Detecting and Localizing Wireless Spoofing Attacks

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— **Wireless networks are vulnerable to spoofing attacks, which allows for many other forms of attacks on the networks. Although the identity of a node can be verified through cryptographic authentication, authentication is not always possible because it requires key management and additional infrastructural overhead. In this paper we propose a method for both detecting spoofing attacks, as well as locating the positions of adversaries performing the attacks. We first propose an attack detector for wireless spoofing that utilizes K-means cluster analysis. Next, we describe how we integrated our attack detector into a real-time indoor localization system, which is also capable of localizing the positions of the attackers. We then show that the positions of the attackers can be localized using either area-based or point-based localization algorithms with the same relative errors as in the normal case. We have evaluated our methods through experimentation using both an 802.11 (WiFi) network as well as an 802.15.4 (ZigBee) network. Our results show that it is possible to detect wireless spoofing with both a high detection rate and a low false positive rate, thereby providing strong evidence of the effectiveness of the K-means spoofing detector as well as the attack localizer.**

I. Introduction

As more wireless and sensor networks are deployed, they will increasingly become tempting targets for malicious attacks. Due to the openness of wireless and sensor networks, they are especially vulnerable to spoofing attacks where an attacker forges its identity to masquerade as another device, or even creates multiple illegitimate identities. Spoofing attacks are a serious threat as they represent a form of identity compromise and can facilitate a variety of traffic injection attacks, such as evil twin access point attacks. It is thus desirable to detect the presence of spoofing and eliminate them from the network.

The traditional approach to address spoofing attacks is to apply cryptographic authentication. However, authentication requires additional infrastructural overhead and computational power associated with distributing, and maintaining cryptographic keys. Due to the limited power and resources available to the wireless devices and sensor nodes, it is not always possible to deploy authentication. In addition, key management often incurs significant human management costs on the network. In this paper, we take a different approach by using the physical properties associated with wireless transmissions to detect spoofing. Specifically, we propose a scheme for both detecting spoofing attacks, as well as localizing the positions of the adversaries performing the attacks. Our approach utilizes the Received Signal Strength (RSS) measured across a set of access points to perform spoofing detection and localization. Our scheme does not add any overhead to the wireless devices and sensor nodes.

By analyzing the RSS from each MAC address using K-means cluster algorithm, we have found that the distance between the centroids in signal space is a good test statistic for effective attack detection. We then describe how we integrated our K-means spoofing detector into a real-time indoor localization system. Our K-means approach is general in that it can be applied to almost all RSS-based localization algorithms. For two sample algorithms, we show that using the centroids of the clusters in signal space as the input to the localization system, the positions of the attackers can be localized with the same relative estimation errors as under normal conditions.

To evaluate the effectiveness of our spoofing detector and attack localizer, we conducted experiments using both an 802.11 network as well as an 802.15.4 network in a real office building environment. In particular, we have built an indoor localization system that can localize any transmitting devices on the floor in real-time. We evaluated the performance of the K-means spoofing detector using detection rates and receiver operating characteristic curve. We have found that our spoofing detector is highly effective with over 95% detection rates and under 5% false positive rates.

Further, we observed that, when using the centroids in signal space, a broad family of localization algorithms achieve the same performance as when they use the averaged RSS in traditional localization attempts. Our

1-4244-1268-4/07/$25.00 ©2007 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE SECON 2007 proceedings.

experimental results show that the distance between the localized results of the spoofing node and the original node is directly proportional to the true distance between the two nodes, thereby providing strong evidence of the effectiveness of both our spoofing detection scheme as well as our approach of localizing the positions of the adversaries.

The rest of the paper is organized as follows. Section II describes the previous research in addressing spoofing attacks, spoofing detection, and the related work in localization. In Section III, we study the feasibility of spoofing attacks and their impacts, and discuss our experimental methodologies. We formulate the spoofing attack detection problem and propose K-means spoofing detector in Section IV. We introduce the real-time localization system and present how to find the positions of the attackers in Section V. Further, we provide a discussion in Section VI. Finally, we conclude our work in Section VII.

II. Related Work

Recently, there has been much active research addressing spoofing attacks as well as those facilitated by adversaries masquerading as another wireless device. We cannot cover the entire body of works in this section. Rather, we give a short overview of traditional approaches and several new methods. We then describe the works most closely related to our work.

The traditional security approach to cope with identity fraud is to use cryptographic authentication. An authentication framework for hierarchical, ad hoc sensor networks is proposed in [1] and a hop-by-hop authentication protocol is presented in [2]. Additional infrastructural overhead and computational power are needed to distribute, maintain, and refresh the key management functions needed for authentication. [3] has introduced a secure and efficient key management framework (SEKM). SEKM builds a Public Key Infrastructure (PKI) by applying a secret sharing scheme and an underlying multicast server group. [4] implemented a key management mechanism with periodic key refresh and host revocation to prevent the compromise of authentication keys. In addition, binding approaches are employed by Cryptographically Generated Addresses (CGA) to defend against the network identity spoofing [5], [6].

Due to the limited resources in wireless and sensor nodes, and the infrastructural overhead needed to maintain the authentication mechanisms, it is not always desirable to use authentication. Recently new approaches have been proposed to detect the spoofing attacks in wireless networks. [7], [8] have introduced a security layer that is separate from conventional network authentication methods. They developed forge-resistant relationships based on packet traffic by using packet sequence numbers, traffic interarrival, one-way chain of temporary identifiers, and signal strength consistency checks to detect spoofing attacks. [9] proposed a lower-layer approach that utilizes properties of the wireless channel at the physical layer to support high-level security objectives such as authentication and confidentiality. The most closely related work to our paper is [10], which proposed the use of matching rules of signalprints for spoofing detection.

Although these methods have varying detection and false alarm rates, none of these approaches provide the ability to localize the positions of the spoofing attackers after detection. Further, our work is novel in that we have integrated our spoofing detector into a real-time localization system which can both detect the spoofing attacks, as well as localize the adversaries in wireless and sensor networks. In addition, we deployed our localization system in a real office building environment which houses our Computer Science Department.

Received signal strength is also employed to detect sybil nodes in wireless sensor networks [11]. However, they did not study how to localize the sybil nodes. [12] utilized signal strength distributions to detect and localize sybil nodes in Vehicular Ad Hoc Networks (VANETs). Their statistical algorithms are closely associated with VANETs.

Finally, a large body of work has developed localization algorithms for wireless and sensor networks. The works that are related to this paper are algorithms using RSS to perform localization, including both fingerprint matching and probabilistic techniques [13]–[15]. In this work we used these schemes to localize the positions of the attackers.

III. Feasibility of Attacks

In this section we provide a brief overview of spoofing attacks and their impact. We then discuss the experimental methodology that we use to evaluate our approach of spoofing detection.

A. Spoofing Attacks

Due to the open-nature of the wireless medium, it is easy for adversaries to monitor communications to find the layer-2 Media Access Control (MAC) addresses of the other entities. Recall that the MAC address is typically used as a unique identifier for all the nodes

on the network. Further, for most commodity wireless devices, attackers can easily forge their MAC address in order to masquerade as another transmitter. As a result, these attackers appear to the network as if they are a different device. Such spoofing attacks can have a serious impact on the network performance as well as facilitate many forms of security weaknesses, such as attacks on access control mechanisms in access points [16], and denial-of-service through a deauthentication attack [17]. A broad survey of possible spoofing attacks can be found in [7], [10].

To address potential spoofing attacks, the conventional approach uses authentication. However, the application of authentication requires reliable key distribution, management, and maintenance mechanisms. It is not always desirable to apply authentication because of its infrastructural, computational, and management overhead. Further, cryptographic methods are susceptible to node compromise– a serious concern as most wireless nodes are easily accessible, allowing their memory to be easily scanned.

It is desirable to use properties that cannot be undermined even when nodes are compromised. We propose to use received signal strength (RSS), a property associated with the transmission and reception of communication (and hence not reliant on cryptography), as the basis for detecting spoofing. Employing RSS as a means to detect spoofing will not require any additional cost to the wireless devices themselves– they will merely use their existing communication methods, while the wireless network will use a collection of base stations to monitor received signal strength for the potential of spoofing.

B. Experimental Methodology

In order to evaluate the effectiveness of our spoofing detection mechanisms, which we describe in the next section, we have conducted experiments using both an 802.11 (WiFi) network, using an Orinoco silver card, as well as an 802.15.4 (ZigBee) network, using a Telosb mote, on the 3rd floor of the Computer Science Department at Rutgers University. The floor size is 200x80ft (16000 ft^2). Figure 1 (a) shows the 802.11 (WiFi) network with 4 landmarks deployed to maximize signal strength coverage, as shown in red squares. The 802.15.4 (ZigBee) network is presented in Figure 1 (b) with 4 landmarks distributed in a squared setup in order to achieve optimal landmark placement [18] as shown in red triangles. The small blue dots in the floor map are the locations used for spoofing and localization tests.

For the 802.15.4 network, we used 300 packet-level RSS samples for each of the 100 locations. We utilized the actual RSS values attached to each packet. We have 286 locations in the 802.11 deployment. Unlike the 802.15.4 data, the RSS values are partially synthetic. We had access to only the mean RSS at each location, but to perform our experiments we needed an RSS value per packet. To generate such data for 200 simulated packets at each location, we used random draws from a normal distribution. We used the measured RSS mean for the mean of the distribution. For the standard deviation, we computed the difference in the RSS from a fitted signal to distance function, and then calculated the standard deviation of the distribution from these differences over all locations. To keep our results conservative, we took the maximum deviation over all landmarks, which we found to be 5 dB.

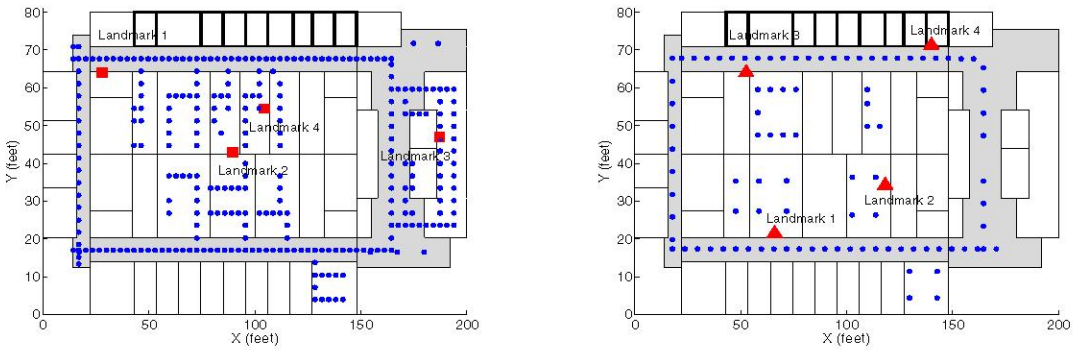
Much work has gone into characterizing the distributions of RSS readings indoors. It has been shown that characterizing the per-location RSS distributions as normal, although not often the most accurate characterization, still results in the best balance between algorithmic usability and the resulting localization error [15], [19].

In addition, we built a real-time localization system to estimate the positions of both the original nodes and the spoofing nodes. We randomly selected points out of the above locations as the training data for use by the localization algorithms. For the 802.11 network, the size of the training data is 115 locations, while for the 802.15.4 network, the size of the training data is 70 locations. The detailed description of our localization system is presented in Section V.

To test our approach's ability to detect spoofing, we randomly chose a point pair on the floor and treated one point as the position of the original node, and the other as the position of the spoofing node. We ran the spoofing test through all the possible combinations of point pairs on the floor using all the testing locations in both networks. There are total 14535 pairs for the 802.11 network and 4371 pairs for the 802.15.4 network. The experimental results will be presented in the following sections for the spoofing detector and the attack localizer.

IV. Attack Detector

In this section we propose our spoofing attack detector. We first formulate the spoofing attack detection problem as one using classical statistical testing. Next, we describe the test statistic for spoofing detection. We then introduce the metrics to evaluate the effectiveness of our approach. Finally, we present our experimental results.



(a) 802.11 network (b) 802.15.4 network

Fig. 1. Landmark setups and testing locations in two networks.

A. Formulation of Spoofing Attack Detection

RSS is widely available in deployed wireless communication networks and its values are closely correlated with location in physical space. In addition, RSS is a common physical property used by a widely diverse set of localization algorithms [13]–[15], [20]. In spite of its several meter-level localization accuracy, using RSS is an attractive approach because it can re-use the existing wireless infrastructure. We thus derive a spoofing attack detector utilizing properties of the RSS.

The goal of the spoofing detector is to identify the presence of a spoofing attack. We formulate the spoofing attack detection as a statistical significance test, where the null hypothesis is:

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

In significance testing, a test statistic T is used to evaluate whether observed data belongs to the null-hypothesis or not. If the observed test statistic $\mathrm{T}^{\mathrm{obs}}$ differs significantly from the hypothesized values, the null hypothesis is rejected and we claim the presence of a spoofing attack.

B. Test Statistic for Spoofing Detection

Although affected by random noise, environmental bias, and multipath effects, the RSS value vector, $\mathbf{s} = \{s_1, s_2, ...s_n\}$ (n is the number of landmarks/access points(APs)), is closely related with the transmitter's physical location and is determined by the distance to the landmarks [15]. The RSS readings at different locations in physical space are distinctive. Each vector $\mathbf{s}$ corresponds to a point in a n-dimensional signal space [21]. When there is no spoofing, for each MAC address, the sequence of RSS sample vectors will be close to each other, and will fluctuate around a mean vector. However, under a spoofing attack, there is more than one node at different physical locations claiming the same MAC address. As a result, the RSS sample readings from the attacked MAC address will be mixed with RSS readings from at least one different location. Based on the properties of the signal strength, the RSS readings from the same physical location will belong to the same cluster points in the n-dimensional signal space, while the RSS readings from different locations in the physical space should form different clusters in signal space.

This observation suggests that we may conduct K-means cluster analysis [22] on the RSS readings from each MAC address in order to identify spoofing. If there are M RSS sample readings for a MAC address, the K-means clustering algorithm partitions M sample points into K disjoint subsets S_j containing M_j sample points so as to minimize the sum-of-squares criterion:

$$J_{min} = \sum_{j=1}^{K} \sum_{\mathbf{s_m} \in S_j} \|\mathbf{s_m} - \mu_j\|^2 \qquad (1)$$

where $\mathbf{s_m}$ is a RSS vector representing the mth sample point and μ_j is the geometric centroid of the sample points for S_j in signal space. Under normal conditions, the distance between the centroids should be close to each other since there is basically only one cluster. Under a spoofing attack, however, the distance between the centroids is larger as the centroids are derived from the different RSS clusters associated with different locations in physical space. We thus choose the distance between two centroids as the test statistic T for spoofing detection,

$$D_c = ||\mu_i - \mu_j|| \qquad (2)$$

with $i, j \in \{1, 2..K\}$. Next, we will use empirical methodologies from the collected data set to determine thresholds for defining the critical region for the significance testing. To illustrate, we use the following definitions, *an original node* P_{org} is referred to as the wireless device with the legitimate MAC address, while *a*

spoofing node P_{spoof} is referred to as the wireless device that is forging its identity and masquerading as another device. There can be multiple spoofing nodes of the same MAC address.

Note that our K-means spoofing detector can handle packets from different transmission power levels. If an attacker sends packets at a different transmission power level from the original node with the same MAC address, there will be two distinct RSS clusters in signal space. Thus, the spoofing attack will be detected based on the distance of the two centroids obtained from the RSS clusters.

C. Determining Thresholds

The appropriate threshold τ will allow the spoofing detector to be robust to false detections. We can determine the thresholds through empirical training. During the off line phase, we can collect the RSS readings for a set of known locations over the floor and obtain the distance between two centroids in signal space for each point pair. We use the distribution of the training information to determine the threshold τ. At run time, based on the RSS sample readings for a MAC address, we can calculate the observed value D_c^{obs}. Our condition for declaring that a MAC address is under a spoofing attack is:

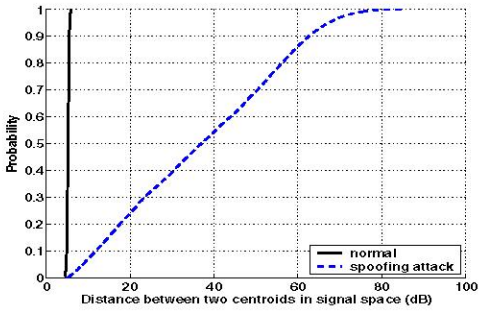
$$D_c^{obs} > \tau. \tag{3}$$

Figure 2 (a) and (b) show the CDF of the D_c in signal space for both the 802.11 network and the 802.15.4 network. We found that the curve of D_c shifted greatly to the right under spoofing attacks, thereby suggesting that using D_c as a test statistic is an effective way for detecting spoofing attacks.

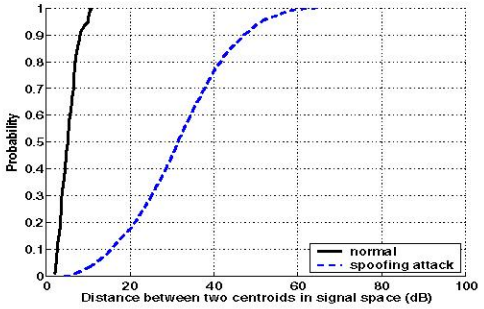
D. Performance Metrics

In order to evaluate the performance of our spoofing attack detector using K-means cluster analysis, we use the following metrics:

Detection Rate and False Positive Rate: A spoofing attack will 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 possible spoofing attacks on the floor. The detection rate is defined as the percentage of spoofing attack attempts that are determined to be under attack. Note that, when the spoofing attack is present, 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}. The detection rate and false positive rate vary under different thresholds.



(a) 802.11 network



(b) 802.15.4 network

Fig. 2. Cumulative Distribution Function (CDF) of D_c in signal space

Receiver 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 a plot of attack detection accuracy against the false positive rate. It can be obtained by varying the detection thresholds. The ROC curve provides a direct means to measure the trade off between false-positives and correct detections.

E. Experimental Evaluation

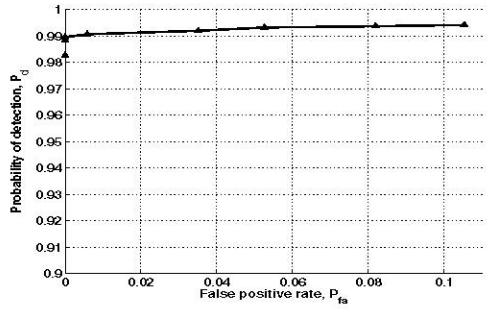
In this section we present the evaluation results of the effectiveness of the spoofing attack detector. Table I presents the detection rate and false positive rate for both the 802.11 network and the 802.15.4 network under different threshold settings. The corresponding ROC curves are displayed in Figure 3. The results are encouraging showing that for false positive rates less than 10%, the detection rates are above 95%. Even when the false positive rate goes to zero, the detection rate is still more than 95% for both 802.11 and 802.15.4 networks.

We further study how likely a spoofing node can be detected by our spoofing attack detector when it is at varying distances from the original node in physical space. Figure 4 presents the detection rate as a function of the distance between the spoofing node and the original node. We found that the further away P_{spoof} is from P_{org}, the higher the detection rate becomes. For the 802.11 network, the detection rate goes to over 90%

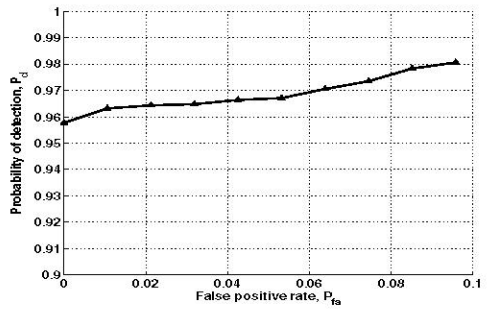
Network, Threshold	Detection Rate	False Positive Rate
802.11, τ = 5.5dB	0.9937	0.0819
802.11, τ = 5.7dB	0.9920	0.0351
802.11, τ = 6dB	0.9884	0
802.15.4, τ = 8.2dB	0.9806	0.0957
802.15.4, τ = 10dB	0.9664	0.0426
802.15.4, τ = 11dB	0.9577	0

TABLE I

DETECTION RATE AND FALSE POSITIVE RATE OF THE SPOOFING ATTACK DETECTOR.



(a) 802.11 network



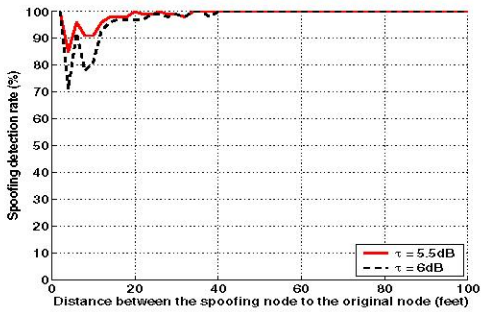
(b) 802.15.4 network

Fig. 3. Receiver Operating Characteristic (ROC) curves

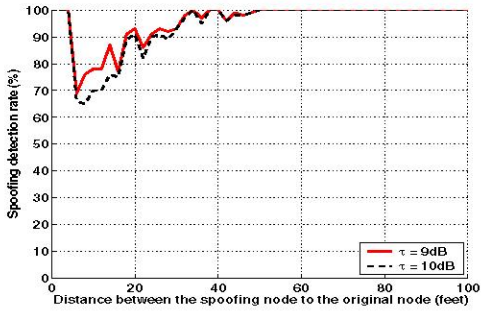
when P_{spoof} is about 13 feet away from P_{org} under τ equals to 5.5dB. While for the 802.15.4 network, the detection rate is above 90% when the distance between P_{spoof} and P_{org} is about 20 feet by setting threshold τ to 9dB. This is in line with the average localization estimation errors using RSS [15] which are about 10-15 feet. When the nodes are less than 10-15 feet apart, they have a high likelihood of generating similar RSS readings, and thus the spoofing detection rate falls below 90%, but still greater than 60%. However, when P_{spoof} moves closer to P_{org}, the attacker also increases the probability to expose itself. The detection rate goes to 100% when the spoofing node is about 45-50 feet away from the original node.

V. LOCALIZING ADVERSARIES

If the spoofing attack is determined to be present by the spoofing attack detector, we want to localize the adversaries and further to eliminate the attackers from the network. In this section we present a real-time localization system that can be used to locate the positions of the attackers. We then describe the localization algorithms used to estimate the adversaries' position. The experimental results are presented to evaluate the effectiveness of our approach.



(a) 802.11 network



(b) 802.15.4 network

Fig. 4. Detection rate as a function of the distance between the spoofing node and the original node.

A. Localization System

We have developed a general-purpose localization system to perform real-time indoor positioning. This system is designed with fully distributed functionality and easy to plug-in localization algorithms. It is built around 4 logical components: Transmitter, Landmark, Server, and Solver. The system architecture is shown in Figure 5.

Transmitter: Any device that transmits packets can be localized. Often the application code does not need to be altered on a sensor node in order to localize it.

Landmark: The Landmark component listens to the packet traffic and extracts the RSS reading for each transmitter. It then forwards the RSS information to the Server component. The Landmark component is stateless and is usually deployed on each landmark or access point with known locations.

Server: A centralized server collects RSS information from all the Landmark components. The spoofing detection is performed at the Server component. The Server summarizes the RSS information such as averaging or clustering, then forwards the information to the Solver component for localization estimation.

Solver: A Solver takes the input from the Server,

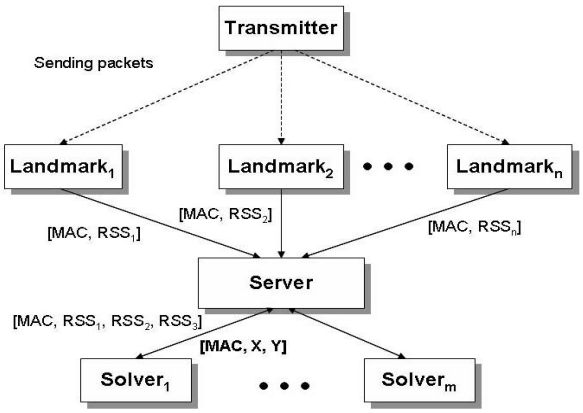


Fig. 5. Localization system architecture

performs the localization task by utilizing the localization algorithms plugged in, and returns the localization results back to the Server. There are multiple Solver instances available and each Solver can localize multiple transmitters simultaneously.

During the localization process, the following steps will take place:

1. A Transmitter sends a packet. Some number of Landmarks observe the packet and record the RSS.
2. Each Landmark forwards the observed RSS from the transmitter to the Server.
3. The Server collects the complete RSS vector for the transmitter and sends the information to a Solver instance for location estimation.
4. The Solver instance performs localization and returns the coordinates of the transmitter back to the Server.

If there is a need to localize hundreds of transmitters at the same time, the server can perform load balancing among the different solver instances. This centralized localization solution also makes enforcing contracts and privacy policies more tractable.

B. Attack Localizer

When our spoofing detector has identified an attack for a MAC address, the centroids returned by the K-means clustering analysis in signal space can be used by the server and sent to the solver for location estimation. The returned positions should be the location estimate for the original node and the spoofing nodes in physical space. Using a location on the testing floor as an example, Figure 6 shows the relationship among the original node P_{org}, the location estimation of the original node L_{org}, the spoofing node P_{spoof}, and the localized spoofing node position L_{spoof}.

In order to show the generality of our localization system for locating the spoofing nodes, we have chosen

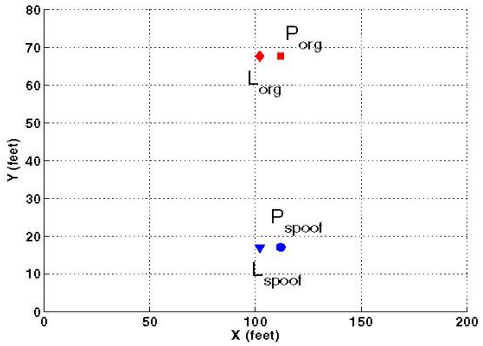


Fig. 6. Relationships among the original node, the spoofing node, and their location estimation through localization system.

two representative localization algorithms using signal strength from point-based algorithms and area-based algorithms.

RADAR: Point-based methods return an estimated point as a localization result. A primary example of a point-based method is the RADAR scheme [13]. In RADAR, during the off line phase, a mobile transmitter with known position broadcasts beacons periodically, and the RSS readings are measured at a set of landmarks. Collecting together the averaged RSS readings from each of the landmarks for a set of known locations provides a radio map. At runtime, localization is performed by measuring a transmitter's RSS at each landmark, and the vector of RSS values is compared to the radio map. The record in the radio map whose signal strength vector is closest in the Euclidean sense to the observed RSS vector is declared to correspond to the location of the transmitter. In this work, instead of using the averaged RSS in the traditional approach, we use the RSS centroids obtained from the K-means clustering algorithm as the observed RSS vector for localizing a MAC address.

Area Based Probability (ABP): Area-based algorithms return a most likely area in which the true location resides. One major advantage of area-based methods compared to point-based methods is that they return a region, which has an increased chance of capturing the transmitter's true location. ABP returns an area, a set of tiles on the floor, bounded by a probability that the transmitter is within the returned area [15]. ABP assumes the distribution of RSS for each landmark follows a Gaussian distribution. The Gaussian random variable from each landmark is independent. ABP then computes the probability of the transmitter being at each tile L_i on the floor using Bayes' rule:

$$P(L_i|\mathbf{s}) = \frac{P(\mathbf{s}|L_i) \times P(L_i)}{P(\mathbf{s})}. \quad (4)$$

Given that the transmitter must reside at exactly one tile satisfying $\sum_{i=1}^{L} P(L_i|\mathrm{s}) = 1$, ABP normalizes the probability and returns the most likely tiles up to its confidence α.

Both RADAR and ABP are employed in our experiments to localize the positions of the attackers.

C. Experimental Evaluation

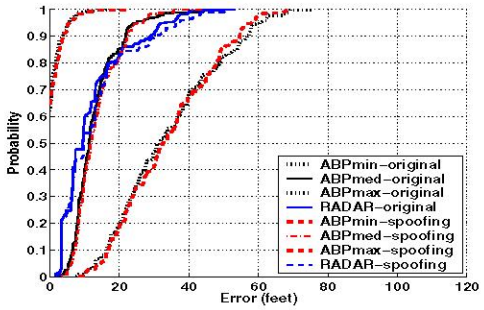
In order to evaluate the effectiveness of our localization system in finding the locations of the attackers, we are interested in the following performance metrics:

Localization Error CDF: We obtain the cumulative distribution function (CDF) of the location estimation error from all the localization attempts, including both the original nodes and the spoofing nodes. We then compare the error CDF of all the original nodes to that of all the possible spoofing nodes on the floor. For area-based algorithms, we also report CDFs of the minimum and maximum error. For a given localization attempt, these are points in the returned area that are closest to and furthest from the true location.

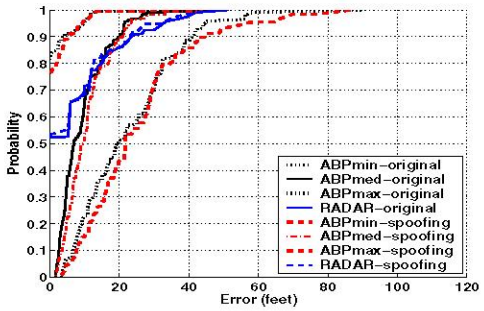
Relationship between the true and estimated distances: The relationship between the true distance of the spoofing node to the original node $||P_{org} - P_{spoof}||$ and the distance of the location estimate of the spoofing node to that of the original node $||L_{org} - L_{spoof}||$ evaluates how accurate our attack localizer can report the positions of both the original node and the attackers.

We first present the statistical characterization of the location estimation errors. Figure 7 presents the localization error CDF of the original nodes and the spoofing nodes for both RADAR and ABP in the 802.11 network as well as the 802.15.4 network. For the area-based algorithm, the median tile error $ABP{-}med$ is presented, as well as the minimum and maximum tile errors, $ABP{-}min$ and $ABP{-}max$. We found that the location estimation errors from using the RSS centroids in signal space are about the same as using averaged RSS as the input for localization algorithms [15]. Comparing to the 802.11 network, the localization performance in the 802.15.4 network is qualitatively better for both RADAR and ABP algorithms. This is because the landmark placement in the 802.15.4 network is closer to that predicted by the optimal and error minimizing placement algorithm as described in [18].

More importantly, we observed that the localization performance of the original nodes is qualitatively the same as that of the spoofing nodes. This is very encouraging as the similar performance is strong evidence that using the centroids obtained from the K-means cluster



(a) 802.11 network



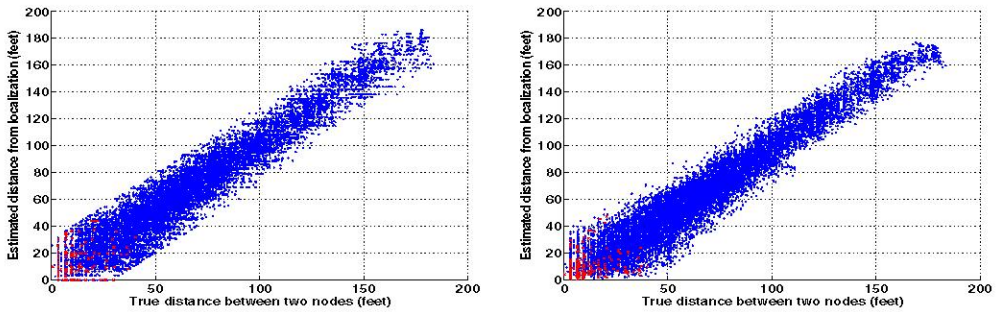
(b) 802.15.4 network

Fig. 7. Localization error CDF across localization algorithms and networks.

analysis is effective in both identifying the spoofing attacks as well as localizing the attackers.

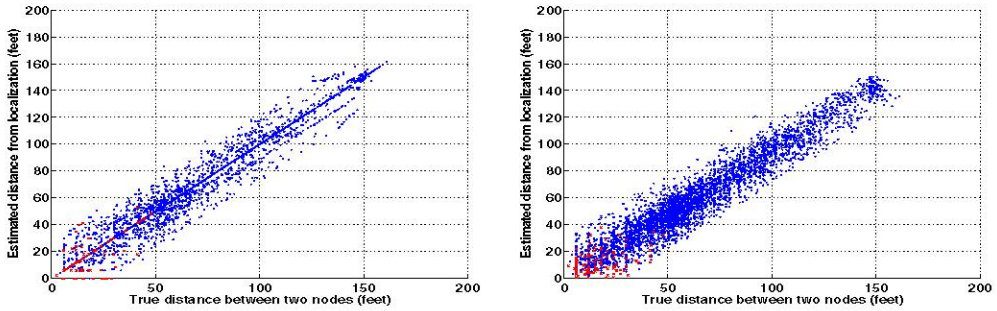
The challenge in localizing the positions of the attackers arises because the system does not know the positions of either the original MAC address or the node with the masquerading MAC. Thus, we would like to examine how accurate the localization system can estimate the distance between P_{org} and P_{spoof}. Figure 8 displays the relationship between $||L_{org} - L_{spoof}||$ and $||P_{org} - P_{spoof}||$ across different localization algorithms and networks. The blue dots represent the cases of the detected spoofing attacks. While the red crosses indicate the spoofing attack has not been detected by the K-means spoofing detector. Comparing with Figure 4, i.e. the detection rate as a function of the distance between P_{org} and P_{spoof}, the results of the undetected spoofing attack cases represented by the red crosses are in line with the results in Figure 4, the spoofing attacks are 100% detected when $||P_{org} - P_{spoof}||$ equals to or is greater than about 50 feet.

Further, the relationship between $||L_{org} - L_{spoof}||$ and $||P_{org} - P_{spoof}||$ is along the 45 degree straight line. This means that $||L_{org} - L_{spoof}||$ is directly proportional to $||P_{org} - P_{spoof}||$ and indicates that our localization system is highly effective for localizing the attackers. At a fixed distance value of $||P_{org} - P_{spoof}||$, the values of $||L_{org} - L_{spoof}||$ fluctuate around the true distance value. The fluctuation reflects the localization errors of



(a) 802.11: RADAR, τ = 6dB

(b) 802.11: ABP, τ = 6dB



(c) 802.15.4: RADAR, τ = 9dB

(d) 802.15.4: ABP, τ = 9dB

Fig. 8. Relationship between the true distance and the estimated distance for the original node and the spoofing node across localization algorithms and networks.

both P_{org} and P_{spoof}. The larger the $||P_{org} - P_{spoof}||$ is, the smaller the fluctuation of $||L_{org} - L_{spoof}||$ becomes, at about 10 feet maximum. This means that if the attacker is farther away from the original node, it is extremely likely that the K-means spoofing detector can detect it. In addition, our attack localizer can find the attacker's position and estimate the distance from the original node to the attacker at about 10 to 20 feet maximum error.

VI. Discussion

So far we have conducted K-means cluster analysis in signal space. Our real-time localization system also inspired us to explore packet-level localization at the server, which means localization is performed for each packet received at the landmarks. The server utilizes each RSS reading vector for localization. Over a certain time period (for example, 60 seconds), for a MAC address there will be a cluster of location estimates in physical space. Intuitively, we think that, during a spoofing attack there will be distinctive location clusters around the original node and the spoofing nodes in physical space. Our intuition was that the clustering results from the per-packet localization would allow the detection and localization of attackers in one step.

However, we found that the performance of clustering packet-level localization results for spoofing detection is not as effective as deriving the centroids in signal space. The relationship between $||P_{org} - P_{spoof}||$ and $||L_{org} - L_{spoof}||$ is shown in Figure 9. Although it also has a trend along the 45 degree line, it shows more uncertainties along the line. Therefore, we believe that given a set of RSS reading samples for a MAC address, working with the signal strength directly preserves the basic properties of the radio signal, and this in turn is more closely correlated with the physical location, and thus working with the RSS values directly better reveals the presence of the spoofing attacks.

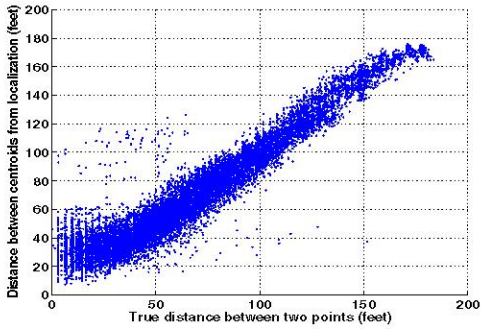


Fig. 9. Packet-level localization: relationship between the true distance and the estimated distance for the original node and the spoofing node when using RADAR in the 802.11 network.

VII. Conclusion

In this work, we proposed a method for detecting spoofing attacks as well as localizing the adversaries

in wireless and sensor networks. In contrast to traditional identity-oriented authentication methods, our RSS based approach does not add additional overhead to the wireless devices and sensor nodes. We formulated the spoofing detection problem as a classical statistical significance testing problem. We then utilized the K-means cluster analysis to derive the test statistic. Further, we have built a real-time localization system and integrated our K-means spoofing detector into the system to locate the positions of the attackers and as a result to eliminate the adversaries from the network.

We studied the effectiveness and generality of our spoofing detector and attacker localizer in both an 802.11 (WiFi) network and an 802.15.4 (ZigBee) network in a real office building environment. The performance of the K-means spoofing detector is evaluated in terms of detection rates and receiver operating characteristic curves. Our spoofing detector has achieved high detection rates, over 95% and low false positive rates, below 5%. When locating the positions of the attackers, we have utilized both the point-based and area-based algorithms in our real-time localization system. We found that the performance of the system when localizing the adversaries using the results of K-means cluster analysis are about the same as localizing under normal conditions. Usually the distance between the spoofing node and the original node can be estimated with median error of 10 feet. Our method is generic across different localization algorithms and networks. Therefore, our experimental results provide strong evidence of the effectiveness of our approach in detecting the spoofing attacks and localizing the positions of the adversaries.

REFERENCES

[1] M. bohge and W. Trappe, "An authentication framework for hierarchical ad hoc sensor networks," in *Proceedings of the ACM Workshop on Wireless Security (WiSe)*, 2003, pp. 79–87.

[2] S. Zhu, S. Xu, S. Setia, and S. Jajodia, "Lhap: A lightweight hop-by-hop authentication protocol for ad-hoc networks," in *Proceedings of the IEEE International Workshop on Mobile and Wireless Network (MWN)*, 2003, pp. 749–755.

[3] B. Wu, J. Wu, E. Fernandez, and S. Magliveras, "Secure and efficient key management in mobile ad hoc networks," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2005.

[4] A. Wool, "Lightweight key management for ieee 802.11 wireless lans with key refresh and host revocation," *ACM/Springer Wireless Networks*, vol. 11, no. 6, pp. 677–686, 2005.

[5] T. Aura, "Cryptographically generated addresses (cga)," *RFC 3972, IETF*, 2005.

[6] E. Kempf, J. Sommerfeld, B. Zill, B. Arkko, and P. Nikander, "Secure neighbor discovery (send)," *RFC 3971, IETF*, 2005.

[7] Q. Li and W. Trappe, "Relationship-based detection of spoofing-related anomalous traffic in ad hoc networks," in *Proceedings of the Third Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, September 2006.

[8] Q. Li and W. Trappe, "Light-weight detection of spoofing attacks in wireless networks," in *Proceedings of the 2nd International Workshop on Wireless and Sensor Network Security (WSNS)*, October 2006.

[9] Z. Li, W. Xu, R. Miller, and W. Trappe, "Securing wireless systems via lower layer enforcements," in *Proceedings of the ACM Workshop on Wireless Security (WiSe)*, 2006.

[10] D. Faria and D. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," in *Proceedings of the ACM Workshop on Wireless Security (WiSe)*, September 2006.

[11] M. Demirbas and Y. Song, "An rssi-based scheme for sybil attack detection in wireless sensor networks," in *Proceedings of the International Workshop on Advanced Experimental Activities on Wireless Networks and Systems*, 2006.

[12] B. Xiao, B. Yu, and C. Gao, "Detection and localization of sybil nodes in vanets," in *Proceedings of the Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS)*, 2006.

[13] 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.

[14] M. Youssef, A. Agrawal, and A. U. Shankar, "Wlan location determination via clustering and probability distributions," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2003, pp. 143–150.

[15] 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.

[16] W. A. Arbaugh, N. Shankar, Y. Wan, and K. Zhang, "Your 802.11 network has no clothes," *IEEE Wireless Communications*, vol. 9, no. 6, pp. 44–51, Dec. 2002.

[17] J. Bellardo and S. Savage, "802.11 denial-of-service attacks: Real vulnerabilities and practical solutions," in *Proceedings of the USENIX Security Symposium*, 2003, pp. 15 – 28.

[18] 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.

[19] A. Haeberlen, E. Flannery, A. Ladd, A. Rudys, D. Wallach, and L. Kavraki, "Practical robust localization over large scale 802.11 wireless networks," in *Proceedings of the Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, September 2004.

[20] T. Roos, P. Myllymaki, H.Tirri, P. Misikangas, and J. Sievanen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, July 2002.

[21] 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.

[22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining Inference, and Prediction*. Springer, 2001.