# Using Classification to Protect the Integrity of Spectrum Measurements in White Space Networks

Omid Fatemieh
University of Illinois
at Urbana-Champaign

Ali Farhadi
University of Illinois
at Urbana-Champaign

Ranveer Chandra
Microsoft Research

Carl A. Gunter
University of Illinois
at Urbana-Champaign

## Abstract

*The emerging paradigm for using the wireless spectrum more efficiently is based on enabling secondary users to exploit white-space frequencies that are not occupied by primary users. A key enabling technology for forming networks over white spaces is distributed spectrum measurements to identify and assess the quality of unused channels. This spectrum availability data is often aggregated at a central base station or database to govern the usage of spectrum. This process is vulnerable to integrity violations if the devices are malicious and misreport spectrum sensing results. In this paper we propose* CUSP*, a new technique based on machine learning that uses a trusted initial set of signal propagation data in a region as input to build a classifier using Support Vector Machines. The classifier is subsequently used to detect integrity violations. Using classification eliminates the need for arbitrary assumptions about signal propagation models and parameters or thresholds in favor of direct training data. Extensive evaluations using TV transmitter data from the FCC, terrain data from NASA, and house density data from the US Census Bureau for areas in Illinois and Pennsylvania show that our technique is effective against attackers of varying sophistication, while accommodating for regional terrain and shadowing diversity.* [1]

## 1 Introduction

The proliferation of smartphones, and a subsequent demand for wireless Internet services, has highlighted the scarcity of spectrum for data communications. CTIA, which includes AT&T and Verizon, recently requested the FCC to grant an additional 800 MHz of spectrum for data communications by 2015 [8]. However, nearly all the spectrum that is ideal for long-range data communications, *i.e.* between 300 MHz and 3 GHz, has been allocated to various primary users.

The FCC white space ruling, which allows unlicensed devices to operate in unused TV spectrum is a significant step towards alleviating this spectrum crunch.[2] Devices determine if a TV channel is not in use at their location before using it to send and receive data. This ruling has met with excitement from industry, academia, and policy makers. The key reason for this excitement is two-fold. White spaces not only provide additional spectrum, they also enable long-range communication since they are in the lower frequencies (below 700 MHz).

An important functionality when forming networks over white spaces is the *aggregation of spectrum availability data* from multiple white space devices. The need for aggregation arises in several contexts. First, nearly all existing standards or proposals for white space networks, *i.e.* CogNeA, IEEE 802.22, IEEE 802.11af and WhiteFi [1, 4, 11], require the white space base station to receive spectrum availability reports from clients and operate on TV channels that are available at all nodes in the network. The spectrum reports from clients can be very diverse, since white space networks are expected to span a radius of up to 100 km [44]. Second, it has been shown that aggregating spectrum sensing data from other devices (also called collaborative sensing) enables white space devices to sense at a higher threshold than when sensing alone. This is very useful since sensing at low thresholds is extremely challenging [22, 48]. Finally, aggregation of spectrum sensing data from white space devices can be used to build a nationwide database of spectrum availability across locations [19]. This is similar to Wi-Fi wardriving data, and can be used for several purposes, for example to improve the accuracy of the white space geo-location database that is being mandated by the FCC [2, 3].

However, a threat to aggregating spectrum sensing reports is that some nodes may maliciously report inaccurate

[1] In the Proceedings of the 18th Annual Network and Distributed System Security Symposium (NDSS '11), San Diego, CA, Feb 2011.

[2] White spaces refer to portions of spectrum that have been allocated to licensed users but are not in use at that time.

data. There may be nodes that seek to *exploit* a spectrum in a given region by falsely reporting that a primary signal is present, or, dually, nodes that seek to *vandalize* a primary by reporting that its signal is *not* present thereby encouraging interference from secondaries. The first attack denies the legitimate users' access to the spectrum and provides exclusive access to attackers, whereas the second attack creates chaos and interference for primary and secondary users.

Countermeasures to prevent mischief are a key enabling technology for white space networks. Existing strategies have focused on an instance of this problem – in the context of collaborative sensing – for the detection of malicious nodes by identifying them as abnormal or outlier nodes within a small 'cell' [27, 37]. If one divides a service region into cells of sufficiently small size, then nodes within a given cell can be expected to give similar readings. If a preponderance of nodes in a given cell provide a reading in a common range, then other readings may be discarded as outliers. Ideally this will prevent malicious nodes from being effective. Unfortunately, this strategy suffers from several drawbacks. First, there is the possibility that a given cell will have a preponderance of malicious nodes. Second, countermeasures that aim to address this key limitation are based on unrealistic assumptions about the distribution of the signal propagation data that are not supported by systematic evidence from the data itself and therefore have limited performance. Third, such countermeasures must typically be tuned by hand, a strategy that is error prone and not easily scalable [19].

This paper presents a new approach called CUSP (for Classification Using Signal Propagation) using which a central aggregation server – a base station or spectrum availability database – can protect against malicious reports of spectrum availability. The key idea is to *let the data speak for itself*. CUSP uses natural signal propagation data in a region to learn a *classifier* that effectively understands the patterns of signal propagation in the region. It can then use the learned classifier to efficiently filter out the malicious spectrum reports as they often represent unnatural propagation patterns. We consider adversaries in three categories: non-collaborating adversaries who act individually, collaborating adversaries who act as a group, and omniscient adversaries who act as a group and possess complete knowledge of the defense mechanism and sensor data, including the data of non-adversaries. CUSP is able to counter malicious reports for all three classes of adversaries by learning classifiers with Support Vector Machines (SVM).

We evaluate the performance of CUSP in detail. We drive our evaluation on predicted propagation data derived from registered digital TV stations and terrain data from the FCC and NASA, as well as house density data from the US Census Bureau. We compute the signal strengths from the TV stations for two regions in the states of Illinois and Pennsylvania. We find that our techniques are quite effective with all three types of attacks, but regional variations have a significant impact that must be properly addressed to assure consistent quality of detection. In particular, areas with hilly terrain and urban activity must be treated in smaller cells. The resulting approach is practical and effective for application in all areas and avoids arbitrary assumptions about models, parameters, and thresholds in favor of direct training data.

More specifically, we make the following contributions.

- We identify and formulate a key threat to white space networks. More specifically, we address attacks in which powerful and coordinated attackers report false spectrum sensing results in order to obtain exclusive spectrum access or create chaos.
- We introduce a novel method to build classifiers from location-tagged signal propagation data. This obviates the need for relying on closed-form formulas, models, parameters, and thresholds when analyzing signal propagation data. Our technique, CUSP, detects misreporting attacks in the process of centrally aggregating spectrum sensing data by building SVM classifiers.
- We present a novel way to evaluate white-space applications using real-world transmitter and terrain data. We show our approach is effective against malicious misreporting attacks and outperforms the state of the art.

## 2 Formulating the Problem

In this section we present some background on white space networks, followed by our assumptions and the attacker model.

### 2.1 Background

On November 4, 2008 (and subsequently on September 23, 2010) the FCC made historic rulings that allowed unlicensed devices to operate over the licensed TV bands. The white space devices may operate on a TV channel as long as it is available at that location. To learn about spectrum availability, the white space devices, also called Cognitive Radios (CRs), may sense the spectrum to detect signals from primary transmitters. The FCC's September 2010 ruling eliminates spectrum sensing as a *requirement* for devices that have geo-location capabilities and can access a new TV band (geo-location) database [2, 3]. However, we believe that spectrum sensing or its variants can be used to improve the performance of white space networks. First, the ruling still supports the operation of sensing-only devices that cannot or do not access the database. Second, the database is

envisioned to be built from propagation models, which are very conservative and are likely to declare many channels (at locations away from the TV transmitters) as occupied even though they are effectively empty; spectrum sensing can provide a more accurate view of spectrum availability. Third, in the case of multiple available channels, the details of spectrum sensing results assist in selecting higher quality channels for operation.

*Energy detection* is the most popular approach for signal detection. This is often attributed to its simplicity and small sensing time (less than 1ms). An energy detector measures the signal power on a target frequency and compares it against a *detection threshold* $\lambda$ to determine whether a primary is present. For example, in the case of primary digital TV (DTV) transmitters, FCC has mandated -114 dBm as the detection threshold [2]. If a specific signature of a signal such as pilot, field sync, or segment sync is known, the more sophisticated *feature detectors* may be employed to detect primary signals. Feature detectors are often more accurate, but are more complex to implement, and require additional information and sensing time (up to 24ms) [22, 29].

Several scenarios in white space networks require the aggregation of spectrum sensing data. For example:

- In order to form a network over the white spaces, the CRs need to periodically report sensing results to a base station. The base station is in charge of collecting the readings from the CRs and determining the areas of primary presence. This centralized approach has been endorsed by the IEEE 802.22 WRAN standard draft [4], CogNeA [1] and recent research publications [11].

- Collaborative sensing refers to the process of combining spectrum sensing results from cognitive radios for the purpose of primary detection. The main benefit of this approach is the mitigation of multi-path fading and shadowing effects, which improves the detection accuracy in highly shadowed environments [22]. In addition, it allows for relaxation of sensitivity requirements at individual CRs [48].

- Crowdsourcing of spectrum reports from white space devices can be used to build a nationwide database of spectrum availability. Such a database can be used to augment the white space geo-location database mandated by the FCC [2] or to learn spectrum usage as part of the recently passed Spectrum Inventory Bill [6].

We use the third scenario for purposes of explanation in the rest of the paper, but CUSP will work without much modifications for the other two scenarios as well.

At a fine-grained level, there exist two broad classes of strategies for combining individual spectrum sensing reports. *Soft-combining* techniques combine raw signal power measurements from CRs, whereas *hard-combining* techniques combine binary decisions from CRs. Note that directly combining individual results happens only within small cells where nodes are expected to provide similar readings. One of the most popular methods for soft-combining is Equal Gain Combining (EGC). In EGC, each node $N_i$ of the $m$ nodes inside a small area periodically provides its signal power measurement $p_i$ to the central server. Assuming a vector of received power observations $(p_1, p_2, \ldots, p_m)$, and a nominal Gaussian model for shadowing and multi-path distribution, EGC is the maximum likelihood detector. It simply averages the power measurements of individual nodes and compares it to a detection threshold $\lambda$. That is, the primary is present if $P_{\text{avg}} = \frac{1}{m}\sum_{i=1}^{m} p_i \geq \lambda$. The threshold $\lambda$ 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 [46]. EGC is known to have near-optimal performance in a diverse set of fading channels with more realistic assumptions [37, 45].

## 2.2 Model & Assumptions

We consider a large network of CRs, each equipped with energy detectors. The choice of energy detection is due to its widespread acceptance and ease of implementation and analysis [9, 29, 44]. For each frequency channel, the outcome of spectrum sensing by node $N_i$ is represented by $p_i$, which is the received signal power at $N_i$. The primary signal faces path loss and shadow fading due to irregular terrain and obstacles such as trees, buildings, walls, and windows. As it will be explained later, the proposed approach does not use or depend on any particular human-formulated signal propagation model as it is designed to only use samples of real propagation data obtained through wardriving or other alternatives (see Section 3.4).

We consider a centralized model for aggregating spectrum sensing reports in which received signal powers from cognitive radios are reported to the central aggregation server, which divides the area to a grid of small cells to facilitate combining individual sensing results. As we will further elaborate, some cognitive radios may be unreliable, malicious, or compromised insider attackers. However, we assume that each node maintains a secure link to the base station for sending spectrum sensing results and that attackers are unable to fabricate nodes or identities arbitrarily ("Sybil" attacks [40]). The secure links can easily be realized using pre-shared keys or a PKI. We also assume that the locations of nodes are reliably known through GPS or other localization techniques and nodes do not misreport their locations. This assumptions is easily achievable in two of the most popular proposed applications of white

space networking that assume fixed nodes with known locations: (1) Internet access for consumer premises using IEEE 802.22 wireless regional area networks [44], and (2) advanced meter infrastructure (AMI) communications [18]. In cases where the network contains untrustworthy mobile devices, secure localization techniques may be employed to assure nodes' locations are not forged [32, 33]. In addition, for our most important attack scenario (exploitation), attackers do not gain any tangible benefit from misreporting their location.

The above assumptions are common for the type of analysis we perform here [15, 19, 37]; if they are violated then additional protective measures are required.

### 2.3 Problem Statement & Attacker Model

We address the problem of detecting malicious reports of spectrum availability in white-space networks. The detection occurs at a central aggregation server. The attacker nodes may act in cooperation to perform a *malicious false reporting* attack in a *cell*. A cell is a small area (*e.g.* 500m $\times$ 500m) that is the unit of combining individual sensing results for primary detection. In principle, the combination can be in the form of taking the average or median of power measurements, majority value of boolean 0/1 decisions, or any other function. We focus on detecting cells in which attackers have either a strong (*e.g.* majority) presence, or regardless of their count are able to 'dominate' the cell and flip the detection outcome. The domination in each of the above combination techniques may take a different form. For example for the EGC rule this involves changing the average signal power from a status indicating primary absence to one indicating primary presence, or vice versa. We call these attacks *exploitation* and *vandalism*, respectively. In exploitation, the attackers aim to deceive the network to abandon the channel to exclusively use it for themselves, whereas in vandalism the main goal is creating chaos on interference for the primaries (and legitimate secondaries). Our goal is to detect such *attacker-dominated* (or *compromised*) cells without necessarily focusing on detecting individual attacker nodes inside them. Without loss of generality, in the rest of this paper we consider the EGC rule as the combining method. Our methods and analysis, however, are easily applicable to other combining methods as well. We also do not consider the sufficiently addressed complementary problem of primary signal emulation attacks in which attackers transmit primary-like signals.

Throughout the paper, we mainly instantiate our approach to the particular case of high-power primary transmitters such as digital TV transmitters with signals spanning up to 150 kilometers. Such signals are often much stronger outside of buildings and to assure non-interference to the primary network, the secondary network must rely on sensing data from outdoor sensors. This is in-line with the envisioned scenarios for 802.22 service providers and AMI. We also briefly discuss the case of low-power primary transmitters such as wireless microphones in Section 5.2.

**Attacker Model.** We consider the following attacker models in this paper. Note that the attackers' behavior should be considered through the lens of a particular cell that the attackers aim to dominate.

**1. Un-coordinated** attackers do not have precise information on the number and power measurements of other legitimate or attacker nodes in the cell. Each attacker node aims to dominate the cell without cooperation with other attackers, if any. This may be due to lack of information, unavailability of communication channels, or to reduce the likelihood of being detected as a result of communicating with peers. In this case, a compromised node that senses a signal power below (above) the detection threshold may falsely report a value such that the average power in the cell changes to a value below (above) the detection threshold. The attacker may use rough estimates of the number and measurements of other nodes for this purpose (for example, for the latter it would be a close value to the attacker's true measurement).

**2. Coordinated** attackers do not know the number and power measurements of the legitimate nodes in the cell, but may roughly estimate them. They do, however, know their own number and measurements, and act according to a coordinated strategy; they collude and use the estimates to calculate the value that each of them should report so that they can dominate the cell and change the detection outcome to a value above (or below) threshold.

**3. Omniscient** attackers are coordinated attackers that know the exact number and measurements of other legitimate users. Therefore, they can simply calculate the exact power levels they should report to change the average power level to a value *slightly* above (or below) threshold, *e.g.* 1dB. This is to reduce the chances of being detected.

**4. Mimicry-capable Omniscient** attackers are omniscient attackers that have the (non-trivial) resources to build a classifier similar to that used in our detection technique. However, we can hide (or simply randomize) the schedule, frequencies, and locations in which we enable the detection scheme. Therefore, before any misreporting attempt the attackers can predict whether our classifier can catch them *if it is enabled* at that particular time, location, and frequency. In the small percentage of cases that they know it cannot detect them (even if enabled), they will misreport according to the omniscient strategy above. Otherwise, they may choose to misreport based on their risk appetite. In any case, if they choose not to misreport, we have achieved our goal of preventing attackers from manipulating the detection outcome. Otherwise, we will detect them as we would have detected

omniscient attackers. Therefore, we do not report separate results for this class of attacks and rely on results for omniscient attackers.

## 3 CUSP: Motivation and Approach

The two problems of detecting individual maliciously false reporting nodes [27, 37] and that of detecting attacker-dominated cells [19] have been mainly formulated as abnormality or outlier detection problems. Despite moderate degrees of success, these approaches suffer from several technical and practical issues. First, they often involve unrealistic assumptions about the models and parameters of signal propagation. Second, the performance of almost all of these methods highly depend on detection threshold parameters which are usually tuned by hand, or depend on the parameters of the signal propagation model. This is impractical, because it requires too much 'conjecturing' and 'manual tuning' for each given region and frequency band of interest. In addition, outlier detection techniques are often very conservative and are not designed for detecting nimble manipulations of data by sophisticated attackers. This limitation is particularly important in the context of spectrum sensing in which there exist natural variations in signal power due to factors such as fading and noise [46].

As an illustrative example consider Figure 1 to be a subset of the area of interest. Each cell is the unit for averaging signal power measurements from sensors to determine primary presence. The average power from the nodes inside a cell are represented by a number (in dBm) in that cell, and the primary detection threshold is -114dBm. Cells $A$ and $B$ are normal, whereas $C$ is dominated by attacker nodes. Therefore, the attacker nodes are able to decrease the average power to -115, which, if undetected, results in a successful vandalism attack. It is tempting to devise heuristics or simple outlier detection techniques based on approximate signal propagation formulas to catch cells like C. For example one may claim the difference between $B$'s average power and its neighbors looks normal since its average is smaller than a threshold $\alpha$, but this is not true for $C$, therefore $C$ is compromised. But 'why is comparing the average distance to $\alpha$ is a good idea?' Why is $C$ suspicious, but $A$ is not? Many other questions may still linger; for example 'how do we know we chose the right threshold', 'how do we know we are not mistaking an attacker-dominated cell with one behind a hill', 'how do we make sure we have taken all the factors into account', or '*can we do better*'?

We believe that we should directly use signal propagation data for this purpose. Leveraging patterns latent in the data will lead to more practical, robust and accurate solutions. The key intuition is to learn the propagation *behavior* of the signal from the observed signal propagation data (we will discuss the practicality of obtaining data later). There

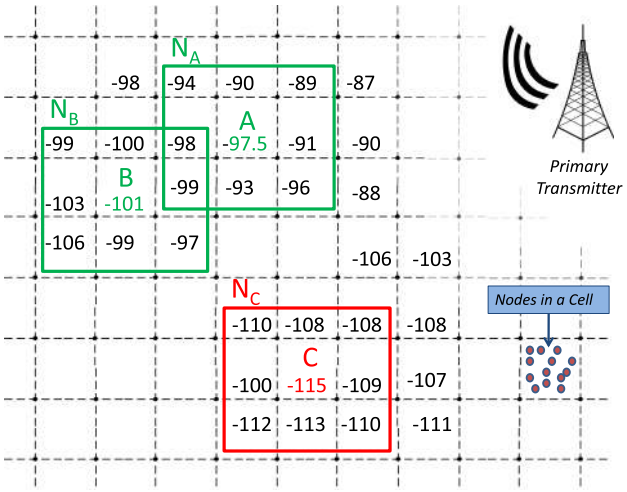


**Figure 1.** Sample grid with normal and attacker-dominated cells.

are patterns in which the signal propagates. We can extract these 'patterns' and utilize them to predict how we expect the signal to behave in the (often large) region of interest. Naturally, the actual behavior of the signal should be similar to what we can predict from the observed propagation patterns. This is mainly because we learn to predict the patterns of propagation from the signal itself. We claim that if the propagation of signal in a given location within the region of interest is not similar to patterns of signal propagation extracted from the same or 'similar' signals in the region of interest, the location should be considered suspicious or un-natural. As a simplistic example, assume we somehow learn that in a particular flat desert, digital TV signals weaker than -70dBm attenuate by at most 5dB every 5 kilometers. Then, a 10dB decrease or an 8dB increase over a three kilometer distance may be considered suspicious, or at least unusual.

We believe that we can spot unnatural propagation of signal in local neighborhoods of adjacent cells by carefully analyzing samples from the actual signals in the same and several different neighborhoods (within the region of interest) in the past. For a given neighborhood, we are now concerned with a new type of question. *Is the propagation of the signal natural in this neighborhood?* Before answering this question, we must define and show how to represent the pattern of signal behavior in a neighborhood of cells. So, we first address the following question. *How to represent the pattern of signal propagation in neighborhoods?*

### 3.1 Representation of Signal Propagation

In order to better understand the patterns by which the signal propagates, we need to define a way to represent them. We start by a simple representation as follows. We consider the *local neighborhood* of any cell $A$

to contain $A$ and its 8 neighboring cells. For example, in Figure 1 the local neighborhood for $A$, $B$, and $C$ are shown and referred to as $N_A$, $N_B$, and $N_C$ respectively. Using this definition for a local neighborhood, we represent a cell $A$ by a 9-element tuple containing the power level in $A$, and the difference in power between $A$ and the rest of the neighbors in a pre-specified order. For example the neighborhood for cell $A$, is represented as $\langle -97.5, -.5, 3.5, 7.5, 8.5, 6.5, 1.5, 4.5, -1.5 \rangle$. We call this the *neighborhood representation* of $A$. Note that the representation can be expanded to include, for example, the neighbors of neighbors of $A$ as well to provide additional context for learning patterns. However, as we will show later, the 9-tuple representation is sufficient for our purposes. This representation provides us with a way to encode the pattern by which the data propagates in this neighborhood. Using this definition for the neighborhood of a cell, our original question can be re-phrased as: *For a given cell, is the propagation of the signal natural in its local neighborhood?*

### 3.2 Using Patterns of Signal Propagation

Let us assume that we have access to reliable power measurements in all of the region of interest. An example for a region would be a 50km by 50km area with a roughly uniform (flat, hilly, etc.) terrain. It is easy to see that the data can be used to create one neighborhood representation for each cell in the region. We refer to each of such representations as an 'example.' Therefore, we can assume access to a large number of such examples representing the 'natural' propagation of signal in local neighborhoods. Also, for now assume that we are magically provided with the neighborhood representation for a sufficiently large and diverse set of 'un-natural' (attacker-dominated) cells.

Having access to representations for patterns of signal propagation as natural and un-natural examples, we believe the best way of approaching our question is to learn the common characteristics in each group and use it to differentiate between natural and un-natural examples. This means that by discovering the key characteristics of signal propagation patterns, we can superimpose a boundary in our space of representations. This boundary works as the decision making module. For a new example, we need to check which side of the boundary the example lies; the natural side or the un-natural side. This is a classic *classification* problem. We have now reduced our problem to a more specific question: *How to cast the problem of detecting attacker-dominated cells as a classification problem*. Before answering this question, we provide an analogy and the background on classification.

A useful analogy to this problem is that of spam detection in email systems: given a set of emails each marked as spam or normal, the goal is to learn the common characteristics among the normal emails, the common characteristics among the spam emails, and characteristics that differentiate between the two groups. Going back to our problem, we would like to discover a model that best describes the behavior of signal, and use it to make predictions about the normalcy of signal propagation in subsequent examples.

### 3.3 Background on Classification

Classification is one of the mainstreams of machine learning and has been widely adopted in many domains ranging from spam email detection [23] and unauthorized spectrum usage [35] to fraud detection [28], object detection [20], and speech recognition [43].

In a binary classification problem we are given a set of *training* examples with their corresponding *labels*, $(\overrightarrow{x_i}, y_i)$ where $\overrightarrow{x_i}$ is the representation of the $i^{th}$ example in the *feature space* and $y_i \in \{1, -1\}$ (yes or no?) is the corresponding binary label. Each example is described by a vector of its attributes which is often called the *feature vector*. For example, in detecting if a person has a significant risk of heart attack, the features can be the blood pressure, cholesterol level, and body mass index. The goal is to predict a binary label for an example for which we do not know the label (*a.k.a.* a *test* example) using the training examples [12]. In the heart attack example, we want to predict whether a person is under a certain risk of heart attack, given her feature vector. We do this by learning the patterns in the features of several different persons with and without the risk of the heart attack.

Looking underneath the surface, a classifier tries to partition the input feature space into regions where positive examples lie versus regions where negative examples lie. The boundary between regions for positive and negative examples is called the *decision boundary*. Training involves learning the decision boundary and classification involves determining on which side of the decision boundary a test example lies. In the simplest case, it is assumed that the decision boundary is a linear function of the input feature vector $\overrightarrow{x}$. Later, we relax this assumption and consider more complex decision boundaries. This linear function usually takes the form of

$$y(\overrightarrow{x}) = \overrightarrow{w}.\overrightarrow{x} + w_0 \tag{1}$$

where $\overrightarrow{w}$ is the *weight vector* and $w_0$ is the *bias* [12]. One might think about the decision boundary as a $(N-1)$-dimensional hyperplane in the $N$-dimensional feature space. The classification is done by determining the side of the hyperplane on which each point in the feature space lies. If $y(\overrightarrow{x}) \geq 0$ then $\overrightarrow{x}$ gets the label 1 and if $y(\overrightarrow{x}) < 0$ it gets the label $-1$.

### 3.4 Casting Attacker-dominated Cell Detection as a Classification Problem

We need to learn a classifier to predict whether a cell seems natural or not. To that end, we represent signal propagation in a local neighborhood of a cell, by the power average in the cell, as well as the 8 numbers representing the difference between the power averages of the cell with its neighbors. We denote these features by $\overrightarrow{x}$. To automatically discover these patterns we search for parameters $\overrightarrow{w}$ and $w_0$ that best explain the training data and provide reliable generalization properties. To be more specific, we are optimizing for $\overrightarrow{w}$ and $w_0$ that, if used for classification, provide the best prediction accuracy over the training data set while not overfit to it. More formally, the prediction of train set label $y$, which takes the form of $\overrightarrow{w}.\overrightarrow{x} + w_0$ should be similar to the actual train set label $y$. At the same time, to avoid too much fine tuning to the train set examples, the size (norm) of the weight vector $\overrightarrow{w}$ should be controlled. One drawback of this model is the assumption of linear separability. Our predictions are linear in the feature space, thus form a linear decision boundary. To be able to model non-linear decision boundaries, we project the data $\overrightarrow{x}$ to higher dimensional spaces where the decision boundaries are linear on that higher dimensional space. Our new predictions take the form of $\overrightarrow{W}.\Phi(\overrightarrow{x}) + W_0$ where $\Phi$ is a mapping to the higher dimensional feature space. We postulate that the decision boundaries in the feature space can be modeled more reliably by quadratic functions, thus modeling $\Phi$ by a quadratic kernel. To be more specific, we are solving the following optimization problem:

$$\min \frac{1}{2}\|\overrightarrow{W}\|^2 + \gamma \sum_{i=1}^{N} \xi_i \qquad (2)$$

$$\text{subject to } y_i(\overrightarrow{W}.\Phi(\overrightarrow{x}) + W_0) \geq 1 - \xi_i \quad \forall i$$

where $N$ is the number of training examples, $\xi_i$ is a collection of non-negative *slack variables* that account for possible misclassifications and $\gamma$ is the *tradeoff factor* between the slack variables and the regularization on the norm of the weight vector $\overrightarrow{W}$. The constraint in this minimization implies that we want our predictions, $\overrightarrow{W}.\Phi(\overrightarrow{x}) + W_0$, to be similar to labels $y_i \in \{1, -1\}$. The objective function works as a regularizer to avoid overfitting to the training data set. We solve this optimization by quadratic programming in dual. This is an example of SVMs [17].

The only parameter that needs to be estimated is $\gamma$. We estimate the $\gamma$ by cross validating it in the validation set, a part of train set which set aside for parameter estimation. This parameter is set using the data itself and there is no need of any assumption about data distribution.

Given a $\overrightarrow{W^*}$ and $\overrightarrow{W}_0^*$, which are the outputs of the Optimization 2, we can predict whether a cell is natural or not by looking at the sign of $\overrightarrow{W^*}.\Phi(\overrightarrow{x}) + W_0^*$.

**Data Collection.** The main remaining question is how to obtain the training examples needed to build the classifier. We argue that *normal* (negative) instances can be obtained in a practical *one-time* process based on a trusted sensor grid. By one-time we mean that in a particular region, we only need to collect signal propagation data once to build the classifier for that region. Once the classifier is built, it can be used forever (or until there is a significant environmental change in the region). A typical strategy for collecting this data is war-driving where a sensor is moved though the region collecting training data as it goes. This data can also be extrapolated by signal propagation models such as Longley-Rice, but our approach does not require the use of any such model. War-driving for collecting spectrum data is similar to the current practice of taking images for street-view capabilities of online map applications in Google and Bing.

An alternative may be realized in the context of 802.22 internet service for residences, as well as the envisioned application of white-spaces for advanced meter communications [18]. In this case, the (one-time) measurements may be collected at the time of deploying radios (meters) at each house by the operator. They may also be collected by a temporary sensor network developed for this purpose alongside the main CR network [42].

Once negative instances are collected, we use a methodology to inject *attacker-dominated* (positive) training instances to incorporate attacker-dominated cells containing attackers of varying degrees of sophistication. For further details please refer to Section 4.1.

### 3.5 A Unified Classifier for each Region

At this point, provided with labeled training examples for a transmitter, we are able to build a classifier that can predict if a cell is attacker-dominated or not. This is still far from practical for the following reasons. First, it requires training and maintaining a classifier for each transmitter. Second, as it will be concretely shown in Section 4, each transmitter may only provide a particular distribution of power levels in the region of interest. This leads to insufficient or non-existent training examples for some power levels, which can lead to low classification accuracy. Given enough training examples for a frequency range (*e.g.* 620-698 MHz for DTVs), we argue that our classifiers are capable of discovering decision boundaries in the feature space which are independent of the transmitter. This is due to the fact that signal propagation is mainly a function of power, propagation environment, and the frequency of transmission. From a practical perspective, this means that we do not need to learn a separate classifier for each transmitter in

the same frequency range. We show this property in Table 3 in the context of six DTV transmitters in Illinois.

We introduce the concept of *Unified Classifiers* that are trained by pooling data from multiple transmitters in such a way that there exist sufficient number of training examples at any power level in the power range of interest. For example for DTV transmitters this range will be between 90 dBm (maximum DTV transmission power) and -130 dBm (weakest signals considered). The new question we are facing is which transmitters to select so that we can ensure sufficient number of examples at any power level; the 'transmitter selection problem.' This problem can be reduced to the set covering problem, which is a well-known NP-Complete problem [47]. We divide the larger power spectrum of interest to a number of smaller power ranges and aim to enforce a lower bound on the number of examples per power range. Our goal is to select the minimum number of transmitters so that we are guaranteed to have at least a fixed number of examples per power range. We greedily select the transmitter that covers the largest number of uncovered power ranges at each stage. This is known to have an approximation ratio of $ln(n)+1$ where $n$ is the number of power ranges [47]. Having selected transmitters that cover the entire power range, we can now learn a classifier from the data from all selected transmitters. This is our unified classifier. We show that we can detect attacker-dominated cells for transmitters we never observe during training. This is of practical significance as one does not need to be concerned with providing information from all the transmitters in a frequency range, or those that may start transmission in future.

Another practical property of our unified classifier is its relative independence to the frequency. We later show that the unified classifier is not considerably sensitive to the frequency change in DTV transmitters in the UHF channels 14-51 ($470 - 698$MHz). This means that our unified classifier is capable of detecting attacker-dominated cells when trained with data from transmitters in different frequency ranges. Therefore, as we will show with evaluations for both Illinois and Pennsylvania, it is sufficient to build one classifier for the entire 470-698 MHz range.

Once the unified classifier detects a cell as compromised, the detection outcome in that cell should be reversed to cancel the attackers' misreporting effect. In cases where the actual power level is important, the power level should be replaced by the average powers reported by the majority of its neighboring cells. This strategy, which is motivated from image smoothing techniques in vision applications, has been validated in the context of white space networks [19]. This strategy, when combined with a multi-resolution deployment of CUSP enables nullifying the affect of attackers at different granularity levels, as well as those that are able to dominate multiple adjacent cells. Alternatively, in the case of using white-spaces for AMI communications, or 802.22 Internet, the firmware for the suspicious devices may be (physically or remotely) examined by the utility or 802.22 service provider.

# 4 Instantiating CUSP

In this section we show how CUSP can be instantiated in a region to provide protection against attacker-dominated cells when aggregating spectrum sensing reports at a central server. To that end, we provide general guidelines as well as specific details for an illustrative environment, namely East-central Illinois. Since it was not practical for us to do wardriving through this region we instead rely on the FCC and NASA databases and the Longley-Rice empirical outdoor signal propagation model to generate sensor data (see [5, 39] for more details). Longley-Rice is endorsed by FCC for determining propagation contours in the TV spectrum and takes into account the effects of terrain as well as transmitter's location, height, and power. For the purpose of these experiments we treat these models as the ground truth provided by sensors and use this to test our method. Note, however, that our method does not rely on any specific choice of a model. Hence if these models have some inaccuracies then we believe that accurate training data and proper application of CUSP will achieve the necessary foundation for integrity protections. We defer the experiments in which we account for additional variations and uncertainties in signal propagation to Section 5.

## 4.1 Environment and Data Collection

We start by considering a 160km $\times$ 160km square area in the flat Midwest area in the US. The following points in (*latitude, longitude*) format define the boundaries of the region: $\langle$ (39.56, -89.4), (41, -89.4), (41, -87.5), (39.56, -87.5) $\rangle$. The area is located in East-central Illinois and mainly consists of rural farmlands and a few small cities with populations under 100,000. Figure 2 depicts this area. We use registered DTV transmitter data from the FCC databases as well terrain data from the NASA database to build our grid-based crowdsourcing data. For any given location we can retrieve the list of nearby DTV transmitters as well as their properties such as channel (frequency), transmission power, and antenna height. We then combine this data with terrain data and use the Longley-Rice propagation model to estimate signal power from each of the DTV transmitters at that location.

**Cell Size and Density.** An important factor when using CUSP in any environment is the cell size and density of sensors (or wardriving samples). To make an informed decision about the cell size and sensor density, the following factors should be taken into consideration. First, the cell


**Figure 2.** Initial evaluation area and the first set of considered DTV transmitters in East-central Illinois.

size must be large enough that about 10 to 20 sensors exist in each cell. Mishra *et al.* [38] show that this many independent sensors provide as much collaborative gain as many more correlated sensors whose collaborative gain is limited by geographical correlation in shadowing. Second, the variation of average signal power in a cell must not be significant (*e.g.* less than 5dB) in order for combining individual reports to be meaningful. Using a similar criteria, Kim *et al.* [29] proposed a maximum radius of 5.6 km for a circular cell for detecting the TV transmitters at the edges of their contours. Third, collaborative sensing often works best when there exists independence in the (shadow) fading among different sensors. Using Godmunson's exponential shadow correlation decay model, it is shown that the maximum sensor density of 3.2 sensors/km$^2$ ensures independence between individual reports [29]. This factor, however, is more a recommendation than a requirement.

Considering an application such as advanced meter communications or 802.22 Internet, one may use the estimate of one sensor per house for spectrum sensing. To that end, we studied house density per square kilometer of the 102 counties in the state of Illinois from the US Census Bureau data [7]. The results show that the least dense county (Pope county) contains 2.5 houses/km$^2$. The 5th percentile of the data is 3.5 and median is 8.5 houses/km$^2$. In view of the discussion above, we opt for the following parameters. We consider base cells of size 2km$\times$2km with the average density of 3.2 sensors per km$^2$. We consider nodes to be uniformly distributed at random. It is known that the actual distribution of the sensors (houses in this case) may not be uniform in real-world, however, for the following reasons we argue that this assumption is reasonable for the evaluations. First, since we take conservative estimates for sensor density, it is likely that in most areas there exist more than the assumed 3.2 sensors per km$^2$. In such cases, the central server can choose from the existing nodes in order to create a relatively uniform distribution. Second, in the rare cases (given the conservative choice of density) that some cells contain less number of sensors, or sensors are closely clustered, the service provider may deploy additional sensing units. For each selected cell, we include the value of the cell's average power (*e.g.* -65) as well the difference of this cell with its immediate neighbors as the features for a normal example. Therefore, a normal example takes the form $\langle -65, 5, -2.5, 0.6, -3, 3, 2, -3, -1.2 \rangle$. Generating attacker instances is a non-trivial problem. The instances have to be general enough to train the classifier in such a way that it is able to detect attacks mounted using unknown strategies with varying fractions of attackers inside the cell. We opt for a randomized approach for generating attacker data in order to provide substantial variations in the training data. For uncoordinated attackers, we replace the actual power in the cell with Rand$(\lambda + 1, \lambda + 10)$ for exploitation attacks and with Rand$(\lambda - 1, \lambda - 10)$ for vandalism attacks, where Rand$(a, b)$ returns a random number between $a$ and $b$ and $\lambda = -114$ is the primary detection threshold. Similarly, we use Rand$(\lambda + 1, \lambda + 5)$, or Rand$(\lambda - 1, \lambda - 5)$ for coordinated attackers. For omniscient attackers, we simply replace the value with $\lambda + 1$ or $\lambda - 1$ for exploitation and vandalism attacks respectively. These attackers are knowledgeable and coordinated, and therefore they can only move the average exactly as much as needed to flip the detection outcome (1dB is the unit of measurements). This minimizes the attacker's chance of being detected.

### 4.2 Initial Evaluation

Of the tens of DTV transmitters in this area, we initially choose six DTV transmitters listed in Table 1 as a representative set. These transmitters are identified in Figure 2 as green antennas. This choice aims to serve two purposes; first, geographical diversity, and second, obtaining a wide range of received power levels across the area. Figure 3 represents the distribution of received signal powers from each of the six transmitters in the area. Later, we use the lessens learned in this section to perform a comprehensive analysis on other transmitters in the area of interest in Illinois, as well as all the transmitters that affect the area of interest in Pennsylvania.

In our first set of experiments, we consider each transmitter separately. For the labeled data for each transmitter, we perform '$K$-fold cross validation,' which is a commonly used technique to evaluate the performance of classifiers. We randomly partition the data into $K$ subsamples. Of the $K$ subsamples, a single subsample is retained as the test data for testing the model, and the remaining $K$ - 1 subsam-

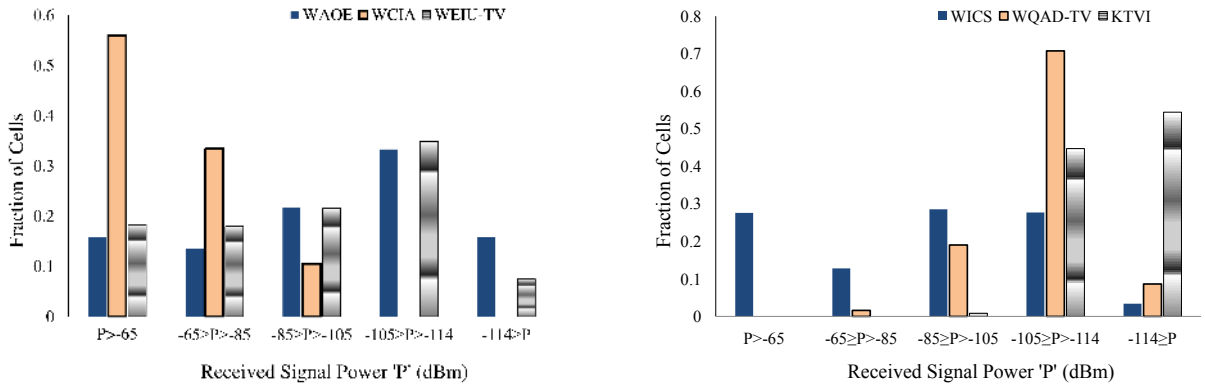


**Figure 3.** Distribution of received signal powers from six DTV transmitters in Illinois.

**Table 1.** Initially-selected DTV transmitters.

| Call Sign | Chan. | Fq. (MHz) | Tx Pow. (kW) |
|---|---|---|---|
| WAOE (MyN) | 39 | 620-626 | 151 |
| WCIA (CBS) | 48 | 674-680 | 1000 |
| WEIU-TV (PBS) | 50 | 686-692 | 255 |
| WICS (ABC) | 42 | 638-644 | 954 |
| WQAD-TV (ABC) | 38 | 614-620 | 1000 |
| KTVI (Fox) | 43 | 644-650 | 1000 |

ples are used as training data. The cross-validation process is then repeated $K$ times (the folds), with each of the $K$ subsamples used exactly once as the validation data. The K results from the folds then are averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. In our experiments we set $K = 10$. The results are summarized in Table 2. Note that these results are obtained with an equal mix among the three attacker models. We will provide further breakdown based on the attack-type later in this section.

### 4.3 Building a Unified Classifier

The results in Table 2 are obtained by considering each transmitter separately. From a practical perspective, it is ideal to use just one classifier. Such a classifier is trained by pooling data from multiple transmitters in a way that there exist sufficient number of training examples at any power level. According to CUSP's greedy method for transmitter selection (Section 3.5), we pick the data from WEIU-TV and KTVI for training the classifier. We test the classifier on the data from the other four transmitters. Table 3 summarizes the performance of the unified classifier. The important outcome is that the unified classifier trained with data from only two transmitters can perform very well on data from four other transmitters.

It is well-known that signal path loss is directly proportional to the logarithm of frequency [41]. However, the approach of considering a unified classifier appears to ignore the difference in path loss between different frequency channels. We argue that in practice, for the limited frequency ranges of our interest, this factor can be ignored in favor of other dominating factors such as the environment and terrain. We show this here and later when we consider a hilly urban/suburban area in Pennsylvania. The success of the unified classifier in detecting attackers in frequencies that differ from its training data (Table 3) only validates this assumption for DTVs in the channels 38-50 (614 - 692 MHz). Ideally it is best to have a unified classifier for up to 100 MHz of spectrum. For example, for the current UHF DTV channels in the US (Channel 14-50; 470-698 MHz), one may consider building three classifiers; one for approximately each 75 MHz of spectrum. However, due to practical considerations such as insufficient data or increased complexity, *we argue in favor of building only one classifier for the entire 470-698 MHz range.* To study this idea, we evaluate the effectiveness of our classifier, which is trained on data from the last third of the UHF DTV spectrum, for detecting attackers operating in frequencies near the first third of the spectrum. For this purpose, we consider the few DTV transmitters in this range in Table 4.

**Table 4.** Three DTV transmitters in the 400 MHz UHF channels.

| Call Sign | Chan. | Fq. (MHz) | Tx Power (kW) |
|---|---|---|---|
| KNLC (IND) | 14 | 470-476 | 891 |
| WAND (NBC) | 18 | 494 - 500 | 347 |
| WYIN (PBS) | 17 | 488-494 | 301 |

The performance of the unified classifier on this data is represented in Table 5. The results approve our statements about the unified classifier.

**Table 2.** Detection accuracy (D.A.) and false positive (F.P.) for six DTV transmitters in Illinois.

| | WAOE | | WCIA | | WEIU-TV | | WICS | | WQAD-TV | | KTVI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | D.A. (%) | F.P. (%) | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. |
| $P > -65$ | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | - | - | - | - |
| $-65 \geq P > -85$ | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 99 | 0 | - | - |
| $-85 \geq P > -105$ | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 99.8 | 0 | 100 | 0 |
| $-105 \geq P > -114$ | 99.1 | 2.2 | - | - | 99.8 | 4.8 | 99.7 | 2 | 99.8 | 1.5 | 98.7 | 2.9 |
| $-114 \geq P$ | 95.3 | 8.7 | - | - | 87.8 | 15 | 87.1 | 8.6 | 95.5 | 11.9 | 99.2 | 2.3 |
| **Overall** | **98.9** | **2** | **100** | **0** | **99** | **3** | **99.5** | **1** | **99.4** | **2.1** | **99** | **2.5** |

**Table 3.** Unified classifier's performance; detection accuracy (D.A.) and false positive (F.P.) for four DTV transmitter using the unified classifier trained with WEIU-TV and KTVI data.

| | WAOE | | WCIA | | WICS | | WQAD-TV | |
|---|---|---|---|---|---|---|---|---|
| | D.A. (%) | F.P. (%) | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. |
| $P > -65$ | 100 | 0 | 99.8 | 0 | 100 | 0 | - | - |
| $-65 \geq P > -85$ | 100 | 0 | 100 | 0 | 99.7 | 0 | 100 | 0 |
| $-85 \geq P > -105$ | 100 | 0 | 100 | 0 | 99.9 | 0 | 100 | 0 |
| $-105 \geq P > -114$ | 99.1 | .9 | - | - | 99.7 | 1.6 | 99.6 | .8 |
| $-114 \geq P$ | 97.3 | 3.2 | - | - | 97 | 2.4 | 95.1 | 7.6 |
| **Overall** | **99.3** | **.8** | **99.9** | **0** | **99.7** | **.5** | **99.3** | **1.3** |

**Table 5.** Unified classifier's performance; detection accuracy (DA) and false positive (FP) for three DTV transmitters in the 400 MHz UHF channels.

| | KNLC | | WAND | | WYIN | |
|---|---|---|---|---|---|---|
| | DA | FP | DA | FP | DA | FP |
| $P > -65$ | - | - | 100 | 0 | 100 | 0 |
| $-65 \geq P > -85$ | - | - | 100 | 0 | 100 | 0 |
| $-85 \geq P > -105$ | 100 | 0 | 100 | 0 | 99.9 | 0 |
| $-105 \geq P > -114$ | 98.8 | 3.4 | 100 | 1.2 | 98.3 | 9 |
| $-114 \geq P$ | 98.6 | 3.1 | - | - | 99.3 | 2.5 |
| **Overall** | **98.7** | **3.2** | **100** | **.1** | **99.2** | **3.3** |

**Effect of Attack-type.** In order to evaluate the effect of attack-type on the performance of our detection scheme, we create test datasets that only include normal examples and attacker examples of one type. We next evaluate these datasets using the unified classifier. We studied the four transmitters in Table 3. The results for WCIA were identical to the results reported earlier for all three attackers. This can be attributed to the fact that the data from this transmitter are mostly far away from $\lambda$ and mostly in the first three power brackets, where detection is very accurate and robust. For the other 3 transmitters, we observed that the results in the first two brackets ($P > -65$ and $-65 \geq P > -85$), are identical to the results in the third bracket ( $-85 \geq P > -105$ ). Therefore, we only report the results in the last three brackets for WAOE, WICS, and WQAD-TV. Figures 4 and 5 report detection accuracy and false positive rates for these transmitters. The results show decreased detection accuracy and increased false positive rates as the attackers gain more sophistication. Overall, the results show that our scheme performs well even against omniscient attackers.

## 5 Stress Test and Comparison

In this section, we extend the initial evaluations in the relatively flat and detection-favorable Illinois environment, to a particularly unfavorable one, *i.e.* urban/suburban areas in hilly Southwest Pennsylvania. To account for additional shadow fading and signal variations in urban/suburban environments (not represented by Longley-Rice), we probabilistically add extra variations to the predicted signal powers. In a subset of our evaluations, where we simulate wireless microphones to compare our work to the state-of-the art, we use the log-distance path loss and log-normal shadow-fading [41] to model signal propagation.

### 5.1 Hilly Urban/Suburban Area: Southwest Pennsylvania

In this section we evaluate the performance of CUSP when instantiated to a hilly urban/suburban area near Pittsburgh in Southwest Pennsylvania. We focus on signal from all DTV transmitters within 150 mile radius of this 20km by 20km area with estimated received powers higher than -130dBm. This results in a list of 37 DTV transmitters. As before, we use the Longley-Rice model to take into consideration the effect of terrain in signal propagation. In addition, in order to represent un-accounted fading and signal variations in urban/suburban environments, we supplement the data with Gaussian variations mean zero and standard deviation $\sigma$ (dB-spread) of up to 6dB. This is in line with the log-normal distribution model commonly used in this

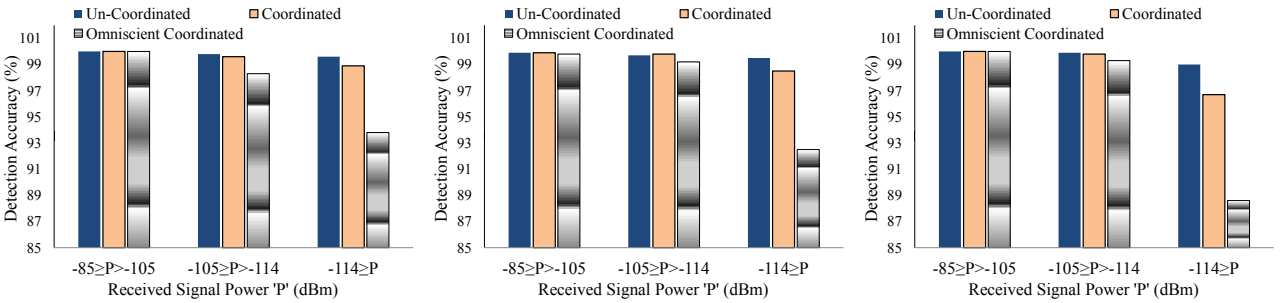


**Figure 4.** Detection accuracy classified by attacker-type for WAOE (left), WICS (center), and WQAD-TV (right).

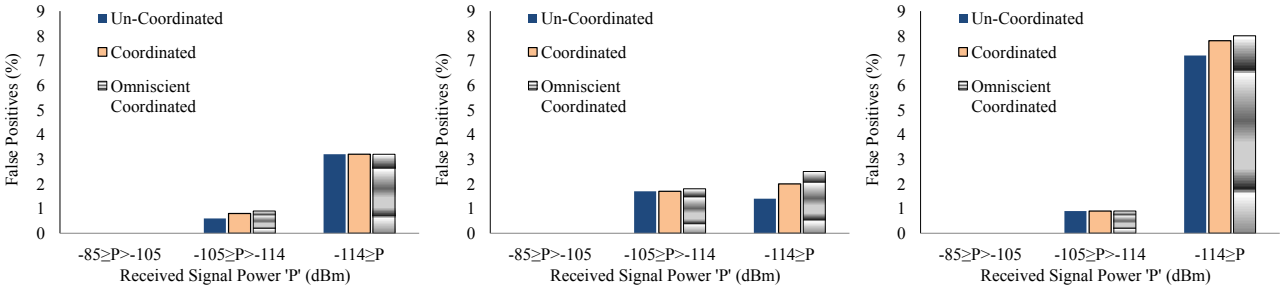


**Figure 5.** False positive rates classified by attacker-type for WAOE (left), WICS (center), and WQAD-TV (right).

context [46]. Figure 6(a) depicts the majority of transmitters affecting this area.

We pool the data from different frequencies to obtain a sufficiently large set of training and testing examples across all power levels. To evaluate the performance of CUSP in cases that it is not practical to use the algorithm in Section 3.5 to carefully choose the training data, we randomly divide the set of transmitters to subsets of size 29 and 8, for training and testing respectively. We call these 29-DTV and 8-DTV data. We train a unified classifier from the 29-DTV data, and test it on the 8-DTV data. The distribution of the received signal powers for training and testing data are provided in Figure 6(b). The cell sizes are 500m by 500m, resulting in a $40 \times 40$ grid of cells. The area is assumed to be populated with sensors at the density of 20 per km$^2$, which is achievable in suburban/urban areas. In particular, this is well below the average house density the Pittsburgh area [7]. The results before adding any additional variations are illustrated in the first column of Table 6.

**Training and Testing Under Different Conditions.** To test the classifier in an extremely unfavorable setting, we add Gaussian variations with mean 0 and standard deviation $\sigma$ to each power measurement in the test data. The classifier, however, remains trained with the data with no added signal variations. Table 6 summarizes the results. It can be seen that despite the significant amount of variation we added to signal propagation data, the classifier still performs reasonably well. As expected, the gradual degradation of performance is explained by the difference of examples that the classifier is trained with and those on which it is being tested. In particular, the relatively high false positive rates at high variation levels reflect the case that some of the variations seem 'too much' to the classifier, and therefore it mistakenly classifies them as malicious.

In general, the effectiveness of our approach can be reduced in environments with considerable natural variations in signal power within short distances. The reduced effectiveness presents itself as lower detection accuracy and higher false positive rates compared to environments in which signal propagation is 'smoother.' This is attributed to the descriptive power of our choice of features; there might be neighborhoods in which the classifier has difficulty differentiating between significant natural variations and an unusual signal propagation pattern created by the false reports of attackers. At a high-level, a remedy would entail modifying the feature space to increase its descriptive power. As an item of future work, we consider adding elevation data to the feature space to improve the classifier's performance (see Section 7). In addition, the cell-size may be optimized for maximized classifier performance.

**Effect of Attack-Type.** Table 6 provides results for an equal mix of the three attack-types (note that we assume each cell is occupied by attackers of one type only). Here, we break the results by the type of attack. Since the false

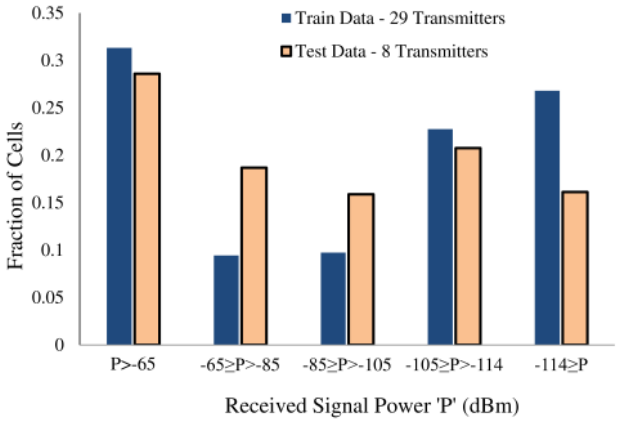


**Figure 6.** (a) Transmitters in parts of Southwest Pennsylvania / East Ohio. (b) Distribution of received signal for the training and testing data in Southwest Pennsylvania.

**Table 6.** Detection accuracy (D.A.) and false positive (F.P.) percentages when variations with dB-spread of $\sigma$ is added to test data from 8 DTVs. The classifier is trained with data from a disjoint set of 29 DTVs with no added variations.

| | Standard Deviation of Added Variations in Test Data | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\sigma = 0$ | | $\sigma = 2$ | | $\sigma = 4$ | | $\sigma = 6$ | |
| | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. | D.A. | F.P. |
| $P > -65$ | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 |
| $-65 \geq P > -85$ | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 |
| $-85 \geq P > -105$ | 99.8 | .5 | 99.9 | .5 | 99.8 | .8 | 99.8 | 1.5 |
| $-105 \geq P > -114$ | 92.7 | 6.8 | 92.2 | 8.3 | 91 | 12 | 89.2 | 17 |
| $-114 \geq P$ | 92.1 | 9 | 92.5 | 9.8 | 92.4 | 15 | 91 | 21 |
| **Overall** | 97.2 | 2.9 | 97.1 | 3.4 | 96.5 | 5.2 | 96.3 | 7.3 |

positive results are similar to those of Table 6, we only provide results for detection accuracy. The results are summarized in Table 7. It can be seen that the classifier provides respectable detection accuracies, even for the most difficult scenarios, that is defending against omniscient attacks in a hilly area with added variations of up to 6dB.

## 5.2 Comparison to State-of-the-Art

The closest piece of related work requires knowledge of the parameters of the log-normal shadowing model in order to detect compromised cells [19]. We are not able to evaluate that approach in our evaluation environment, since their approach only works with the assumption using the log-distance path loss and log-normal shadow-fading. In order to provide a fair comparison, we evaluate our approach in an environment similar to that of the related work. The signal power at node $N_i$ is written as $p_i = p_t - (10 \log_{10} r_i^\alpha + S_i)$ where $p_t$ is the transmit power of the primary, $r_i$ is the distance from $N_i$ to the primary transmitter, $10 \log_{10} r_i^\alpha$ represents the path loss with exponent $\alpha$ (typically $2 < \alpha < 4$), and $S_i \sim N(\mu_s, \sigma^2)$ is the loss due to shadow-fading. $\mu_s$ is often considered to be 0, and the dB-spread $\sigma$ independent of the distance to the transmitter (typically $2 \leq \sigma \leq 6$). Therefore we have $p_i \sim N(\mu(r), \sigma^2)$, where $\mu(r) = p_t - (10 \log_{10} r_i^\alpha + \mu_s)$.

Please note that the simulation setup and parameters are chosen based on the related work and we simply replicate them here on a larger scale. The simulation environment is an 8192m $\times$ 8192m area in which secondary users are deployed uniformly at random with the density of 0.0008 per square meter. The area is divided into $64 \times 64 = 4096$ square cells of size 128m $\times$ 128m each. Therefore, the expected number of nodes per cell is about 13. Depending on the scenario, primary transmitters with power ranging from 17dBm to 20dBm are placed at different locations in this area to represent wireless microphone primaries. The detection threshold is $\lambda = -74$dBm, $\alpha = 3$ and the standard deviation for the fading and shadowing process, $\sigma = 3$ (in dB scale).

The results are summarized in Table 8. It can be seen that our approach outperforms the outlier-based approach in terms of detection accuracy, however this comes at the cost of moderate false positive rates. Note that our approach does not use any information about the nature or specification of signal propagation model, whereas the out-

**Table 7.** Breakdown by attacker type; detection accuracy (D.A.) when variations with dB-spread of $\sigma$ is added to the test data from 8 DTVs. The classifier is trained with data from a disjoint set of 29 DTVs with no added variations. Uncoordinated, coordinated, and omniscient attacks are represented by UC, CO, and OM.

| | Standard Deviation of Added Variations in Test Data | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma = 0$ | | | $\sigma = 2$ | | | $\sigma = 4$ | | | $\sigma = 6$ | | |
| | Type of Attacker | | | | | | | | | | | |
| | UC | CO | OM | UC | CO | OM | UC | CO | OM | UC | CO | OM |
| $P > -65$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| $-65 \geq P > -85$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| $-85 \geq P > -105$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99 | 100 | 100 | 99 |
| $-105 \geq P > -114$ | 97 | 93 | 88 | 97 | 93 | 88 | 95 | 91 | 88 | 93 | 89 | 87 |
| $-114 \geq P$ | 92 | 87 | 84 | 92 | 87 | 84 | 91 | 85 | 84 | 89 | 85 | 84 |
| **Overall** | 98 | 96 | 95 | 98 | 96 | 95 | 97 | 96 | 95 | 97 | 95 | 94 |

lier detection approach requires knowledge of $\lambda$, $\alpha$, and primary powers. In addition, the related work requires setting two thresholds, which we set according to the authors' suggested values. In other words, we outperform that approach in a setting in which it had been designed and tuned.

**Table 8.** Outlier vs Non-linear SVM.

| | Fraction of Cells | **Outlier** | | **Non-Linear SVM** | |
|---|---|---|---|---|---|
| | | D.A. | F.P. | D.A. | F.P. |
| $P > -55$ | .02 | 81 | 0 | 100 | 0 |
| $-55 \geq P > -65$ | .04 | 95 | 0 | 100 | 0 |
| $-65 \geq P > -74$ | .14 | 67 | 0 | 95 | 7.4 |
| $-74 \geq P > -80$ | .29 | 85 | 0 | 96.7 | 7.6 |
| $-80 \geq P > -85$ | .30 | 99 | 0 | 100 | 0 |
| $-85 \geq P$ | .19 | 100 | 0 | 100 | 0 |
| **Overall** | 1 | 89.0 | 0 | 98.3 | 3.5 |

## 6 Related Work

Most prior work in the context of white space networks considers identifying individual attackers within a cell as part of collaborative sensing. Such approaches are not capable of detecting cells that are dominated by attackers. Min *et al.* [37] group sensors in a neighborhood to clusters (cells), and exclude or minimize the effect of abnormal sensor reports using shadow fading correlation-based filters. However, it fails to detect attackers that constitute more than $1/3$ of the population of the nodes in a cell. Kaligineedi *et al.* [27] address a similar problem by pre-filtering outlying sensing data, and a strategy to assign trust factors to nodes for weighting measurements and potentially omitting some nodes. In addition to the general problems enumerated with outlier-detection techniques, the attacker model is too simplistic and falls short in cases where attackers constitute a large fraction of nodes in a cell, or employ sophisticated misreporting strategies. Chen *et al.* [15] propose a weighted, reputation-based data fusion technique based on the sequential probability ratio test. Their approach only considers hard 0/1 decisions from each node, requires prior knowledge about the false positive and false negative ratios at each node, and cannot detect attacker-dominated regions.

The problem considered by Fatemieh *et al.* [19] is similar to ours. They start 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, their mechanism entails *corroboration* and *merging* of neighboring cells in a hierarchical structure to identify cells with outlier aggregates. Their solution, however, requires fairly accurate knowledge about the signal propagation formula and parameters. In addition, the paper does not introduce any systematic approach to tuning some of the detection thresholds for the distance-based outlier detection, and therefore it requires manual tuning.

Chen *et al.* [16] consider primary user emulation attacks in which an attacker may modify the air interface of a radio to mimic a primary transmitter signal's characteristics, thereby causing legitimate secondary users to erroneously identify the attacker as a primary user. They propose LocDef, which utilizes both signal characteristics and location of the transmitter to verify primary transmitter signals. An alternative is using cryptographic and wireless link signatures to authenticate primary users' signal in presence of attackers that may mimic the same signal. Liu *et al.* [36] achieve this by using a helper node close to a primary user to enable a secondary user to verify cryptographic signatures carried by the helper node's signals and then obtain the helper node's authentic link signatures to verify the primary users signals. We consider this problem to be complementary to the problem we address.

There exists a rich body of related work on this topic in the sensor network security literature. We consider the following to be the most relevant ones. Wagner introduced *resilient aggregation* [49], where he studies resilience of various aggregators to malicious nodes in an analytical framework based on statistical estimation theory and robust statistics. However, his work is limited to small regions and does not consider attacker detection as we do. Zhang *et al.* [50]

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. Hur *et al.* [26] 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.' They also do not consider uncertainties in the data. An avid reader may refer to following list for additional resources in the related are of secure data aggregation in wireless sensor networks [10, 14, 21, 24].

Insider attacker detection in wireless networks is another area of related work. This problem has been explored in a general setting [13, 25, 51] as well as more specific contexts such as insider jammers [31]. As an illustrative example in the general context of sensor networks, Liu *et al.* [34] 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.* [30] 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.

## 7 Conclusions

Aggregation of spectrum sensing data at base stations or spectrum availability databases plays a key enabling role in the deployment and success of white-space networks. This approach opens avenues for attacks in which devices falsely declare an occupied spectrum as available or vice versa. In this paper we presented CUSP, a new technique for detecting such attacks while aggregating spectrum sensing data from white space devices spanned over large regions. Our approach uses classification techniques based on SVMs with quadratic kernels to learn to differentiate between natural and un-natural signal propagation patterns in the region of interest. We evaluated the performance of CUSP using real-world transmitter, terrain, and sensor density data from two regions in the US. We showed that CUSP can achieve high detection accuracies even in the most unfavorable situations, *i.e.* hilly urban/suurban areas with significant amounts of additional signal uncertainty.

**Multi-Resolution Analysis.** In the future, we will enhance the approach to detect attacker-dominated cells at different resolutions. A high-resolution view entails dividing existing cells to smaller cells, whereas a low-resolution view allows for considering a set of neighboring cells as one cell. This enables detecting attackers at a fine level, coarse level, or those that are able to dominate multiple adjacent cells.

**Elevation Data as Features.** We will add elevation data as features to the training and testing data. This will provide the classifier with more information to learn and decide whether an observed signal propagation pattern is natural. Our preliminary experiments with this approach show improvements of performance in areas with irregular and hilly terrain.

## Acknowledgements

We thank the anonymous NDSS reviewers for their valuable comments. This work was supported in part by HHS 90TR0003-01, NSF CNS 09-64392, NSF CNS 09-17218, NSF CNS 07-16626, NSF CNS 07-16421, NSF CNS 05-24695, and grants from the MacArthur Foundation, Boeing Corporation, and Lockheed Martin Corporation. The views expressed are those of the authors only.

## References

[1] CogNea: Cognitive Networking Alliance. *http://www.cognea.org/*.

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

[3] FCC, Second Memorandum Opinion and Order, ET Docket No FCC 10-174, September 2010.

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

[5] Microsoft Research WhiteFi Service. *http://whitespaces.msresearch.us/*.

[6] S. 649: Radio Spectrum Inventory Act. *http://www.govtrack.us/congress/bill.xpd?bill=s111-649*.

[7] US Census Bureau. *http://www.census.gov*.

[8] CTIA: The Wireless Association Files Ex Parte to FCC to Request More Spectrum. *http://www.ctia.org/media/press/body.cfm/prid/1866*, 2009.

[9] 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.

[10] 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.

[11] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh. White space networking with wi-fi like connectivity. *SIGCOMM Comput. Commun. Rev.*, 39(4):27–38, 2009.

[12] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[13] 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.

[14] H. Chan, A. Perrig, B. Przydatek, and D. Song. Sia: Secure information aggregation in sensor networks. *J. Comput. Secur.*, 15(1):69–102, 2007.

[15] R. Chen, J.-M. Park, and K. Bian. Robust distributed spectrum sensing in cognitive radio networks. *IEEE INFOCOM 2008*, pages 1876–1884, April 2008.

[16] 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.

[17] C. Cortes and V. Vapnik. Support-vector networks. In *Machine Learning*, pages 273–297, 1995.

[18] O. Fatemieh, R. Chandra, and C. A. Gunter. Low cost and secure smart meter communications using the tv white spaces. *ISRCS '10: IEEE International Symposium on Resilient Control Systems*, August. 2010.

[19] O. Fatemieh, R. Chandra, and C. A. Gunter. Secure collaborative sensing for crowdsourcing spectrum data in white space networks. *DySPAN '10: IEEE International Dynamic Spectrum Access Networks Symposium*, April. 2010.

[20] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints), 2009.

[21] 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.

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

[23] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1-3):389–422, 2002.

[24] L. Hu and D. Evans. Secure aggregation for wireless networks. *SAINT-W '03: Proceedings of the 2003 Symposium on Applications and the Internet Workshops (SAINT'03 Workshops)*, page 384, 2003.

[25] 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.

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

[27] 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.

[28] H. Kim, S. Pang, H. Je, D. Kim, and S. Bang. Pattern classification using support vector machine ensemble. pages II: 160–163, 2002.

[29] H. Kim and K. G. Shin. In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? In *MobiCom '08: Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 14–25, New York, NY, USA, 2008. ACM.

[30] 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.

[31] 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.

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

[33] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust statistical methods for securing wireless localization in sensor networks. *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 12, 2005.

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

[35] S. Liu, Y. Chen, W. Trappe, and L. J. Greenstein. Aldo: An anomaly detection framework for dynamic spectrum access networks. In *INFOCOM*, pages 675–683, 2009.

[36] Y. Liu, P. Ning, and H. Dai. Authenticating primary users' signals in cognitive radio networks via integrated cryptographic and wireless link signatures. In *IEEE Symposium on Security and Privacy*, 2010.

[37] A. Min, K. Shin, and X. Hu. Attack-tolerant distributed sensing for dynamic spectrum access networks. In *Network Protocols, 2009. ICNP 2009. 17th IEEE International Conference on*, pages 294–303, Oct. 2009.

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

[39] R. Murty, R. Chandra, T. Moscibroda, and V. Bahl. Eliminating the need for low threshold spectrum sensing in white space networks. *Microsoft Research Technical Report MSR-TR-2010-127*, September 2010.

[40] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. *IPSN '04: Proceedings of the 3rd international symposium on Information processing in sensor networks*, pages 259–268, 2004.

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

[42] N. Shankar, C. Cordeiro, and K. Challapali. Spectrum agile radios: utilization and sensing architectures. *IEEE DySPAN '05*, pages 160–169, Nov. 2005.

[43] R. Solera-Ure na, D. Martín-Iglesias, A. Gallardo-Antolín, C. Peláez-Moreno, and F. Díaz-de María. Robust asr using support vector machines. *Speech Commun.*, 49(4):253–267, 2007.

[44] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell. Ieee 802.22: the first cognitive radio wireless regional area network standard. *Comm. Mag.*, 47(1):130–138, 2009.

[45] A. Taherpour, Y. Norouzi, M. Nasiri-Kenari, A. Jamshidi, and Z. Zeinalpour-Yazdi. Asymptotically optimum detection of primary user in cognitive radio networks. *IET Communications*, 1(6):1138–1145, 2007.

[46] 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.

[47] V. V. Vazirani. *Approximation Algorithms*. Springer, March 2004.

[48] 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.

[49] 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.

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

[51] 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.