Performing Joint Learning for Passive Intrusion Detection in Pervasive Wireless Environments

Jie Yang*, Yong Ge†, Hui Xiong†, Yingying Chen*, Hongbo Liu*

*Dept. of ECE, Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030
{jyang, yingying.chen, hliu3}@stevens.edu

† MSIS Department, Rutgers University
110 Washington St., Newark, NJ 07102
yongge@pegasus.rutgers.edu, hxiong@rutgers.edu

Abstract—**Recent years have witnessed increasing interests in passive intrusion detection for wireless environments, e.g., asset protection in industrial facilities and emergency rescue of trapped people. Most previous studies have focused primarily on exploiting a single intrusion indicator, such as moving variance, for capturing an intrusion pattern at a time. However, in real-world, there are many intrusion patterns which may be only detectable by combining different intrusion indicators and performing detection jointly. To this end, we propose a joint intrusion learning approach, which has the ability in combining the detection power of several complementary intrusion indicators and detects different intrusion patterns at the same time. We developed the GREEK algorithm, which utilizes grid-based clustering over K-neighborhood to effectively diagnose the presence of intrusions. Further, we show that the performance of intrusion detection can be enhanced by utilizing the collaborative detecting efforts among multiple transmitter-receiver pairs. To validate the effectiveness of the joint intrusion learning method, we conducted experiments in a real-office environment using an IEEE 802.15.4 (Zigbee) network. Our experimental results provide strong evidence of the effectiveness of our joint learning approach in performing passive intrusion detection with a minimized false positive rate.**

I. INTRODUCTION

The widespread deployment of wireless communication systems creates unprecedented opportunities to change the computing paradigm in pervasive wireless environments. It is possible to capture the environmental changes as long as wireless networks have been deployed with sufficient density and equipped with enough sensing capabilities. Regardless of whether wireless infrastructures are used just for communications or as the basis for actual responses, the wireless sensing data may be dual-used for intrusion detection in wireless environments. For example, in real-world, there are increasing availability of pervasive wireless infrastructures in industrial facilities, office buildings, transportation infrastructures, and military battlefields. These pervasive wireless infrastructures can be used to assist a broad array of applications, such as intrusion detection in industrial facilities for asset protection, identification of people trapped in a fire building during emergency evacuation, and battlefield protection.

Indeed, in this paper, we focus on exploiting Received Signal Strength (RSS) obtained from the existing wireless infrastructures for performing intrusion detection when the intruders or objects do not have any radio devices attached to them. This is also known as passive intrusion detection [1].

Although there have been extensive studies on intrusion detection [2], [3] and wireless localization [4]–[7], most of the existing intrusion detection techniques utilize video, pressure, ultrasound or infrared, which are either expensive, label intensive or require pre-deployment of specialized hardware, and are thus not easily deployed for unscheduled tasks and may not be scalable. On the other hand, although lots of current wireless localization schemes can reuse existing wireless infrastructures to perform localization, these schemes require the target object to carry a radio device or actively participate in localization.

Different from the above mentioned work in detecting and localizing intrusion objects that either requires specialized infrastructure setup or relies on communication devices attached on objects, an alternative method on device-free passive wireless localization [8] has shown the feasibility of using radio signal dynamics for object detection in wireless environments. However, existing work on device-free passive intrusion detection has focused primarily on exploiting the detection power of a single intrusion indicator, such as the moving variance of RSS values. Besides, their work focuses on detecting intrusion events in a controlled environment. It has the ability in capturing one intrusion pattern when the intruders are moving around. They cannot detect the type of events when intruders are static. For instance, an intruder is standing and hiding, or a person is trapped in a fire building.

In real-world, there are many intrusion patterns such as standing, hiding, and moving towards a certain direction, which may be only detectable by employing the complementary detection power of different intrusion indicators jointly. To this end, in this work we identify different intrusion patterns and propose a joint intrusion learning approach by developing an algorithm called GREEK (Grid-based clustering over K-neighborhood), which has the ability in combining the detection power of several complementary intrusion indicators and enhance the performance of intrusion detection. In addition, our joint learning approach explores to profile the environmental uncertainties and detect the environmental changes caused by intrusion activities using collaborative efforts among multiple transmitter-receiver pairs.

To validate the effectiveness of the proposed approach, we conducted experiments in a real office environment using an IEEE 802.15.4 (Zigbee) network. Our experimental results show that GREEK is effective in diagnosing the presence of intrusions by performing grid-based clustering. Further, the

false positive rate can be minimized through collaborative detection from multiple transmitter-receiver pairs. Another interesting result of our analysis is that the multi-pair collaboration strategy can help to identify problematic wireless devices, which report unreliable data. Thus, our experimental results provide strong evidence for the effectiveness of our joint intrusion learning method.

The rest of this paper is organized as follows. We first put our work in the context of current research in Section II. We next discuss the feasibility of passive intrusion detection with problem overview and present our experimental methodology and threat model in Section III. Section IV describes the pattern profiling of different intrusions. We then present our joint learning scheme in Section V. In Section VI, we validate our approach and show the results of our experiments in the real office environment. Finally, we conclude our work in Section VII.

II. RELATED WORK

Different technologies can be used to detect intruders in various environments including corporate, civilian, and military. There have been active work using video, pressure, ultrasound, or infrared. [2] utilized video-based algorithms to analyze sequences of images captured by cameras and to track moving people. The video-based or surveillance-based technology is expensive, label intensive (when analyzing the video), and fails in dark or none-line-of-sight environments. Moreover, people being tracking may raise privacy concerns. [3] deployed air pressure sensors under the floor to detect the footsteps of people and build people-profile based on footsteps, and ultravision [9] produces ultra-sensor as motion detection sensors. These techniques require careful deployment of sensors, involve high cost, and are thus not easily scalable.

Further, [7] proposed methods relying on ultrasonic Time-of-Arrival (ToA) or Time-Difference-of-Arrival (TDoA) between ultrasound and RF signal to perform both static and mobile object localization. The sensor networks using ultrasound to conduct localization and possible object tracking require specialized localization infrastructure and each target object to carry a wireless device, e.g., a transmitter or a receiver. [10] used an infrared localization infrastructure to achieve location estimation. However, the infrared technology also needs a specialized localization infrastructure, which has a short range and requires dense deployment. The localization techniques can be further classified based on ranging methodology. Range-based algorithms involve distance estimation to access points using the measurement of various physical properties such as RSS [4], [11]–[13], ToA [6] and TDoA [7], whereas range-free algorithms [14] use coarser metrics to place bounds on candidate positions. All of these schemes require the target object to carry a radio device or actively participate in localization. In addition, there are several emerging commercial products for indoor localization or intrusion prevention [15]. However, these products usually require specific chip sets and operating systems.

The device-free localization is a new concept to localize and track target objects without carrying radio devices or actively participate in the localization process. [16] proposed to use the signal dynamic property between the static environment and the dynamic environment to conduct transceiver-free object tracking in wireless sensor networks. [17] deployed a Radio Tomographic Imaging (RTI) system, which used large number of wireless nodes to image passive objects within a wireless network. The works that are most closely related to ours are [1], [8]: [8] demonstrated the feasibility of device-free passive localization in a controlled environment using the moving average or variance of RSS to detect the events and proposed to conduct localization based on passive radio map construction. Whereas [1] performed passive event detection in real environments using the moving variance and resulted in a low precision. However, their detection capability is limited as only single intrusion indicator is exploited. Our work is novel in that we proposed a joint intrusion learning approach, which is generic and can combine the detection power of complimentary intrusion indicators. Our approach is highly flexible to incorporate new indicators, and consequently can maximize the performance of passive intrusion detection.

III. FEASIBILITY OF INTRUSION LEARNING

In this section, we first provide an overview of passive intrusion detection using RSS changes. We then discuss about our experimental methodology and the threat model.

A. Problem Overview

Passive intrusion detection based on the RSS data collected in pervasive wireless environments is especially attractive as it reuses the existing wireless environmental data without requiring a specialized infrastructure such as surveillance camera-based intrusion detection methods [2], [18]. The advantage is that no specialized infrastructure needs to be purchased and installed ahead of time, which may involve high cost and remain idle most of the time. Furthermore, under emergency situations such as emergency evacuation in a fire building, it is critical to detect and locate people trapped inside the building in a timely manner. However, it is hard, if not impossible, to install expensive intrusion detection systems in every building.

Therefore, it is desirable if we can reuse the environmental sensing data, such as RSS, which not only provides tremendous cost savings as the collected sensing data can be dual-used for intrusion learning, but also is available at any time for performing detection analysis. Further, even if the specialized hardware infrastructure has already been installed for important asset protection, the wireless environmental data can assist to refine the process of intrusion detection.

Although the radio signal is affected by reflection, refraction, shadowing and scattering, the RSS at wireless devices should be relatively stable if there is no movement or changes in wireless environments. On the other hand, the wireless environment will be affected if there is a presence of intrusions, for instance, an intruder standing or walking in a wireless environment will absorb, reflect, and diffract some of the transmitted power.
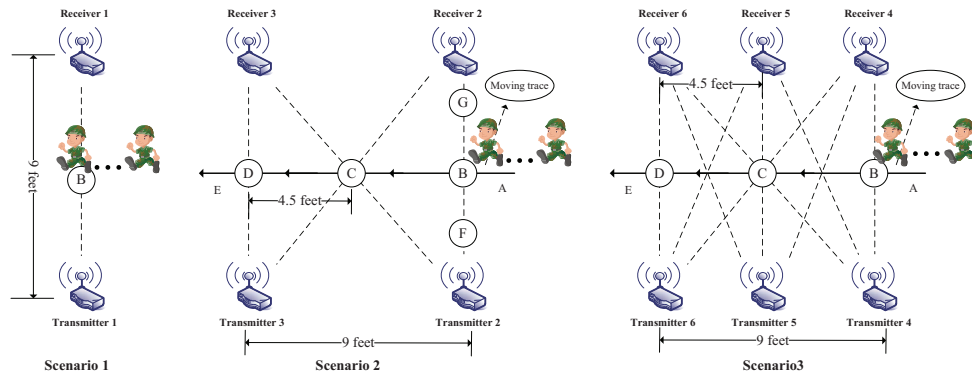
Fig. 1.   Our experimental setup when one or more intruders are present in the system.

Consequently, the RSS at wireless devices will be impacted and results in changes of RSS values. Therefore, based on the changes of RSS at wireless devices, it is possible to detect intrusion in wireless environments.

However, performing passive intrusion detection is challenging as the intruders usually do not carry any radios and are not cooperative. We identified the main issues that need to be addressed in order to make it feasible to perform passive intrusion detection in wireless environments. First, the collected RSS data is affected by the noisy environment such as the interference from other signals and measurement errors due to inaccurate off-the-shelf devices. Thus, it is desirable to filter out the measurement errors and clean up the noise mixed in the actual readings in order to facilitate effective intrusion learning. Furthermore, to perform intrusion learning, one important capability is to profile the RSS data and derive meaningful patterns for further learning analysis. For instance, the derived intrusion patterns should indicate whether the intruder is standing or walking, and to which direction the intruder is walking. Finally, by utilizing the different characteristics captured by intrusion patterns, the passive intrusion detection scheme should be able to effectively differentiate those intrusions based on different profiles. To address these issues, we present our pattern profiling approach and our detection scheme based on joint learning in the following sections.

B. Experimental Methodology & Threat Model

We conducted experiments in an uncontrolled wireless environment in the Burchard building at Steven Institute of Technology. The size of the lab is 15 feet by 18 feet. In our experiments, we used a Zigbee (802.15.4) network, which operates at the 2.4GHz frequency range. The wireless devices we used are the Tmote Sky motes with MSP430 microcontrollers, RF chip CC2420 and monopole antennas. We used the default transmission power which is 0 dBm for RF chip CC2420 in all the experiments. We deployed the motes under three different topologies as shown in Figure 1. The motes configured as transmitters broadcasted beacons periodically with 100 ms interval. Whereas the motes configured as receivers recorded transmitter's ID, RSS and time stamp of each beacon packet, and then forwarded these information to a central server.

In this study, we explore two representative types of intrusion events: *static* and *moving*. We define a static event when an intruder breaks in the area of interest and moves from one position to another, at each position the intruder stands still for a certain period of time. Whereas a moving event is defined for an intruder walking or running across the area of interest.

Since the passive intrusion learning relies on detecting the changes of wireless environments affected by the presence of intruders, the topology of the wireless infrastructure and the density of wireless devices may impact the effectiveness of learning. To analyze the impact we set up three experimental scenarios with different topology layouts and device densities. In our experiments, the time interval between two consecutive intrusion events in a series of events is around 180 seconds. We note that there can be multiple intruders present in the system. Since multiple intruders will cause more changes in wireless environments and have bigger impact on RSS readings, the detection of the presence of multiple intruders is easier than an individual intruder. We will focus on to present the results of individual intruder in the rest of our work. The detailed experimental setup of each scenario and behavior of the intruder are described below.

Experimental Scenario 1. In this scenario, there are one transmitter and one receiver in the area of interest. The distance between the transmitter and the receiver is 9 feet. This scenario may represent a low density environment in office buildings since there is just one transmitter-receiver pair which represents the wireless link between one wireless device and an access point. The receiver recorded packets for approximately 1560 seconds from the transmitter. There are three intrusion instances during this time period. For each instance, the intruder came in and stood at the center of the transmitter-receiver pair for about 120 seconds.

Experimental Scenario 2. We increased the density of the devices in this scenario, which may represent the typical density in an office building environment in which there are many wireless devices communicate with access points. There are two transmitters and two receivers deployed at four corners of the 9 feet by 9 feet square area. There are four transmitter-receiver pairs in total. Two receivers recorded packets for approximately 2400 seconds from two transmitters. There are nine intrusion instances including five static cases and four

moving cases during this time period. For each static intrusion instance, the intruder stood at different locations (shown as B, C, D, G and F in Figure 1) for about 120 seconds, whereas the intruder went across the experimental area for each observed moving instance.

Experimental Scenario 3. In this scenario, there are three transmitters and three receivers. The distance between two adjacent transmitter and receiver is 4.5 feet. There are nine transmitter-receiver pairs in total. The duration of this experiment is about 1800 seconds including seven intrusion instances in total with three static cases and four moving cases. The intruder stood for 120 seconds at three different positions (B, C, and D) for each static instance and went across the experiment area for each moving case. We envision there will be an increasing density of wireless devices deployed in our environments as the wireless networks become more pervasive. Thus, this set up with higher device density can help to analyze the impact of device density on diagnosing passive intrusion. In addition, wireless devices are usually not uniformly deployed. For instance, wireless devices (e.g., sensor nodes) can be deployed in a higher density in the sensitive area for asset protection and at the entrance or exit of the facility.

IV. PATTERN PROFILING

In this section, we present our method for data cleansing. We then describe our approach for intrusion pattern profiling.

A. Data Cleansing

The RSS measurements collected from wireless environments are from the spatial and temporal domain and are thus correlated to the position of the device and the time that the measurement is performed. Determining changes from RSS measurements are the basis for intrusion pattern profiling. However, the observed RSS can be very noisy due to interferences from other signals and measurement inaccuracy inherited from off-the-shelf wireless devices, especially in an uncontrolled environment. Thus, it is desirable to smooth out the noise and filter out the measurement errors, whereas keeping the true reflection of the spatial and temporal environment within the data measurements for intrusion pattern profiling.

In this study, we use the Alpha-trimmed Mean Filter [19] to perform data cleansing. Comparing to the other filters such as mean filter which works well with impulsive noise and median filter which works with Gaussian noise, Alpha-trimmed Mean Filter works well with multiple types of noise, such as combination of impulsive and Gaussian noise. In addition, Alpha-trimmed Mean Filter outperforms other nonlinear filters in simplicity and noise reduction when the data distribution contains some outliers, which is especially suitable for cleaning the RSS data. The idea is that rather than averaging the entire data samples observed in a time window, a few data samples are removed (trimmed) and the remainder are averaged by Alpha-trimmed Mean Filter. The data samples which are removed are the most extreme values, both low and high, with an equal number of data samples
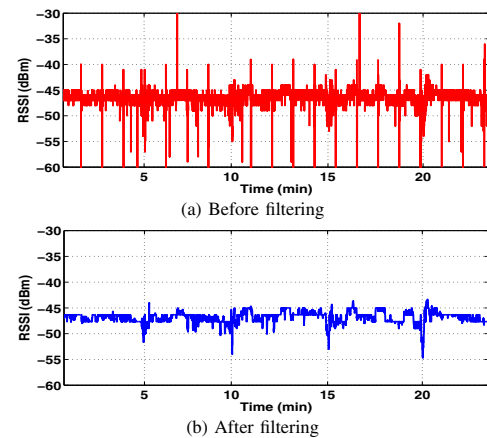

Fig. 2. Using Alpha-trimmed Mean Filter for RSS measurement cleansing with $n = 17, \alpha = 0.2$.

dropped at each end (i.e., symmetric trimming). The number of data values which are dropped from the average is controlled by the trimming parameter α, which assumes values between 0 and 0.5. Let $\{s(i), s(i+1), ..., s(i+n-1)\}$ be a set of n RSS sample values observed in each time window from one transmitter-receiver pair. We sort the RSS sample values in ascending order:

$$s_1(i) \leq s_2(i) \leq ... \leq s_n(i) \qquad (1)$$

where $s_1(i)$ is the minimum RSS value and $s_n(i)$ is the maximum value in the above RSS data set. Then, the output of the alpha-trimmed mean filter is

$$f(i; \alpha) = \frac{1}{n - 2\lceil \alpha n \rceil} \sum_{j=\lceil \alpha n \rceil + 1}^{n - \lceil \alpha n \rceil} s_j(i) \qquad (2)$$

where $\lceil \alpha n \rceil$ denotes nearest integers greater than or equal to αn and α indicates the percentage of the trimmed RSS samples, $0 \leq \alpha < 0.5$.

Figure 2 presents the noise suppression on RSS measurements over a period of 20 minutes from our experiments. There are four intrusion instances embedded in this measurement segment. It shows the RSS readings before and after applying the Alpha-trimmed Mean Filter. In Figure 2(a), the intrusion instances and environmental noise are mixed together. It is hard to determine the existence of intrusions. After applying Alpha-trimmed Mean Filter, we clearly observed that the RSS readings dropped at 5, 10, 15 and 20 minutes respectively, which corresponded to four possible intrusion instances during this time period. During our experiments, we found that the Alpha-trimmed Mean Filter can greatly reduce the noise presented in the raw measurement and is particularly effective in the presence of impulse noise.

B. Deriving Intrusion Patterns

By utilizing the data after filtering, we can explore various profiles to describe different intrusion patterns. In passive intrusion detection, it is essential to differentiate intrusion activities from random environmental changes. The critical property that a pattern profiling approach exhibits is that

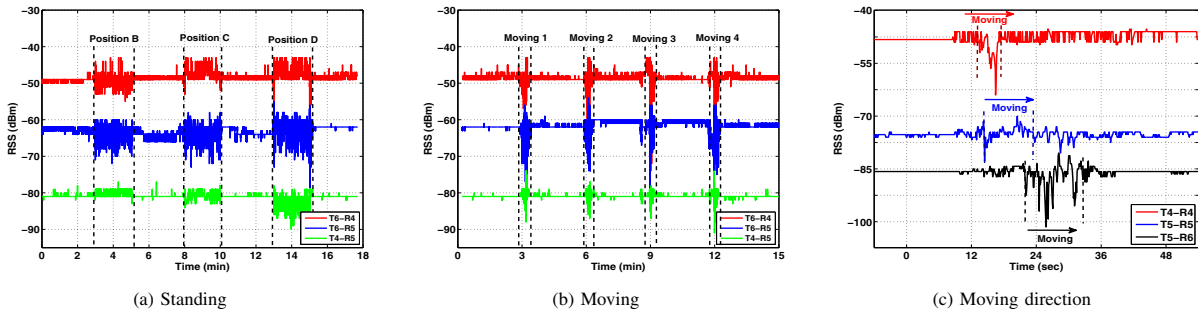| (a) Standing | (b) Moving | (c) Moving direction |

Fig. 3.   Pattern profiling of different intrusion cases.

it can drive unclear or complicated situations into separate categories, which make it possible for further analysis based on categorized information. This largely helps for passive intrusion learning as we can systematically detect the intrusion and define its characteristics.

Presence of Intrusions - Standing or Walking. As a first step, we need to detect the presence of intrusions and further to differentiate whether the intruder is standing or walking around, which may help to determine the purpose of intrusion. Figure 3 presents the results of two experiments under *Experimental Scenario 3* (as shown in Figure 1). In the first experiment, the experimenter stood at positions B, C, and D respectively and spent two minutes at each position. Whereas in the second experiment, the experimenter walked from position A to position E four times.

The three patterns in Figure 3 (a) represent the RSS readings for three transmitter-receiver pairs, *T6-R4*, *T6-R5*, and *T4-R5* respectively when the experimenter stood at positions B, C, and D respectively. In order to examine the changes of RSS clearly, we shifted the RSS readings by 15 dBm for *T6-R5* and by 30 dBm for *T4-R5*. We observed that there is an obvious change in RSS readings when the experimenter walked in and stood within the experimental area. Further, the results of the second experiment in Figure 3 (b) show that there is an obvious RSS pattern change for each moving instance. The key observation is that the RSS patterns when the experimenter is static are different from those when the experimenter is walking around. These results indicate that different RSS profiles can be established to distinguish the moving patterns of intruders.

Moving Direction. When the intruder is moving around, determining the moving direction of the intruder is also an important task in our exploration as the resulting pattern can help to direct further defense strategies, e.g., turning on the surveillance camera in one part of the floor or directing the law enforcement officers to follow the direction that the intruders go to. Figure 3 (c) presents the RSS readings for three transmitter-receiver pairs, *T4-R4*, *T5-R5*, and *T5-R6* respectively when the experimenter walked from position A toward position E. In order to examine the changes of RSS clearly, we shifted the RSS readings by 20 dBm for *T5-R5* and by 35 dBm for *T5-R6*. By combining the RSS readings from multiple sources, i.e., multiple transmitter-receiver pairs *T4-R4*, *T5-R5* and *T5-R6*, we can determine the moving direction

of the experimenter based on the moving pattern delay in time series. The moving direction can be further calculated as the positions of receivers are usually known and the locations of the transmitters can be localized easily using the traditional localization methods [4], [20].

V. INTRUSION LEARNING SCHEME

In this section, we first describe a declustering effect observed in our experimental study when intrusions are present. We then present our algorithm by performing grid-based clustering over K-neighborhood, which captures intrusion effects. Finally, we discuss our intrusion detection strategy utilizing multi-pair collaboration on top of grid-based clustering.

A. Declustering Effect

Figure 4 shows the relationship between the mean of RSS values and the variance of RSS values when intrusions are present. We applied a sliding window over the cleaned data to compute the mean and the variance of RSS readings within each time window. The window size is 10 seconds. In the figure, the small triangles indicate normal situations without intrusion, whereas the small squares and small circles represent situations when an intruder is either standing or walking across the experimental area respectively.

A clear observation is that the points in Figure 4 have a clustering effect under normal situations, whereas the presence of intrusions result in a declustering effect of points. In other words, the points are more spread when there are standing or moving intruders. Further, in Figure 4 we observed that the presence of intrusions is obvious with the combining view of the mean and variance of RSS. Thus, we call the mean and variance of RSS readings as *intrusion indicators*, which can be used to diagnose intrusion activities jointly.

B. Grid-based Clustering over K-neighborhood (GREEK)

To capture the declustering effect under intrusion, we developed an algorithm called *GREEK*, grid-based clustering over K-neighborhood, which combines the detection power of different intrusion indicators to perform a joint learning of intrusions.

Algorithm Overview. Since a grid is an efficient way to organize a set of data points [21], GREEK breaks the d-dimensional space of d different intrusion indicators into grid cells, and then performs clustering based on the density of
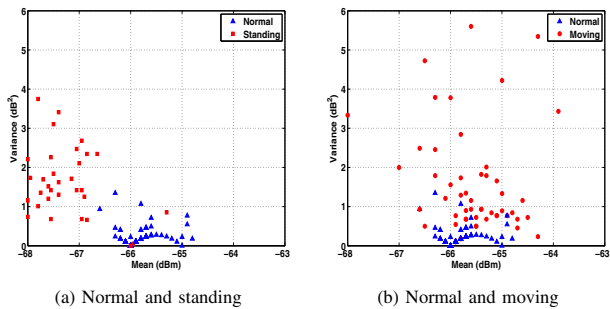
(a) Normal and standing   (b) Normal and moving

Fig. 4. Comparison of the clustering effect in normal conditions without intrusion and the declustering effect under intrusions.

(a) Data points under normal conditions   (b) Clustering results of data with intrusion

Fig. 5. Illustration of GREEK performing grid-based clustering in *Experimental Scenario 1*.

data points in each grid cell. The objective of GREEK is to capture the declustering effect by partitioning the data into two clusters so that one cluster has a higher density (i.e., the normal data without intrusion) and the other has a lower density (i.e., the data with intrusion). GREEK consists of two main steps: *density calculation using grid* and *cluster forming incorporating K-neighborhood*.

Algorithm Flow. Suppose there are d different intrusion indicators, which construct the d-dimensional data space $\mathcal{S}$. Let $D = \{S_1, S_2, ..., S_n\}$ be the input of data and $S_i = \{s_{i1}, s_{i2}, ..., s_{id}\}$, where s_{ij} is the value of the ith data point of the jth intrusion indicator. GREEK partitions the whole data space $\mathcal{S}$ into non-overlapping grids by partitioning each dimension into N equal length intervals. The intersection of one interval from each dimension (i.e. intrusion indicator) forms one grid, which can be denoted by the form $\{g_1, g_2, ..., g_d\}$ where $g_j = [r_j, l_j)$ is one interval of jth dimension. The data point S_i falls into a grid if $r_j \leq s_{ij} < l_j$ for all j (i.e. $j = 1, 2, ..., d$). All data points are placed into grids based on this simple criteria.

Definition 1. A grid is a **core grid** if the number of points fall into the grid exceeds the density threshold τ, which can be derived empirically.

Definition 2. A grid is a **border grid** if the number of points falling into the grid is less than τ. And the grid is within the K closest neighborhood (i.e. K-neighborhood) of a core grid. In this study, we set K to 8.

Definition 3. A grid p is **directly reachable** from a grid q if p belongs to q's K-neighborhood and q is a core grid.

Definition 4. A grid p is **reachable** from a grid q if there exist some intermediate grids $p_1, p_2, ..., p_n$, $p_1 = p, p_n = q$ such that $p_{(i+1)}$ is directly reachable from p_i.

Definition 5. A grid p is **connected** to a grid q if there is a grid o such that both p and q are reachable from o or if p and p are all core grids and the distance between these two grids is less than the Manhattan distance D_M.

Definition 6. A **grid cluster** is a maximal set of connected grids.

Finally, GREEK partitions all the data points that fall into the grid cluster into a dense cluster (with higher density) and the rest of the points into a sparse cluster (with a lower density).

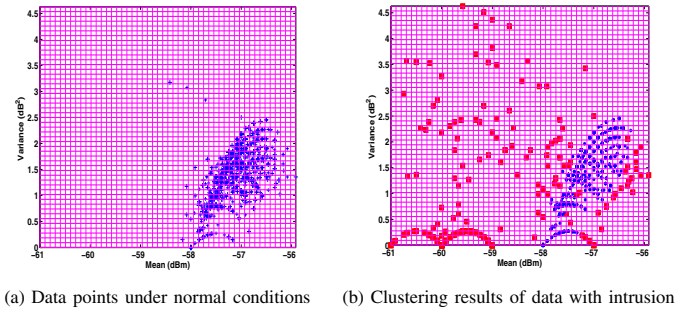**Example.** We used a two-dimensional space of intrusion

indicators to illustrate how GREEK might operate. The two intrusion indicators we use to validate GREEK are: the average mean and the variance of RSS. The grid size is set to 50 and there are 250 grids in total. We derived the density threshold τ and the distance threshold D_M based on our empirical study using data without intrusion events. Figure 5 (a) shows the data points of the mean and variance of RSS under normal conditions and Figure 5 (b) presents the results of clustering using GREEK for *Experimental Scenario 1*. We observed that the experimental data points have been partitioned into two clusters with different densities in Figure 5 (b). In particular, we found that points from normal data without intrusion have been placed in the dense cluster (represented by asterisk dots), whereas points of intrusion data have been placed in the sparse cluster (shown in square dots). The key observation here is that the cluster formulation by GREEK is consistent with the distribution of data points presented in Figure 5 (a), indicating that clustering based on grid density is feasible to capture intrusion effects.

C. Detection Strategy

Once the data points are partitioned into two clusters by GREEK, these two clusters will have significantly different density when there are intrusions present. In other words, when intrusions are present, the sparse cluster produced by GREEK represents the points affected by intrusion, whereas the dense one representing points not affected by intrusion (i.e., normal conditions).

To measure whether the density distributions of the two partitioned clusters are significantly different, we use statistical hypothesis testing where the null hypothesis is defined as:

$$\mathcal{H}_0 : \textit{two clusters are statistically the same.}$$

We perform a student t-test [22] on the two clusters obtained from GREEK to calculate the p-value of the test statistic. The goal is to see whether the resulted p-value is less than the significant level[1]. If the p-value is larger than the significant level, the null hypothesis is accepted.

In particular, suppose the two clusters obtained from GREEK are $C_1 = \{p_1, p_2, ..., p_n\}$ and $C_2 = \{q_1, q_2, ..., q_m\}$. The value of the test statistic t of these two clusters can be

---

[1]In this paper, we use standard R package for student t-test.

calculated by:

$$t = \frac{\overline{C}_1 - \overline{C}_2}{S_{\overline{C}_1 - \overline{C}_2}}, \qquad (3)$$

where

$$S_{\overline{C}_1 - \overline{C}_2} = \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}. \qquad (4)$$

$\overline{C}_1$ and $\overline{C}_2$ are the means of cluster C_1 and C_2, respectively. S_1 and S_2 are the variances of cluster C_1 and C_2, respectively. Further, the distribution of the test statistic t is approximated as an ordinary Student's t distribution with the degrees of freedom calculated using Welch-Satterthwaite equation:

$$D.F. = \frac{(S_1^2/n + S_2^2/m)^2}{(S_1^2/n)^2/(n-1) + (S_2^2/m)^2/(m-1)}. \qquad (5)$$

Given the value of the test statistic t and the degrees of freedom, we obtain the p-value from the student's t distribution. By comparing p-value with the significant level, the decision can be made on whether to accept the null hypothesis or not.

If the p-value is less than the given significant level, indicating that the density distributions of the two clusters are significantly different from each other, the null hypothesis is rejected and the presence of intrusions is declared.

Multi-pair Collaboration. Furthermore, intrusion detection relying on the data from only one transmitter-receiver pair may provide weak evidence of the presence of intrusions and trigger a false positive. On the other hand, as wireless networks become more pervasive, the wireless devices can be deployed with sufficient density such that the same intrusion event can be observed from the data collected from multiple transmitter-receiver pairs. Thus, to reduce false positive and increase the detection accuracy, we exploit using the observations from multiple transmitter-receiver pairs to collaboratively determine the presence of intrusions. Integrating multi-pair collaboration with GREEK can enhance the reliability of intrusion detection, especially in noisy environments.

In multi-pair collaborative learning, we define an *alert* when one transmitter-receiver pair has p-value below the significant level. We declare the presence of intrusions when the number of alerts among all available transmitter-receiver pairs during one examining time period exceeds the learning threshold. The learning threshold in our strategy is adjustable based on the total number of available transmitter-receiver pairs in the area of interest.

VI. EXPERIMENTAL RESULTS

In this section, we present our experimental results obtained under *Experimental Scenario 1, 2,* and *3*. We use mean and variance as our intrusion indicators to validate the joint learning scheme. We note that our joint learning approach is generic and can incorporate other intrusion indicators easily.

A. Evaluation Using t-test

We first study the results of t-test for statistical hypothesis testing. Tables I, II and Figure 6 present the t-test results of various intrusion events from *Experimental Scenario 1, 2,* and

Events	Type of Intrusion	P-value
Event 1	Static	0.005316
Event 2	Static	0.000506
Event 3	Static	0.000217
No Event	NaN	0.4249

TABLE I
t-TEST RESULTS FOR EXPERIMENTAL SCENARIO 1.

3 respectively. By examining the t-test results of static events in all three scenarios, from Table I we observed that all of the p-values are much smaller than the significant level (i.e. 0.05) for static intrusion events in *Experimental Scenario 1*. In addition, we observed lower p-values for static events in Table II and Figure 6 consistently for all the transmission-receiver pairs (except *T6-R6*) when comparing to the significant level. This indicates that the partitioned two clusters have significantly different distributions (i.e., the declustering effect) and there are intrusions present in the wireless environments. Further, under normal conditions, the p-values in both Table I and II are much larger than the significant level, indicating that the partitioned two clusters do not have significant difference, and thus there is no intrusion present in the system.

Furthermore, from Tables II and Figure 6, we have similar observations for moving events where most of the p-values are much lower than the significant level when experimenters are walking around in the experimental area. These results are encouraging as they indicate that our detection strategy based on hypothesis testing using t-test is feasible in diagnosing the presence of intrusions.

We also found that p-values of certain transmitter-receiver pairs are above the significant level when intrusions are present. For instance, in Table II, *T2-R3* of events 6 and 8, *T2-R2* of event 8, *T3-R3* of events 8 and 9, and *T5-R5* of event 4 (in Figure 6) have p-values larger than 0.05. This may be due to signal interference or random noise in the environment, which causes these transmitter-receiver pairs failed to observe the presence of intrusion. Further in Figure 6, we observed that *T6-R6* of all seven events have p-values above 0.05. This indicates that the *T6-R6* pair may have deficiencies in devices and is unreliable to report data.

However, the presence of the intrusion can still be detected by using the multi-pair collaborative strategy in our joint learning scheme. By using the multi-pair collaborative approach,

Events	Type of Intrusion	P-value of each pair			
		T2-R3	T2-R2	T3-R3	T3-R2
Event 1	Static B	0.0001	0.0132	0.0178	0.0096
Event 2	Static C	0.0001	0.0188	0.0237	0.0102
Event 3	Static D	3.297e-05	6.426e-06	0.0423	0.0069
Event 4	Static F	7.913e-05	6.94e-07	0.0229	0.0070
Event 5	Static G	7.593e-05	5.819e-06	0.0213	0.0048
Event 6	Moving	0.969	0.0001	0.0182	0.0085
Event 7	Moving	0.0033	2.835e-06	0.0056	0.0437
Event 8	Moving	0.3275	0.5006	0.0697	0.0186
Event 9	Moving	0.0033	0.0021	0.0584	0.0044
No Event	NaN	0.1813	0.1491	0.5582	0.6302

TABLE II
t-TEST RESULTS FOR EXPERIMENTAL SCENARIO 2.

(a) P-values   (b) Zoomed-in P-values
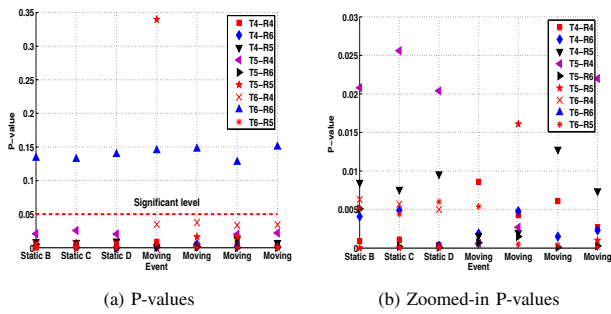
Fig. 6. *t*-test results of *experimental scenario 3*.

(a) Experimental scenario 2   (b) Experimental scenario 3

Fig. 7. False positive rate and detection rate when using multi-pair collaboration for intrusion detection in *experimental scenarios 2* and *3*.

we evaluate the p-values of all the transmission-receiver pairs in the same examination time period and report the detection of intrusion when at least one pair has p-value less than the significant level. That is, even when some transmitter-receiver pairs failed to observe the presence of intrusions, other pairs in the close-by neighborhood can complement the detection function, and consequently maximize the detection power.

B. Effectiveness

Performing passive intrusion learning accurately is challenging as we need to differentiate an intentional intrusion scenario from noisy environments. If not handled properly, the disturbance caused by noises would easily trigger false detection. Thus, for a scheme to be effective for passive intrusion learning, it is crucial to minimize the false positive rate, while achieving high detection rate. In this section, we study the effectiveness of our multi-pair collaborative strategy in terms of false positive rate and the corresponding intrusion detection rate.

Figure 7 presents the false positive rate and the detection rate when using different number of alerts to determine the presence of intrusions for *Experimental Scenario 2* and *3* respectively. We observed that when increasing the number of alerts, which work collaboratively to determine the presence of intrusions, the false positive rate goes down to 0% quickly with 2 alerts for *Scenario 2* and 3 alerts for *Scenario 3*. Further, we found that in *Scenario 2* the detection rate is about 89% when the number of alerts is set to 2 and 3, whereas in *Scenario 3*, the detection rate keeps at 100% when the number of alerts ranges from 1 to 7. Comparing to [1], which resulted in a low precision of about 0.3, our approach achieves a much higher precision by performing joint learning. In particular, the precision is 0.9 when alerts is 2 and 3 in *Scenario 2* and 1.0 when alerts ranges from 3 to 7 in *Scenario 3*. Thus, the key observation here is that using multi-pair collaborative detection can significantly reduce the false positive rate, while keeping high detection rate, indicating that detection using multiple transmitter-receiver pairs is highly effective in differentiating intentional intrusions from random environmental changes.

C. Impact of Device Density

We next examine the impact of wireless device density on the performance of our detection scheme. Based on the results in Figure 7, we found that the performance of intrusion
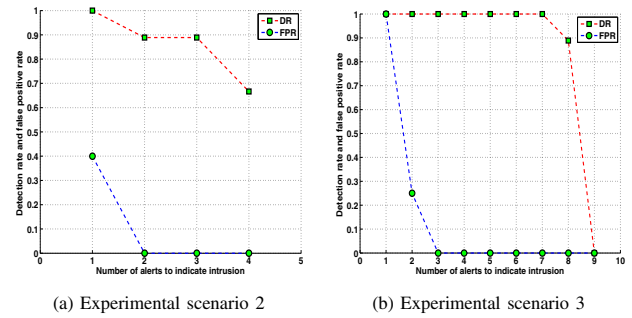
detection in *Experimental Scenario 3* outperforms that in *Experimental Scenario 2*. In *Scenario 3*, the density of wireless devices is higher than that in *Scenario 2*. This indicates that when intrusion is present in the area of interest, there are potentially more wireless devices (transmitter-receiver pairs) in *Scenario 3*, which are affected by the presence of intruders, than those in *Scenario 2*. Consequently, when diagnosing the presence of intrusions, there are more wireless devices that can work together in *Scenario 3* to perform intrusion detection collaboratively.

On the other hand, our multi-pair collaborative strategy also brings another dimension of knowledge of identifying problematic wireless devices. For instance, we observed that the transmitter-receiver pair *T6-R6* in Figure 6 consistently has high p-values, which is different from its neighboring pairs, indicating that the wireless devices, *T6* and *R6*, are not reliable during data collection and may have hardware deficiencies. Our future work will further quantify the relationship between the detection power and the density of wireless density.

D. Differentiating Intrusion Events

Detecting the presence of intrusions in the system provides first order information towards defending against them. Learning the different intrusion patterns allows the system to further determine the appropriate defense strategies in the next step. For example, differentiating the intruder is hiding in a place or moving around will enable the next step of action to either capture the intruder or follow him; or learning a person is trapped in a fire building will allow the firefighters to determine the best rescue strategy. Based on the results in Figure 3, we observed that the moving intrusion instances tend to result in larger variance of RSS when compared with the static intrusion instances. This suggests that it is feasible to use the moving variance to differentiate intrusion patterns in passive intrusion detection.

Figure 8 presents our results of accumulated moving variance for each intrusion event in *Experimental Scenario 2* and *3* respectively. For each transmitter-receiver pair, we used a sliding time window of 10 seconds to calculate the moving variance, we then average the computed moving variance over all the time windows during the examining time period for each pair. The moving variance shown in Figure 8 is the accumulated results by adding up the averaged moving variance from each pair. We observed that the values of variance of

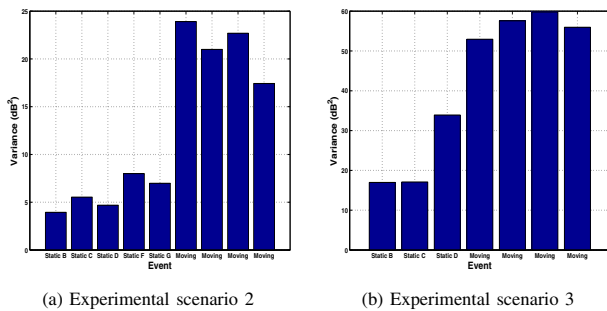(a) Experimental scenario 2   (b) Experimental scenario 3

Fig. 8. The accumulated moving variance of each event under different experimental scenarios.

static events are much smaller than those of moving events. This is because intruders moving around cause changes in wireless environments constantly, which results in higher value of variance. Therefore, by using accumulated moving variance in our joint learning scheme, we can effectively differentiate intrusion patterns.

VII. CONCLUSION

In this paper, we proposed to perform joint learning for detecting intrusion when intruders do not carry any wireless devices (e.g., intrusion to corporate assets or people trapped in a fire building). Our joint intrusion learning approach combines the detection power of complementary intrusion indicators and has the capability to detect different intrusion events in wireless environments. In particular, we utilized the Received Signal Strength (RSS) from the existing wireless infrastructure and exploited to use the changes of RSS caused by intrusions for diagnosing the presence of intrusions. We profiled environmental uncertainties through data cleansing and intrusion pattern derivation. We developed the grid-based clustering over K-neighborhood (GREEK) algorithm, which captures the declustering effect in intrusion indicators when intrusions are present. Additionally, our detection strategy utilizing multi-pair collaboration can enhance the reliability of intrusion detection under noisy environments.

We conducted experiments in a real office environment using an 802.15.4 (Zigbee) network. We evaluated the performance of our joint intrusion learning approach using false positive rate and detection rate. Our experimental results provide strong evidence of the feasibility of performing joint learning for passive intrusion detection. Moreover, the results show that our strategy of using collaborative efforts across multiple transmitter-receiver pairs can complement the detection function and maximize the detection power with a minimum false positive rate (zero percent). Finally, an interesting observation is that the collaborative detection strategy can also bring another dimension of knowledge of identifying problematic wireless devices, which report wrong signal readings.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Moussa and M. Youssef, "Smart devices for smart environments: Device-free passive detection in real environments," in *The 2nd IEEE Workshop on Intelligent Pervasive Devices*, March 2009.

[2] Q. Cai and J. K. Aggarwal, "Automatic tracking of human motion in indoor scenes across multiple synchronized video streams," in *Proceedings of the Sixth International Conference on Computer Vision*, 1998.

[3] J. O. Robert and D. A. Gregory, "The smart floor: a mechanism for natural user identification and tracking," in *Proceedings of the CHI 200 Conference on Human Factors in Computing Systems*, 2000.

[4] 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.

[5] 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.

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

[7] 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, pp. 32–43.

[8] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*, Aug 2007.

[9] "Ultravision Corporation," white paper available at http://www.ultravisionsecruity.com.

[10] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, Jan. 1992.

[11] K. Kleisouris, Y. Chen, J. Yang, and R. P. Martin, "The impact of using multiple antennas on wireless localization," in *Proceedings of the Fifth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, June 2008.

[12] J. Yang and Y. Chen, "A theoretical analysis of wireless localization using RF-based fingerprint matching," in *Proceedings of the Fourth International Workshop on System Management Techniques, Processes, and Services (SMTPS)*, April 2008.

[13] 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)*, 2005, pp. 91–98.

[14] T. He, C. Huang, B. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes in large scale sensor networks," in *Proceedings of the Ninth Annual ACM International Conference on Mobile Computing and Networking (MobiCom'03)*, 2003.

[15] "Airtight networks," white paper available at http://www.airtightnetworks.net.

[16] D. Zhang, J. Ma, Q. Chen, and L. M. Ni, "An rf-based system for tracking transceiver-free objects," in *Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2007.

[17] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," in *Tech Report*, Sep. 2008.

[18] D. B. Yang, H. H. Gonzalez-Banos, and L. J. Guibas, "Counting people in crowds with a real-time network of simple image sensors," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003.

[19] J. Bednar and T. Watt, "Alpha-trimmed means and their relationship to median filters," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 1, pp. 145–153, Feb 1984.

[20] 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.

[21] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2005.

[22] G. Casella and R. L. Berger, *Statistical Inference*. Belmont, California: Duxbury Press, 1990.