Attack Detection in Wireless Localization

Yingying Chen, Wade Trappe, Richard P. Martin
{yingche,rmartin}@cs.rutgers.edu, trappe@winlab.rutgers.edu
Department of Computer Science and Wireless Information Network Laboratory
Rutgers University, 110 Frelinghuysen Rd, Piscataway, NJ 08854

***Abstract*— Accurately positioning nodes in wireless and sensor networks is important because the location of sensors is a critical input to many higher-level networking tasks. However, the localization infrastructure can be subjected to non-cryptographic attacks, such as signal attenuation and amplification, that cannot be addressed by traditional security services. We propose several attack detection schemes for wireless localization systems. We first formulate a theoretical foundation for the attack detection problem using statistical significance testing. Next, we define test metrics for two broad localization approaches: multilateration and signal strength. We then derived both mathematical models and analytic solutions for attack detection for any system that utilizes those approaches. We also studied additional test statistics that are specific to a diverse set of algorithms. Our trace-driven experimental results provide strong evidence of the effectiveness of our attack detection schemes with high detection rates and low false positive rates across both an 802.11 (WiFi) network as well as an 802.15.4 (ZigBee) network in two real office buildings. Surprisingly, we found that of the several methods we describe, all provide qualitatively similar detection rates which indicate that the different localization systems all contain similar attack detection capability.**

I. Introduction

Obtaining accurate positions of nodes in wireless and sensor networks is important because the location of sensors is a critical input to many higher-level networking tasks. Recent research efforts have resulted in a plethora of algorithms to localize sensor nodes, a fraction of which have been covered in recent surveys [1]–[3]. As more location-dependent services are deployed, they will increasingly become tempting targets for malicious attacks. Unlike traditional systems, the localization infrastructure is sensitive to non-cryptographic attacks, and these cannot be addressed using traditional security services. We have found that the performance of the localization algorithms degrades significantly under physical attacks, for example, when signals are attenuated, amplified, or reflected by an adversary [4].

Compromised localization results are a serious threat because of their impact on applications, and it is thus desirable to detect the presence of localization attacks. In this paper, we examine the problem of detecting attacks on wireless localization. We present a general formulation for attack detection using statistical significance testing and then build tests that are applicable to broad classes of multilateration and signal strength-based methods, as well as several other test statistics that are unique to a variety of different localization algorithms.

Multilateration is a popular localization approach that uses Least Squares (LS) techniques to perform localization [2], [5]–[8], and has the desirable property of supporting mathematical analysis, in part because LS-based regression has well-known statistical descriptions when operating near ideal conditions. By examining Linear Least Squares (LLS), we build a mathematical model and derive an analytic solution for attack detection using the residuals of an LLS regression. We show that attack detection using LLS is easy to conduct and is suitable for both single-hop and multi-hop ranging methods because it is independent of the ranging modality used by the localization system.

On the other hand, many signal strength based algorithms [3], [9] rely on either statistical inference or machine-learning in the context of scene matching to perform localization, and consequently do not yield closed-form solutions. However, for algorithms based on signal strength, we found that the minimum distance between an observation and the database of signal strength vectors is a good test statistic to perform attack detection. One key advantage of our approach for signal strength based methods is that the detection phase can be performed before localization.

To evaluate the effectiveness of our attack detection mechanisms we first present experimental results illustrating the feasibility of physical attacks on localization. We then conducted a trace driven evaluation using both an 802.11 (WiFi) network as well as an 802.15.4 (ZigBee) network in two real office buildings. In particular, we applied signal strength attenuation and amplification, using a linear attack model obtained from our experiments, to the Received Signal Strength (RSS) readings collected from these two office buildings. We evaluated the performance of our attack detection schemes using detection rates and receiver operating characteristic curves. Our experimental results provide strong evidence of the effectiveness of our attack detection schemes with high detection rates, over 95%, and low false positive rates, often under 5%. Surprisingly, we found that most of the attack detection schemes provide qualitatively similar performance. This shows that the different localization systems have similar attack detection capabilities.

The rest of the paper is organized as follows. Section II discusses previous research in localization and attack detection. We study the feasibility of attacks and present our experimental methodologies in Section III. We present our generalized theoretical formulation for the attack detection problem in Section IV. We next derive an analytic solution for attack detection using Least Squares in Section V. Using common features for attack detection in signal strength based algorithms is presented in Section VI. We study the test statistics that are specific to a variety of

different algorithms in Section VII. Then, we provide a discussion in Section VIII. Finally, we conclude in Section IX.

II. Related Work

There has been much activity towards developing localization systems for wireless and sensor networks. In this section, we first give a broad overview and then focus on activity related to secure localization.

Localization approaches can be categorized using various taxonomies. Range-based algorithms involve distance calculation to landmarks with known positions using the measurement of various physical properties like RSS [3], [9], Time Of Arrival (TOA) [5] and Time Difference Of Arrival (TDOA) [10]. Range-free algorithms use coarser metrics such as connectivity [11] or hop counts [6] to place bounds on node positions. Another classification method relates how a node is mapped to a location. Lateration approaches [2], [5]–[8], use distances to landmarks, while angulation uses the angles from landmarks. Scene matching strategies [3], [9], [12] use a function that maps observed radio properties to locations on a pre-constructed radio map or database. Finally, a third dimension of classification extends to aggregate or singular algorithms. Aggregate approaches [11], [13] use collections of many nodes in the network in order to localize (often by flooding), while localization of a node in singular methods only requires it to communicate to a few landmarks. In this paper, we focus our work on two broad localization mechanisms: multilateration and signal strength. Multilateration clearly applies to both single and multi-hop range-based approaches, while signal strength can be applied to a wider variety of both range-based and scene matching algorithms.

There has been considerably less work on the problem of ensuring the trustworthiness of wireless localization. Cryptographic threats on localization, such as device spoofing, can be addressed through traditional security services, e.g. authentication. Non-cryptographic threats, such as physically perturbing the environment, require different strategies. [14], [15] proposed distance bounding protocols for verification of node positions. [16] proposed the Verifiable Multilateration mechanism which is based on the distance bounding protocols for secure position computation and verification. [17] uses hidden and mobile base stations to localize and verify location estimates. [18] uses both directional antennas and distance bounding to achieve security. [7], [19], [20] tries to eliminate attack effects and still provide accurate localization. The closest works to this paper are [19], [21]. A general location anomaly detection scheme is described in [21] that relied on the neighbor information to detect inconsistencies. However, it assumes a highly dense network where the positions of the nodes follow a Gaussian distribution, which is contrary to the structure of many deployed systems where much lower densities are typical. Our proposed LLS approach is more general than the ARMMSE approach in [19]. Our approach provides a broader choices of detectors than that work.

Our work is unique in that we have formulated location attack detection as a general statistical significance testing problem. We showed how the test statistics come naturally out of the localization algorithms themselves without additional assumptions. In addition, our work differs from most previous research in that we experimentally validated our approaches using real networks deployed in two different buildings.

III. Feasibility of Attacks

In this section we provide background on how attackers can impact the localization system. We next discuss the feasibility of conducting these attacks on signal strength, and provide the experimental methodology that we use to evaluate our attack detection mechanisms later in this paper.

A. Localization Attacks

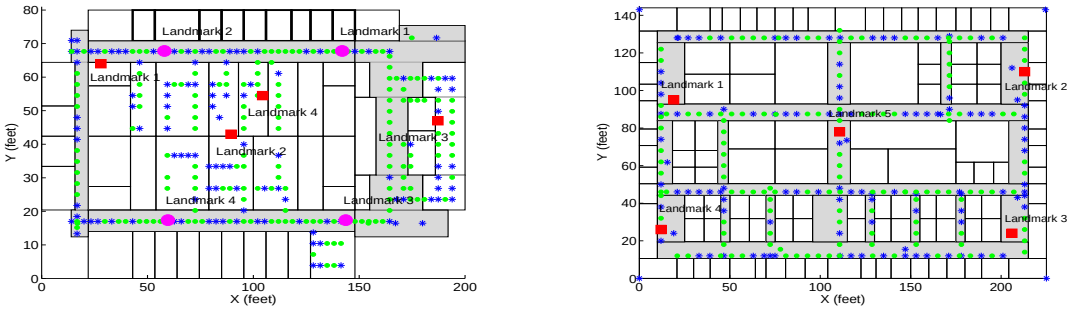
Localization mechanisms are built upon different ranging modalities, such as RSS, TOA, AOA, and hop count. These all rely on the measurement of the physical properties of the wireless system. Adversaries can apply non-cryptographic attacks against the measurement processes, bypassing conventional security services, and as a result can affect the localization performance. For example, wormhole attacks tunnel through a faster channel to shorten the observed distance between two nodes [22]. An attenuation attack would decrease the radio range, and thus potentially lengthen the hop-count. Compromised nodes may delay response messages to disrupt distance estimation [7]. RSS readings can be altered due to attenuation or amplification of the signal strength by an adversary [4]. A broad survey of the potential non-cryptographic attacks that are unique to localization can be found in [7].

B. Signal Strength Attacks

We choose to use RSS as the ranging modality for localization algorithms. An adversary can attack the wireless node directly or compromise the landmarks involved in localization by attenuating or amplifying the signal strength readings. Based on our experimental attacks using real materials, we will use the linear attack model [4] (i.e. a material causes a constant percentage power loss independent of distance) as shown in Figure 1 to describe the effect of an attack on the RSS readings at the wireless device or at the landmarks. As presented in the figure, these attacks are easy to conduct with low cost materials. The linear relationship implies that it is easy for an adversary to control the effect of an attack on the observed signal strength by appropriately selecting different materials.

C. Experimental Methodology

In order to study the generality of our attack detection approaches, we have conducted experiments in two office buildings, one is the 3rd floor of the Computer Science building at Rutgers University (CoRE) as shown in Figure 2 (a) and the other is in a floor of an industrial



(a) CoRE (b) Industrial Lab

Fig. 2. Layout of the experimental floor

research lab (Industrial Lab) as presented in Figure 2 (b). In Figure 2 (a), the experiments are performed for both an 802.11 (WiFi) network as well as an 802.15.4 (ZigBee) network. For the 802.11 (WiFi) network, there are 4 landmarks shown in red squares deployed in a collinear manner to maximize signal strength coverage. While for the 802.15.4 (ZigBee) network, there are 4 landmarks shown in magenta circles placed in a square set to maximize localization accuracy [23]. For experiments conducted in the industrial lab, as depicted in Figure 2 (b), we only used an 802.11 (WiFi) network with 5 landmarks. The small green dots are the localization testing points and the small blue stars are the training points. We will present the results of our experiments for each of the proposed attack detection cases in its associated section in this paper. Across all experiments, we have performed a trace-driven evaluation by either attenuating or amplifying RSS readings collected from these two buildings.

IV. Generalized Attack Detection Model

In this section we first propose a general formulation for the localization attack detection problem. We then introduce metrics for evaluating the effectiveness of our approaches.

A. Localization Attack Detection

In general, the error of a localization algorithm is defined as the distance between the true location $\mathbf{x} = [x, y]^T$ and the estimated location $\hat{\mathbf{x}}$, $D_{err} = \|\mathbf{x} - \hat{\mathbf{x}}\|$. We found in prior work that under physical attacks, the localization error D_{err} increases significantly [4]. However, D_{err} is not directly available during run-time, and the challenge in attack detection is to devise strategies for detecting localization attacks that do not use localization errors.

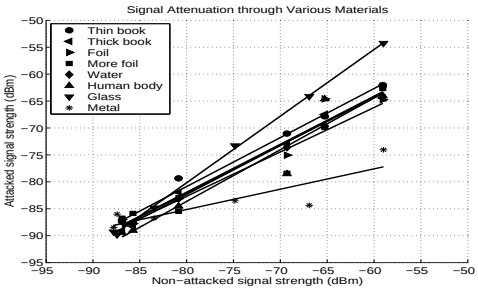


Fig. 1. Linear attack model on received signal strength for various media.

We propose to formulate location attack detection as a statistical significance testing problem, where the null hypothesis is

$$\mathcal{H}_0 : \text{normal (no attack).}$$

In significance testing, a test statistic $\mathbf{T}$ is used to evaluate whether observed data belongs to the null-hypothesis or not. For a particular significance level, α (defined as the probability of rejecting the hypothesis if it is true), there is a corresponding *acceptance region* Ω such that we declare the null hypothesis valid if an observed value of the test statistic $\mathbf{T^{obs}} \in \Omega$, and reject the null hypothesis if $\mathbf{T^{obs}} \notin \Omega$ (i.e. declare an attack is present if $\mathbf{T^{obs}} \in \Omega^c$, where Ω^c is the *critical region* of the test). In our attack detection problem, the region Ω and decision rule is specified according to the form of the detection statistic $\mathbf{T}$ (for example, when using distance in signal strength space for $\mathbf{T}$, the decision rule becomes comparison against a threshold), and rejection of the null hypothesis corresponds to declaring the presence of an attack.

B. Effectiveness

In order to evaluate the effectiveness of our attack detection methods, we will utilize the following performance metrics:

Cumulative Distribution Function (CDF): The CDF of the test statistic $\mathbf{T}$ provides the sensitivity of $\mathbf{T}$ under attack. Based on the CDF, we can study the feasibility of using $\mathbf{T}$ for attack detection.

Detection Rate (DR): An attack may cause the significance test to reject $\mathcal{H}_0$. We are thus interested in the statistical characterization of the attack detection attempts over all the localization attempts. The Detection Rate is defined as the percentage of localization attempts that are determined to be under attack, i.e.:

$$DR = \frac{N_{attack}}{N_{total}} \tag{1}$$

where N_{total} is the total number of localization attempts and N_{attack} is the number concluded under attack by detection. Note that when the signal is attacked, the detection rate corresponds to the probability of detection P_d, while under normal (non-attack) conditions it corresponds to the probability of declaring a false positive P_{fa}. We will examine DR as a function of the attack strength.

Receiving Operating Characteristic (ROC) curve: To evaluate an attack detection scheme we want to study the false positive rate P_{fa} and probability of detection P_d together. The ROC curve is usually used to measure the tradeoff between false-positives and correct detections. The ROC curve is a plot of attack detection accuracy against the false positive rate. It can be obtained by varying the detection thresholds.

V. Using Least Squares

In this section we provide mathematical analysis for attack detection in multilateration algorithms. We first provide background in using LS to perform localization. Next, based on the properties of the LLS estimator, we define an attack detection scheme that utilizes regression residuals, and give an analytic formulation to specify the acceptance region Ω. Finally, the experimental results are presented to evaluate the effectiveness of the detection scheme.

A. Localization

To perform localization with LS requires 2 steps: ranging and lateration.

Ranging Step: Recent research has seen a host of variants on the ranging step such as RSS, TOA, TDOA, and hop count. Our attack detection approach works with any ranging modality.

Lateration Step: From the estimated distances d_i and known positions (x_i, y_i) of the landmarks, the position (x, y) of the localizing node can be found by finding $(\hat{x}, \hat{y})$ satisfying:

$$(\hat{x}, \hat{y}) = arg \min_{x,y} \sum_{i=1}^{n} [\sqrt{(x_i - x)^2 + (y_i - y)^2} - d_i]^2 \quad (2)$$

where n is the total number of landmarks. We call solving the above problem *Nonlinear Least Squares*, or NLS. Solving the NLS problem requires significant complexity and is difficult to analyze. We may approximate the NLS solution and linearize the problem [23] into the system $\mathbf{Ax} = \mathbf{b}$, where:

$$\mathbf{A} = \begin{pmatrix} x_1 - \frac{1}{n}\sum_{i=1}^{n} x_i & y_1 - \frac{1}{n}\sum_{i=1}^{n} y_i \\ \vdots & \vdots \\ x_n - \frac{1}{n}\sum_{i=1}^{n} x_i & y_n - \frac{1}{n}\sum_{i=1}^{n} y_i \end{pmatrix} \quad (3)$$

and

$$\mathbf{b} = \frac{1}{2} \begin{pmatrix} (x_1^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2) + (y_1^2 - \frac{1}{n}\sum_{i=1}^{n} y_i^2) \\ -(d_1^2 - \frac{1}{n}\sum_{i=1}^{n} d_i^2) \\ \vdots \\ (x_n^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2) + (y_n^2 - \frac{1}{n}\sum_{i=1}^{n} y_i^2) \\ -(d_n^2 - \frac{1}{n}\sum_{i=1}^{n} d_i^2) \end{pmatrix}. \quad (4)$$

Note that $\mathbf{A}$ is described by the coordinates of landmarks only, while $\mathbf{b}$ is represented by the distances to the landmarks together with the coordinates of landmarks. We call the above formulation of the problem *Linear Least Squares*, or LLS. The estimate of $\mathbf{x} = [x, y]^T$ is done via

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \quad (5)$$

In addition to its computational advantages, the LLS formulation allows for tractable statistical analysis, as we shall now see.

B. The Residuals

In practice, there are estimation errors from the ranging step. The LLS formulation can be refined as a linear regression, $\mathbf{b} = \mathbf{Ax} + \mathbf{e}$, where $\mathbf{e}$ corresponds to model errors. The localization result is then $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, and the fitted values $\hat{\mathbf{b}}$ corresponding to the observed values $\mathbf{b}$ are given by

$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}[(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}] = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \quad (6)$$

Further, we define the vector of residuals $\hat{\mathbf{e}}$ as

$$\hat{\mathbf{e}} = \mathbf{b} - \hat{\mathbf{b}} = [1 - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T]\mathbf{b}. \quad (7)$$

When the regression model is performing well we may assume that the model errors are Gaussian [24], [25]. Under this assumption, the residuals also follow a Gaussian distribution, $\mathbf{N}(\mu, \mathbf{\Sigma})$, since the residuals are a linear combination of the elements of $\mathbf{b}$ and $\mathbf{e}$. Here, μ is the mean vector and $\mathbf{\Sigma}$ is the covariance matrix. We choose the residuals $\hat{\mathbf{e}}$ as the test statistic $\mathbf{T}$, and will build our attack detection scheme by using the statistical properties of $\hat{\mathbf{e}}$ when LLS is operating in a desirable performance regime.

C. The Detection Scheme

The LLS attack detection is performed after localization. The residuals are correlated Gaussian random variables and the multivariate Gaussian distribution of $\hat{\mathbf{e}}$ can be expressed as:

$$f(\hat{\mathbf{e}}) = \frac{1}{(\sqrt{2\pi})^n |\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\hat{\mathbf{e}}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}}-\mu)}. \quad (8)$$

In order to determine whether the location result is compromised by adversaries, we perform attack detection through significance testing. We can define an acceptance region in $\hat{\mathbf{e}}$ space by

$$\Omega = \{\hat{\mathbf{e}} : Pr(\{\mathbf{T} : (\mathbf{T} - \mu)^T\mathbf{\Sigma}^{-1}(\mathbf{T} - \mu) > (\hat{\mathbf{e}} - \mu)^T\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}} - \mu)\}) > \alpha\}.$$

In practice, after performing localization using LLS, we have an observed value of residuals $\hat{\mathbf{e}}^{\mathbf{obs}}$. Testing the null hypothesis, we can decide that the localization is under attack if the probability $P = 1 - M < \alpha$, where

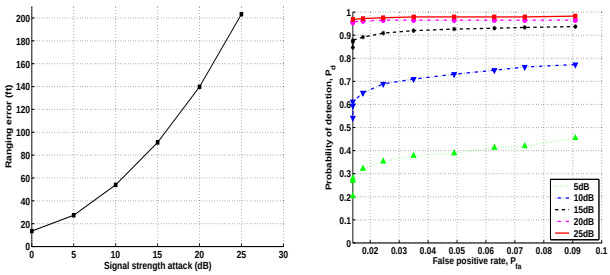
$$M = \frac{1}{(\sqrt{2\pi})^n |\mathbf{\Sigma}|^{\frac{1}{2}}} \int ... \int_E e^{-\frac{1}{2}(\hat{\mathbf{e}}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}}-\mu)} d\hat{e}_1 ... d\hat{e}_n \quad (9)$$

and E is the integration region defined by $(\hat{\mathbf{e}} - \mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}} - \mu) \leq X^2$ with

$$X^2 = (\hat{\mathbf{e}}^{\mathbf{obs}} - \mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}}^{\mathbf{obs}} - \mu).$$

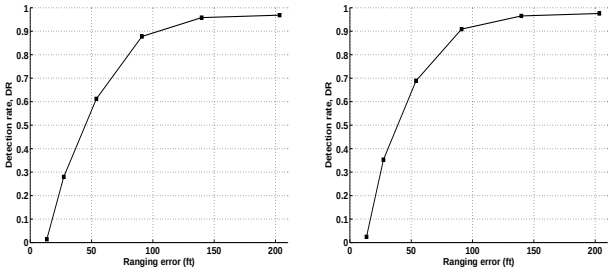
We can express the term

$$\begin{aligned} (\hat{\mathbf{e}} - \mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{e}} - \mu) &= (\hat{\mathbf{e}} - \mu)^{\mathbf{T}}\mathbf{D}^{\mathbf{T}}\mathbf{D}(\hat{\mathbf{e}} - \mu) \\ &= (\mathbf{D}(\hat{\mathbf{e}} - \mu))^{\mathbf{T}}(\mathbf{D}(\hat{\mathbf{e}} - \mu)) \\ &= \mathbf{y}^{\mathbf{T}}\mathbf{y}. \end{aligned} \quad (10)$$



(a) Ranging errors (b) ROC curves

Fig. 3. CoRE 802.11: (a) Ranging errors under the signal strength attacks (b) LLS residuals: Receiver Operating Characteristic (ROC) curves



(a) LOA: $\alpha = 0.01$ (b) LOA: $\alpha = 0.05$

Fig. 4. CoRE 802.11, LLS residuals: effectiveness of attack detection

Substituting $\mathbf{y} = \mathbf{D}(\hat{\mathbf{e}} - \mu)$ into Equation (9), we get

$$M = \frac{1}{(\sqrt{2\pi})^n} \int ... \int_{E'} e^{-\frac{1}{2}\mathbf{y}^\mathbf{T}\mathbf{y}} dy_1 ... dy_n \qquad (11)$$

with E' defined by $\mathbf{y}^\mathbf{T}\mathbf{y} \leq X^2$. Based on the calculation of the integral (see the Appendix), we get

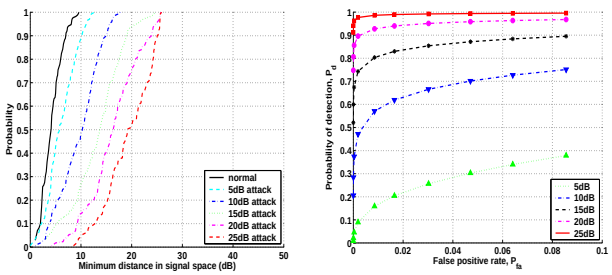
$$M = \frac{\Gamma(n/2, X^2/2)}{\Gamma(n/2)} \qquad (12)$$

where Γ is the incomplete gamma function. We then further obtain the probability by $P = 1 - M$. Based on the definition in Section IV, if the probability is sufficiently low, i.e. $P < \alpha$, then $\hat{\mathbf{e}}^{\mathbf{obs}}$ belongs to the critical region Ω^c and we can conclude that the location result is under attack.

D. Experimental Evaluation

In this section we present the evaluation of the effectiveness of the attack detection scheme. We chose RSS as the ranging modality and performed signal strength attacks according to the experimental methodologies described in Section III.

The average ranging error as a function of the severity of signal strength attacks is shown in Figure 3(a). We know that the relationship between the RSS error and the ranging error is multiplicative with distance [23]. Even small random perturbation in RSS readings can cause large ranging errors due to this multiplicative factor. We observed this effect in Figure 3(a); the ranging error increases super-linearly to attack severity. Figure 4 presents DR vs. the ranging errors when tested against significance level $\alpha = 0.01$ and $\alpha = 0.05$. We found that under a normal situation, where the ranging errors are less than 15 feet, the false alarm probability, P_{fa}, is less than 1.5% and 2.5% for $\alpha = 0.01$ and $\alpha = 0.05$ respectively. Large signal strength



(a) CDF (b) ROC curves

Fig. 5. CoRE 802.11: (a) Cumulative Distribution Function (CDF) of minimum distance D_s in signal space. (b) Minimum distance D_s: Receiver Operating Characteristic (ROC) curves

attacks, greater than 15dB, can cause ranging errors larger than 90 feet, and then the detection rates are more than 90%. These results strongly indicate that using residuals in LS as a test statistic for attack detection is effective.

Further, the ROC curves in Figure 3(b) show that for false positive rates less than 10%, the detection rates are above 90% and close to 99% when the attack strength increases to 20dB and 25dB. This shows that if the adversary wants to cause a large localization error, it is almost certain that our attack detection mechanism will detect it. For small attacks of less than 5dB, the detection rates are about 40%. In this case, it is difficult to distinguish whether the anomaly in the test statistic is caused by attacks or by measurement errors since the RSS readings can fluctuate around 5dB due to environmental effects. However, for such small attacks, because the resulting impact on the final localization result was shown to be small [4], the consequences of failing to detect such attacks would likely be small as well.

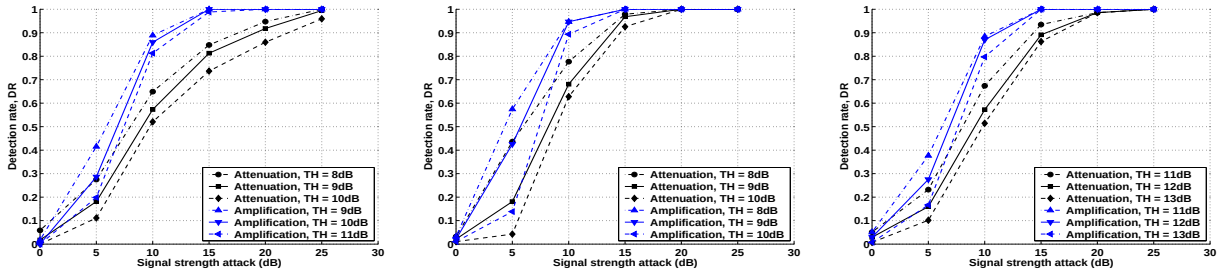
VI. Distance In Signal Space

RSS is a common physical property used by a widely diverse set of algorithms. For example, most scene matching approaches utilize the RSS, e.g. [1], [9], and many multilateration approaches [26] use it as well. In spite of its several meter-level accuracy, using the RSS is an attractive approach because it can re-use the existing wireless infrastructure — this feature presents a tremendous cost savings over deploying localization-specific hardware. In this section, we thus derive an attack detection scheme applicable to any signal strength based localization system.

A. Overview

All of the above algorithms take a vector $\mathbf{s}$ of n RSS readings to (or from) n landmarks for the node to be localized. Note that $\mathbf{s}$ corresponds to a point in a n-dimensional signal space [4]. Under normal conditions, the RSS vectors obtained from the physical positions in a floor form a surface S in the n-dimensional signal space; we can think of this surface as comprising 'valid' points in signal space. Due to measurement noise, multipath effects, and unknown biases, $\mathbf{s}$ will fluctuate around this idealized RSS surface.

A localization attacker would perturb the correct $\mathbf{s}$ to produce a corrupted n-dimensional RSS vector $\mathbf{s}'$. In signal



(a) CoRE 802.11 (b) CoRE 802.15.4 (c) Industrial 802.11

Fig. 6. Minimum distance in signal space $\mathbf{D_s}$: attack detection across different networks and buildings.

space, $\mathbf{s}'$ will be moved away from the ideal surface constructed by the correct RSS vectors. The stronger the attack, the more likely the vector $\mathbf{s}'$ will be distant from the RSS surface. We thus choose the minimum distance to the surface S, i.e. $D_s = \min\{\|\mathbf{s}' - \mathbf{s}^j\| : \text{where } \mathbf{s}^j \in S\}$, as the test statistic for signal strength based attack detection. The key advantage of this approach is that the attack detection is independent of the localization algorithms and can be performed before the localization process.

Although it is possible to devise a statistical model for D_s based on models for normal measurement errors, in this section we shall take a different approach and apply empirical methodologies from training data to determine thresholds for defining the critical region.

B. Finding Thresholds

Choosing an appropriate threshold τ will allow the detection scheme to be robust to false detections. In order to obtain the thresholds, we don't need to know the exact RSS surface in the signal space (in practice, it is hard to determine and exhibits discontinuities due to wall boundaries). Instead, we can obtain the thresholds through empirical training. During the offline phase, we can collect the RSS vectors for a set of known positions over the floor and construct a radio map. During the localization phase, we get an observed vector $\mathbf{s^{obs}}$, and we can then determine whether the $\mathbf{s^{obs}}$ is being attacked by calculating the D_s using the pre-constructed radio map.

We define that if

$$D_s > \tau, \tag{13}$$

the signal strength readings are under attack. We use the distribution of the training data to help decide on the thresholds. Figure 5 (a) shows the CDF of the D_s in signal space. We found that the curve of D_s shifted to the right under signal strength attacks, especially for larger attacks, thereby suggesting that we can use D_s as a test statistic for detecting attacks, and also that we can use the non-attacked CDF to obtain τ for a given α value.

C. Experimental Evaluation

We next present the evaluation of the effectiveness of using minimum distance D_s for attack detection. Figure 6 presents the Detection Rate under different threshold (TH) levels as a function of signal strength attacks for both the 802.11 and the 802.15.4 networks in CoRE and the 802.11 network in the Industrial Lab. Figure 5 (b) is the corresponding ROC curves under signal attenuation attacks for the 802.11 network in CoRE. We found that, in general, the effectiveness of the attack detection scheme is similar across the different networks and buildings. Interestingly, we found that the performance of the attack detection scheme under signal amplification attacks is uniformly better than those for signal attenuation attacks, although the shapes of the DR curves are similar. Because of the higher detection rates under amplification attacks, we do not present additional amplification results in the remainder of the paper. All these results are highly encouraging because they show our methods are quite general and do not depend on a specific network or environment.

Further, we observed that the DR under the 802.15.4 network in CoRE outperformed the DR under the 802.11 networks in both CoRE and Industrial Lab for the signal attenuation attacks as well as the signal amplification attacks. For attack strengths of 15dB or larger, the DR in the 802.15.4 network is over 95% and equals 100% when attack severity reaches 20dB and larger. We believe that the better landmark placement for localization [23] of the 802.15.4 network can account for its higher detection rates, although further investigation of this effect is required.

VII. Other Test Statistics

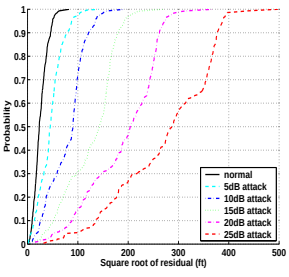
In this section, we examine algorithm-specific test statistics, which use properties specific to a particular localization algorithm. We have chosen a representative set of diverse algorithms. For the multilateration category, we investigate the NLS algorithm, while for signal strength based algorithms, we study both Area Based Probability (ABP) and Bayesian Networks (BN) algorithms. Detailed descriptions of these can be found in [3], [23], [26].

A. Nonlinear Least Squares (NLS)

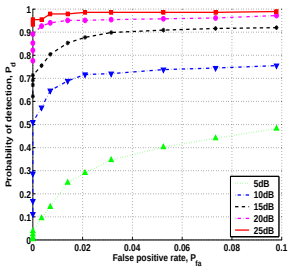
As presented in Section V, NLS is a multilateration algorithm that tries to satisfy the condition shown in Equation (2). The estimated $(\hat{x}, \hat{y})$ is the solution that minimizes the Sum of Squared Errors $\mathcal{E}^2$:

$$\mathcal{E}^2 = \sum_{i=1}^{n} [\sqrt{(x_i - \hat{x})^2 + (y_i - \hat{y})^2} - d_i]^2. \tag{14}$$

We define a test statistic $\mathcal{E} = \sqrt{\mathcal{E}^2}$ because $\mathcal{E}$ will likely increase under the attack. The CDF of $\mathcal{E}$ presented in Figure 7 (a) confirms that the $\mathcal{E}$ grows rapidly with the attack severity. Figure 7 (b) and Figure 8 show that the



(a) CDF



(b) ROC curves

Fig. 7. CoRE 802.11, NLS: (a) Cumulative Distribution Function (CDF) of $\mathcal{E}$. (b) $\mathcal{E}$: Receiver Operating Characteristic (ROC) curves

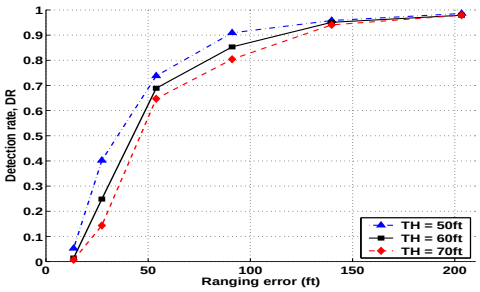


Fig. 8. CoRE 802.11, NLS using $\mathcal{E}$: effectiveness of attack detection

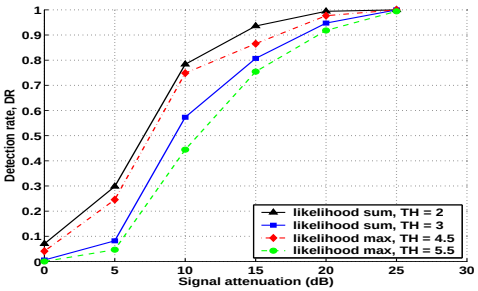


Fig. 9. CoRE 802.11, ABP: effectiveness of attack detection

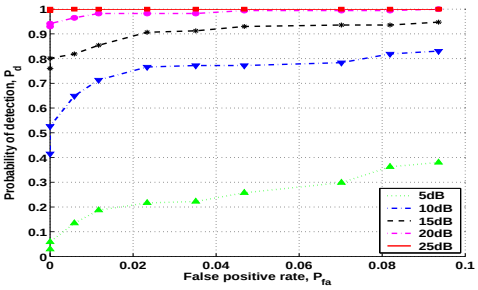


Fig. 10. CoRE 802.11, ABP: Receiver Operating Characteristic (ROC) curves

performance of attack detection when using $\mathcal{E}$ for the 802.11 network in CoRE is comparable to that using residuals in Section V. The thresholds are also obtained from training.

B. Area Based Probability (ABP)

Turning to signal strength based algorithms, ABP is an area-based algorithm that uses Bayes' Rule to return an area which has the highest likelihood of capturing the true location [3]. ABP divides the floor into a set of tiles. The total likelihood that the wireless node resides at each tile is calculated using:

$$P = \prod_{i=1}^{n} P_i \tag{15}$$

where n is the total number of landmarks and P_i is the likelihood of observing the measured RSS reading at landmark i which is usually modeled as a Gaussian random variable. The total likelihood is calculated at each tile, and the returned location estimation is either a region whose likelihood is above a certain level, or is the tile with the maximum likelihood.

When under attack, the corrupted RSS readings reduce the set of likely positions on the floor to localize a node. We found that the highest tile-likelihood denoted as $likelihood_{max}$ decreases significantly under attack, as well as the sum of the likelihoods over all the tiles, $likelihood_{sum}$. We explored both $likelihood_{sum}$ and $likelihood_{max}$ as test statistics. The thresholds are learned from the training data by taking the negative log of the values of the highest likelihood and the sum of the likelihoods.

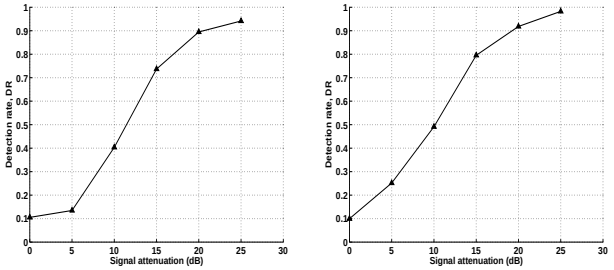
The effectiveness of using $likelihood_{sum}$ and $likelihood_{max}$ for attack detection in ABP are presented in Figure 9 and Figure 10. We found that using $likelihood_{sum}$ under threshold equal to 2 had better performance than others in detecting larger attacks, but on the other hand resulted in slightly higher false positive rates around 7%.

C. Bayesian Networks (BN)

Another representative signal strength based algorithm, BN, uses Bayesian networks [26]. BN uses Bayesian statistical inference to predict the probability distribution of the unknown positions. BN uses a Monte-Carlo sampling technique (Gibbs sampling) to compute the full joint-probability distribution for not just the position coordinates, but also for every random variable in the Bayesian network. Without an attack, the contribution from each landmark to the full joint-probability distribution is almost uniform. Under an attack, we found that the contribution from each landmark can become significantly reduced as the attack severity increases. Thus, we can use the fraction of contribution to the joint probability as a test statistic in BN.

Another method we explored is to use the probability likelihood because the conditional probability distribution of the coordinates in BN relies on the prior and the likelihood. We observed that under an attack, the value of the likelihood became significantly smaller. During the sampling process, the calculation of the likelihood uses the same approach as in Equation (15). Because the absolute value of the likelihood is very small, we take the negative log of the likelihood and use it as a test statistic for attack detection in BN.

Figure 11 shows the effectiveness of using the fraction of contribution and the likelihood for attack detection in BN. The detection rates are over 90% for attack strength of 20dB or larger. The false positive rates are about 10%. Comparing the absolute performance of these two methods with the other schemes we proposed in this paper, the performance of these two methods is qualitatively worse.



(a) (b)

Fig. 11. CoRE 802.11, BN: (a) Using fraction of contribution of each landmark for attack detection with threshold = 0.15. (b) Using likelihood in Bayesian inference for attack detection with threshold = 0.25.

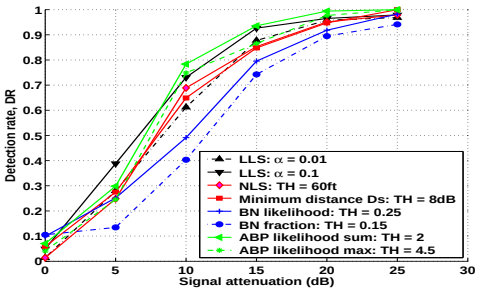


Fig. 12. CoRE 802.11: Comparison between generic and specific test statistics for attack detection.

VIII. Discussion

Comparing all of our detection schemes, Figure 12 shows the DR as a function of the signal attenuation attacks for the 802.11 network in the CoRE building. Surprisingly, we found that the performance of all the schemes provided qualitatively similar detection rates, although utilizing the residuals in LLS and the sum of likelihoods in ABP slightly outperformed the others, while using the fraction of contribution and the likelihood in BN underperformed the others.

Based on these similar performance characteristics, it is advantageous to use the minimum distance in the signal space D_s for signal strength based algorithms. Since the attack detection can be performed prior to the localization process and thus results in localization computation cost savings under attack. Additionally, the attack detection performance under the 802.15.4 network when using D_s outperforms the 802.11 network with 100% detection rate for large attacks as shown in Figure 6.

Moving to examine the relationship between attack detection and localization error, Figure 13 shows the DR when using residuals in LLS for attack detection, and the localization errors under the corresponding signal attacks with different localization algorithms. The figure shows that detection rates are more than 90% for attack strength equal to or greater than 15dB, and at this attack strength the average localization error is about 35ft.

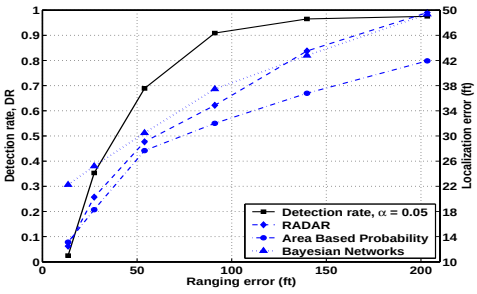


Fig. 13. CoRE 802.11: Relationships among Detection Rate (DR), ranging error, and localization error.

The above result is quite encouraging, as it shows that an attacker cannot cause gross localization errors without there being a very high probability of detection (>95%). In the case of RSS, with mean errors of 10-15 ft [3], an attacker can not cause errors of about 2-3 times over the average error without a very high probability of detection. Even for detection rates as low as 50%, the attacker's position error is limited to about 20 ft.

IX. Conclusion

In this work, we analyzed the problem of detecting non-cryptographic attacks on wireless localization. We proposed a theoretical foundation by formulating attack detection as a statistical significance testing problem. We then concentrated on test statistics for two broad localization approaches: multilateration and signal strength. For multilateration that uses Linear Least Squares, we derived a closed-form representation for the attack detector. Further, for localization schemes that employ signal strength, we showed that by utilizing the signal strength as a common feature, the minimum Euclidean distance in the signal space can be used as a test statistic for attack detection independent of the localization process. Further, we derived additional test statistics for a selection of representative localization algorithms.

We studied the effectiveness and generality of our attack detection schemes using a trace-driven evaluation involving both an 802.11 (WiFi) network and an 802.15.4 (ZigBee) network in two real office buildings. We evaluated the performance of our attack detection schemes in terms of detection rates and receiver operating characteristic curves. Our experimental results provide strong evidence of the effectiveness of our attack detection schemes with high detection rates, over 95% and low false positive rates, often below 5%. Also, our approach is generic across a diverse set of algorithms, networks, and buildings. Interestingly, we found that the performance of the different attack detection schemes are more similar than different. This result shows that different localization systems have similar attack detection capabilities, and consequently that system designers can focus on using algorithms that provide the highest localization accuracy rather than having to tradeoff position accuracy against attack detection abilities.

Appendix
Computing the Probability Mass M

We have simplified the probability mass M of a multivariate Gaussian distribution as

$$M = \frac{1}{(\sqrt{2\pi})^n} \int \dots \int_{E'} e^{-\frac{1}{2}\mathbf{y}^\mathbf{T}\mathbf{y}} dy_1 \dots dy_n$$
$$= \frac{1}{(\sqrt{2\pi})^n} \int \dots \int_{E'} e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2} dy_1 \dots dy_n \quad (16)$$

with E' defined by $\mathbf{y^T y} \leq X^2$. Changing to polar coordinates, we get

$$\begin{aligned} M &= \frac{1}{(\sqrt{2\pi})^n} \int_0^X \int_0^{2\pi} \int_0^{\pi} ... \int_0^{\pi} [e^{-\frac{r^2}{2}} r^{n-1} dr d\phi_1 \\ &\quad sin\phi_2 d\phi_2 ... sin^{n-2}\phi_{n-1} d\phi_{n-1}] \\ &= \frac{1}{(\sqrt{2\pi})^n} \int_0^X e^{-\frac{r^2}{2}} r^{n-1} dr \times \int_0^{2\pi} d\phi_1 \\ &\quad \times \prod_{i=2}^{n-1} \int_0^{\pi} sin^{i-1}\phi_i d\phi_i \\ &= \frac{2}{(\sqrt{\pi})^{n-2}} \times A_{r,n} \times \prod_{i=2}^{n-1} B_i \end{aligned} \quad (17)$$

with

$$A_{r,n} = \frac{1}{(\sqrt{2})^n} \int_0^X e^{-\frac{r^2}{2}} r^{n-1} dr$$

and

$$B_i = \int_0^{\pi} sin^{i-1}\phi_i d\phi_i.$$

Using $v = r^2/2$, we have

$$A_{r,n} = \frac{1}{2} \int_0^{\frac{X^2}{2}} e^{-v} v^{\frac{n-2}{2}} dv = \frac{1}{2} \times \Gamma(\frac{n}{2}, \frac{X^2}{2}) \quad (18)$$

where Γ is the incomplete gamma function. Since

$$B_i = \beta(\frac{i}{2}, \frac{1}{2}) = \frac{\Gamma(\frac{i}{2})}{\Gamma(\frac{i+1}{2})} \times \sqrt{\pi}. \quad (19)$$

Through further simplification, we can get

$$\prod_{i=2}^{n-1} B_i = (\sqrt{\pi})^{n-2} \times \frac{1}{\Gamma(\frac{n}{2})}. \quad (20)$$

Hence, substituting Equations (18) and (20) into (17), we obtain the probability mass

$$M = \frac{\Gamma(n/2, X^2/2)}{\Gamma(n/2)}.$$

REFERENCES

[1] R. Battiti, M. Brunato, and A. Villani, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," University of Trento, Informatica e Telecomunicazioni, Technical Report DIT-02-086, Oct. 2002.

[2] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Comput. Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[3] E. Elnahrawy, X. Li, and R. P. Martin, "The limits of localization using signal strength: A comparative study," in *Proceedings of the First IEEE International Conference on Sensor and Ad hoc Communcations and Networks (SECON 2004)*, Oct. 2004, pp. 406–414.

[4] Y. Chen, K. Kleisouris, X. Li, W. Trappe, and R. P. Martin, "The robustness of localization algorithms to signal strength attacks: a comparative study," in *Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS)*, June 2006, pp. 546–563.

[5] P. Enge and P. Misra , *Global Positioning System: Signals, Measurements and Performance*. Ganga-Jamuna Pr, 2001.

[6] D. Niculescu and B. Nath, "Ad hoc positioning system (APS)," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2001, pp. 2926–2931.

[7] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*, 2005, pp. 91–98.

[8] K. Chintalapudi, A. Dhariwal, R. Govindan, and G. Sukhatme, "Ad hoc localiztion using ranging and sectoring," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2004.

[9] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2000, pp. 775–784.

[10] N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*, Aug 2000.

[11] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from mere connectivity," in *Proceedings of the Fourth ACM International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc)*, Jun 2003.

[12] M. Youssef, A. Agrawal, and A. U. Shankar, "WLAN location determination via clustering and probability distributions," in *Proceedings of IEEE PerCom'03*, Fort Worth, TX, Mar. 2003.

[13] L. Doherty1, K. S. J. Pister, and L. ElGhaoui, "Convex position estimation in wireless sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2001.

[14] S. Brands and D. Chaum, "Distance-bounding protocols," in *Proceedings of the Workshop RTBPTheory and Application of Cryptographic Techniques on Advances in Cryptology*, 1994, pp. 344–359.

[15] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proceedings of the 2003 ACM workshop on wireless security*, 2003, pp. 1–10.

[16] S. Capkun and J. P. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 2005.

[17] S. Capkun and J. Hubaux, "Securing localization with hidden and mobile base stations," to appear in Proceedings of IEEE Infocom 2006.

[18] L. Lazos, R. Poovendran, and S. Capkun, "Rope: robust position estimation in wireless sensor networks," in *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, 2005, pp. 324–331.

[19] D. Liu, P. Ning, and W. Du, "Attack-resistant location estimation in sensor networks," in *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, 2005.

[20] D. Liu and P. Ning and W. Du, "Detecting malicious beacon nodes for secure location discovery in wireless sensor networks," in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS 05)*, June 2005, pp. 609–619.

[21] W. Du, L. Fang, and P. Ning, "Lad: Localization anomaly detection for wireless sensor networks," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 05)*, April 2005.

[22] Y. Hu, A. Perrig, and D. Johnson, "Packet leashes: a defense against wormhole attacks in wireless networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 2003.

[23] Y. Chen, J. Francisco, W. Trappe, and R. P. Martin, "A practical approach to landmark deployment for indoor localization," in *Proceedings of the Third Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, September 2006.

[24] S. Weisberg, *Applied Linear Regression*. Wiley Series in Probability and Mathematical Statistics, 2005.

[25] A. Krishnakumar and P. Krishnan, "On the accuracy of signal strength-based location estimation techniques," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2005.

[26] D. Madigan, E. Elnahrawy, R. Martin, W. Ju, P. Krishnan, and A. S. Krishnakumar, "Bayesian indoor positioning systems," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2005, pp. 324–331.