Exploiting Dispersive Power Gain and Delay Spread for Sybil Detection in Industrial WSNs: A Multi-Kernel Approach

Qihao Li, Michael Cheffena

***Abstract*—Industrial wireless sensor networks (IWSNs) promote innovations in industry such as structural status mapping, instrument fault diagnosing and oriented automation system associating, among others. However, due to the shared nature of the wireless propagation environment, the emerging sensor nodes (SNs) with wireless properties are vulnerable to external malicious attacks. Security threats, especially Sybil attacks, impose great difficulties in fulfilling quality requirements of industrial applications. What's more complicated is that the harsh industrial environment brings about new challenges which can degrade the accuracy of detecting Sybil threats. In this paper, we focus on how to detect the malicious packets transmitted from Sybil attackers without adding extra authentication overhead into the transmission frame. We develop a Multi-Kernel-based Expectation Maximization (MKEM) scheme to detect Sybil attacks in IWSNs. Instead of directly investigating the radio resource of SNs, we produce channel-vectors which are extracted from the power gain and delay spread of the channel impulse response obtained from the received packets to represent each SN. Specifically, a kernel-oriented method is designed to discriminate the malicious packets from benign ones without establishing a pre-defined database of channel features of all SNs. Meanwhile, we allocate different kernel weights to the proposed kernels and combine them to improve the discrimination ability of the scheme. Moreover, a kernel parameter optimization method is developed to regulate each kernel weight and parameter to reduce the effects of transmission impairments in IWSNs. To avoid poor detection accuracy when the number of Sybil attackers increases, we use Gap statistical analysis method to verify and Expectation Maximization (EM) method to summarize the detection results. The simulation results show that the proposed MKEM scheme can achieve high accuracy on detecting malicious packets transmitted from Sybil attackers from benign ones, and tolerate the effects of transmission impairments in the industrial environment. Moreover, MKEM scheme can guarantee the detection accuracy even if the number of Sybil attackers increases.**

***Index Terms*—Industrial WSNs, Sybil detection, Power Delay Profile, Kernel Optimization, Multiple Kernels.**

I. Introduction

INDUSTRIAL Wireless Sensor Networks (IWNSs) is a cybernetic system which builds up a tight coordination among different computational physical elements [1]–[4]. However, the shared nature of the wireless medium makes IWSNs vulnerable to identity-based attack, as attackers gather useful identity information during passive monitoring and utilize them to launch attacks. In particular, **Sybil attacks**, which may incur data distortion or even malfunction of the whole system, play like a typical harmful attack against sensor networks [5]–[7]. In Sybil attacks, a malicious sensor node (SN) can manipulate large numbers of pseudo packets forging different identities to reduce the reliability of the networks [8]. In IWSNs, these adversarial intrusions may result in harmful and disastrous accidents. For instance, in an industrial storage room, the temperature in storage areas for flammable materials should be kept below their burning point. Versatile SNs can be installed in storage containers to monitor the temperate. All the measurement results can be gathered by a handle controller (HC) carried by workers. If an attacker confuses a SN, in Sybil attacks, s/he can claim a large number of SNs, either by impersonating other legal SNs or claiming false identities. After that, the attacker can send several false alarms to the HC and then destroy the dependability of the sensor networks. If there were no periodical and dependable alarm from SNs, the leakage of flammable materials would go unnoticed, which could contaminate the industrial environment and even endanger the public safety. As a special kind of denial-of-service attack, Sybil attacks seriously endanger the dependability of IWSNs.

Some research efforts have been made on Sybil detection in traditional WSNs. The Sybil attack in WSNs is systematically analyzed in [9] in terms of attack taxonomies. Generally, the defensive scheme against Sybil attack can be categorized into two methods: *key pre-distribution* [10] and *radio-resources testing* [11]. In the *key pre-distribution* method, the key paired identification and key validation verification are proposed to improve the data confidentiality. However, to ensure the network scalability, portability and interoperability among various industrial applications, uncomplicated association protocols and frame structures are applied for IWSNs. These intentions impair the capability of the cryptographic authentication favoring external key management and framework overhead. The conventional key pre-distribution methods based on heavy authentication protocols are not suitable to be employed in IWSNs. Thus, in this paper, we detect the Sybil attackers by exploring the *radio-resource testing* method, especially, by measuring the channel features of the Sybil attackers.

In the measurement-based Sybil detection method, the channel characteristics, such as the strength of the received signal strength indicator (RSSI), the received channel state information, the variation of the received signal phase, etc.,

Manuscript received Jun. 28, 2018; revised Sep. 08, 2018; accepted Jan. 27, 2019. The associate editor coordinating the review of this paper and approving it for publication was M. Xiao.

Q. Li and M. Cheffena was with the Faculty of Engineering, Norwegian University of Science and Technology Teknologivn. 22, Gjøvik N-2815, Norway e-mail: {qihao.li,micheal.cheffena}@ntnu.no.

are used to discriminate physical terminals. Although this detection method avoids the disadvantage of the preceding key pre-distribution method, it leaves some other challenging security issues for the new industrial environment. First, it is difficult to detect Sybil attackers without establishing a pre-labeled database of channel features of each SN in the network. In most cases, Sybil attackers are detected by assuming that the channel features of Sybil attackers can be measured in advance as predefined labels which are used to represent attackers. However, due to the open source and dynamic deployment nature of IWSNs, it is practically impossible to establish a standard database to record all the channel feature information for each SN. Second, the propagation impairments, impulse noise and interference effects in the harsh industrial environment may reduce the reliability of the measurement-based Sybil detection methods. Specifically, industrial environments may contain a lot of steel, metals, and machinery, which produce multipath fading and interference effects to deteriorate the wireless propagation performance. These fading dips can be far down below the proper threshold in a short period and degrade the received signal strength by as much as $30 - 40$ dB [12]–[14]. Moreover, electrical motors, cranes, and vehicles can also have an impact on communication systems. These interferences are composed of random high energy spikes which occur randomly. They are not like Gaussian noise and affect the sensor network differently [12], [15]. Furthermore, a large group of multipath propagation signals may be received with the same signal phases after a long-distance transmission. A large received signal power will then be detected, which implies that the receiver may be next to the transmitter. In addition, different path-loss exponents and shadowing variances that are affected by scatterers in the channel may result in dissimilar received signal power. If the SNs are surrounded by scatterers composed of different materials, some traditional radio-source testing method may not be applicable for IWSNs. Also note that the vibrating scatterers and moving people bring further changes to the phase and Doppler frequency of the received signal. Third, some smart attackers can make their behaviors plausible by adjusting their radio resources. This may reduce the Sybil detection accuracy as the Sybil SNs may forge several pseudonymous entities by claiming the differences of their channel features. Even worse, when the number of malicious SNs is increased, the network would be filled with pseudonymous entities with different channel features. Since IWSNs are mostly deployed in the industrial environment, SNs in the networks tend to provide similar propagation channel features. Consequently, there is high chance that these bogus entities belonging to different adversaries may be detected as benign SNs. Given these challenges of Sybil detection, we develop a new Sybil detection scheme to fulfill the dependable requirement in IWSNs.

In this paper, we propose a Multi-Kernel-based Expectation Maximization (MKEM) scheme to detect the Sybil attacks in IWSNs. We discriminate the Sybil SNs from benign ones by considering a fuzzy c-means algorithm without establishing a database of pre-recorded channel feature information of Sybil attackers. We reduce the influence of channel impairments on the Sybil detection results by combining multiple kernels and properly optimizing the kernel parameters. In addition, we modulate the detected number of the malicious SNs and increase the accuracy of identifying the Sybil attacker by investigating the expectation maximization algorithm when the number of malicious SNs is increased. We have preliminarily studied the propagation channel features of a SN and proposed a power gain and delay spread (PGDS) scheme to detect the Sybil attacks in [5]. In the PGDS scheme, we extract the power gain and delay spread from the channel impulse and combine them as a channel-vector to represent the related SNs. However, the PGDS scheme cannot satisfy the dependable requirement of IWSNs when the effect of noise and interference is increased. To improve the dependability of Sybil attack detection, a multi-kernel scheme is proposed afterward in [16]. This scheme can achieve high accuracy in identifying the packets sent by Sybil attackers and tolerate the time-varying attenuation in an industrial environment. However, the multi-kernel scheme loses its dependability when the number of Sybil attackers increase and fails to change its kernel parameters to adopt to a new environment.

In this paper, the MKEM scheme is proposed to address these problems. Specifically, the main contributions of this paper are as follows.

- **First**, by mapping the obtained channel-vectors into a higher dimensional Hilbert space, unsupervised fuzzy c-means method composed with multiple kernels is designed to distinguish between benign and malicious packets which are received by the SNs throughout the industrial wireless environment.
- **Second**, we produce a kernel parameter selection method (KPS) to iteratively optimize the parameters in the proposed kernels. Since we can use a kernel function to describe the difference between channel-vectors, we enlarge the included angles between the channel-vectors in different clusters and decrease them within the same clusters to improve the noise reduction ability of the proposed scheme. In addition, we build up the multi-kernel structure by linearly combining all the proposed kernels with different weights. By appropriately calculating the weights of different kernels, we ensure the MKEM Sybil detection scheme can adjust to different industrial propagation environment.
- **Third**, we investigate the Gap statistics analysis to detect the malicious SNs through the obtained number of classification from the multi-kernel method. We further provide an iterative procedure, expectation maximization (EM) algorithm, to optimize the mean and variance of incomplete data sets with respect to the latent variables. When the number of the malicious SNs is increased, the EM algorithm can improve the detection results with the obtained channel-vectors pre-labeled by the kernel method.

The remainder of this paper is organized as follows. In Sec. II, we review the related works on channel-based Sybil detection schemes. Then, we present the network and attack models in Sec. III. In Sec. IV the detailed MKEM scheme is

provided. Finally, we evaluate the performance of our scheme in Sec. V and conclude the paper in VI.

II. Related Works

Wireless transmission in IWSNs may allow various attacks such as eavesdropping, modification and fabrication, making it liable to harmful and disastrous accidents in industrial environments caused by adversarial intrusions [17], [18]. Sybil attack, which is firstly presented and studied for sensor networks in [19], refers to the behavior of deceiving the network with large numbers of multiple forged identities by either impersonating other benign SNs or claiming fake identities. To protect against this injury in the industrial environment, an invasion detection scheme is necessary to be designed. A plethora of research efforts have been made on the measurement-based Sybil detection for conventional sensor networks [20]–[22]. Their main objective is to maximize the disparity between different channel responses by exploring the spatial variability of the wireless channel. For example, [10] directly exploits the RSSI and conducts the ratios between RSSIs received in pre-ascertained controllers from the SNs that are under assessment. Then, these detection results are analyzed by comparing the proportions in different timestamps. However, since the RSSI is influenced by the topologies and structures of the industrial buildings, the decision error is likely to be increased if the RSSI without any calibration is utilized [15]. In addition, the material properties, size and distribution of the scatterers also have an impact on the path-loss exponent of the wireless propagation channel such that the RSSI may vary with the elapse of time and variation of the surrounding environments [23]. The spatial correlation features of RSSIs are further discussed in [24] and [25] by evaluating the probability distribution variance of RSSI ratios between SNs and landmarks. The detection results are obtained by minimizing assessment errors, which are the results of mistaking RSSIs from the same physical location for those coming from different locations. Besides, a RSSI-based cooperated detection scheme is designed in [26] via recording the RSSI observations from neighboring SNs. The recorded RSSIs are further compared among neighboring SNs by exchanging information so that the malicious SNs can be detected. Given the complex characteristics of the practical channel, [8] proposes a channel-based generalized likelihood ratio test to ascertain the Sybil attacks via considering the decorrelation of the channel response in the space. By doing this, similar channel-vectors extracted from the received packets with different identities can be argued to be at the same location. Likewise, by exploring the radio channel information, [27] extends the utility of the hypothesis test method and formulates the interaction between the benign users and their adversaries as a zero-sum authentication game to detect the spoofing attacks. Specifically, [27] designs a Dyna-Q learning method to achieve the optimal test threshold in the hypothesis test for a dynamic radio environment. In [28], a semi-supervised learning approach is considered to explore the threats with a joint probability distribution over all pre-defined labels of benign as well as Sybil SNs. The proposed scheme meditates the local information around the SNs according to the knowledge of vertices and edges in the network. Tags of pre-defined SNs are spread to the remaining SNs and further used to rate malicious SNs.

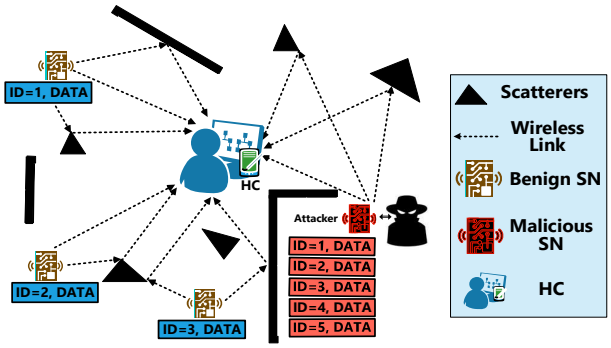


Fig. 1: Network Model

As can be seen from the above discussions, in the specific industrial wireless propagation environment, various properties of the multiple scatters may increase the effect of the propagation impairments, which may influence the detection accuracy if the proposed detection schemes merely utilize the information of RSSI or channel features. In addition, impulse noise recurred at short intervals and interference effects caused by vibrating scatterers may also have an impact on the detection results. Moreover, the detection schemes based on a supervised learning approach are hard to be realized. It is undesirable to provide pre-defined labels and pre-explored features to both benign and malicious SNs under time-varying channel properties and uncertain characteristics of the attackers. Furthermore, although several current learning-based works can detect the attackers by maximizing a proposed reward objective function through trying actions in different situations, a partially observed set of system states, their transition probability, and actions which can affect the states are still necessary to be attained before achieving the goal. Considering all these, we introduce a MKEM Sybil detection scheme that is built on the constituted power gain and delay spread extracted from the channel power delay profile. A kernel-based clustering method is employed to minimize the multipath distortion and interference effects in the industrial environments.

III. System Model

A. Network Model

IWSNs react to the SN data by issuing control signals via a controller to physical components. As shown in Fig. 1, the network consists of two primal ingredients: **SNs** and a **HC**. Carrier-sense multiple access with collision avoidance (CSMA/CA) are used to ensure their association [29]. The scatterers, made of different materials, are distributed uniformly in the environment. They reflect the signal according to their material properties [15], which can cause heavy multipath propagation that impair the received signal. To satisfy the network hierarchy and keep per-terminal cost low, SNs only have a limited storage and computational resources so that tamper-resistant hardware is un-occupied [19]. We assume that SNs are artificially deployed in the environment and cannot

move after deployment. Moreover, SNs can monitor their surroundings functionally and transmit the obtained data to the HC. HC is a wireless receiver, which can harvest and inquire information from the SNs. Furthermore, to diminish the Doppler shift effects, we assume that the HC stays in a place where the SNs' packet transmission quality is guaranteed.

B. Attacker Model

Since sensor networks usually operate in an unattended and open environment, malicious SNs can gather profitable information and launch attack. We assume that instead of ruling the channel behavior by re-broadcasting the eavesdropped packets through varying transmission power and delay, a passive eavesdropping is defined, which means that the malicious SNs can only listen and record the critical information but not re-broadcast it to the HC. Malicious SN can occupy the wireless channel by disobeying the CSMA/CA channel access modality such that normal SNs scan a busy channel and stop transmitting the packets. We describe the attacks as below:

- **Sybil attack**: SNs are compromised by Sybil attackers and manipulate fake identities to be part of the legitimate SNs. Because of the un-achievable memory storages and computational resources, SNs barely satisfy the requirements on the cryptographic authentication with additional infrastructure overhead and key management components. Therefore, it is hard for the network to promise the confidentiality such that the bogus packets labeled with multiple fake identities are imperceptibly received by the HC. With numerous counterfeit packets, Sybil attacker can defeat group-based voting techniques and reverse the measurement results [13].
- **Sybil attack hidden in multipath propagation**: The received packets are influenced by the effects of multipath propagation, noise and interference. The same attacker can produce various malicious packets with different channel features. By analyzing the channel characteristics, attacker can hide the specific channel features of the malicious packets in the harsh propagation environment.
- **Increase the number of malicious SNs**: A malicious SN can influence the network by increasing the number of transmitted packets with different bogus identities. However, the similar channel features of these packets make it easier to discriminate them from the benign packets. Smart attackers can produce more malicious SNs with different channel features to increase the difficulty of Sybil detection. In this paper, we assume that all the attackers are smart enough to build more malicious SNs in IWSNs.

C. Design Goals

In this paper, we aim to propose a detection scheme in which the HC can detect the malicious packet transmitted from the Sybil attackers without extending the communication overheads and computational resources for authentication. Specifically, our design goals can be summarized as follows. **First**, the proposed scheme should be able to detect the Sybil attackers without establishing a pre-labeled database for all the channel features of the SNs. **Second**, the proposed scheme should have the ability to reduce the effects of the multipath propagation, impulse noise and interference in the industrial environment. **Third**, the detection scheme should guarantee the accuracy requirements to discriminate the Sybil attackers from the benign SNs, even if the received packets are contaminated by the propagation impairment.

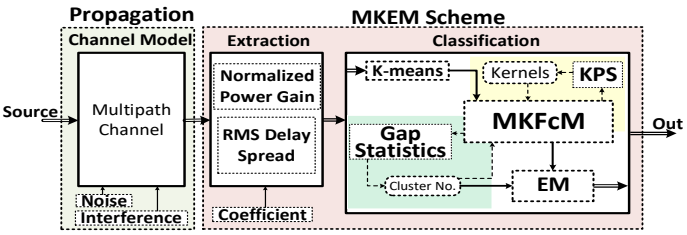


Fig. 2: Overview of the MKEM Scheme.

IV. Proposed Scheme

A. Overview

As depicted in Fig. 2, the MKEM scheme consists of extraction and classification blocks. The inputs are signals transmitted by each SN. The channel model is built up with attenuation factor, delay, Doppler frequency and phase from each propagation path. Impulse noise and interference effects are also considered in the model. All the coefficients, e.g., path-loss exponents, shadowing variances, cluster decays and ray decays, are measured from the real-life industrial environment. The inputs of the MKEM scheme are the impulse responses of the propagation channel, which may be affected by mulipath fading, impulse noise, and interference effects.

In the **Extraction** block, the power gain and delay spread are extracted from the received data sets and are combined into a two-dimension channel-vectors. Then, these channel-vectors are put into the classification block. In the **Classification** block, the channel-vectors are firstly classified by **K-means** clustering, which is an unsupervised learning method used to explore hidden patterns from unlabeled data sets. The grouped data sets and predicted cluster number are then brought into the **Multiple Kernel Fuzzy c Means (MKFcM)** block for further classification. After that, we provide **Kernel Parameters Selection (KPS)** block to optimize the combinatorial coefficient and kernel component parameters such that the group data sets with similar characteristics are mapped into the same area and dissimilar ones are in different areas. The number of clusters is extendedly optimized in the **Gap statistics** block by comparing the differences between the channel-vectors and the expectation of the corresponding appropriate references within the same cluster according to the grouped data set processed by MKFcM algorithm. Since the channel-vectors with the similar features are partitioned into one cluster, we detect the Sybil attack by comparing the obtained cluster number with the total amounts of the identities of the received packets. Finally, the optimized cluster number and grouped data sets with latent labels are transferred to the **Expectation-Maximization (EM)** block.

B. Data Extraction

In this section, we compose the source data set by extracting the propagation channel features from the received packets. The characteristics of the multipath fading channel are due to reflections of the signal from local scatterers around the transmitter SNs. By approximating the power delay profile, we export the power gain and delay spread from the channel impulse response which is evaluating the auto-correlation function of the Pseudo Noise (PN) sequences in the received packets. We define the channel-vector at time t as

$$\boldsymbol{x}(t) = \left(\frac{g_m^2(t)}{\gamma^2 + \left((1-\gamma)\cdot \frac{g_m(t)}{\sigma_m(t)}\right)^2}, \frac{\sigma_m^2(t)}{\gamma^2 + \left((1-\gamma)\cdot \frac{g_m(t)}{\sigma_m(t)}\right)^2} \right) \quad (1)$$

where $g_m(t)$ is the normalized power gain from the m-th scatterer, $\sigma_m(t)$ is the Root Mean Square (RMS) which describes the difference between the obtained delay spread values, γ is an adapter factor balance the importance of the power gain during the classification. The mathematical details of the power gain and delay spread are given in our previous work reported in [5]. Let $i = 1, \ldots, N$ be the number of packets received during the sampling time slot. Let $\boldsymbol{x}_i$ be a data point deployed on a R^2 plane constituted by $g_m(t)$ and $\sigma_m(t)$, where $\boldsymbol{x}_i = \boldsymbol{x}_i(t) \in \Xi \subseteq R^2$. Our objective is to optimize the estimation of the pattern clustering based on the knowledge of the extracted channel-vectors $\boldsymbol{x}_i$ in an environment subject to propagation impairments.

C. K-Means Clustering

As shown in Fig. 2, we use K-means clustering method to pre-classify the obtained channel-vectors. The channel-vectors $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ in the original feature space Ξ are partitioned into C distinct clusters corresponding to the ID announced in the packet header. The original feature space Ξ can thus be reduced into several smaller disjoint subspaces which are used as the inputs of the proposed MKFcM method. Specifically, we initialized the K-means algorithms by selecting the centroids, $V = \{\boldsymbol{v}_1^{(\eta)}, \ldots, \boldsymbol{v}_C^{(\eta)}\}$, uniformly at random from the channel-vector set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ in order that the quality of final clustering results can be improved. The centroid $\boldsymbol{v}_c$ is obtained by calculating the mean of the channel-vectors in the new clusters, which is given by $\boldsymbol{v}_c^{(\eta+1)} = \frac{1}{N_c} \sum_{\boldsymbol{x}_i \in \boldsymbol{v}_c^{(\eta)}} \boldsymbol{x}_i$ where N_c is the number of channel-vectors in the cluster c. Until all the channel-vectors are re-allocated to the nearest $\boldsymbol{v}_c$, these pre-clustering solutions of the channel-vectors and corresponding cluster labels are then imported into MKFcM for further clustering.

D. Multi-Kernel Fuzzy c-Means Clustering

As the channel-vectors are extracted from the packet transmitted through the industrial wireless propagation channel, they may be subject to different channel impairments. If we detect the Sybil attacks via investigating these distorted channel-vectors, it may be hard to propose a proper nonlinear function to achieve the requirement of the detection accuracy without pre-labeled samples. In order to reduce the effects of channel impairments, we produce a kernel method to map the channel-vector set X from the original feature space Ξ into a much higher dimensional Hilbert space H. Thus, we denote by $\phi : \Xi \to H$ the function which maps Ξ into Hilbert space. Particularly, in the Hilbert space, we move the channel-vectors with similar features into the same place and separate the ones with different features into disparate space to discriminate the malicious packets from benign ones. We combine these mapping functions with different weights to move the channel-vectors in the Hilbert space. By iteratively optimizing the kernel weights, we can gradually move the channel-vectors to improve the Sybil detection accuracy.

Considering K such kinds of mappings, we denote the set of all mapping functions by $\Phi = \{(\phi_1(\boldsymbol{x}), \ldots, \phi_K(\boldsymbol{x})) : \phi_k(\boldsymbol{x}) \in \Xi, \text{ for } k = 1, \ldots, K\}$. Additionally, denote by $\phi'(\boldsymbol{x})$ the combination of $\phi_k(\boldsymbol{x})$ with different weights, which is defined as $\phi'(\boldsymbol{x}) = \sum_{k=1}^{M} w_k \phi_k(\boldsymbol{x})$, where $w_k \geq 0$ is the kernel weight of the k-th mapping function $\phi_k(\cdot)$.

Let $\kappa : \Xi \times \Xi \to R$ be the kernel function, which is the inner product of two channel-vectors. Denote by $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle$ the kernel function, where $\forall \boldsymbol{x}, \boldsymbol{y} \in \Xi$, and $\langle , \rangle$ is the inner product of Hilbert space. Let $\boldsymbol{\kappa} = \{(\kappa_1(\boldsymbol{x}_i, \boldsymbol{x}_j), \ldots, \kappa_M(\boldsymbol{x}_i, \boldsymbol{x}_j)) : \kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j) \in H, \text{ for } k = 1, \ldots, K\}$ be the Mercer kernels with respect to the inner product of two mappings, which is defined as $\kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi_k(\boldsymbol{x}_i)^T \phi_k(\boldsymbol{x}_j)$.

Let Ψ be a new set of independent mappings from the linear combination of $\phi_k(\boldsymbol{x}_i)$, which is defined as $\Psi = \{(\psi_1(\boldsymbol{x}), \ldots, \psi_K(\boldsymbol{x})) : \psi_k(\boldsymbol{x}) \in \Xi, \text{ for } k = 1, \ldots, K\}$. $\psi_k(\boldsymbol{x})$ is an all-zero vector except for $\phi_k(\boldsymbol{x}_i)$ at the kth position. The linear combination of $\psi_k(\boldsymbol{x})$ with different weights can be defined as:

$$\begin{aligned} &\sum_{k=1}^{M} w_k \psi_k(\boldsymbol{x}) \\ &= w_1 \begin{bmatrix} \phi_1(\boldsymbol{x}) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + w_2 \begin{bmatrix} \mathbf{0} \\ \phi_2(\boldsymbol{x}) \\ \vdots \\ \mathbf{0} \end{bmatrix} + \cdots + w_M \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \phi_M(\boldsymbol{x}) \end{bmatrix} \end{aligned} \quad (2)$$

On this basis, we obtain two specific properties of the multiplication between different $\psi_k(\boldsymbol{x})$ that $w_k \psi_k(\boldsymbol{x}_i)^T w_{k'} \psi_{k'}(\boldsymbol{x}_j) = w_k^2 \phi_k(\boldsymbol{x}_i)^T \phi_k(\boldsymbol{x}_j)$, if $k = k'$; and $w_k \psi_k(\boldsymbol{x}_i)^T w_{k'} \psi_{k'}(\boldsymbol{x}_j) = 0$, if $k \neq k'$.

Denote by $\{\boldsymbol{x}_{ic}\}_{i=1,\ldots,N_c}$ the set of channel-vectors obtained from K-means clustering, where $c = 1, \ldots, C$, C is the number of clusters, N_c is the number of training channel-vectors in cluster c. Denote by $\kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) = \exp\left(-\frac{\|\boldsymbol{x}_{ic} - \boldsymbol{x}_{jc}\|^2}{2\sigma_k^2}\right)$ the kernel function, where $\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc} \in R^2$, and σ_k is the corresponding parameter. The kernel function κ has two important properties [30]:

- $\kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) = 1$. The norm of every channel-vector in the feature space is 1;
- $0 < \kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) \leq 1$. The cosine value of two training channel-vectors $\boldsymbol{x}_{ic}$ and $\boldsymbol{x}_{jc}$ in the feature space is greater than 0 and less than or equal to 1, and it determines the similarity between these two samples. The relation can be

Algorithm 1 Kernel Parameter Selection

Input: Observation set $\{\boldsymbol{x}_{ic}\}_{i=1,\ldots,N_c,c=1,\ldots,C} \subset R_d$
Output: $\boldsymbol{w}, \boldsymbol{\sigma}$
1: Let $\boldsymbol{\sigma}^{old}$ be a randomly selected starting vector
2: $J(\boldsymbol{w},\boldsymbol{\sigma}) \leftarrow 1 - \mathcal{S}_c(\boldsymbol{w},\boldsymbol{\sigma}) + \mathcal{D}_c(\boldsymbol{w},\boldsymbol{\sigma})$
3: **repeat**
4: Calculate $\mathcal{A}_k(\sigma_k^{old})$ and $\mathcal{B}_k(\sigma_k^{old})$, where $k = 1,\ldots,M$
5: **for** each $k \in \{1,\ldots,M\}$ **do**
6: $w_k \leftarrow \left(\sum_{k'=1}^{M} \frac{\mathcal{B}_k(\sigma_k)-\mathcal{A}_k(\sigma_k)}{\mathcal{B}_{k'}(\sigma_{k'})-\mathcal{A}_{k'}(\sigma_{k'})}\right)^{-1}$, $H_0 \leftarrow \nabla^2 J(\sigma_\eta, w_k)$
7: **while** $|\nabla J(\sigma_\eta, w_k)| > \epsilon$ **do**
8: $s_\eta \leftarrow -H_\eta^{-1}\nabla J(\sigma_\eta, w_k)$, $\sigma_{\eta+1} = \sigma_\eta + \alpha s_\eta$, $y_\eta \leftarrow \nabla J(\sigma_{\eta+1}, w_k) - \nabla J(\sigma_\eta, w_k)$
9: $H_{\eta+1} \leftarrow H_\eta + \frac{y_\eta y_\eta^T}{y_\eta s_\eta} - \frac{H_\eta s_\eta s_\eta^T H_\eta}{s_\eta^T H_\eta s_\eta}$
10: **end while**
11: $\sigma_k \leftarrow \sigma_\eta$
12: **end for**
13: Set $\boldsymbol{w}^{old} \leftarrow \boldsymbol{w}$ and $\boldsymbol{\sigma}^{old} \leftarrow \boldsymbol{\sigma}$
14: **until** $J(\boldsymbol{w},\boldsymbol{\sigma}) - J(\boldsymbol{w}^{old},\boldsymbol{\sigma}^{old}) < \epsilon$
15: **return** $\boldsymbol{w}, \boldsymbol{\sigma}$

shown as $\kappa_k(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc},\sigma_k) = \phi_k(\boldsymbol{x}_{ic})^T\phi_k(\boldsymbol{x}_{jc}) = \frac{\phi_k(\boldsymbol{x}_{ic})^T\phi_k(\boldsymbol{x}_{jc})}{\|\phi_k(\boldsymbol{x}_{ic})\|\|\phi_k(\boldsymbol{x}_{jc})\|} = \cos\angle(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc})$.

It shows that we can use kernel function to describe the difference between channel-vectors. Specifically, we can map the similar channel-vectors into same area by decreasing the included angle; and discriminate two different channel-vectors by enlarging the included angle. Additionally, a proper σ_k should be calculated to help optimize the discrimination [31].

First, the mean of the included angle between similar channel-vectors in the same class is

$$\mathcal{S}_c(\boldsymbol{\sigma}) = \frac{1}{\sum_{c=1}^{C} N_c^2} \sum_{c=1}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_c} \kappa(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc},\boldsymbol{\sigma}) \tag{3}$$

where $\kappa(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc},\boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2\kappa_k(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc},\sigma_k)$. We can iteratively calibrate σ_k to make $\mathcal{S}_c(\sigma_k)$ close to 1. Reform (3) as $\mathcal{S}_c(\boldsymbol{w},\boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2\mathcal{A}_k(\sigma_k)$, where $\mathcal{A}_k(\sigma_k)$ is

$$\mathcal{A}_k(\sigma_k) = \frac{1}{\sum_{c=1}^{C} N_c^2} \sum_{c=1}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_c} \kappa_k(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc},\sigma_k) \tag{4}$$

Second, the mean of the included angle of the channel-vectors between different clusters is

$$\mathcal{D}_c(\boldsymbol{\sigma}) = \frac{1}{\sum_{c=1}^{C}\sum_{\substack{c'=1\\c'\neq c}}^{C} N_c N_{c'}} \sum_{c=1}^{C}\sum_{\substack{c'=1\\c'\neq c}}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_{c'}} \kappa(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc'},\boldsymbol{\sigma}) \tag{5}$$

where $\kappa(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc'},\boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2\kappa_k(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc'},\sigma_k)$. Thus, σ_k should be determined such that $\mathcal{D}_c(\sigma_k)$ approaches 0. Reform (5) as $\mathcal{D}_c(\boldsymbol{w},\boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2\mathcal{B}_k(\sigma_k)$, where $\mathcal{B}_k(\sigma_k)$ is

$$\mathcal{B}_k(\sigma_k) = \frac{1}{\sum_{c=1}^{C}\sum_{\substack{c'=1\\c'\neq c}}^{C} N_c N_{c'}} \sum_{c=1}^{C}\sum_{\substack{c'=1\\c'\neq c}}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_{c'}} \kappa_k(\boldsymbol{x}_{ic},\boldsymbol{x}_{jc'},\sigma_k) \tag{6}$$

Considering that $\mathcal{S}_c(\boldsymbol{w},\boldsymbol{\sigma})$ should be closed to 1 and $\mathcal{D}_c(\boldsymbol{w},\boldsymbol{\sigma})$ should be closed to 0, we define the objectivity function with respect to w_k as

$$\begin{aligned} \underset{\boldsymbol{w},\boldsymbol{\sigma}}{\text{argmin}} \quad & 1 - \mathcal{S}_c(\boldsymbol{w},\boldsymbol{\sigma}) + \mathcal{D}_c(\boldsymbol{w},\boldsymbol{\sigma}) \\ \text{s.t.} \quad & (4),(6),\sum_{k=1}^{M} w_k = 1, \quad w_k \geq 0 \end{aligned} \tag{7}$$

Solving Eq. 7, we can obtain the close-form of weights w_k as

$$w_k = \left(\sum_{k'=1}^{M} \frac{\mathcal{B}_k(\sigma_k) - \mathcal{A}_k(\sigma_k)}{\mathcal{B}_{k'}(\sigma_{k'}) - \mathcal{A}_{k'}(\sigma_{k'})}\right)^{-1} \tag{8}$$

And we consider quasi-Newton algorithm to find the minimum value of σ_k with the obtained w_k. The Kernel Parameter Selection (KPS) algorithm is summarized in Alg. 1.

According to the aforementioned kernel features, we cluster the channel-vectors by estimating the centers within all the channel-vectors. We further evaluate which cluster the channel-vectors belong to by calibrating the probability with respect to the distance from the channel-vectors to that cluster center [32]. Hence, the objective function is

$$\begin{aligned} \underset{\boldsymbol{w},\boldsymbol{u},\boldsymbol{v}}{\text{argmin}} \quad & \sum_{i=1}^{N}\sum_{c=1}^{C} u_{ic}^m \|\psi(\boldsymbol{x}_i) - \boldsymbol{v}_c\|^2 \\ \text{s.t.} \quad & \psi(\boldsymbol{x}_i) = \sum_{k=1}^{M} w_k\psi_k(\boldsymbol{x}_i), \quad \sum_{c=1}^{C} u_{ic} = 1, \quad u_{ic} \geq 0 \end{aligned} \tag{9}$$

where $\boldsymbol{v}_c$ is the center of the c-th cluster in the implicit feature space, $\boldsymbol{w} = (w_1, w_2, \ldots, w_M)^T$ is a vector consisting of weights, $\boldsymbol{u}$ is an $N \times C$ membership matrix whose elements are the memberships u_{ic}, and $\boldsymbol{v}$ is an $L \times C$ matrix whose columns correspond to cluster centers. Solving (9), we have

$$\boldsymbol{v}_c = \sum_{i=1}^{N}\sum_{i'=1}^{N} \frac{u_{ic}^m}{u_{i'c}^m}\psi(\boldsymbol{x}_i) = \sum_{i=1}^{N} \hat{u}_{ic}\psi(\boldsymbol{x}_i) \tag{10}$$

where $\hat{u}_{ic} = \sum_{i'=1}^{N}(u_{ic}^m/u_{i'c}^m)$ is the normalized membership. (9) can be further rearranged as $\|\psi(\boldsymbol{x}_i) - \boldsymbol{v}_c\|^2 = \sum_{k=1}^{M} \alpha_{ick} w_k^2$, where the coefficient α_{ick} can be written as

$$\begin{aligned} \alpha_{ick} = \kappa_k(\boldsymbol{x}_i,\boldsymbol{x}_i) - 2\sum_{j=1}^{N} \hat{u}_{jc}\kappa_k(\boldsymbol{x}_i,\boldsymbol{x}_j) \\ + \sum_{j=1}^{N}\sum_{j'=1}^{N} \hat{u}_{jc}\hat{u}_{j'c}\kappa_k(\boldsymbol{x}_i,\boldsymbol{x}_j') \end{aligned} \tag{11}$$

Algorithm 2 Multi-Kernel Fuzzy c Means.

Input: Observation set $\{\boldsymbol{x}_{ic}\}_{i=1,\ldots,N,c=1,\ldots,C} \subset R^d$, kernels with $\{\kappa_k\}_{k=1,\ldots,M}$ and weights for the kernels $\{w_k\}_{k=1,\ldots,M}$
Output: $\{label_i\}_{i=1,\ldots,N}$
Let u_{ic}^m be a randomly selected starting vector
$m \leftarrow 2$, $\hat{u}_{ic}^{old} \leftarrow \sum_{i'=1}^{N}(u_{ic}^m/u_{i'c}^m)$, $\alpha_{ick}^{old} \leftarrow (11)$
repeat
  for each $i \in \{1,\ldots,N\}$ **do**
    $\boldsymbol{w}, \boldsymbol{\sigma} \leftarrow KPS(\boldsymbol{x}_{ic})$ (**Alg**. 1)
    Calculate $\sum_{k=1}^{M} \alpha_{ick}^{old} w_k^2$
    $$u_{ic} = \left(\sum_{c'=1}^{C} \left(\frac{\sum_{k=1}^{M} \alpha_{ick}^{old} w_k^2}{\sum_{k=1}^{M} \alpha_{i'ck}^{old} w_k^2} \right)^{\frac{1}{m-1}} \right)^{-1}$$
    $$\hat{u}_{ic}^{old} \leftarrow \sum_{i'=1}^{N}(u_{ic}^m/u_{i'c}^m),\ \alpha_{ick}^{old} \leftarrow (11),\ label_i \leftarrow \underset{c}{\text{argmax}} \sum_{i=1}^{N}\sum_{c=1}^{C} u_{ic}^m \left(\sum_{k=1}^{M} \alpha_{ick} w_k^2 \right)$$
  end for
until $J(\boldsymbol{u}, \boldsymbol{\alpha}) - J(\boldsymbol{u}^{old}, \boldsymbol{\alpha}^{old}) < \epsilon$
return $\{label_i\}_{i=1,\ldots,N}$

Substitute (10) and (11) into (9). We have the close-form of membership u_{ic} as

$$u_{ic} = \left(\sum_{c'=1}^{C} \left(\frac{\sum_{k=1}^{M} \alpha_{ick} w_k^2}{\sum_{k=1}^{M} \alpha_{i'ck} w_k^2} \right)^{\frac{1}{m-1}} \right)^{-1} \tag{12}$$

E. Gap statistical analysis

When the number of malicious SNs is increasing, the network would be filled with pseudonymous entities with different channel features. If the new malicious SNs are initiated after determining the number of clusters of the channel-vectors, the malicious packets may be highly detected as benign ones. In this section, we propose a Gap statistical analysis method to remove the mistakes in detecting the number of clusters.

Our data $\{\boldsymbol{x}_j^{(c)}\}_{j=1,\ldots,N;c=1,\ldots,C}$, consist of C features measured on N independent observations. Let $d_{jj'} = \sum_{c=1}^{C} \|\boldsymbol{x}_j^{(c)} - \boldsymbol{x}_{j'}^{(c)}\|$ be the distance between channel-vectors j and j'. Suppose that we have clustered the data into C clusters $N_1, \cdots, N_C$, with N_c denoting the indices of observations in cluster c. Let $D_c = \sum_{j,j' \in N_c} d_{jj'}$ be the sum of the pairwise distance for all points in cluster c, and set $W_c = \sum_{c=1}^{C} \frac{1}{2N_c} D_c$ to be the pooled within-cluster sum of squares around the cluster means. We further compare the differences between the $\log(W_c)$ and corresponding the expectation of the appropriate references. Hence we define

$$Gap_n(c) = E_n^*\{\log(W_c)\} - \log(W_c) \tag{13}$$

where E_n^* denotes expectation under a sample of size n from the reference distribution.

Algorithm 3 Gap Statistic Analysis.

Input: Observation set $\{\boldsymbol{x}_i\}_{i=1,\ldots,N} \subset R^d$
Output: The number of cluster C
for each $b \in \{1,\ldots,B\}$ **do**
  $\boldsymbol{x}_i^*(:,b) \leftarrow$ generate reference data from $\boldsymbol{x}_i$
end for
for each $C \in \{1,\ldots,N\}$ **do**
  $\{\boldsymbol{x}_{ic}\}_{c\in\{1,\ldots,C\}} = MKFcM(\boldsymbol{x}_i, C)$ (**Alg**. 2)
  $$W_C = \sum_{c=1}^{C} \frac{1}{2N_c} \sum_{i=1}^{N_c}\sum_{i'=1}^{N_c} \|\boldsymbol{x}_{ic} - \boldsymbol{x}_{i'c}\|^2$$
  for each $b \in \{1,\ldots,B\}$ **do**
    $\{\boldsymbol{x}_{ic}^*(:,b)\}_{c\in\{1,\ldots,C\}} = MKFcM(\boldsymbol{x}_i^*(:,b), C)$
    $$W_C^*(:,b) = \sum_{c=1}^{C} \frac{1}{2N_c^*} \sum_{i=1}^{N_c^*}\sum_{i'=1}^{N_c^*} \|\boldsymbol{x}_{ic}^*(:,b) - \boldsymbol{x}_{i'c}^*(:,b)\|^2$$
  end for
  $E\{\log(W_C^*)\} \leftarrow (\frac{1}{B}) \sum_b \log(W_C^*(:,b))$
  $Gap(C) = E\{\log(W_C^*)\} - \log(W_C)$
  $$sd(C) \leftarrow \left((\tfrac{1}{B}) \sum_b \{\log(W_C^*(:,b)) - E\{\log(W_C^*)\}\}^2 \right)^{\frac{1}{2}}$$
  $s(C) \leftarrow sd(C)\sqrt{1 + \frac{1}{B}}$
end for
for each $C \in \{1,\ldots,N\}$ **do**
  if $Gap(C) \geq Gap(C+1) - s(C+1)$ **then**
    return C;
  end if
end for

The $E_n^*\{\log(W_c)\}$ is determined by the average of B copies $\log(W_c^*)$ with respect to the Monte Carlo samples $x_1^*, \cdots, x_n^*$ from the corresponding reference distribution. To this end, we show the standard deviation of the B copies $\log(W_c^*)$ as $sd(c)$. Accounting additionally for the simulation error in $E_n^*\{\log(W_c)\}$ results in the quantity $s_c = \sqrt{1 + 1/B}\, sd_c$, where $sd_c = [(1/B)\sum_{b=1}^{B}\{\log(W_{cb}^*) - (1/B)\sum_{b=1}^{B}\log(W_{cb}^*)\}^2]^{1/2}$. Using this we choose the cluster size $\hat{c}$ to be the smallest c such that $Gap(c) \geq Gap(c+1) - s_{c+1}$. The details of the Gap statistic proceeds in Alg 3.

F. Expectation Maximum Clustering

We introduce a complete data set $\boldsymbol{z} = \{\boldsymbol{c}, \boldsymbol{x}\}$ with $\boldsymbol{c} = [c_{11}, \cdots, c_{1N}, \cdots, c_{C1}, \cdots, c_{CN}]^T$ being a vector of $C \times N$ latent variables whose values tell us which mixture components have generated the corresponding measurements. More specifically, we define $c_{ik} \in 1,\ldots,C$ and set $c_{ik} = k$ if x_{ik} is generated by the corresponding mixture component $p_c(\boldsymbol{x}; \boldsymbol{\theta})$. The complete data log-likelihood function is easily expressed by $\ln(p(\boldsymbol{x}, \boldsymbol{c}; \boldsymbol{\theta})) = \ln\left(\prod_{i=1}^{N}\prod_{k=1}^{C} p(\boldsymbol{x}_{ik}, c_{ik}; \boldsymbol{\theta})\right)$. The idea behind the EM criterion is to estimate the unknown parameters iteratively in two steps: an expectation (E)-step and a maximization (M)-step. In the first step, statistical expectation of the complete data log-likelihood is taken with respect to the conditional probability of the latent variables. In the second step, the conditional expectation obtained above is maximized with respect to the parameters of interest. The

two steps iterate until a predetermined convergence condition is met. Given a *prior* parameter estimate $\boldsymbol{\theta}^{(\eta)}$, we show in the sequel the work-flow of the proposed EM algorithm on the $(\eta+1)$-th iteration.

E-step: Let us first define the conditional expectation of the complete data log-likelihood. Consider $h(\boldsymbol{C},\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a normal variable that we wish to adjust and $\boldsymbol{C}$ is a random variable governed by the probability $Pr\left\{\boldsymbol{C}|\boldsymbol{X};\boldsymbol{\theta}^{(\eta)}\right\}$. Considering the expectation of $h(\boldsymbol{C},\boldsymbol{\theta})$ with respect to $\boldsymbol{X}$ and $\boldsymbol{\theta}^{(\eta)}$, we have the objectivity function as

$$Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)}\right)=E_{\boldsymbol{C}}[h(\boldsymbol{C},\boldsymbol{\theta})|\boldsymbol{X},\boldsymbol{\theta}^{(\eta)}] \tag{14}$$

We evaluate the expectation of the log-likelihood function of the complete-data with respect to the latent variable $\boldsymbol{C}$ given the observation data X and current normal variable $\boldsymbol{\theta}^{(\eta)}$. Then, (14) can be re-written as

$$\begin{aligned}Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)}\right)=&\sum_{i=1}^{N}\sum_{k=1}^{C}\ln\left(p(\boldsymbol{x}_i|k,\boldsymbol{\theta})\right)\tilde{p}(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})\\&+\sum_{i=1}^{N}\sum_{k=1}^{C}\ln\left(p(k|\boldsymbol{\theta})\right)\tilde{p}(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})\end{aligned} \tag{15}$$

where $\tilde{p}(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})$ can be computed by means of Bayes' rule as follows:

$$\tilde{p}(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})=\frac{p(\boldsymbol{x}_i|k,\boldsymbol{\theta}^{(\eta)})p(k|\boldsymbol{\theta}^{(\eta)})}{\sum_{k}p(\boldsymbol{x}_i|k,\boldsymbol{\theta}^{(\eta)})p(k|\boldsymbol{\theta}^{(\eta)})} \tag{16}$$

M-step: We maximize $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)})$, with respect to the vector parameter $\boldsymbol{\theta}$, and thus obtain on the $(\eta+1)$-th iteration. We firstly solve the maximum problem to obtain the result of $p(k|\boldsymbol{\theta})$ on the $(\eta+1)$-th iteration:

$$\underset{p(k|\boldsymbol{\theta})}{\text{argmax}}Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)}\right)\quad\text{s.t.}\sum_{k=1}^{M}p(k|\boldsymbol{\theta})=1 \tag{17}$$

Solving (17), we have

$$p(k|\boldsymbol{\theta})=\frac{1}{N}\sum_{i=1}^{N}p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)}) \tag{18}$$

$$\boldsymbol{\mu}_k^{(\eta+1)}=\frac{\sum_{i=1}^{N}\boldsymbol{x}_i p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})}{\sum_{i=1}^{N}p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})} \tag{19}$$

$$\boldsymbol{\Sigma}_k^{(\eta+1)}=\frac{\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T(\boldsymbol{x}_i-\boldsymbol{\mu}_k)p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})}{\sum_{i=1}^{N}p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})} \tag{20}$$

The details of the EM clustering proceeds in Alg 4.

G. Summary

In this section, we summarize in detail the proposed MKEM Sybil detection scheme. As mentioned in the system model, we developed this detection scheme from the radio resource-testing method. To build up the observation space of all packets, we investigate the channel features of each SN by exploring the power gain and delay spread. Then, these channel-vectors influenced by the channel impairments are put into the classification parts of the detection scheme.

Algorithm 4 Expectation Maximum Clustering.

Input: Observation set $\{\boldsymbol{x}_i\}_{i=1,\ldots,N}\subset R^d$; labels obtained from (**Alg**. 2), $\{label_i^{(old)}\}_{i=1,\ldots,N}$; number of clusters obtained from (**Alg**. 3), C.

Output: $\{label_i\}_{i=1,\ldots,N}$

$\{label_i^{(old)}\}_{i=1,\ldots,N}\leftarrow$ (**Alg**. 2), $C\leftarrow$ (**Alg**. 3), $\eta\leftarrow 1$

for each $k\in\{1,\ldots,C\}$ **do**

  $N_k\leftarrow\sharp\{label_i^{(old)}==k\}$

  $p(k|\boldsymbol{x}_i,\boldsymbol{\theta}^{(\eta)})\leftarrow\frac{N_k}{N}$, $p(k|\boldsymbol{\theta}^{(\eta+1)})\leftarrow$ (18), $\boldsymbol{\mu}_k^{(\eta+1)}\leftarrow$ (19), $\boldsymbol{\Sigma}_k^{(\eta+1)}\leftarrow$ (20)

end for

$\eta\leftarrow\eta+1$

repeat

  $label_i^{(old)}\leftarrow k=\max\left(Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)}\right)\right)$

  $p(k|\boldsymbol{\theta}^{(\eta+1)})\leftarrow$ (18), $\boldsymbol{\mu}_k^{(\eta+1)}\leftarrow$ (19), $\boldsymbol{\Sigma}_k^{(\eta+1)}\leftarrow$ (20), $Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta+1)}\right)\leftarrow$ (15)

until $Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta+1)}\right)-Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta)}\right)<\epsilon$

$label_i\leftarrow k=\max\left(Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(\eta+1)}\right)\right)$

return $\{label_i\}_{i=1,\ldots,N}$

In the MKEM scheme, since each obtained packet is represented by a channel-vector, we use K-means method to briefly classify the channel-vectors into different clusters. The obtained grouped channel-vectors are further put into the MKFcM method. After mapping these grouped results into the Hilbert space, the MKFcM method can gradually optimize the combinatorial weights and kernel parameters with KPS method to improve the classification performance. As a result, if a group of the channel-vectors of the related packets with different identities is mapped into the same area, it means that these channel-vectors appear similar propagation features, which implies that these packets are transmitted from Sybil attackers. Since we can modify the combinational weights and kernel parameters, we can produce different mapping function for MKFcM such that we can reduce the effects of the propagation impairments and improve the detection accuracy in different industrial environments. When a smart attacker produces more malicious SNs and increases the number of Sybil attackers, the group number of classified channel-vectors may be changed. The Gap statistical analysis method is used to verify and renew the group number to increase the detection accuracy when the number of malicious SNs is increased. Once the MKFcM method obtains the classification results, it passes the newly grouped channel-vectors into the EM method for further verification. The EM method is also used to reduce the influence on detection accuracy when the malicious SNs are increased. As aforementioned, we suppose that the channel-vectors may be close to each other if they are extracted from the packets transmitted from the same SN. It is because that their channel-vectors are determined according to the propagation environment around the related SN. Meanwhile, these channel-vectors may follow a similar probability distribution. On this basis, we can use the EM method to verify the classification results with the channel-

TABLE I: Simulation Setting

Parameter		
Notation	**Description**	**Value**
f	Frequency	**2.4GHz**
f_c	Sampling frequency	**10Hz**
T	Simulation time	**300s**
M	The number of scatterers	**100**
p_t	Transmit power	Adaptive
F	Receiver noise figure	**11dBm**
B	Receiver noise bandwidth	**5MHz**
T_n	Noise temperature	**290K**
p_{01}	Channel state probability from good to bad	**0.005**
p_{10}	Channel state probability from bad to good	**0.1**
R	Ratio of noise power in the bad and good channel	**100**
N	Number of interfering signals	**7**
J	Total interference power	**4dBm**
K	Ricean K-factor of multipath component	**5dBm**
η	Relative permittivity of the reflecting surface at 2.4GHz	$\mathbf{1-j802}$
n	Path-loss exponent	**1.72** [16]
σ	Shadowing variation	**3.76** [16]
γ	Adaptor factor	**0.2**

vectors obtained from the MKFcM method, especially, when the number of malicious SNs are increasing. After detecting the malicious packets in EM method, we can find those malicious packets which are classified into one group with mass numbers of packets. Finally, we detect the packets from Sybil attackers. We can rebuild the channel impulse response based on the related channel-vectors and mark them for future verification.

V. Performance Evaluation

A. Simulation Setup

We propose a time-varying channel model taking into account the multipath propagation, Doppler effects, and signal phase deviations. Additionally, the unprecedented impact of the impulse noise and interference effects are considered in the performance of the Sybil detection. In order to indicate the typical burst nature of the impulse noise, we introduce a two-state Markov process in our channel model, which is shown in Fig. 3. In [15], the noise is defined as $n(t) = b(t) \cdot w(t) + (1 - b(t)) \cdot k(t)$, for $t \in \{1, \ldots, T\}$, where $b(t)$ is a random variable with state space, $\{0, 1\}$. Let $n_b(t)$ be $b(t) = 0$ and $n_g(t)$ be $b(t) = 1$ as the good channel state with AWGN and bad channel state with impulse noise, respectively. We define $w(t)$ and $k(t)$ as zero-mean Gaussian distributed processes with different variances, which are differentiated by R and given as: $p_{n_g(t)|b(t)=1}(n(t)) \sim N(0, \sigma^2)$ and $p_{n_b(t)|b(t)=0}(n(t)) \sim N(0, R\sigma^2)$.

The transition probability of the two-state Markov process is $p_{i,j} = P\{b(t+1) = i | b(t) = j\}$, for $i, j \in \{0, 1\}$.

We define the channel model as the sum of the power gain reflected from the different scatterers in the channel. Considering the properties of time-varying and multipath propagation, we have:

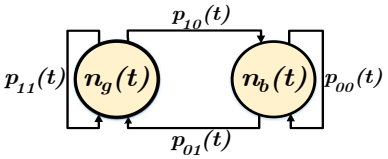


Fig. 3: Two-State Markov Process

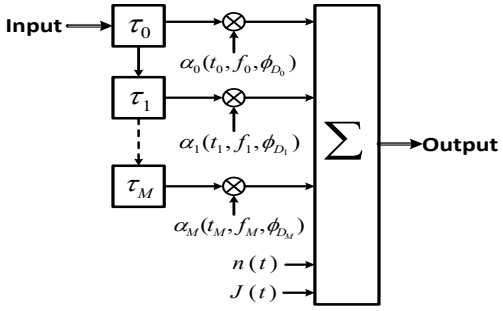
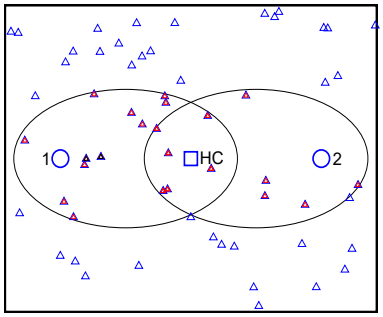


Fig. 4: Time-delay Wireless Channel Model

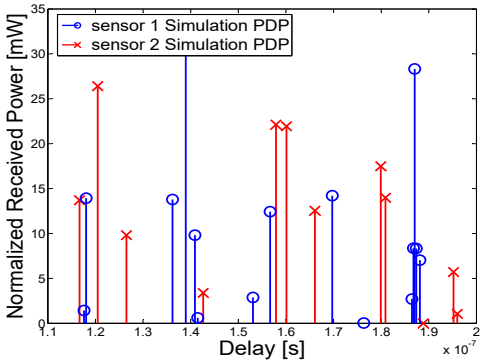
$$
\begin{aligned}
h(t) &= \sum_{m=0}^{M} \{[A(t - \tau_m(t)) \cdot \frac{\lambda\sqrt{G_a}}{4\pi d_m} \cdot e^{-j\frac{2\pi d_m}{\lambda}}] \\
&\qquad \cdot e^{j(2\pi f_c \tau_m(t) - \phi_{D_m}(t))}\} + n(t) + J(t) \\
&= \sum_{m=0}^{M} \alpha_m(t) \cdot \delta(t - \tau_m(t)) \\
&\qquad \cdot e^{j(2\pi f_m(t) - \phi_{D_m}(t))} + n(t) + J(t) \quad (21)
\end{aligned}
$$

where $\tau_m(t)$ is the delay caused by the m-th scatterer at time t, G_a is the antenna gain and d_m is the distance from the SN to the HC through the mth scatter. α_m is the power gain factor from the m-th scatterer. $\phi_{D_m}(t)$ is the summation of phase resulted from the movement/displacement of the m-th scatterer. f_c is the carrier frequency and f_m is the Doppler shift at the m-th moving scatterer. $J(t)$ is the interference effects which is given as $J(t) = \sum_{n=0}^{N-1} \sqrt{2}\beta_n e^{j\phi_n(t)}$, where β_n is the n-th interference power and $n(t)$ is the aforementioned impulse noise. The corresponding channel simulator is shown in Fig. 4. See [15] for details on the channel model.

The simulation setting of the wireless channel is presented in Table. I. We configure the SNs and the controller to operate at 2.4 GHz. 100 scatterers are uniformly deployed in a 100×100 map. The path-loss exponent and shadowing variance are obtained from empirical measurements of an industrial environment [16]. The SNs randomly deploy in the map and provide one-hop transmission with the HC. All the SNs are initialized with transmission power of 0 dB. We evaluate the performance of MKEM scheme in MATLAB. First, a discrete-time event producer is designed to randomly produce packets like SNs in the network. These packets only contain a group of impulse signals and will be transmitted through the proposed channel model. Then, we extract channel-vectors from the obtained packets to do the classification. Specifically, when a Sybil attack takes place during the simulation, only the malicious SN can adaptively change its transmission power. We provide a two-SN case example to promote better per-



(a) Two-SN Layout



(b) Power Delay Profile

Fig. 5: Two SNs Case: Position Layout and PDP

ceivability on the simulation environment and the observation data (channel-vectors) extracted from the channel model.

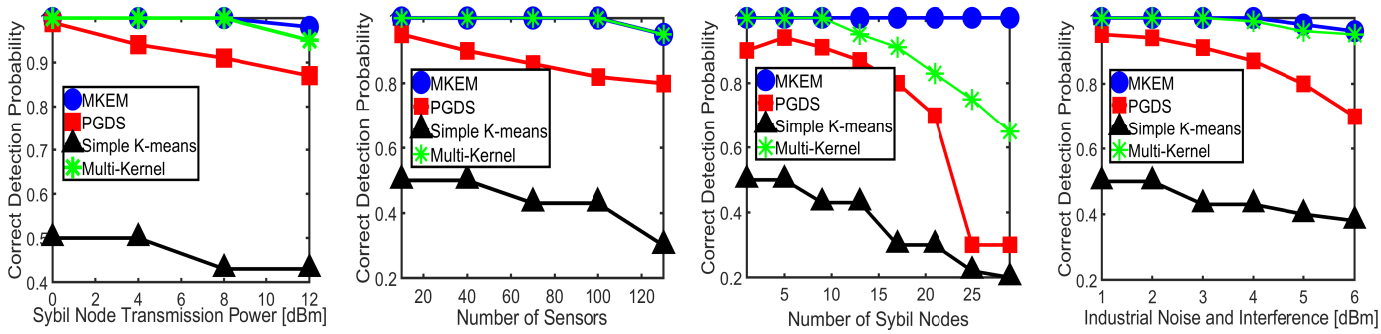
We show a simulation sketch with two SNs and a ground floor occupied by numerous uniformly distributed scatterers in Fig. 5(a). The received signal is affected by the scatterers. For the sake of straightforwardness, we introduce elliptical scattering zones for analysing the model to investigate the wireless channel shown in Fig. 5(a), so that the scatterers leading to the multipath fading are only considered within the elliptical zones. The SN and the HC are located at the foci of the ellipse. The Power delay profile (PDP) in Fig. 5(b) is obtained based on the channel simulator defined in Fig. 4. Fig. 5(b) shows that the propagation signals from SN 1 come to each scatterer with distorted power and time delay. Then, signals arrive via reflections from different scatterers and experience the gradually time-varying fading again. The power delay of SN 1 and SN 2 follow the same scatterer decay mode, as described in Saleh-Valenzuela (SV) model [23]. Fig. 5(b) shows that in industrial environment, multipath power gain, impulse noise and interference effects, on one side, may distort the signal (e.g., deep fading effect); on the other side, they may also enhance the received power on the HC side (due to constructive summation of the multipath components). It also indicates that the RSSI-based Sybil detection scheme may increase detection error in this scenario. Moreover, the delays between different signal powers also illustrate the density of the scatterers around each SN in the ellipse domain.

B. Simulation Results

In this paper, we use the Sybil detection probability (DP), false positive rate (FPR) and false negative rate (FNR) to evaluate the performance of the proposed MKEM scheme. The Sybil DP is the probability of precisely detecting the malicious packets in the received packets. Both FNR and FPR are used to evaluate the detection ability of MKEM scheme to distinguish between the packets sent from the Sybil attackers and those from the benign SNs. The details of FPR and FNR can be found in [5]. The variables employed in this evaluation include the transmission power of Sybil attackers, the number of SNs, the number of Sybil attackers and the effects of industrial noise and interference. We evaluate the scheme by changing one of the variables while keeping others fixed for each time. When the number of SNs and the number of Sybil attackers are not changed as variables, they will be set as 50 and 5, respectively. Then, following the steps declared in the Sec IV-A, we extract the channel-vector based on $\boldsymbol{x} = [\frac{g_m^2(t)}{\gamma^2+(1-\gamma)^2(g_m(t)/\tau_m)^2}, \frac{\tau_m^2}{\gamma^2+(1-\gamma)^2(g_m(t)/\tau_m)^2}]$ with respect to $\gamma = 0.2$. Additionally, the K-means method, PGDS scheme in [5] and Multi-Kernel scheme in [16] are used to compare with the performance of MKEM scheme. The simulation results are shown as follows.

1) Detection Probability: Fig. 6(a) shows that the Sybil DP of the MKEM scheme is high and it remains high even if the transmission power of the Sybil SNs is increased. As explained in Sec. 6(a), the DP is assessed with channel-vectors, which are constructed by normalizing the power gain obtained from the PDP. The channel-vectors can minimize the impact of power gain by configuring the coefficient. In addition, instead of ensuring the number of cluster by minimizing the within cluster sum of square (WCSS), we estimate the cluster number by studying the difference of WCSS between the reference samples and the determined ones with the Gap Statistics algorithm. The number of clusters is the value that makes the vector direction of the Gap statistics start to descend for the first time. Subsequently, we can know the number of the SNs in the network. This also explains why MKEM scheme has a better Sybil detection performance than the multi-kernel scheme, PGDS scheme and simple K-means algorithm, as implied in Fig. 6(a).

Fig. 6(b) shows that the Sybil DP remains 100% until the number of SNs increases to 100, and then starts to decrease if the number of SNs continuously increased. As mentioned in the MKEM scheme, we use the PDP characteristics of the packets which are received from the SNs deployed under diversified spatial environment to represent the uniqueness of the SNs. In this paper, we evaluate the MKEM scheme by uniformly deploying the SN around the HC. Therefore, if we increase the number of SNs, we may also increase the probability that a malicious SN is placed next to the benign SNs. In this case, the malicious SN may produce packets with similar channel features as benign SNs because of the similar PDP characteristics. Moreover, the proposed MKEM scheme shows better Sybil detection results than conventional PGDS scheme. This is because the proposed scheme additionally proposes kernel method to optimize the classification of different SNs. In addition, since we use RBF kernel to classify the channel-vectors, we transfer all the channel-vectors into a higher dimension Hilbert space and make them start from the



(a) DP vs. the power of Sybil attackers

(b) DP vs. the number of SNs

(c) DP vs. the number of Sybil Attackers

(d) DP vs. Noise+Interference

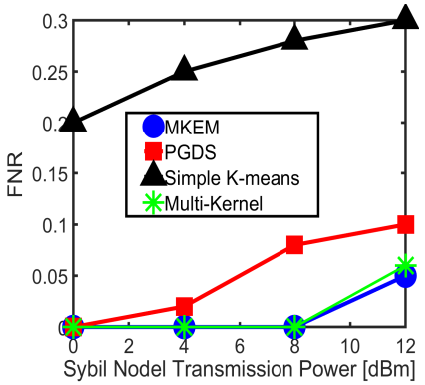
Fig. 6: Correct Detection Probability

centroid of a sphere. Thus, we can regulate the included angles between these channel-vectors to move them into different clusters and further maximize these included angles to increase the distance among clusters. As a result, the Sybil DP result of the MKEM scheme will not change too much when the number of SNs is increasing. In Fig. 6(b), it also shows that, in Sybil DP, the MKEM scheme provides the same evaluation results as multi-kernel scheme. The reason is that the extending Gap statistics analysis and EM algorithm are used to iteratively verify the number of clusters and summarize the classification results, which will not change the Sybil DP results if the number of SNs are increased.

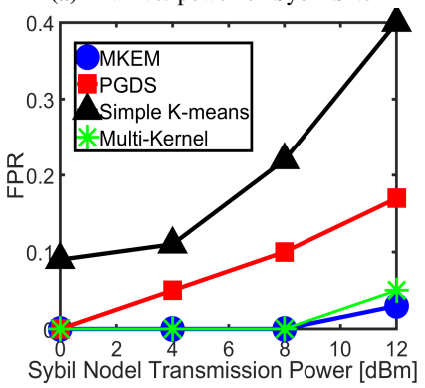
In Fig. 6(c), it shows that by increasing the number of Sybil SNs, the Sybil detection probability of the proposed MKEM scheme remains in 100%. The reasons are manifold. Firstly, the detection results are obtained from the Gap statistics method which compares the differences between the mean of WCSS and the corresponding expectation of the appropriate references. Secondly, the clustering criterions of the WCSS are determined by optimized kernel parameters of a soft clustering algorithm, with the assistance of the EM algorithm. This allows the scheme to effectively classify the channel-vectors with similar characteristics into the same cluster. All these two reasons make MKEM scheme achieve better performance than the multi-kernel scheme. When the number of malicious SNs increases, the diversity of the channel features will decrease since the number of the SNs is fixed. This reduces the impact of interference from other benign SNs and increases the accuracy of the Sybil detection. This should apply to the conventional PGDS scheme as well. However, the figure shows an obvious reduction in the Sybil detection accuracy of the conventional PGDS scheme when the amount of Sybil SNs increases. This is because the conventional scheme is based on the improved K-means clustering principle which measures the WCSS that is difficult to classify the channel-vectors into the determined clusters. When the number of Sybil SNs increase, it will forcibly separate the channel-vectors from the same Sybil SNs and put into different clusters in order to achieve the clustering purpose. The result is that when the number of Sybil SNs is small, the Sybil detection accuracy can be promised through an invisible hard clustering. Whereas, in the case of increasing the number of the Sybil SNs, instead of considering all the channel-vectors in the MKEM, conventional PGDS only deliberates the WCSS, which accounts for the reducing Sybil detection accuracy.

Fig. 6(d) shows that the MKEM scheme maintains a high Sybil detection accuracy even if we increase the industrial impulse noise and interference effects. This is because the proposed MKEM scheme maps the channel-vectors from a lower dimension into a higher dimensional space defined by the kernel method. Considering the noise and interference effects in the channel-vectors, we further utilizes fuzzy clustering method to softly cluster all the channel-vectors by iteratively updating their corresponding probability of each cluster set. Both kernel method and fuzzy clustering algorithm minimize the impact of the noise and interference and thus enhance the Sybil detection accuracy. These advantages make MKEM and multi-kernel scheme present similar evaluation results which are better than the PGDS and simple K-means schemes. In addition, if the noise and interference effects continue to increase, the appropriate channel-vectors will be hardly generated due to the lower signal-to-noise ratio, which leads to a decrease in the Sybil detection accuracy.

2) FNR and FPR: Fig. 7 shows that the FNR and FPR increase slightly when increasing the transmission power of the Sybil SNs. The reason is that the normalized power gains do not reshape substantially the corresponding probability distribution through multipath propagation channel. Besides, we calibrate the adaptive factor to enhance the function of the RMS delay spread, making it able to better identify the environment characteristics around the SNs. This help minimize the impact of the power gain on the channel-vector. The figure also shows that compared with the conventional PGDS scheme, the proposed MKEM scheme shows a better performance with the increased transmission power. This is due to the fact that the MKEM scheme also employs the EM algorithm to further process the channel-vector sets dealt with by the MKFcM clustering. According to the EM algorithm, we consider the SN identities of the channel-vectors as the latent variable (the missing information in the received packets), which is exported from the MKFcM clustering. We measure the expectation on the likelihood function of the condition probability of the channel-vector with respect to the latent variable. Since we consider the latent variable

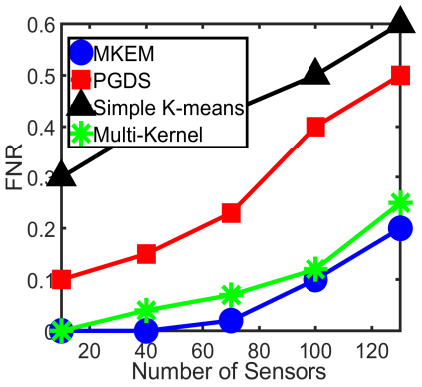


(a) FNR vs. power of Sybil SNs

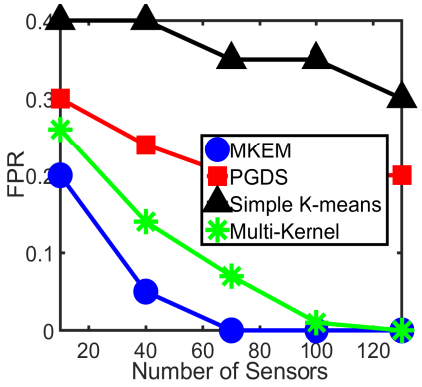


(b) FPR vs. power of Sybil SNs

Fig. 7: The Impacts of the Transmission Power of Sybil SNs



(a) FNR vs. the number of SNs

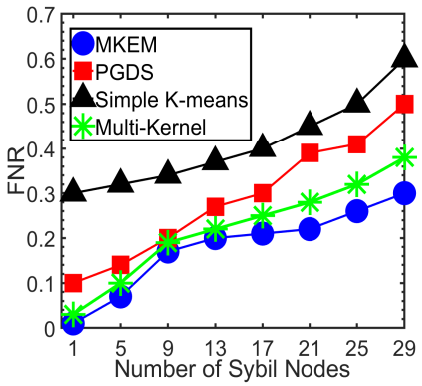


(b) FPR vs. the number of SNs

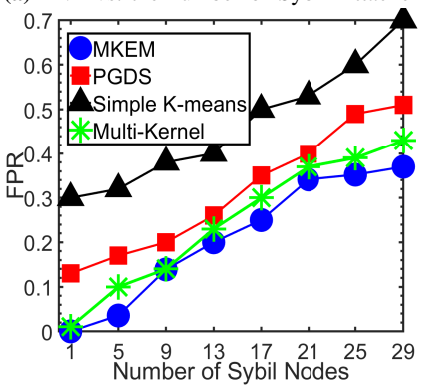
Fig. 8: The Impacts of the Number of SNs

as the unknown variable, we ascertain the expectation by providing the likelihood function of the latent variable and the corresponding condition probability function of the latent variable with respect to the channel-vector. Then, we calculate the mean and variance of the channel-vectors related to the current clusters in a way that we can iteratively update the cluster composition according to the obtained likelihood parameters until convergence is achieved. According to (15), the likelihood function of the latent variables is determined by the condition probability of the channel-vectors with respect to the given latent variables. The advantage of EM algorithm also makes MKEM scheme perform better than multi-kernel scheme when the transmission power of malicious SN is continuously increased. In addition, the simulation results of FNR and FPR tend to stay in 0%, because the channel model is proposed in line with the ray tracing results. Even though the transmitting signal phases are randomly initialized, the spatial differences of the SNs may be amplified by extracting the channel-vectors based on the constitution of the power gain and delay spread. Moreover, the probability distribution of the normalized power gain in the channel-vectors keep stable despite magnifying the power gain. This explains why the FNR and FPR increase slightly in the simulation. Furthermore, we employ half Bayesian methods in measuring the latent variables. Consequently, the MKEM shows a better performance than do the conventional one.

Fig. 8(a) shows that the FNR increase significantly with increasing number of SNs, whereas FPR changes a little. As introduced above, the FNR is defined as the probability of incorrectly identifying the Sybil packets as the benign packets. When there are large numbers of SNs uniformly deployed in the map, both malicious and benign SNs will share the same surrounding environment. As a result, the channel-vectors from either Sybil or benign SNs may present similar characteristics. In addition, each iteration in the EM algorithm relies strongly on the results of previous iterations. If one of the channel-vectors from the Sybil packets is misinterpreted as benign one, other channel-vectors with same channel features are likely to be clustered as benign ones according to the iteration mechanism in the EM algorithm. Moreover, the MKEM scheme is proposed by importing the clustering results of the MKFcM algorithm into the EM algorithm, where the exportations from the EM algorithm are not fed back to the MKFcM. Hence, the FNR is increased because of the shortage of the EM algorithm. However, in the Fig. 8(a), the MKEM scheme still achieves better performance than multi-kernel scheme. The reason is that the Gap statistics analysis help to improve the classification results in MKEM scheme. In contrast to the FNR, Fig. 8(b) shows that the FPR decrease when the number of SNs are magnified. As the adversary SN deliveries numerous bogus packets by pretending to be multiple SN with fake identities, it has to occupy the channel to transmit large number of packets in a wireless propagation environment. Accordingly, normal SNs around the Sybil SNs tend to scan a busy channel condition rather than a free delivery state. This makes the channel-vectors received from the benign SNs provide significant differences with the bogus packets. Consequently, the packets from benign SNs tend to

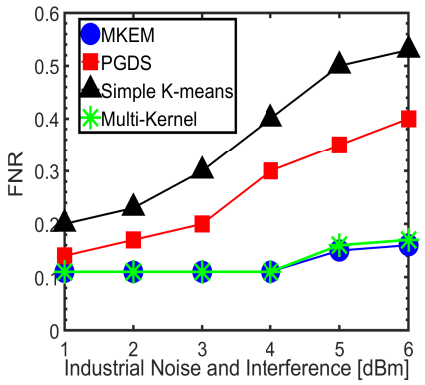


(a) FNR vs. the number of Sybil Attacker

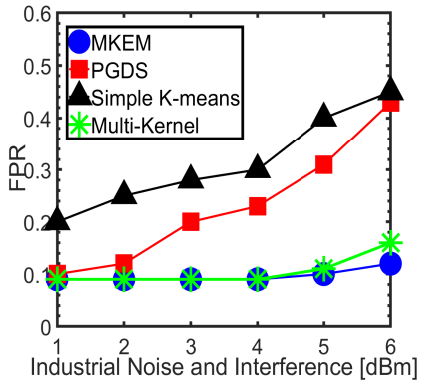


(b) FPR vs. the number of Sybil Attacker

Fig. 9: The Impacts of the Number of Sybil Attacker



(a) FNR vs. Noise and Interference



(b) FPR vs. Noise and Interference

Fig. 10: The Impacts of the Impulse Noise and Interference

be hardly mis-clustered as Sybil packets.

In Fig. 9, we show that the FPR rapidly increase to 50% after increasing the number of Sybil SNs to half of the total number of SNs in the network. When Sybil SNs take up to half of the network and scatterers in different places, the channel-vectors will be multiplied and diversified. This will lead to an increase in the possibility of mis-clustering channel-vectors from the benign SNs as those from Sybil ones during the clustering. In contrast to the FPR, Fig. 9(a) shows that the FNR decrease when the number of adversary SNs are magnified. The reason is that the increasing adversary SNs tend to provide numerous packets with similar channel characteristics. This reduces the quantity of the packets from benign SNs and further cuts down the diversity of the whole received packets, such that the bogus packets are easier to be figured out with increasing number of adversary SNs. In addition, in Fig. 8(a) and Fig. 9(b), both FNR and FPR increase while enlarging the networks size. However, this simulation is designed to be conducted in the unchanged scatterer distribution environment. That is, we set the simulation in a 100×100 map and keep the surrounding obstacles unmoved during the simulation. If we enlarge the network size, we will obtain a higher probability of the malicious SNs and benign SNs sharing similar surrounding scenario. Since the proposed MKEM scheme is based on the uncorrelated channel response due to spatial variance, the FNR and FPR will become worse when the probability of sharing similar surrounding environment increases. In both Figs. 8(a) and 9(b), the MKEM scheme achieves better detection performance than all the other schemes.

Fig. 10 shows that the FNR and FPR remain below 20% despite magnifying the impulse noise and interference effects. The impact of noise and interference effects on the clustering is diminished when we ascend the dimension of the channel-vectors into a new dimension space according to the kernel method and further amplify the included-angle between the channel-vectors having different channel features. However, in the case where the signal-to-noise ratio is decreased by increasing the noise and interference effects, the FNR and FPR will increase because the channel-vectors are hard to be reconstructed from the received signals.

VI. Conclusion

In this paper, we propose a MKEM scheme to detect the Sybil attacks in IWSNs by exploring the correlation between channel impulse responses from various spatial SNs. The simulation results show that the proposed scheme is capable of detecting malicious packets transmitted from the Sybil attackers without establishing a pre-labeled database of the channel features of all SNs. Moreover, the MKEM scheme can achieve high detection accuracy even if the wireless propagation channel is influenced by the noise and interference effects in the industrial environment. Furthermore, when the malicious SNs adaptively change their transmission power or increase their number, the MKEM scheme can still guarantee the detection accuracy. In future works, by studying more characteristics of the industrial wireless channel, we intend to produce a new unsupervised learning approach by combining

the idea of neural network and kernel methods to explore the hidden features of the Sybil attackers in IWSNs.

ACKNOWLEDGMENT

The work has been partly supported by the research grant from the Regional Research Fund of Norway (RFF).

REFERENCES

[1] X. Cao, P. Cheng, J. Chen, and Y. Sun, "An online optimization approach for control and communication codesign in networked cyber-physical systems," *IEEE Trans. on Ind. Informat.*, vol. 9, no. 1, pp. 439–450, 2013.

[2] G. Hackmann, W. Guo, G. Yan, Z. Sun, C. Lu, and S. Dyke, "Cyber-physical codesign of distributed structural health monitoring with wireless sensor networks," *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 25, no. 1, pp. 63–72, 2014.

[3] M. Cheffena, "Industrial wireless communications over the millimeter wave spectrum: opportunities and challenges," *IEEE Commun.*, vol. 54, no. 9, pp. 66–72, 2016.

[4] L. Xiao, X. Wan, and Z. Han, "Phy-layer authentication with multiple landmarks with reduced overhead," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 3, pp. 1676–1687, 2018.

[5] Q. Li, K. Zhang, M. Cheffena, and X. Shen, "Exploiting dispersive power gain and delay spread for sybil detection in industrial wsns," in *IEEE/CIC Proc. of ICCC*, 2016, pp. 1–6.

[6] K. Zhang, X. Liang, R. Lu, and X. Shen, "Sybil attacks and their defenses in the internet of things," *IEEE J. on Internet of Things*, vol. 1, no. 5, pp. 372–383, 2014.

[7] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. on Autom. Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[8] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based detection of sybil attacks in wireless networks," *IEEE Trans. Inf. Forens. Security*, vol. 4, no. 3, pp. 492–503, 2009.

[9] J. K. Tugnait, "Wireless user authentication via comparison of power spectral densities," *IEEE J. on Sel. Areas in Commun.*, vol. 31, no. 9, pp. 1791–1802, 2013.

[10] M. Demirbas and S. Youngwhan, "An rssi-based scheme for sybil attack detection in wireless sensor networks," in *IEEE Symp. on WoWMoM*, 2006, pp. 565–570.

[11] S. Ruj, A. Nayak, and I. Stojmenovic, "Pairwise and triple key distribution in wireless sensor networks with applications," *IEEE Trans. on Comput.*, vol. 62, no. 11, pp. 2224–2237, 2013.

[12] T. Olofsson, A. Ahlén, and M. Gidlund, "Modeling of the fading statistics of wireless sensor network channels in industrial environments," *IEEE Trans. on Signal Process.*, vol. 64, no. 12, pp. 3021–3034, 2016.

[13] Q. Xiong, Y. Liang, K. Li, and Y. Gong, "An energy-ratio-based approach for detecting pilot spoofing attack in multiple-antenna systems," *IEEE Trans. on Inf. Forens. and Security*, vol. 10, no. 5, pp. 932–940, 2015.

[14] Q. Li, K. Zhang, M. Cheffena, and X. Shen, "A measurement-based boundary estimation approach for localization in industrial wsns," in *IEEE Proc. of ICC*, 2017, pp. 1–6.

[15] M. Cheffena, "Industrial wireless sensor networks: channel modeling and performance evaluation," *J. on Wireless Commun. and Netw. EURASIP*, pp. 1–8, 2012.

[16] Q. Li, K. Zhang, M. Cheffena, and X. Shen, "Channel-based sybil detection in industrial wireless sensor networks: a multi-kernel approach," in *IEEE Proc. of Globecom*, 2017, pp. 1–6.

[17] S. Shin, T. Kwon, G. Jo, Y. Park, and H. Rhy, "An experimental study of hierarchical intrusion detection for wireless industrial sensor networks," *IEEE Trans. on Ind. Informat.*, vol. 6, no. 4, pp. 744–757, 2010.

[18] J. Tang, H. Wen, L. Hu, H. Song, G. Zhang, F. Pan, and H. Liang, "Associating mimo beamforming with security codes to achieve unconditional communication security," *IET Communications*, vol. 10, no. 12, pp. 1522–1531, 2016.

[19] J. Newsome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor networks: analysis & defenses," in *IEEE Symp. on IPSN*, 2004, pp. 259–268.

[20] K. Islam, W. Shen, and X. Wang, "Wireless sensor network reliability and security in factory automation: A survey," *IEEE Trans. on Syst. Man and Cybern. Part C (Appl. and Rev.)*, vol. 42, no. 6, pp. 1243–1256, 2012.

[21] J. Tang, H. Wen, K. Zeng, L. Hu, and S. Chen, "Achieving unconditional security for mimo-ban under short blocklength wiretap code," in *Proc. of VTC*, 2017, pp. 1–5.

[22] J. Tang, H. Wen, H. Song, and F. Pan, "Combining mimo beamforming with security codes to achieve unconditional communication security," in *IEEE/CIC Proc. of ICCC*, 2015, pp. 105–109.

[23] D. Sexton, M. Mahony, M. Lapinski, and J. Werb, "Radio channel quality in industrial wireless sensor networks," in *Proc. of SIC*, 2005, pp. 88–94.

[24] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 24, no. 1, pp. 44–58, 2013.

[25] Y. Chen, J. Yang, W. Trappe, and R. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. on Veh. Technol.*, vol. 59, no. 5, pp. 2418–2434, 2010.

[26] Y. Liu, D. Bild, R. Dick, M. Mao, and D. Wallach, "The mason test: A defense against sybil attacks in wireless networks without trusted authorities," *IEEE Trans. on Mobile Computing*, vol. 14, no. 11, pp. 2376–2391, 2015.

[27] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "Phy-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. on Vehicular Tech.*, vol. 65, no. 12, pp. 10 037–10 047, 2016.

[28] Z. Gong, M. Frank, and P. Mittal, "Sybilbelief: A semi-supervised learning approach for structure-based sybil detection," *IEEE Trans. on Inf. Forens. and Security*, vol. 9, no. 6, pp. 976–987, 2014.

[29] I. S. Association, "Ieee standard for local and metropolitan area networks — part 15.4: Low-rate wireless personal area networks (lr-wpans)," 2011.

[30] B. Kuo, H. Ho, C. Li, C. Hung, and J. Taur, "A kernel-based feature selection method for svm with rbf kernel for hyperspectral image classification," *IEEE J. of Sel. Topics. in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 317–326, 2014.

[31] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[32] H. Huang, Y. Chuang, and C. Chen, "Multiple kernel fuzzy clustering," *IEEE Trans. on Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, 2012.

[33] W. Dong and X. Liu, "Robust and secure time-synchronization against sybil attacks for sensor networks," *IEEE Trans. on Ind. Informat.*, vol. 11, no. 6, pp. 1482–1491, 2015.

[34] M. Alsheikh, S. Lin, D. Niyato, and H. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.

Qihao Li (M'16) received the M.Sc. degree in Information and Communication Technology from University of Agder, Norway, in 2013. He is currently pursuing his Ph.D. degree in Norwegian University of Science and Technology (NTNU), Norway. His current research interests include industrial wireless sensor network, optimal control and optimization, wireless network security and localization. He served as the a member of Technical Program Committee for IEEE Globecom' 19, IEEE ICC' 19, EuCAP' 19, IEEE Globecom' 18, IEEE ICC' 18, IEEE CIC ICCC' 18, BDEC-SmartCity' 18, IEEE Globecom' 17, IEEE CIC ICCC' 17.

Michael Cheffena received his M.Sc. degree in electronics and computer technology from the University of Oslo, Norway in 2005 and his Ph.D. degree from the Norwegian University of Science and Technology (NTNU), Trondheim, in 2008. In 2007, he was a visiting researcher at the Communications Research Center, Canada. From 2009 to 2010, he conducted a postdoctoral study at the University Graduate Center, Kjeller, Norway, and at the French Space Agency, Toulouse. Currently, he is a full professor at NTNU, Gjøvik. His research interests include modeling and prediction of propagation radio channels, signal processing and medium access control (MAC) protocol design.