Channel-based Sybil Detection in Industrial Wireless Sensor Networks: a Multi-kernel Approach

Qihao Li*, Kuan Zhang†, Michael Cheffena* and Xuemin (Sherman) Shen†
*Faculty of Technology, Economy and Management, Norwegian University of Science and Technology
Teknologivn. 22, Gjøvik N-2815, Norway
†Department of Electrical and Computer Engineering, University of Waterloo
200 University Avenue West, Waterloo, N2L 3G1, Canada
Email: {qihao.li,micheal.cheffena}@ntnu.no; {k52zhang, xshen}@bbcr.uwaterloo.ca

Abstract—**Industrial Wireless Sensor Networks (IWSNs) integrate various types of sensors to measure and control industrial production. However, the unattended open environment makes IWSNs vulnerable to malicious attacks, such as Sybil attacks, which may degrade the network performance. In addition, multipath distortion, impulse noise and interference effects in the harsh industrial environment may influence the accuracy of attack detection. In this paper, we propose a Sybil detection scheme based on power gain and delay spread analysis by exploiting the spatial variability from their channel responses. Specifically, we utilize channel-vectors to represent the sensor features based on the power gain and delay spread extracted from channel response. Furthermore, we develop a kernel-oriented method to distinguish Sybil attackers from benign sensors by clustering the channel-vectors. In addition, to alleviate the impact of industrial noise and interference effects, we design a multi-kernel based fuzzy c-means method to map the extracted channel-vectors into a new feature space such that the dispersive effects on the channel-vectors can be reduced. We also propose a parameter selection method to optimize the employed kernels. The simulation results show that the proposed multi-kernel scheme can achieve high accuracy in detecting the packets from Sybil attackers, and tolerate the dispersive attenuation and interference effects in the industrial environments.**

I. Introduction

Industrial Wireless Sensor Networks (WSNs) provide an integrated platform to incorporate sensor networks with intelligent industrial automation systems, which bring benefits to the distributed environment monitoring, instrument fault diagnosing and multiple-system cooperating [1]. For example, in an industrial storage room, various products are automatically conveyed to their related trunks such that they can be packaged and transported to customers. Versatile sensors are installed in each storage trunk to track the corresponding position, real-time regional temperature and trunk saturation status. The sensing information is gathered by a handle controller acquiring the storage condition. After analyzing and processing this collected information, the handle controller feedbacks with control signallings to the automatic terminals. Therefore, IWSNs facilitate the flexible automated manufacturing by enabling reliable data collection and transmission even in areas difficult or impossible to reach.

However, the unattended open environment of wireless communication makes IWSNs vulnerable to identity-based attack, since attackers may gather important identity information during passive monitoring and utilize them to launch attacks, such as, **Sybil attacks**. A Sybil sensor may manipulate a large number of fake packets by forging different identities, such that the aggregated reading results may be modified [2], resulting in harmful and disastrous accidents in industrial environments. For example, an attack on sewage control system in Queensland leads to leakage of millions liters of untreated sewage into a stormwater drain. The attacker takes control of several sewage pumping stations and interacts the sensors to falsely feedback with normal states. Without periodical and dependable alarm from the sensors, the leakage of flammable liquid and harmful gases could contaminate the industrial environment and endanger the public. Therefore, it is necessary to detect the Sybil attacks and eliminate them from the IWSNs.

Research efforts have been put on Sybil detection in traditional WSNs [3]. The defense schemes against Sybil attack can be categorized into two types: *radio resource testing* [4] by exploring the resource characteristics of various physical terminals; and *key pre-distribution* [5] which utilizes key paired identification and key validation verification. However, the limited computational resources of sensors and time-delay validation of packets in WSNs, may decrease the capability of cryptographic protocols. To ensure the detection on Sybil attack, radio resource test becomes a preferable detection method in WSNs. The uniqueness of the wireless propagation channel of the transmitter can be used to discriminate the Sybil attackers without exchanging authentication messages. Nevertheless, it is challenging for channel-based Sybil detection schemes to maintain their precision in the harsh industrial environment due to the following issues. Firstly, it is difficult to distinguish attackers by comparing the received signal power which are attenuated by the scattering environment. Secondly, impulse noise and interference may decrease the signal-to-noise ratio and distort the envelop of the received signal such that the specific channel features of Sybil attackers may be deformed during transmission [1]. Thirdly, vibrating scatters bring further magnifications to phase and frequency of the received signal to affect the features extraction. These challenging issues motivate us to further improve the accuracy of channel-based Sybil detection scheme in a harsh industrial environment.

In this paper, we propose a kernel-based Sybil detection scheme in IWSNs based on the channel-vectors exploited from the power gain and delay spread. Specifically, the contributions of this paper are twofold.

- First, we propose an unsupervised kernel-oriented method to distinguish benign and malicious sensors by mapping the extracted channel-vectors into a higher dimensional Hilbert space. For the sake of discriminating the malicious sensors, we develop Kernel Parameters Selection (KPS) method to optimize the parameters in the proposed kernels by enlarging the included angles between the channel-vectors in different clusters and further decreasing them within the same clusters.
- Second, considering an affine combination of the proposed kernels, we propose a Multiple Kernel Fuzzy c Means (MKFcM) method to improve the anti-interference capability of the Sybil detection performance in the industrial environment. In addition, as an unsupervised learning algorithm, MKFcM fairly evaluates all the observations with determined clusters such that each observation has a chance to compare with others.

The remainder of this paper is organized as follows. The related works are presented in Sec. II. We present the network and attack models in Sec. III. We propose the detailed multi-kernel scheme in Sec. IV. Finally, we conclude this paper in Sec. V

II. Related Work

A plethora research efforts have been put on the radio resource test to detect the Sybil attacks in the conventional WSNs [3]. In [6], with channel response de-correlativity, the spatial variability of wireless channel is considered to detect Sybil attack based on deriving a generalized likelihood ratio test. In [7], attackers have autonomy and flexible control over their behaviors. The interactions between benign and malicious users are derived as a zero-sum authentication game with channel parameters optimized by Q-learning method. The characteristic of the Sybil attackers can also be pre-defined according to the experience of their behaviors. [8] explores the contacts and pseudonym changing behaviors of the Sybil attackers in mobile social networks. By exploring the difference in the contact rate distribution between the benign and malicious users, a semi-supervised learning method based on hidden Markov model is proposed to detect the collusion of mobile users. The Sybil attacks are further categorized into three types and discussed upon their defense schemes in [2]. [9] employs the variance of power distribution between the received signal strength in the same and different places. A signal-strength-based localization scheme is further introduced to detect malicious user positions. Besides, discrimination can also be determined based on the information which is obtained by negotiating with neighbors. A cooperated detection scheme is proposed in [10] by gathering neighbor transmission powers. Signalprints are initiated by selecting a subset of the received power observations which are shared among all the neighbors. In [11], users associate with each other and derive the signalprints by exploiting the similarity trust relationship among neighbors. There are also some other Sybil detection methods based on the unsupervised or semi-supervised learning approach. [12] proposes a fuzzy C-means method to softly cluster the probability of a user being malicious or benign by examining the deterministic sensing reports in a period. In [13], Sybil detection is determined by exploring the existence of users and their connection information. A semi-supervised learning approach is considered with a joint distribution probability over all pre-defined labels of benign or malicious users. The local rule according to the probability distribution is further used to classify and rank the remaining users by adopting loopy belief propagation algorithm.

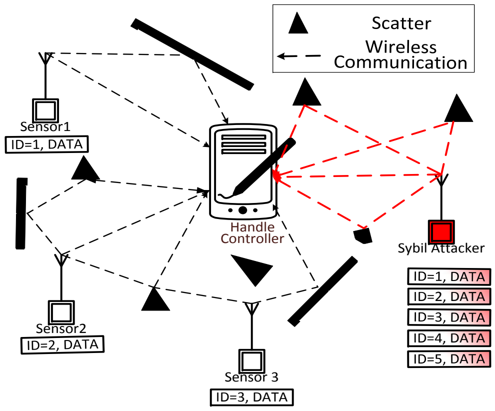


Fig. 1: Network Model

Different from the existing work, we propose a multi-kernel based Sybil detection scheme to explore the hidden features of unsupervised packets from Sybil attackers. The proposed multi-kernel scheme can exploit the spatial variability from the extracted channel-vectors and tolerate the noise effects in clustering the channel-vectors from Sybil attackers.

III. System Model

A. Network Model

The network model consists of **Sensors** and **Handle Controllers (HC)**, where carrier-sense multiple access with collision avoidance (CSMA/CA) is adopted to ensure the association, as shown in Fig. 1. The network operates in a manufacturing environment fulfilled with the scatters consisted by various kinds of materials such that heavy multipath attenuation takes a significant impact on the received signal. The scatters are uniformly distributed in the environment. In order to keep per-terminal cost low, sensors suffer limited storage and computational resources so that tamper-resistant hardware is un-occupied. Sensors are artificially positioned in the environment and fixed after the deployment. Moreover, sensors functionally perceive their surrounding environment and forward the acquired data to the controller. HC is a wireless receiver, which can harvest and inquire information from the sensors. HC keeps stabilization to promise the interaction to the sensors within its communication range such that the time-varying fading effects can be diminished.

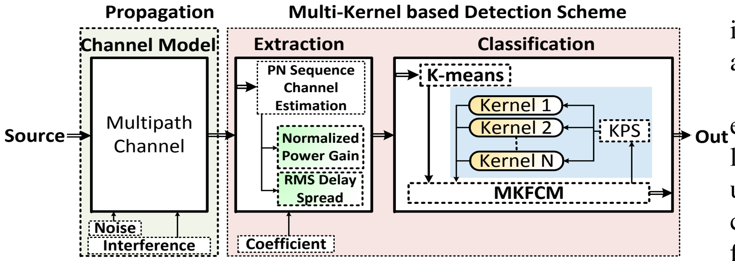


Fig. 2: Overview of the Multi-Kernel Scheme

B. Attacker Model

A malicious sensor may gather profitable information and launch attack, since sensor networks usually operate in an unattended and open environment. Instead of ruling the channel behavior by re-broadcasting the eavesdropped packets with varying transmission power and delay, the malicious sensors can only listen and record the critical information but not re-broadcast it to the HC. We describe the attacks as below:

- **Sybil Attack**: Sensors are compromised by attackers and then manipulate fake identities to pretend to be multiple sensors to launch a Sybil attack. Because the sensors hardly employ the cryptographic authentications in consideration of the limited memory storages and computational resources, the network cannot promise the confidentiality such that the bogus packets labeled with multiple fake identities are received by the HC.
- **Various spatial Sybil attacks**: Sensors deployed in different places are compromised by attackers and then produce fake identities to launch Sybil attacks. The fake packets are transmitted from different places such that dissimilar channel features can be extracted to decrease the detection accuracy.
- **Sybil attack hidden in heavy interference effects**: Since the received packets are distorted by industrial noise and interference effects, the channel features extracted from packets transmitted by same malicious sensor may have different appearances. Malicious sensor may detect the channel state and hide the attacks in heavy interference effects.

IV. The Proposed Scheme

A. Overview

The multi-kernel scheme consists of two steps, extraction and classification, as shown in Fig. 2. The channel model is built up with attenuation factor, delay, Doppler frequency and phase from each propagation path. Impulse noise and interference are also considered in the model. The inputs of the multi-kernel scheme are the packets transmitted through the wireless channel, which may be affected by multipath fading, impulse noise, and interference effects.

In the extraction step, the channel impulse responses are estimated by evaluating the auto-correlation function of the Pseudo Noise (PN) sequences in the received packets. The power gain and delay spread are further extracted from the impulse responses and combined into a two-dimension channel-vectors. The channel-vectors containing the hidden clustering information between the Sybil attackers and benign sensors are imported into the classification block.

In the classification step, the channel vectors are firstly explored by **K-means** clustering, which is an unsupervised learning method used to roughly explore hidden patterns from unlabeled data sets. The grouped data sets and predicted cluster number are then brought into the **MKFcM** block for further classification. The MKFcM algorithm alleviates the effects of impulse noise and interference by mapping the grouped data set from the data space into Hilbert space based on the multiple linear combinatorial kernel components. Furthermore, we provide **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.

B. Data Extraction

In this section, we derive the channel-vector observations by extracting the propagation channel features from the received packets. Since the signal propagation distance and the spatial variance can be used to analyze the difference between Sybil and benign sensor nodes, we define power gain and delay spread as the clustering features. Specifically, we export the channel gain and delay spread from the channel impulse response which can be easily obtained by approximating the auto-correlation function of the PN sequences in the received packets. Moreover, we denote the *power gain* as $\alpha_m(t-\tau_m)$ and the *delay spread* as τ_m in the sample slot, where $m \in \{1, \cdots, M\}$ is the number of the scatters. Firstly, we normalize the power gain as the first part of the extracted vector, which is given by $g_m(t) = \frac{\alpha_m(t-\tau_m)}{\sum_{\tau_m} \alpha_m(t-\tau_m)}$. Secondly, root mean square (rms) of the delay spread is indicated as the second part of the extracted vector, which is given by $\sigma_m(t) = \sqrt{\overline{\tau_m^2(t)} - (\overline{\tau_m(t)})^2}$, where $\overline{\tau_m^2(t)} = \frac{\tau_m^2 \alpha_m^2(t-\tau_m)}{\sum_m \alpha_m^2(t-\tau_m)}$. We can investigate the topology distribution characteristics of the scatters around the sensors. The import data set to the classification block can be defined as the channel-vector, which is given as a two-dimension data set constituted by $g_m(t)$ and $\sigma_m(t)$ in the sample time slot.

However, in order to deceive the controllers, a clever malicious sensor may change the transmission power when it attempts to claim different identities. In this case, we reconstruct the channel-vector and further provide an adapter factor γ to balance the importance of the power gain during the classification. To this end, we define the channel-vector at time t as

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

We give the data set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, where the data point $\boldsymbol{x}_i = \boldsymbol{x}_i(t) \in \Xi \subseteq R^2$, $i = 1, \ldots, N$, N is the number of packets received in the sampling time slot. $\boldsymbol{x}_i$ is a data point deployed on a R^2 plane characterizing by the power gain and delay spread. The objective is to optimize the estimation of the pattern clustering based on the knowledge of the extracted channel-vectors $\boldsymbol{x}_i$ under an environment fulfilled with multipath attenuation, impulse noise and interference effects.

C. Multi-Kernel Fuzzy c-Means Clustering

Since the extracted channel features may be distorted by the noise and interference, we have to develop a nonlinear function to cluster the Sybil attackers from the benign sensors in the original space. However, it is too complex to find out such a function without pre-defined samples. We propose a unsupervised MKFcM method to map the channel-vector set X from the data space $\Xi \subseteq \mathcal{R}^2$ into Hilbert space H by a function $\phi : \Xi \rightarrow H$ [14]. After this reproduction in the kernel Hilbert space, the channel-vectors are more easily to be clustered and more robustness to handle the noise. Generally, we consider the kernel function instead of discussing the mapping function ϕ, which is given by $\kappa(\boldsymbol{x},\boldsymbol{y}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \Xi$, where $\kappa : \Xi \times \Xi \rightarrow R$ and $\langle , \rangle$ is the inner product of Hilbert space. Additionally, the objective of ascending the dimension is to diminish the influence of the noise in the obtained channel-vectors. We transit the channel-vectors with similar features into the same space in the proposed Hilbert space and separate the different vectors into the disparate space. In order to flexibly move the vector, we provide various scalar multiplications with respect to different vector additions while ascending the dimension. We iteratively optimize the parameters in the kernel function and scalars with respect to the multiplication to achieve better clustering. Considering M such kinds of mappings, i.e., $\Phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$, we re-write the $\boldsymbol{x} \in R^2$ with respect to the transform function ϕ_k as $\phi_k(\boldsymbol{x})$. If we let $\{\kappa_1, \kappa_2, \ldots, \kappa_M\}$ be the Mercer kernels corresponding to these implicit mappings, we have the inner product with respect to k-th kernel transition as $\kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi_k(\boldsymbol{x}_i)^T \phi_k(\boldsymbol{x}_j)$. To combine these kernels, we first list some necessary Mercer kernels' properties in the following.

Theorem: Let κ_1 and κ_2 be kernels over $\Xi \times \Xi$, $\Xi \subseteq R^p$. Let function $\psi : \Xi \rightarrow R^p$

- $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \kappa_1(\boldsymbol{x}_i, \boldsymbol{x}_j) + \kappa_2(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a kernel.
- $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \alpha \kappa_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a kernel, when $\alpha > 0$.

The proof of these properties can be referred to [15].

Ensuring that the resulted kernel still satisfies Mercer's condition, we consider a nonnegative combination of these mappings, ϕ', i.e., $\phi'(\boldsymbol{x}) = \sum_{k=1}^{M} w_k \phi_k(\boldsymbol{x})$, where $w_k \geq 0$ is the kernel weight of the k-th mapping function $\phi_k(\cdot)$. As these implicit mappings do not necessarily have the same dimensionality, we construct a new set of independent mappings $\Psi = \{\psi_1, \ldots, \psi_M\}$ from the direct linear combination of the original mappings ϕ as [16]:

$$\sum_{k=1}^{M} \psi_k(\boldsymbol{x}) = \begin{bmatrix} \phi_1(\boldsymbol{x}) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \phi_2(\boldsymbol{x}) \\ \vdots \\ \mathbf{0} \end{bmatrix} + \cdots + \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \phi_M(\boldsymbol{x}) \end{bmatrix} \tag{2}$$

In addition, these mappings form a new set of orthogonal bases that $\psi_k(\boldsymbol{x}_i)^T \psi_k(\boldsymbol{x}_j) = \kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\psi_k(\boldsymbol{x}_i)^T \psi_{k'}(\boldsymbol{x}_j) = 0$, if $k \neq k'$. Suppose that $\{\boldsymbol{x}_{ic}\}_{i=1,\ldots,N_c}$ are the set of training channel-vectors in the cluster c (sensor c), where N_c is the number of training channel-vectors in cluster c, $c = 1, \ldots, C$, and C is the number of clusters. The kernel is defined as $\kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) = \exp\left(-\frac{\|\boldsymbol{x}_{ic} - \boldsymbol{x}_{jc}\|^2}{2\sigma_k^2}\right)$, 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 [15]:

- $\kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) = 1$. 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.

Two properties are further described as follows: 1) channel-vectors in the same cluster should be mapped into the same area, and 2) channel-vectors in different clusters should be mapped into different areas. A proper parameter σ_k should be found, such that $\kappa(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \boldsymbol{\sigma}) \approx 1$ and $\kappa(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc'}, \boldsymbol{\sigma}) \approx 0$ if $c \neq c'$. Two criterias are proposed for measuring separability in the feature space. First, the mean distances of channel-vectors belonging to the same cluster 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 $\kappa(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2 \kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k)$. The parameter σ_k should be optimized such that $\mathcal{S}_c(\sigma_k)$ is close to 1. We reform (3) and further simplify the expression as

$$\mathcal{S}_c(\boldsymbol{w}, \boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2 \mathcal{A}_k(\sigma_k) \tag{4}$$

$$\mathcal{A}_k(\sigma_k) = \frac{1}{\sum_{c=1}^{C} N_c^2} \sum_{c=1}^{C} \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} \kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc}, \sigma_k) \tag{5}$$

Second, the mean distances of the channel-vectors belonging to different clusters is

$$\mathcal{D}_c(\boldsymbol{\sigma}) = \frac{1}{\sum_{c=1}^{C} \sum_{\substack{c'=1 \\ c' \neq c}}^{C} N_c N_{c'}} \sum_{c=1}^{C} \sum_{\substack{c'=1 \\ c' \neq c}}^{C} \sum_{i=1}^{N_c} \sum_{j=1}^{N_{c'}} \kappa(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc'}, \boldsymbol{\sigma}) \tag{6}$$

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. We further reform (6) and further simplify the expression as

$$\mathcal{D}_c(\boldsymbol{w}, \boldsymbol{\sigma}) = \sum_{k=1}^{M} w_k^2 \mathcal{B}_k(\sigma_k) \tag{7}$$

$$\mathcal{B}_k(\sigma_k) = \frac{1}{\sum_{c=1}^{C} \sum_{\substack{c'=1 \\ c' \neq c}}^{C} N_c N_{c'}} \sum_{c=1}^{C} \sum_{\substack{c'=1 \\ c' \neq c}}^{C} \sum_{i=1}^{N_c} \sum_{j=1}^{N_{c'}} \kappa_k(\boldsymbol{x}_{ic}, \boldsymbol{x}_{jc'}, \sigma_k) \tag{8}$$

Hence, the optimal σ_k can be obtained by solving the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{w}, \boldsymbol{\sigma}}{\text{argmin}} \quad & 1 - \mathcal{S}_c(\boldsymbol{w}, \boldsymbol{\sigma}) + \mathcal{D}_c(\boldsymbol{w}, \boldsymbol{\sigma}) \\ \text{s.t.} \quad & (4), (7) \text{ and } \sum_{k=1}^{M} w_k = 1, \quad w_k \geq 0 \end{aligned} \tag{9}$$

We first fix the $\boldsymbol{\sigma}$ to find the optimal kernel weight w_k

Algorithm 1: Kernel Parameter Selection

```
Data: Observation set $\{\boldsymbol{x}_{ic}\}_{i=1,\dots,N_c,c=1,\dots,C} \subset R_d$
Result: $\boldsymbol{w}, \boldsymbol{\sigma}$
Let $\boldsymbol{\sigma}^{old}$ be a randomly selected starting vector
$J(\boldsymbol{w}, \boldsymbol{\sigma}) \leftarrow 1 - \mathcal{S}_c(\boldsymbol{w}, \boldsymbol{\sigma}) + \mathcal{D}_c(\boldsymbol{w}, \boldsymbol{\sigma})$
repeat
    Calculate $\mathcal{A}_k(\sigma_k^{old})$, where $k = 1, \dots, M$
    Calculate $\mathcal{B}_k(\sigma_k^{old})$, where $k = 1, \dots, M$
    for each $k \in \{1, \dots, M\}$ do
        $w_k \leftarrow \left(\sum_{k'=1}^{M} \frac{\mathcal{B}_k(\sigma_k) - \mathcal{A}_k(\sigma_k)}{\mathcal{B}_{k'}(\sigma_{k'}) - \mathcal{A}_{k'}(\sigma_{k'})}\right)^{-1}$
        $H_0 \leftarrow \nabla^2 J(\sigma_\eta, w_k)$
        while $|\nabla J(\sigma_\eta, w_k)| > \epsilon$ do
            $s_\eta \leftarrow -H_\eta^{-1} \nabla J(\sigma_\eta, w_k)$
            $\sigma_{\eta+1} = \sigma_\eta + \alpha s_\eta$
            $y_\eta \leftarrow \nabla J(\sigma_{\eta+1}, w_k) - \nabla J(\sigma_\eta, w_k)$
            $H_{\eta+1} \leftarrow H_\eta + \frac{y_\eta y_\eta^T}{y_\eta s_\eta} - \frac{H_\eta s_\eta s_\eta^T H_\eta}{s_\eta^T H_\eta s_\eta}$
        end
        $\sigma_k \leftarrow \sigma_\eta$
    end
    Set $\boldsymbol{w}^{old} \leftarrow \boldsymbol{w}$ and $\boldsymbol{\sigma}^{old} \leftarrow \boldsymbol{\sigma}$
until $J(\boldsymbol{w}, \boldsymbol{\sigma}) - J(\boldsymbol{w}^{old}, \boldsymbol{\sigma}^{old}) < \epsilon$
Output: $\boldsymbol{w}, \boldsymbol{\sigma}$
```

with Lagrange multiplier λ with respect to the constraint $\sum_{k=1}^{M} w_k = 1$, and then define the Lagrange function as $J_\lambda = 1 - \sum_{k=1}^{M} w_k^2 \mathcal{B}_k(\sigma_k) + \sum_{k=1}^{M} w_k^2 \mathcal{A}_k(\sigma_k) + \lambda \left(\sum_{k=1}^{M} w_k - 1\right)$. We take its derivatives with respect to the weights and set them to zero to obtain the close-form of weights w_k

$$w_k = \left(\sum_{k'=1}^{M} \frac{\mathcal{B}_k(\sigma_k) - \mathcal{A}_k(\sigma_k)}{\mathcal{B}_{k'}(\sigma_{k'}) - \mathcal{A}_{k'}(\sigma_{k'})}\right)^{-1} \tag{10}$$

And we consider quasi-Newton algorithm to find the minimum value of σ_k with the obtained w_k The proposed kernel parameter algorithm is summarized in Alg. 1.

A nonnegative linear expansion of the based in Ψ, i.e., $\psi(\boldsymbol{x}) = \sum_{k=1}^{M} w_k \psi_k(\boldsymbol{x})$, which maps data to an implicit feature space, is investigated below. 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. Hence, the objective function is

$$\begin{aligned} \underset{\boldsymbol{w},\boldsymbol{u},\boldsymbol{v}}{\text{argmin}} \quad & \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ic}^m \|\psi(\boldsymbol{x}_i) - \boldsymbol{v}_c\|^2 \\ \text{s.t.} \quad & \psi(\boldsymbol{x}_i) = \sum_{k=1}^{M} w_k \psi_k(\boldsymbol{x}_i), \quad \sum_{c=1}^{C} u_{ic} = 1, \quad u_{ic} \geq 0 \end{aligned} \tag{11}$$

where $\boldsymbol{v}_c$ is the center of the c-th cluster in the implicit feature space and can be defined as $\sum_{i=1}^{N} \sum_{i'=1}^{N} \frac{u_{ic}^m}{u_{i'c}^m} \psi(\boldsymbol{x}_i)$, $\boldsymbol{w} = (w_1, w_2, \dots, w_M)^T$ 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. The hidden labels of the channel-vectors are acquired by finding the clusters with the maximum weight $\boldsymbol{u}$ with respect to the distance from the channel-vectors to that cluster center. If there are packets with multiple different IDs from the same cluster, these packets may be delivered by Sybil attackers. We further compare these IDs with IDs in other clusters to figure out which sensor node is malicious. The MKFcM algorithm is summarized in Alg. 2.

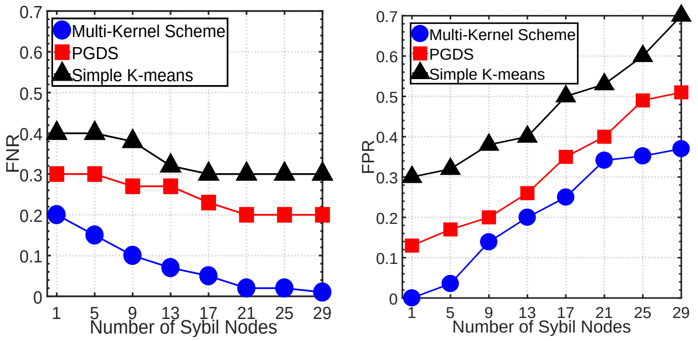
Algorithm 2: Multi-Kernel Fuzzy c Means.

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

V. Performance Evaluation

We evaluate the performance in terms of false positive rate (FRP) and false negative rate (FNR). FPR refers to the probability of mistaking a packet from benign sensors for a bogus one, while FNR refers to the probability of regarding a packet from the Sybil attackers as a benign one. The details of FPR and FNR are formulated in [17]. The variables employed in the scheme include the number of Sybil nodes and the industrial impulse noise with interference effects. The wireless propagation channel model is designed according to [17]. Following the steps explained in the Sec. IV-A, we extract the channel-vector with respect to $\gamma = 0.2$. In addition, we compare the performance of the multi-kernel scheme with the PGDS scheme proposed in [17].

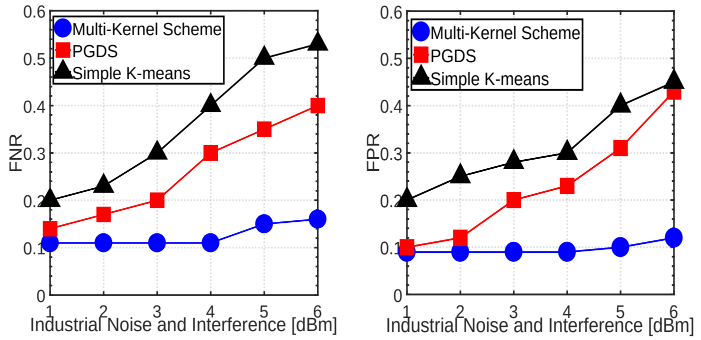
Fig. 3 shows that FPR rapidly increase to 50% after increasing the number of malicious sensors to half of the total number of sensors in the network. When malicious sensors take up to half of the network and deploy in different places, the channel-vectors are multiplied and diversified. This may increase the possibility of mis-clustering channel-vectors from the benign sensors as those from Sybil ones during the clustering. In contrast to FPR, Fig. 3a shows that FNR decreases when the number of adversary sensors are magnified. The reason is that the numerous packets with similar channel characteristics are



(a) FNR vs. the number of Sybil Attackers

(b) FPR vs. the number of Sybil Attackers

Fig. 3: The Impacts of the Number of Sybil Attackers



(a) FNR vs. Noise and Interference

(b) FPR vs. Noise and Interference

Fig. 4: The Impacts of the Impulse Noise and Interference

provided by increasing the number of malicious sensors. This reduces the quantity of the packets from benign sensors and further cuts down the diversity of the whole received packets, such that the bogus packets are easier to be discriminated from the whole received packets. In addition, in Fig. 3b, FPR increases while enlarging the networks size. The simulation is designed to be conducted in an unchanged scatter 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 can obtain a higher probability that the malicious and benign sensors share a similar surrounding scenario. Since the proposed multi-kernel scheme is based on the uncorrelated channel response due to spatial variance, FPR increases when more channel-vectors show resemblance in channel characteristics.

Fig. 4 shows that FNR and FPR remain below 20% despite magnifying the impulse noise and interference effects. This is because trying to reduce the impact of noise and interference effects, we ascend the dimension of the channel-vectors into a new dimension space according to the kernel method and further intensify the de-correlation of channel-vectors between different clusters. However, in the case that the noise and interference effects are magnified to affect the signal-to-noise ratio of the received packets, FNR and FPR increase because the channel-vectors are hard to be constructed.

VI. Conclusion

In this paper, we have proposed a novel Sybil detection scheme in IWSNs by exploiting the correlation between channel responses from various spatial sensors. The proposed unsupervised multi-kernel scheme has investigated the impacts of the multipath fading, impulse noise and interference effects on the Sybil detection accuracy in the industrial environments. An affine combination of kernels which are optimized by parameter selection method is employed with fuzzy c-means algorithm to discriminate the channel-vectors from Sybil attackers. The noise in the distorted channel-vectors can be alleviated by mapping the channel-vectors into Hilbert space. The simulation results have shown that in the harsh industrial environment, the multi-kernel scheme increases detection accuracy with an anti-interference capability. In the future work, we intend to investigate the hidden features of the Sybil attackers in the IWSNs and propose an unsupervised learning approach to study the observations with hidden features.

References

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

[2] 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.

[3] I. Butun, S. Morgera, and R. Sankar, “A survey of intrusion detection systems in wireless sensor networks,” *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 1, pp. 266–282, 2014.

[4] 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.

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

[6] L. Xiao, L. Greenstein, N. 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.

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

[8] K. Zhang, X. Liang, R. Lu, K. Yang, and X. Shen, “Exploiting mobile social behaviors for sybil detection,” in *Proc. of INFOCOM*, 2015, pp. 271–279.

[9] 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.

[10] Y. Liu, D. Bild, R. Dick, Z. 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.

[11] G. Wang, F. Musau, S. Guo, and M. Abdullahi, “Neighbor similarity trust against sybil attack in p2p e-commerce,” *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 26, no. 3, pp. 824–833, 2015.

[12] L. Li and C. Chigan, “Fuzzy c-means clustering based secure fusion strategy in collaborative spectrum sensing,” in *Proc. of IEEE ICC*, 2014, pp. 1355–1360.

[13] 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.

[14] L. Chen, C. P. Chen, and M. Lu, “A multiple-kernel fuzzy c-means algorithm for image segmentation,” *IEEE Trans. on Syst. Man and Cybern. Part B (Cybern.)*, vol. 41, no. 5, pp. 1263–1274, 2011.

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

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

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