum data*, are combined in order to obtain a more accurate picture of the primary's presence [17]. This information can be used to govern the usage of spectrum by the secondary network(s).

The centralized approach to collaborative sensing lends itself well to the concept of *crowdsourcing*. The term is formally defined as 'the act of taking a job traditionally performed by a designated agent and outsourcing it to an undefined, generally large group of people in the form of an open call [20].' This concept has been adopted in many contexts, ranging from open software testing competitions [6] to investigating the expenses of members of parliament in the UK [4]. In our context, this can be realized as the act of gathering and combining sensing results from a large group of nodes that may be unreliable, untrustworthy, or even malicious.

Crowdsourcing of collaborative sensing in a large scale network raises a number of security threats. An attacker with access to a few compromised CRs can perform an *exploitation attack* in which they provide sensing reports that falsely declare the presence of a primary signal in a particular frequency. This may cause the secondary network to falsely believe the frequency is occupied and abandon it. The attacker can then exclusively use that frequency band. Alternatively, in a *vandalism attack* it can hide the presence of primary transmitters in a frequency to create unwanted interference, and thus perform a denial of service attack. We study the above *malicious false reporting* attacks in a large cognitive radio network. This is a challenging problem for the following reasons. First, due to the 'spatial variability' and uncertainties in the primary signal from factors such as shadowing, there is a lack of 'common ground truth' in the measured quantity among geographically separated CRs. This makes it easier for the attackers to hide under the natural variations of the measured quantity. Second, due to the open and easily recon-

[1] In the 4th IEEE Symposia on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10), Singapore, April 2010.

[2] Spectrum data is also envisioned to be collected from a separate *sensor network* deployed for this purpose alongside the main CR network [13], [37].

figurable nature of CRs, they are more prone to compromise and, once compromised, capable of more diverse misbehavior. For example, they may employ provably optimal strategies to evade basic detection techniques. This makes this problem much more difficult than finding faulty or misconfigured radios whose misbehavior is more evident.

Prior work on this problem focuses on strategies that work in small regions where a common ground truth is viable, and attackers constitute a small fraction of the secondaries or use unsophisticated strategies [12], [23], [31], [32]. In this paper we propose viewing the area of interest for detecting primary presence (or absence) as a grid of square cells and use it to identify and disregard false reports. Each cell is a unit of collaborative sensing and each individual sensing report is a real-valued measurement of primary signal power. The proposed mechanism starts by identifying outlier measurements inside each cell and 'punishing' them. The punishment is in the form of exclusion or a low weight assignment in the proposed weighted aggregation process. Subsequently, the mechanism entails *corroboration* and *merging* of neighboring cells in a hierarchical structure to identify cells with outlier aggregates. This can be used as a sign of significant malicious node presence in a cell. Our solution uses a simple model based on exponential decay and log-normal distribution to account for the uncertainties in signal propagation. In particular, we provide a novel framework for quantifying the expected legitimate variations in measurements. This reduces the likelihood of inaccurate classification of valid measurements as outliers.

We use simulations to evaluate the effectiveness of the proposed approach against attackers with varying degrees of knowledge and intelligence. The attacker model ranges from *naive* attackers that know nothing about their neighboring nodes and the defense mechanism to *smart omniscient* attackers with complete knowledge about the number and measurements of their neighboring nodes, as well as the detection strategy and parameters. The results show that depending on the attacker-type and the distance from primary to the region of interest, in the worst case we can nullify the effect of up to 41% of attackers nodes. This figure is as high as 100% for areas that are not near the border of primary's protection region.

The rest of the paper is organized as follows. Section II describes our setting and problem statement. Sections III and IV present the proposed approach and a framework for setting the introduced threshold parameters. Section V presents the first part of our simulation study. Section VI provides an extended version of our protocol and the second part of our simulation study. Sections VII and VIII describe related work and conclude the paper.

II. Setting and Problem Definition

In this section we first provide background information on collaborative sensing in the context of cognitive radio networks. Next, we describe our setting, assumptions, and problem statement.

A. Collaborative Sensing

A common approach to detecting a primary transmitter is *energy detection* [16]. In energy detection, the output signal of a bandpass filter with bandwidth W is squared and integrated over the observation interval T. The output, P, would be the measure of primary signal's presence, which can be compared with a *detection threshold* λ to decide whether a licensed user is present or not. If some features of the primary signal such as carrier frequency, bit rate, or modulation type are known, the more sophisticated *feature detectors* may be employed to detect primary signals. This comes at the cost of increased complexity, but enables the CR to more accurately identify and discriminate between sources of the received energy [17], [25]. In collaborative sensing, spectrum sensing results from CRs are incorporated for primary detection. This provides several benefits. First, it allows for relaxation of sensitivity requirements at individual CRs [39]. Second, it allows for mitigation of multi-path fading and shadowing effects, which improves the detection probability in highly shadowed environments [16].

In *centralized collaborative sensing*, which has been included in the IEEE 802.22 standard draft [3], the CRs report results to a base station on a periodic or on-demand basis. The base station or a centralized database is in charge of collecting the readings from the CRs and determining the areas of primary presence [7]. There exist two groups of strategies for combining individual reports. *Soft-combining* techniques combine raw signal power measurements from CRs, whereas in *hard-combining* techniques a 0/1 decision (and optionally false alert and missed detection ratios) from each CR is considered. In this paper, we use an instance of soft-combining for energy detectors based on *maximum likelihood estimation* [38]. This technique is presented in Section III.

An alternative model is *distributed collaborative sensing* in which individual sensing measurements are exchanged with the neighbors, and primary presence is determined by the network without relying on a base station. For example in [41] the authors envision an ad-hoc CR network in which each CR performs spectrum sensing and reports a 0/1 binary detection outcome to all the nodes in its range. Each CR considers a frequency band to be available for communication if neither itself nor any of the nodes in its range detect a primary signal in that band. A pair of nodes within each others' range that aim to communicate choose from the intersection of frequencies available to each of them.

B. The Setting

We consider a network of CRs (also referred to as secondary user, secondary node, or node) distributed over a large area; much larger than the coverage area of a primary transmitter. For simplicity, we consider the entire spectrum to be a collection of disjoint, equally sized, adjacent frequency channels. Without loss of generality, we only focus on detecting primary presence in one channel. We consider the nodes to use energy detectors due to their simplicity, efficiency, and widespread use [25].

The outcome of sensing by node N_i is p_i, which represents an estimate of the received primary power at node N_i. In dB, this is written as $p_i = p_t - (10\log_{10} r_i^{\alpha} + S_i + M_i)$ where p_t is the transmit power of the primary signal, r_i is the distance from N_i to the primary transmitter, $10\log_{10} r_i^{\alpha}$ represents the signal attenuation with exponent α (typically $2 < \alpha < 4$), and S_i and M_i are losses due to shadowing and multipath fading. We adopt the log-normal shadowing model [35] and therefore consider S_i and M_i to follow a Gaussian distribution ($S_i + M_i \sim N(\mu_s, \sigma^2)$) on the dB scale. Therefore we have $p_i \sim N(\mu(r), \sigma^2)$, where $\mu(r) = p_t - (10\log_{10} r_i^{\alpha} + \mu_s)$. For simplicity of analysis, unless otherwise noted, we consider μ_s to be 0 and σ to be independent of the distance to the transmitter [38].

The network infers the presence of primary users in a centralized fashion by crowdsourcing spectrum data. Collaborative sensing is often used with a specified set of nodes that are considered to be trustworthy. In crowdsourcing, however, this specified set of nodes is 'everyone.' Therefore, they may be unreliable, malicious, or compromised insider attackers. We assume the primary remains active at durations much longer than the period between consecutive spectrum sensings and that there exists a secure end-to-end connection, such as a TLS tunnel, between each participating CR and the base station. We also assume that the base station is not threatened by *Sybil* attacks [34], in which malicious CRs would create a large number of fake entities to obtain a disproportionately large influence. This is perhaps because of the difficulty of faking multiple link layer addresses or transmitters. Or one can take the dual view that we aim to demonstrate a method that forces adversaries to discover and deploy a practical Sybil attack. We also assume that the location of CRs is difficult or undesirable to fake. The problem of secure localization and that of primary emulation, in which a secondary actually transmits primary signals, are orthogonal to our problem and have been considered in the literature [13], [28], [29].

C. Problem Definition

We address the problem of secure collaborative sensing for crowdsourcing spectrum data in presence of malicious nodes in the 'crowd.' Faulty nodes that may unknowingly report false or inaccurate readings (due to software or hardware errors) are a subset of our problem space and we do not treat them separately.

A malicious node is one that is under control of an attacker. The attacker may have a number of compromised nodes under control and can make them act in cooperation for executing attacks on collaborative sensing. The attacker may aim for one of the following objectives: (1) Exploitation: the attacker makes the network falsely believe that an empty channel is currently occupied by primary incumbents in some part of the network. Under energy detection, this can be achieved by falsely reporting a primary signal measurement greater than λ. (2) Vandalism: the attacker makes the network falsely believe that a channel that is occupied by primary is available for communication. It effect, the attacker tries to hide the presence of incumbents. Under energy detection, this can be achieved by falsely reporting a primary signal measurement less than λ.

Our goal is to develop a mechanism by which despite the existence of malicious nodes, the base station can identify areas of incumbent presence/absence with high accuracy. Prime examples of incumbent communication are TV transmission and wireless microphones.

III. Hierarchical Approach

In this section we first introduce the soft-combining technique for collaborative sensing based on maximum likelihood estimation. Our approach is based on this method of collaboration among nodes. Next we provide a 'basic approach' which incorporates the basic ideas in the proposed scheme. We extend the basic approach to a 'weighted approach' which is the main protocol we use to evaluate our solution in Section V.

A. Maximum Likelihood Detector

Consider a square grid consisting of $n \times n$ square cells. Each cell is the basic unit of collaborative sensing and we call it a *level 0 cell*, or simply *cell*. The dimensions of a level 0 cell C are denoted by $r_0 \times r_0$. Consider a level 0 cell containing m nodes. Periodically, each node N_i in this cell provides its signal power measurement p_i to the central base station. Given a vector of received power observations $(p_1, p_2, \ldots, p_m)$ for this cell, a maximum likelihood (ML) detector would determine the primary presence by averaging the power measurements of individual nodes and comparing it to detection threshold λ [17], [38]:

$$\text{Primary is } \begin{cases} \text{Present}, & \text{if } P_{\text{avg}} = \frac{1}{m}\sum_{i=1}^{m} p_i \geq \lambda \\ \text{Absent}, & \text{otherwise.} \end{cases} \quad (1)$$

λ is determined based on the power of the transmitter and the radius around it, r, that needs to be protected. This is done such that the probability of missed detection stays below a threshold (*e.g.* .95), while the probability of false alerts are minimized. λ can be determined for a cell with m nodes as follows. If each measurement at distance r is distributed according to a normal distribution with mean $p_r = p_t - (10\log_{10} r^{\alpha})$ and standard deviation σ, we have $P_{\text{avg}} \sim N(p_r, \frac{\sigma^2}{m})$. We can determine λ such that:

$$\mathbf{Pr}(P_{\text{avg}} \geq \lambda) = .95 \Rightarrow \lambda = \frac{\sigma}{\sqrt{m}} Q^{-1}(.95) + p_r \quad (2)$$

where Q is the standard Gaussian distribution tail function and Q^{-1} is its inverse.

B. Basic Approach

It is easy to show that a few malicious nodes that report extremely high or low measurements can significantly skew the average in Equation 1, and thus alter the detection outcome. To that end, we propose a hierarchical structure for reducing or eliminating the effect of maliciously misreporting nodes. At the lowest level of the hierarchy (level 0) there exist level 0 cells. At higher levels of the hierarchy, each level l cell constitutes the area covered by b level $l-1$ cells that are

adjacent. b is called the *branching factor* of the hierarchy and we assume $\sqrt{b}$ is an integer greater than one. Figure 1 provides an illustration for $b = 4$.

In simple words, our scheme first aims to detect outlier measurements inside a cell by peer comparisons. If the attackers compromise a large fraction of nodes in a cell, they effectively take over the call and may no longer be detectable as outliers in the cell. Therefore, we use the hierarchy to compare each cell's average with its neighbors to detect if it is unexpectedly high or low. This corroboration allows our protocol to identify 'outlier cells' with significant attacker presence.

Consider a level 0 cell C_j that contains m secondary nodes. We define a *dispute threshold* for level 0, d^0, as the maximum acceptable difference between the measurements of two nodes inside that cell. In Section IV we provide a disciplined mechanism for deriving the dispute thresholds. As it will be shown, the dispute thresholds may vary for different cells. At level 0, pairwise comparisons between measurements of individual nodes are performed inside each cell. In each pairwise comparison between nodes N_i and N_j, if the difference is greater than d^0, N_i and N_j's dispute counts, c_i and c_j, are increased by one. After all pairwise comparisons, if $\frac{c_i}{m}$ is greater than or equal the *outlier threshold* for level 0, τ_0 $(0 < \tau_0 < 1$, *e.g.* $\tau_0 = .75)$, the node is flagged as an *outlier* and is excluded in the primary presence calculation in Equation 1. In other words, for a node *not* to be an outlier, at least a fraction $(1 - \tau_0)$ of the nodes in its cell should be within its d^0 distance. This method for outlier detection is an instance of the well-known *distance-based outlier detection* techniques in the literature. Formally, an object o in a data set D is a distance-based outlier with parameters *pct* and *dmin* if at least a fraction *pct* of the objects in D lie at a distance grater than *dmin* from o [18].

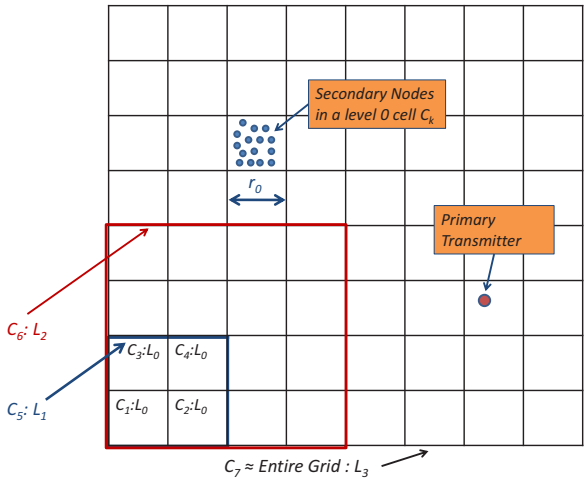

Fig. 1. Cells of different levels in a hierarchy with branching factor $b = 4$. $C_i : L_j$ denotes cell C_i at level j.

The higher levels of the hierarchy are formed as follows. A collection of b adjacent level 0 cells form a $r_1 \times r_1$ level 1 cell, where $r_1 = \sqrt{b} r_0$. At this step, after discarding outliers, the average signal presence at each of the consisting level 0 cells is calculated. The b resulting averages are compared in a pairwise fashion, and at each comparison, the dispute count is increased for a level 0 cell that has a difference greater than the dispute threshold for level 1, d^1, with a neighboring cell. Again, after all comparisons, if a cell's dispute count divided by the number of cells (b) is greater than the outlier threshold ratio τ_1 $(0 < \tau_1 < 1$, *e.g.* $\tau_1 = .75)$, the cell is flagged as an outlier and its result is considered unacceptable. The same procedure (averaging and neighbor comparison) is applied for up to l_{max} levels, and at each level if a cell j is flagged as outlier, all the cells it contains are flagged as 'indeterminate' for which primary presence cannot be accurately determined. For example in Figure 1, if C_5 is an outlier, the primary presence at C_1, C_2, C_3, and C_4 is indeterminate. For indeterminate cells, we consider primary presence to be difficult to tell, in which case an alternative source of information or method should be used for decision-making. For example, if there exist out-of-band mechanisms for establishing high trust in a subset of nodes, we can rely only on the measurements of the (few but trusted) nodes in that region. We do not explore this particular method in this paper, and leave it as an item of future work. Therefore, in our first set of simulations (see Section V) we simply report these cells as indeterminate. However, in Section VI, once we provide other means (based on median) to identify indeterminate cells, we propose and evaluate a simple method based on the average of 8 surrounding cells for primary detection in indeterminate cells.

The following Propositions state the limits that the basic approach imposes on exploitation attacks. Similar results can be derived for vandalism attacks.

Proposition 1: Consider a level 0 cell with dispute threshold d^0 and outlier threshold τ_0 under an exploitation attack. Let $\alpha < (1 - \tau_0)$ be the fraction of compromised nodes. If the average power of the un-compromised nodes and the average power including compromised nodes are denoted by m and m', under the basic approach we have: $m' \leq m + 2d^0\alpha$.

Proposition 2: Consider level i cells $C_l, \ldots, C_{l+b}$ with averages $m_l, \ldots, m_{l+b}$ that constitute the level $i + 1$ cell C_t with dispute threshold d^{i+1}. Let the outlier threshold $\tau_j = (b-1)/b$ for all level $j > 0$ cells. Consider a level i cell $C_n \in \{C_l, \ldots C_{l+b}\}$ under an exploitation attack. In order for C_n with the attacker influenced average m'_n to stay undetected as an outlier under the basic approach, the following property should hold: $m'_n \leq \max_{k \in \{l,\ldots,l+b\}-\{n\}}(m_k) + d^{i+1}$.

As an example for exploitation, consider a level 0 cell C_k (dispute threshold $= d^0_k$) with a fraction α of attackers. Assume the conditions of Propositions 1 and 2 hold, and $l_{max} = 2$. C_k is in level 1 cell C_l (dispute threshold $= d^1_l$). For ease of exposition, we represent all level 0 cells (excluding C_k) that are in C_l by $C_{k+1}, \ldots, C_{k+b-1}$. Also assume that C_l is in level 2 cell C_m (dispute threshold $= d^2_m$). For ease of exposition, we represent all level 1 cells (excluding C_l) that are in C_m by $C_{l+1}, \ldots, C_{l+b-1}$. Propositions 1 and 2 provide the following constraints on the attacker influenced average

for C_k, denoted by m'_k:

(1) $m'_k \leq m_k + 2d_k^0\alpha$,
(2) $m'_k \leq \max_{k+1\leq i\leq k+b-1}(m_i) + d_l^1$,
(3) $m'_k \leq b \max_{l+1\leq i\leq l+b-1}(m_i) + bd_m^2 - \sum_{i=k+1}^{k+b-1} m_i$.

C. Weighted Approach

The basic approach may result in flagging a number of nodes, level 0 cells, and higher level cells as outliers. The outlier nodes are excluded in the averaging for their respective cells. Likewise, the outlier cells are excluded in the averaging at higher levels, and the primary presence status in them is considered indeterminate. We propose using the results of the basic approach to assign and update weights to individual nodes over time (at the end of each round). In a level 0 cell C_i, each node N_j is assigned a weight w_j such that $\sum_{N_j\in C_i} w_j = 1$. In a cell with m nodes, each node's weight is initialized to $\frac{1}{m}$. We do not assign weight to cells. At level 0, the weighted sum of node's measurements is compared to the detection threshold:

$$\text{Primary is } \begin{cases} \text{Present}, & \text{if } \sum_{i=1}^{m} w_i p_i \geq \lambda \\ \text{Absent}, & \text{otherwise.} \end{cases} \quad (3)$$

Algorithm 1 Determine Level 0 Cell Status

Input: Level 0 cell C
lowCount $\leftarrow 0$; *highCount* $\leftarrow 0$
for each $C_i \in \big(Ancestors(C)\cup\{C\}\big)$ s.t. Level$(C_i) \leq l_{max}$
  if *LowOutlier*(C_i) **then**
    lowCount ++
  else if *HighOutlier*(C_i) **then**
    highCount ++
  end if
if *highCount* + *lowCount* > 1 **then**
  UpdateWeights$(C$,'conflicted')
else if *highCount* $== 1$ **then**
  UpdateWeights$(C$,'high')
else if *lowCount* $== 1$ **then**
  UpdateWeights$(C$,'low')
else // Neither C nor any of its ancestors is an outlier
  UpdateWeights$(C$,'neutral')
end if

Outlier detection is performed similar to the basic scheme. The only difference is that cells (not nodes) that are flagged as outliers can be assigned a 'low' or 'high' label; if the average value at an outlier cell is considered too low compared to its peers, it is flagged as a *low-outlier*, otherwise it is a *high-outlier*. After all the outlier detection and averaging is performed (starting from level 0, up to level l_{max}), Algorithms 1 and 2 are used to update the weights of nodes for the next round. In these algorithms, functions LowOutlier(C) (HighOutlier(C)) are considered to return 'true' if C is a low-outlier (high-outlier) cell.

Algorithm 2 UpdateWeights (C, *status*)

Input: Level 0 cell C, and *status* $\in$ {'conflicted', 'high', 'low', 'neutral'}
switch (*status*)
  case 'conflicted': **return**
  case 'high':
    sort the nodes in C based on power measurement
    cut the weights of the last 25% by half and equally distribute it to others in C
  case 'low':
    sort the nodes in C based on power measurement
    cut the weights of the first 25% by half and equally distribute it to others in C
  case 'neutral':
    cut the weights of the outlier nodes in C by half and equally distribute it to others in C
end switch

IV. Dispute Threshold Calculation

The dispute thresholds introduced in Section III aim to define maximum 'reasonable' differences between the observed signal powers among nodes (or averaged measurements among cells), beyond which the differences are highly questionable. Deriving the thresholds entails identifying and analyzing the sources of such power differences. The observed signal strength p (in dBm) at a secondary node is determined by the power of the transmitted signal p_t minus losses in power due to (1) attenuation at a distance r from the transmitter $l(r)$, (2) shadowing S, and (3) multi-path fading M, that is $p = p_t - (l(r) + S + M)$ [38]. Therefore, in order to characterize the differences, we need to study the effects of these three factors. We study the problem of determining thresholds at two different levels: (1) *Intra-cell* dispute thresholds (d^0) that are used to compare individual power measurements between nodes in a level 0 cell (2) *Inter-cell* dispute thresholds (d^i, $1 \leq i \leq l_{max}$) that are used to compare averaged measurements from each of the level $i-1$ cells contained in a level i cell.

A. Intra-cell Dispute Thresholds

Consider honest nodes N_i and N_j in a level 0 cell at distances r_i and r_j from the primary transmitter. Without loss of generality assume $r_j > r_i$. Therefore, $r_j = r_i + \Delta r_{i,j}$ ($0 < \Delta r_{i,j} \leq \sqrt{2}r_0$). Our goal is to find a value d^0 such that with high probability (*e.g.* 0.9) we have: $p_i - p_j \leq d^0$. Assuming independent, identically distributed (i.i.d.) Gaussian shadowing and fading at both nodes we have $p_i \sim N(p_t - 10\log_{10}(r_i^\alpha) - \mu_s, \sigma^2)$ and $p_j \sim N(p_t - 10\log_{10}(r_j^\alpha) - \mu_s, \sigma^2)$. Therefore we obtain the distribution of the difference as:

$$p_i - p_j = N(10\alpha\log_{10}\frac{r_i + \Delta r_{i,j}}{r_i}, 2\sigma^2)$$

For a fixed r_i, choosing $\Delta r_{i,j} = \sqrt{2}r_0$ maximizes the mean of the distribution. However, since we do not know the exact

location of the transmitter, we do not know r_i. In an ideal world where α is accurately known, and there is no loss due to shadowing and fading, one can use $p_i = p_t - 10\log_{10}(r_i^\alpha)$ to obtain r_i. We propose the following approach to estimate r_i in a more realistic environment where α is not accurately known and the effect of shadowing and fading is not negligible.

In order to reduce the uncertainty due to α, we take a conservative approach by taking the value of α that creates the largest attenuation from r_i to $r_i + \sqrt{2}r_0$. This is achieved by assuming a large α (*e.g.* $\alpha = 4$). In addition, the signal power, p_i, may have faced shadowing and fading. Therefore, p_i may not be the most valid choice for determining r_i. Since the size of a level 0 cell is relatively small compared to the distance to the transmitter, the average power reported by the nodes inside a cell may seem as an obvious candidate to estimate p_i. This average, however, is highly vulnerable to excessively high (or low) reports by malicious or deeply faded nodes. Therefore we opt for using the robust statistic of median [40] of the reported powers inside the cell for determining a conservative estimate of r_i. For a level 0 cell C, if p_{rep} is the representative power of this cell, and r_{rep} is the representative distance from this cell to the transmitter, we have:

$$p_{rep} = \text{median}(p_j), \text{ for all nodes } N_j \text{ in level 0 cell } C$$
$$p_{rep} = p_t - 10\log_{10}(r_i^\alpha) \Rightarrow r_i \sim r_{rep} = 10^{\frac{p_t - p_{rep}}{10\alpha}}$$

Therefore, if we aim to determine d^0 such that $\mathbf{Pr}\big(p_i - p_j < d^0\big) > .9$, we have:

$$p_i - p_j \sim N(10\alpha\log_{10}\frac{r_{rep}+\sqrt{2}r_0}{r_{rep}}, 2\sigma^2)$$
$$\mathbf{Pr}\big(p_i - p_j \geq d^0\big) \leq .1 \Rightarrow$$
$$Q\big(\frac{d^0 - 10\alpha\log_{10}\frac{r_{rep}+\sqrt{2}r_0}{r_{rep}}}{\sqrt{2}\sigma}\big) = .1$$
$$\boxed{d^0 = \sqrt{2}\sigma Q^{-1}(.1) + 10\alpha\log_{10}\frac{r_{rep}+\sqrt{2}r_0}{r_{rep}}}$$

where Q is the standard Gaussian tail probability function. Note that using this scheme, the dispute thresholds for different cells will likely be different. For future use, we denote $10\alpha\log_{10}\frac{r_{rep}+\sqrt{2}r_0}{r_{rep}}$ for a level 0 cell C_k by $\Delta\mu_k^{rep}$. We introduce the notation of d_k^0 to represent the dispute threshold for a level 0 cell C_k. We generalize this notation to represent the dispute threshold and average power for a level i cell C_k by d_k^i and p_k^i.

B. Inter-cell Dispute Thresholds

For simplicity, we first provide details on how d_k^1, the dispute threshold for a level 1 cell C_k, is calculated. Then we generalize the obtained result to higher levels. Consider a hierarchy with branching factor b. After outlier nodes are detected, and (weighted) averages for level 0 cells are calculated, we advance to level 1. At level 1, we perform pairwise comparisons between averages provided by each of the b level 0 cells contained in C_k, identify and leave-out outliers, and average the values of the rest to be passed to level 2. Consider two neighboring level 0 cells C_i and C_j (in C_k), with corresponding computed average powers p_i^0, and p_j^0. Assume there are m nodes in each cell. We have:

$$p_i^0 \sim N(\mu_i, \frac{\sigma^2}{m}),\ p_j^0 \sim N(\mu_j, \frac{\sigma^2}{m})$$
$$p_i^0 - p_j^0 \sim N(\mu_i - \mu_j, \frac{2\sigma^2}{m})$$

Ideally, if we were absolutely sure about the integrity of the majority of the nodes in each of the cells C_i and C_j, we could have replaced μ_i and μ_j by the averages of the corresponding cells. However, either of the cells may be populated by a large number of malicious nodes in a way not detectable at level 0. Hence, either of the averages could be highly skewed. As a result, using the difference between the sample averages is not a safe way to determine the probability distribution of the difference. Otherwise, very high dispute thresholds may be created that allow attackers to hide their presence. Besides, for simplicity, we are interested in using only one dispute threshold for each level 1 cell (as opposed to one dispute threshold for each level 0 pair). We employ a similar strategy to the intra-cell case and estimate $\mu_i - \mu_j$ by: $\Delta\mu_k^{rep} = \sqrt{b} \times \text{median}(\Delta\mu_i^{rep})$, for all level 0 cells $C_i \in C_k$. We can generalize this method to any level greater than 0. Therefore, we have $p_i^0 - p_j^0 \sim N(\Delta\mu_k^{rep}, \frac{2\sigma^2}{m})$. If we aim to determine d_k^1 such that $\mathbf{Pr}\big(p_i^0 - p_j^0 < d_k^1\big) > .9$, we obtain:

$$\mathbf{Pr}\big(p_i^0 - p_j^0 \geq d_k^1\big) \leq .1 \Rightarrow$$
$$Q\big(\frac{d_k^1 - \Delta\mu_k^{rep}}{\frac{\sqrt{2}\sigma}{\sqrt{m}}}\big) = .1$$
$$\boxed{d_k^1 = \frac{\sqrt{2}\sigma}{\sqrt{m}}Q^{-1}(.1) + \Delta\mu_k^{rep}}$$

It can be easily shown that the same argument could be used for higher layers of hierarchy. Therefore, if we represent the dispute threshold for a level i cell C_k by d_k^i, we have:

$$\boxed{d_k^i = \frac{\sqrt{2}\sigma}{\sqrt{b^{i-1}m}}Q^{-1}(.1) + \Delta\mu_k^{rep}}$$

where $\Delta\mu_k^{rep} = \sqrt{b} \times \text{median}(\Delta\mu_j^{rep})$, for all level $i-1$ cells $C_j \in C_k$.

Note that the dispute thresholds do not depend on the detection threshold, λ. It is easy to verify that as we go up in the hierarchy, the mean of the distribution for determining the dispute threshold is increased, while its standard deviation is decreased. This stems from the fact that the mean of the distribution mainly represents variation due to signal power attenuation over distance, whereas the standard deviation represents variations due to shadowing, which (as expected) is reduced as a result of aggregating increasing number of individual measurements.

V. Simulation Study (Part 1)

In this section we first provide the simulation setup used for evaluating the proposed scheme. The attacker model, results, and a brief analysis of the results are followed.

A. Simulation Setup

The simulation environment is a $4096\text{m} \times 4096\text{m}$ area in which secondary users are deployed uniformly at random with the density of 0.0008 per square meter. The branching factor, b, is 4. The size of each level 0 cell is $128\text{m} \times 128\text{m}$, creating a total of 1024 level 0 cells. Therefore, the expected number of nodes per cell is about 13. A primary transmitter with transmission power of 50mW (17 dBm) is located at the center of the area to represent a wireless microphone [10]. We consider a circular area with radius 1000m around the primary as the area that needs to be protected. This represents the area in which the primary signal must be detected with high probability. In particular, we require that primary signal be detectable by collaborating nodes in a level 0 cell with probability greater than .95 (max false negative rate of 5%). Using the formulation in Equation 2, this translates to the detection threshold of $\lambda = -74.4\text{dBm}$. We set the attenuation exponent, α, to 3 [38], and the standard deviation for the fading and shadowing process, σ, to 3 (in dB scale) [17]. The dispute threshold for each cell is determined based on the framework proposed in Section IV. The outlier threshold for level 0 cells, τ_0, is 0.6, and for all $i > 0, \tau_i = \frac{b-1}{b} = .75$.

B. Attack Scenarios

We first study exploitation attacks. We pick two cells outside the protection radius of the primary transmitter. First cell is selected randomly in such a way that is located at a distance marginally greater than the protection radius. This choice helps us gauge the worst-case performance of our protocol. We call this cell the *borderline-outside* cell. Next we randomly select another cell with the constraint that it is located at about two times the protection radius of the transmitter. We call this cell the *well-outside* cell. In each scenario, the attacker has compromised a certain fraction of the nodes inside the cell. Compromised nodes work in cooperation to report values higher than their true measurements to change the detection outcome. For a given cell and attack type, we vary the fraction of compromised nodes and study the results. The compromised nodes' behavior is according to one of the following models.

- **Naive Attackers** do not have any information about the number or measurements of others in their cell. They only know λ, and simply report measurements that are a fixed amount greater than λ.
- **Average Attackers** do not know the exact number or measurements of others. They know true measurements of themselves, the 'expected' number of nodes per cell, and λ. Assuming similar measurements by non-compromised nodes, they report measurements such that the estimated cell average is a few decibels (*e.g.* 4 dB) over λ. This to guarantee that if they underestimate the total number of nodes, they still succeed.
- **Smart Omniscient Attackers** know the number and measurements of all nodes in their cell, and λ. Using this information, they report measurements such that the final average for the cell is slightly over λ and not greater than $(p_{\text{avg}} + d^0)$. Here, p_{avg} is the average power of the cell assuming honest reports, and d^0 is the dispute threshold for the cell. This helps attackers reduce the chances of being detected as outliers at level 0.

For vandalism attacks, similar to exploitation scenarios, we pick two cells inside the protection radius of the primary transmitter. One cell is randomly selected from cells at a distance marginally smaller than the protection radius of the primary transmitter. We call this cell the *borderline-inside* cell. We randomly select another cell located at about half the protection radius away from the transmitter. We call this cell a *well-inside* cell. Attacker strategies are defined similar to the exploitation case, except here they aim to lower the average power measurement below λ for their cell.

C. Results

Figures 2, 3, and 4 (pictures on left) depict the measured average and final detection outcome for exploitation attacks. Results are collected after running the simulations for enough number of runs so that the weights (and thus the final outcomes) are stabilized. Note that in all graphs the y-axis represents the weighted and outlier-excluded average of the power of the nodes in the cell. For indeterminate cells (low, high, or conflicted based on Algorithm 1; represented by a '$\times$'), we do not provide any disambiguation solution in this section. Later, in Section VI, we introduce and evaluate one such solution. The results from an unmodified ML detector are provided in the captions for comparison.

It can be seen that for the well-outside cell, none of the attacker models can succeed. As a commonly observed pattern (except for smart attackers), when attackers constitute small fractions of the population in the cell, they are detected as outliers, and their weights are reduced. Therefore, they cannot move the cell average above the threshold. Once the attackers gain enough population to meaningfully increase the average without being individually detected, the entire cell is detected as an outlier at higher levels (level 1 here).

The picture is not as rosy in the case of the borderline-outside cell. It can be seen that once attackers obtain enough population (23% to 35% depending on the attacker model), they are able to successfully flip the detection outcome. This is not a surprise, and is in fact a direct consequence of our uncertainty model. In other words, once the mean of the distribution is close to λ, very few compromised nodes with reported measurements that *are* acceptable by the uncertainty model can move the average and flip the outcome without being detected. Note that such measurements could have come from a valid distribution, and thus been legitimate.

Figures 2, 3, and 4 (pictures on right) depict the measured average and final detection outcomes for vandalism attacks. The results from an unmodified ML detector are provided in the captions for comparison. Since the results and analysis are similar to the exploitation attacks we do not discuss them in detail and defer further analysis to the extended version of this paper.

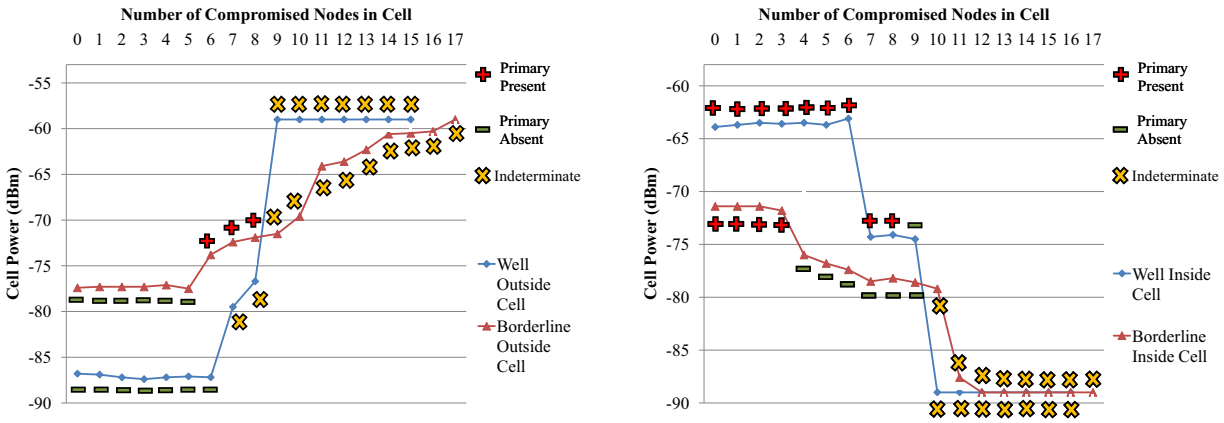

Fig. 2. (left) Exploitation by +15dB Naive Attackers. ML detector is beat when 7 (4) nodes are compromised in the well-outside (borderline-outside) cell. (right) Vandalism by -15dB Naive Attackers. ML detector is beat when 7 (3) nodes are compromised in the well-inside (borderline-inside) cell.

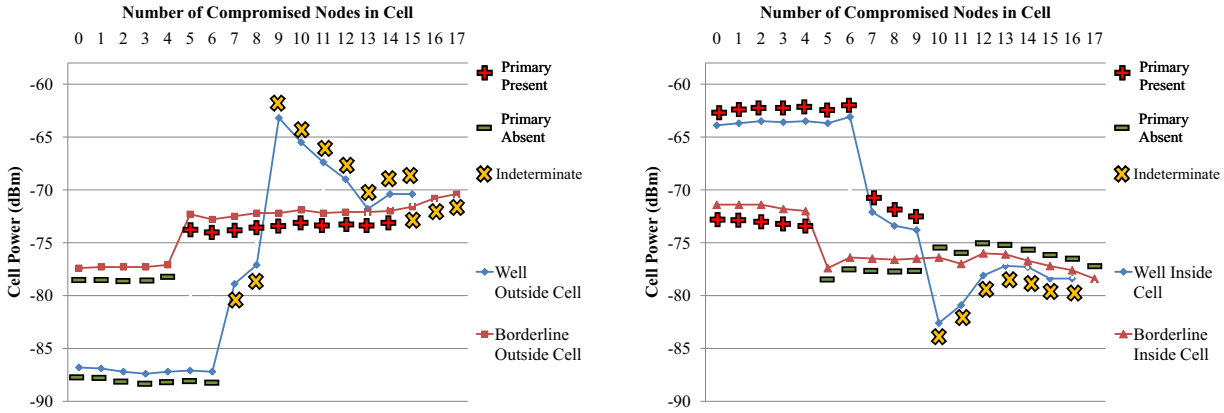

Fig. 3. (left) Exploitation by Average Attackers. ML detector is beat when one node is compromised in both the well-outside and borderline-outside cases. (right) Vandalism by Average Attackers. ML detector is beat when one node is compromised in both the well-inside and borderline-inside cases.

VI. Extensions & Simulation Study (Part 2)

The simulation results in Section V show that in areas where the average (mean) of signal power is close to the detection threshold, a modest fraction of compromised nodes in a cell can change the outcome of spectrum sensing without being detected. This is due to the difficulty of distinguishing between legitimate variations in signal power and slightly skewed false reports by attackers. Therefore, the attackers succeed by effectively 'hiding' under the 'acceptable' measurement variations. In this section we propose extending our solution by using median as a safeguard, in conjunction with mean, for secure primary detection. We show that our solution achieves a desirable mix of accuracy (from mean), and robustness (from median).

A. Median: A Safeguard for Collaborative Sensing

An alternative estimator for signal power in a cell is the median of measurements. The median of a sample is known to be robust to outliers. The median, however, has the disadvantage that it does not use all the data available in the sample, and therefore is often not as accurate as the mean [40]. For a normal distribution, it is well known that the sample mean is the most 'efficient' estimator, that is no other unbiased statistic for estimating μ can have smaller variance. The efficiency of median, measured as the ratio of the variance of the mean to the variance of the median, depends on the sample size $m = 2n + 1$ as $\frac{4n}{\pi(2n+1)}$, which tends to the value $2/\pi \approx .63$ as m becomes large [24]. So, we can consider the following distribution for the median power in a cell: $P_{\text{med}} \sim N(\mu, \frac{\pi\sigma^2}{2m})$. Therefore, similar to Equation 2, in order to use median for primary detection we can derive the threshold λ' such that the probability of missed detection stays below a certain value (*e.g.* .95):

$$\lambda' = \frac{\sqrt{\pi}\sigma}{\sqrt{2m}} Q^{-1}(.95) + p_r \tag{4}$$

where Q^{-1} is the inverse of standard Gaussian distribution tail function. Note that since $Q^{-1}(.95) < 0$ we have $\lambda' < \lambda$.

The next question that arises is how we can integrate median into our existing approach. To that end, we propose a framework based on the following principles: (1) safety (in terms of causing interference to primaries) is not compromised, and (2)

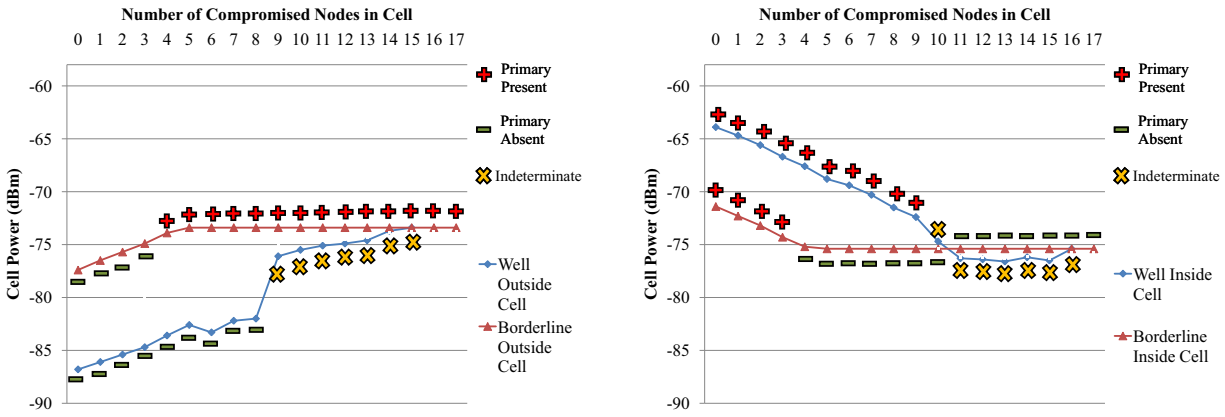

Fig. 4. (left) Exploitation by Smart Omniscient Attackers. ML detector is beat if one node is compromised in both the well-outside and borderline-outside cases. (right) Vandalism by Smart Omniscient Attackers. ML detector is beat if one node is compromised in both the well-inside and borderline-inside cases.

a reasonable combination of efficiency (*i.e.* mean) and robustness (*i.e.* median) is achieved. In a given cell, we first perform the hierarchical grid-based scheme proposed in Section III. If the status of the cell is 'neutral' (see Algorithm 1), then we perform the following *additional* operations. Consider P_{med} and P_{avg} to be the median and weighted mean of the power measurements. We have the following four cases:

1) $P_{\text{med}} >= \lambda'$ and $P_{\text{avg}} >= \lambda$: Since both estimators agree on the positive outcome, we consider primary signal to be **present**.
2) $P_{\text{med}} < \lambda'$ and $P_{\text{avg}} < \lambda$: Since both estimators agree on the negative outcome, we consider primary signal to be **absent**.
3) $P_{\text{med}} >= \lambda'$ and $P_{\text{avg}} < \lambda$: There exists a conflict; primary is present based on the median, but is absent based on the mean. Considering the importance of not causing interference to primary users, we disregard the potential optimality of the outcome from mean and opt for the conservative choice of declaring primary **present**. This choice is expected to reduce the chances of successful vandalism attacks, but may increase the chances of mistakenly declaring a borderline-outside cell as occupied (due to the relative inefficiency of median).
4) $P_{\text{med}} < \lambda'$ and $P_{\text{avg}} >= \lambda$: There exists a conflict; primary is present based on the mean, but is absent based on the median. The difference in opinions may be caused by an exploitation attack, or simply a legitimate inaccuracy by either of the two estimators. Given the previous choices, if we go with the mean's decision, we are effectively taking the decision to be the 'or' of the two. This choice has the drawback that would not make exploitation attacks any harder to launch. On the other hand, if we go with the median's decision, we are effectively ignoring mean in all the four cases, which is not desirable. Since we know that $\lambda' < \lambda$, the mean and median are at least separated by $\lambda - \lambda'$. This may be a sign of anomaly (*e.g.* an exploitation attack). We propose considering this cell as **indeterminate** and using the average power of the 8 neighboring cells (and compare it to λ) to determine the cell's status. We propose to use a similar disambiguation technique for **indeterminate** cells from Section III ('conflicted', 'low,' or 'high' in Algorithm 1). For the cells at the border of the area of interest (that do not have 8 neighbors), we consider the status to stay indeterminate.

B. Simulation Study (Part 2)

We first study the effect of using median in conjunction with mean in *absence* of attackers. This evaluation is done in terms of false positive and false negative rates.

TABLE I
THE NUMBER OF FALSE POSITIVES AND FALSE NEGATIVES.

Algorithm	False Positives	False Negatives
Hierarchical Average-Based (Section III)	16	10
Extended Median-Based (Section VI)	49	0

Consider any cell that is *entirely* outside the no-talk radius of the primary transmitter. If either of our approaches mistakenly declare primary to be present in this cell, we count this as a false positive. Similarly, consider a cell that is (in part) in the no-talk radius of the primary. If either of our approaches mistakenly declare primary to be absent for this cell, we count this as a false negative. We measure false positive and false negative rates in two cases: (1) when only the average-based framework in Section III is used, and (2) when it is combined with the median-based framework introduced in this section. The results are summarized in Table I. The table shows the number of false positives and false negatives for the final decision (after disambiguating indeterminate cells) for both approaches. The total number of cells is 1024. It can be seen that in the absence of attackers the extended approach provides an extra level of safety. This comes at the cost of higher false positive rates.

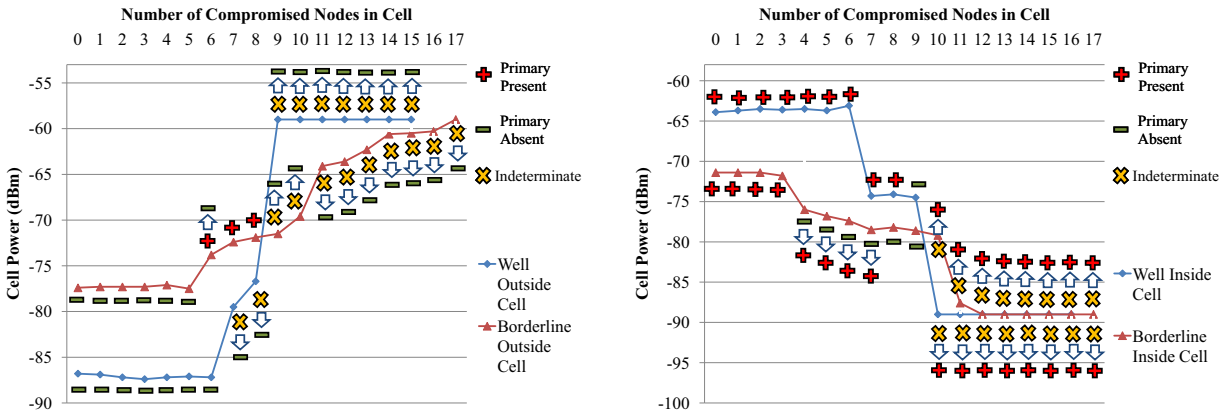

Fig. 5. Exploitation (left) and Vandalism (right) by Naive Attackers. Arrows represent change of final detection outcome based on extensions in Section VI.

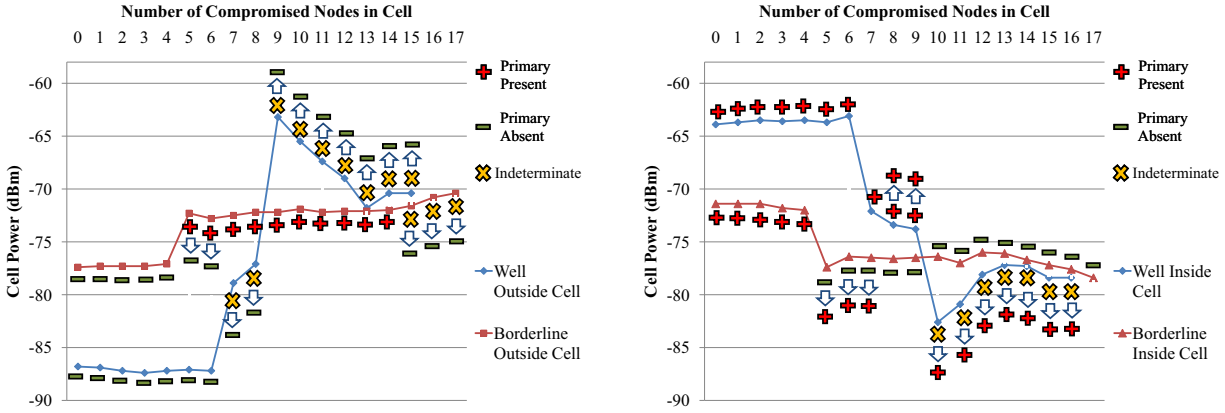

Fig. 6. Exploitation (left) and Vandalism (right) by Average Attackers. Arrows represent change of final detection outcome based on extensions in Section VI.

Next, we study the effectiveness of the extensions in this section against exploitation and vandalism attacks. For this purpose, we run the same experiments as in Section V with the added extensions in this section. Figures 5, 6, and 7 represent the results for Naive, Average, and Smart Omniscient attackers respectively. The new changes are represented by arrows. In particular, arrows originating from a '+' or '−' represent scenarios for which cases (3) or (4) apply, that is when the mean and median do not agree. The symbol at the head of the arrow represents the final decision. Arrows originating from a $\times$ represent cases that are considered indeterminate based on the hierarchical scheme in Section III, and the sign at the head of the arrow represents the final outcome after neighbor averaging rule.

The results show that for the well-inside (well-outside) case, in almost all scenarios, our solution completely nullifies the effect of attackers. For borderline-inside (borderline-outside) case, the attackers need to compromise at least 47% (41%) of nodes to be able to succeed. Note that these ratios are higher for the cases of less sophisticated attackers. The difference between the results for exploitation and vandalism can be explained by our conservative approach that prioritizes safety (non-interference) over security. Overall, the results show a considerable improvement over the original grid-based hierarchical scheme.

VII. Related Work

Kaligineedi *et al.* [23] introduce methods to detect malicious users that provide false measurements in collaborative sensing. The proposal includes pre-filtering of outlying sensing data, and a strategy to assign trust factors for weighting measurements and potentially omitting some nodes. Our solution is similar to their work in that it is based on outlier detection and weighting mechanisms. However, their proposal does not account for spatial variability of spectrum availability and only focuses on detection in a small region. Their solution falls short in cases where attackers constitute a large fraction of nodes in a cell. In addition, their approach and evaluation only considers simple 'always yes' and 'always no' attackers and unlike us does not consider sophisticated attackers.

Chen *et al.* [12] also consider malicious false reporting in collaborative sensing. They propose a weighted, reputation-based data fusion technique based on the sequential probability ratio test. The proposed scheme depends on apriori knowledge of the reporting values of radios given the true state of the world. It also does not account for spatial variability of

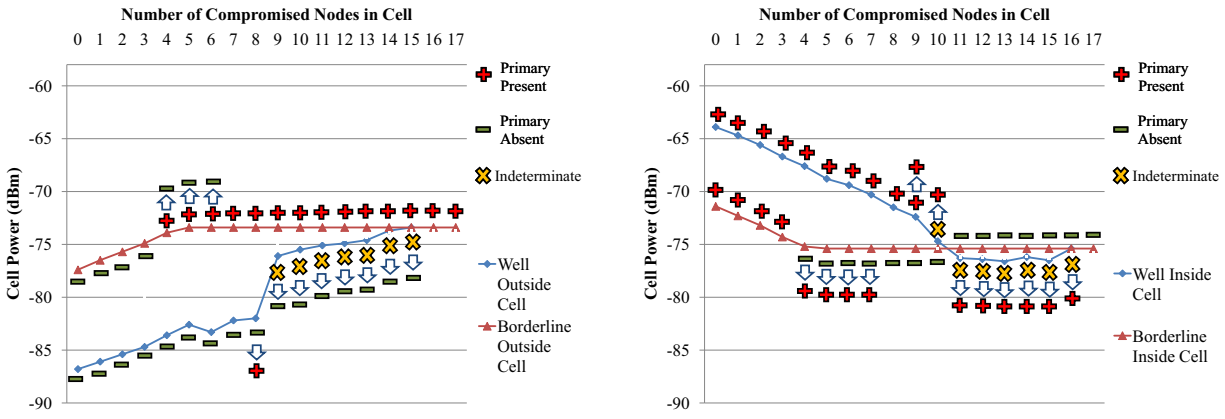

Fig. 7. Exploitation (left) and Vandalism (right) by Smart Omniscient Attackers. Arrows represent change of final detection outcome based on extensions in Section VI.

spectrum availability and only focuses on detection in a small region. In addition, the proposed mechanism is designed for hard-combining collaboration techniques, whereas we consider soft-combining techniques. Min *et al.* [31] propose grouping sensors in close proximity into clusters, and use shadow fading correlation-based filters to exclude or minimize the effect of abnormal sensor reports in detecting digital TV primaries. Their work elegantly considers shadow-fading correlation among nearby sensors, however, it only works when attackers constitute less than $1/3$ of the nodes in a cluster, and is not able to detect regions that are dominated by attackers.

A number of proposals motivate and identify various security issues in cognitive radio networks [11], [14], [36]. Although the attacks we consider, or a variation of them, are mentioned in these works and high-level ideas for mitigating them are proposed, none of them provide detailed solutions for addressing them.

A general area of related work is the broad subject of secure data aggregation in wireless sensor networks. A comprehensive survey on secure data aggregation can be found in [8]. Among the extensive body of work in this area, the following are the most relevant to this work. Wagner coined the term *resilient aggregation* [40], where he studies resilience of various aggregators to malicious nodes in an analytical framework based on statistical estimation theory and robust statistics. We benefit from the problem setting as well as the analysis technique in parts of this work, however, his work is limited to small regions and does not consider outlier-detection or combination of estimators as we do. Hur *et al.* [22] propose a trust-based framework in a grid in which each sensor builds trust values for neighbors and reports them to the local aggregator. Our work is similar to this work in that it is based on a grid. Their solution, however, does not provide a global view for a centralized aggregator, and also cannot identify compromised 'regions.' In addition, their work does not consider statistical propagation models and uncertainties in the data. Zhang *et al.* [42] propose a framework that identifies readings not statistically consistent with the distribution of readings in a cluster of nearby sensors. Their proposal, however, is local, that is only works for a small region. For example, it is not able to handle situations where attacker can compromise a large fraction of the nodes in a cluster. It also assumes the data comes from a distribution in the time domain, which in not a valid assumption in our domain.

Ganeriwal *et al.* [15] propose a reputation-based trust framework, where each sensor maintains a local reputation and trust for its neighbors. This work is very general, and is mainly focused on local decision making at each sensor. It is also local and peer to peer, meaning that the reputation is typically considered to be updated based on the quality of pairwise interactions between nodes.

Insider attacker detection in wireless networks is another area of related work. This problem has been explored in a general setting [9], [21], [43] as well as more specific contexts such as insider jammers [27]. As an illustrative example in the general context of sensor networks, Liu *et al.* [30] propose a solution in which each node builds a distribution of the observed measurements around it and flags deviating neighbors as insider attackers. This work is again local and peer to peer and does not work in areas with more than 25% attackers. Krishnamachari *et al.* [26] consider fault tolerant event region detection in sensor networks using a Bayesian framework. This work differs from our work in that it only considers faulty nodes that are not necessarily malicious, the faulty nodes are assumed to be uniformly spread, and the nodes itself participates in the detection process.

VIII. Conclusions and Future Work

In this paper, we studied malicious false reporting attacks on collaborative sensing in the form of crowdsourcing spectrum data. We provided a solution that is applicable to large regions where no single ground truth is viable at all places. Our solution uses outlier detection at two levels: (1) intra-cell among individual CR measurements and (2) inter-cell by corroboration among cells in a hierarchical structure. The results

are used in a weighted detection mechanism, in conjunction with a median-based framework, to eliminate or lower the effect of the attackers. We provided a novel framework for deriving the dispute thresholds for outlier detection based on the underlying propagation and uncertainty model of the signal power. We provided analytical and simulation results to quantify the extent to which attackers can succeed. The attackers in the simulations ranged from ones with very little sophistication, to those with complete knowledge about their neighbors and the detection mechanism (who use it to avoid detection). Our results showed that in cases where attackers are not near the border of the primary's protection area, we can detect and fully eliminate the effect attackers in a particular region. For our worst-case scenarios, that is cells that are close to the border of primary's protection area in which attackers employ smart strategies, we can nullify the effect of up to 41% of attackers nodes.

In future, we will provide further analytical results and a disciplined framework for quantifying the percentage of compromised nodes that can flip the detection outcome as a function of distance to the transmitter. We will also consider more sophisticated uncertainty models (*e.g.* spatially correlated shadowing). We will also consider cases in which multiple primaries co-exist.

IX. Acknowledgements

We thank Amir Nayyeri for his valuable comments. This work was supported in part by NSF CNS 09-17218, NSF CNS 07-16626, NSF CNS 07-16421, ONR N00014-08-1-0248, NSF CNS 05-24695, and grants from the MacArthur Foundation, Boeing Corporation, and Lockheed Martin. The views expressed are those of the authors only.

References

[1] FCC, ET Docket No 03-222 Notice of proposed rule making and order, December 2003.

[2] FCC, ET Docket No FCC 08-260, November 2008.

[3] IEEE 802.22 WRAN WG on Broadband Wireless Access Standards. *www.ieee802.org/22*.

[4] Investigate your MP's expenses. *http://mps-expenses.guardian.co.uk/*.

[5] Shared Spectrum Company. *http://www.sharedspectrum.com/*.

[6] uTest. *http://www.utest.com/*.

[7] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Comput. Netw.*, 50(13):2127–2159, 2006.

[8] H. Alzaid, E. Foo, and J. G. Nieto. Secure data aggregation in wireless sensor network: a survey. *AISC '08: Proceedings of the sixth Australasian conference on Information security*, pages 93–105, 2008.

[9] J. Branch, B. Szymanski, C. Giannella, R. Wolff, and H. Kargupta. In-network outlier detection in wireless sensor networks. *ICDCS 2006*, pages 51–51, 2006.

[10] G. Buchwald, S. Kuffner, L. Ecklund, M. Brown, and E. Callaway. The design and operation of the ieee 802.22.1 disabling beacon for the protection of tv whitespace incumbents. *IEEE DySPAN '08*, Oct. 2008.

[11] J. Burbank. Security in cognitive radio networks: The required evolution in approaches to wireless network security. *CrownCom '08*.

[12] R. Chen, J.-M. Park, and K. Bian. Robust distributed spectrum sensing in cognitive radio networks. *IEEE INFOCOM 2008*, April 2008.

[13] R. Chen, J.-M. Park, and J. Reed. Defense against primary user emulation attacks in cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 26(1):25–37, Jan. 2008.

[14] T. Clancy and N. Goergen. Security in cognitive radio networks: Threats and mitigation. *CrownCom '08*.

[15] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava. Reputation-based framework for high integrity sensor networks. *ACM Trans. Sen. Netw.*, 4(3):1–37, 2008.

[16] A. Ghasemi and E. Sousa. Collaborative spectrum sensing for opportunistic access in fading environments. *IEEE DySPAN '05*, Nov. 2005.

[17] A. Ghasemi and E. Sousa. Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. *Communications Magazine, IEEE*, 46(4):32–39, April 2008.

[18] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA, second edition, 2006.

[19] S. Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, Feb. 2005.

[20] J. Howe. The Rise of Crowdsourcing. *Wired Magazine*, 2006.

[21] Y.-a. Huang and W. Lee. A cooperative intrusion detection system for ad hoc networks. *SASN '03: Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks*, pages 135–147, 2003.

[22] J. Hur, Y. Lee, S.-M. Hong, and H. Yoon. Trust management for resilient wireless sensor networks. *ICISC*, pages 56–68, 2005.

[23] P. Kaligineedi, M. Khabbazian, and V. Bhargava. Secure cooperative sensing techniques for cognitive radio systems. *ICC '08: IEEE International Conference on Communications*, pages 3406–3410, May 2008.

[24] J. Kennedy and E. Keeping. *Mathematics of Statistics*. Van Nostrand, Princeton, NJ, USA, 3rd edition, 1962.

[25] H. Kim and K. G. Shin. In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? In *MobiCom '08*, New York, NY, USA, 2008. ACM.

[26] B. Krishnamachari and S. Iyengar. Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Transactions on Computers*, 53(3):241–250, March 2004.

[27] L. Lazos, S. Liu, and M. Krunz. Mitigating control-channel jamming attacks in multi-channel ad hoc networks. *WiSec '09: ACM conference on Wireless network security*, 2009.

[28] L. Lazos and R. Poovendran. Serloc: Robust localization for wireless sensor networks. *ACM Trans. Sen. Netw.*, 1(1):73–100, 2005.

[29] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust statistical methods for securing wireless localization in sensor networks. *IPSN '05: 4th intl. symp. on information processing in sensor networks*, 2005.

[30] F. Liu, X. Cheng, and D. Chen. Insider attacker detection in wireless sensor networks. *IEEE INFOCOM 2007*, pages 1937–1945, May 2007.

[31] A. Min, K. Shin, and X. Hu. Attack-tolerant distributed sensing for dynamic spectrum access networks. In *ICNP '09: 17th IEEE International Conference on Network Protocols*, Oct. 2009.

[32] S. Mishra, A. Sahai, and R. Brodersen. Cooperative sensing among cognitive radios. *ICC '06: IEEE International Conference on Communications*, 4:1658–1663, June 2006.

[33] J. Mitola. Cognitive radio an integrated agent architecture for software defined radio. *Ph.D Thesis, KTH Royal Institute of Technology*, 2000.

[34] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. *IPSN '04: 3rd international symposium on Information processing in sensor networks*, 2004.

[35] T. Rappaport. *Wireless Communications: Principles and Practice*. IEEE Press, New York, 1996.

[36] A. Sethi and T. Brown. Hammer model threat assessment of cognitive radio denial of service attacks. *IEEE DySPAN '08*, Oct. 2008.

[37] N. Shankar, C. Cordeiro, and K. Challapali. Spectrum agile radios: utilization and sensing architectures. *IEEE DySPAN '05*, Nov. 2005.

[38] R. Tandra, A. Sahai, and S. Mishra. What is a spectrum hole and what does it take to recognize one? *IEEE Magazine Special Issue on Cognitive Radio*, 97(5):824–848, May 2009.

[39] E. Visotsky, S. Kuffner, and R. Peterson. On collaborative detection of tv transmissions in support of dynamic spectrum sharing. *IEEE DySPAN '05*, pages 338–345, Nov. 2005.

[40] D. Wagner. Resilient aggregation in sensor networks. *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 78–87, 2004.

[41] Y. Yuan, P. Bahl, R. Chandra, P. A. Chou, J. I. Ferrell, T. Moscibroda, S. Narlanka, , and Y. Wu. KNOWS: Kognitive Networking Over White Spaces. *IEEE DySPAN '07*, 2007.

[42] W. Zhang, S. Das, and Y. Liu. A trust based framework for secure data aggregation in wireless sensor networks. 1:60–69, Sept. 2006.

[43] Y. Zhang and W. Lee. Intrusion detection in wireless ad-hoc networks. *MobiCom '00: 6th annual international conference on Mobile computing and networking*, pages 275–283, 2000.