IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON SPECIAL SECTION ON PROXIMITY SERVICE (PROSE) CHALLENGES AND APPLICATIONS

Received November 30, 2018, accepted December 12, 2018, date of publication December 19, 2018,
date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2888693

Clone Detection Based on Physical Layer Reputation for Proximity Service

FEI PAN[1], (Student Member, IEEE), ZHIBO PANG[3], (Senior Member, IEEE), MING XIAO[4], (Senior Member, IEEE), HONG WEN[2], (Senior Member, IEEE), AND RUN-FA LIAO[1]

[1]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China
[3]ABB Corporate Research, 72226 Västerås, Sweden
[4]KTH-Royal Institute of Technology, 10044 Stockholm, Sweden

Corresponding author: Hong Wen (70000413@qq.com)

This work was supported by the National Key R&D Program of China under Grant 2018YFB0904900 and Grant 2018YFB0904905.

ABSTRACT Proximity-based service (ProSe) provides direct communications among smart sensor nodes in proximity which aims at reserving resource consumption and alleviating the load in base stations, which is a promising solution for smart sensor systems that possess limited computing and energy resources. During the ProSe direct communications, most of the prior art security methods are usually provided by the ProSe function and are based on complex cryptography. However, despite the computing complexity, it is difficult for cryptographic methods to detect clone attack which is a common kind of attack in sensor systems. Clone nodes feature different physical positions but claim colliding IDs with captured nodes. Thus, clone nodes can be detected by spatial differences, in particular, by the surveillance of physical layer channel state information (CSI). However, CSI is not absolute static due to the random noise in wireless propagation environment. Accordingly, the detection accuracy varies with the stability of CSI. To address this challenge, we take the first attempt to introduce physical layer reputation and then elaborate the physical layer reputation based clone detection protocol to detect clone attack in multiple scenarios. The proposed protocol significantly improves the detection rate and false alarm rate and it is validated both by simulations and realizations.

INDEX TERMS Clone detection, proximity service, reputation based detection, smart sensor network.

I. INTRODUCTION

Smart sensor networks normally consist of low cost and resource constrained sensor nodes and these nodes are always applied in depopulated zones, barely under supervision. Consequently, the security performance of the network is highly constrained and it is prone to various attacks, for example, malicious node attacks. In active researches, reputation is introduced to detect malicious nodes. The core idea of CONFIDANT protocol [1] and OCEAN protocol [2] is to keep neighbor nodes under surveillance and acquire reputations from their communication behaviors. If the reputation is lower than a certain threshold, attack alarm will be triggered. In protocols as CORE [3] and BRSN [4], reputation acquisition method combines the first-hand and second-hand information, whereby scattering watchdogs across the network. However, slandering is a common weakness among those protocols. Prior art researches have tried to address this challenge by degrading reputations of slander nodes as [5], or by depending on extra reputations of evaluation behavior as [6]. Nonetheless, the first-hand observation of a node is based on the assertions from itself, namely, based on the integrity of the node itself. The second-hand observations are trustworthy only when no slandering occurs. It is also difficult to guarantee the credibility of evaluation reputation. Therefore, low evidence is a ubiquitous demerit of these kinds of reputation based schemes.

The protocols discussed above are committed to general malicious node attacks, including clone attacks. Some researchers have devoted themselves to concentrating on clone attack detection. Clone attack is one of the crucial malicious node attacks. To launch clone attack, the adversary first captures a legitimate node and sneaks all the confidential information. In the next moment, he is able to scatter clone nodes across the network by claiming the same identity of the captured node. The foundation of clone detection are colliding identities and different locations. To elaborate, clone nodes possess different locations from the captured nodes, although they all claim the same identity.

TABLE 1. Comparison of clone detection protocols.

Protocol	Core Idea	Demerits
CONFIDANT [1], **OCEAN** [2]	Based on first-hand upper layer reputation.	Vulnerable to dishonest nodes.
CORE [3], **BRSN** [4]	Based on first-hand and second-hand reputation.	Vulnerable to slandering.
Reference [5]	Degrades reputations of slander nodes.	Vulnerable to false judgement of slander nodes.
Reference [6]	Based on extra reputations of evaluation behaviors.	Vulnerable to false judgement of evaluation behaviors.
Node-To-Network Broadcasting [7]	Floods location information.	Communication load increases significantly with the increasing of network scale.
Deterministic Multicast [7]	Broadcasts location information to pre-determined witness nodes.	Sensitive to failed witness nodes.
Randomized Multicast [7]	Broadcasts location information to random selected witness nodes.	Constrained by witness selection probability.
LSM [7]	Broadcasts location information to witness nodes on pre-determined witness route lines.	Constrained by the design of witness route.
RAWL [8]	Selects witness nodes by random walking.	Constrained by the distribution of witness nodes.
LSCD [9], **ERCD** [10]	Selects witness nodesby non-staight witness routing.	Constrained by witness routing algorithm.
RF based detection [19]	Identifies nodes by RF fingerprints.	Affected by environment, e.g., temperature and humidity.
CSI based detection [17], [18]	identifies nodes by CSI.	CSI is sensitive to channel quality.

Parno *et al.* [7] proposes four approaches to detect clone attack which are known as Node-To-Network Broadcasting, Deterministic Multicast, Randomized Multicast and Line-Selected Multicast (LSM). Preliminarily, Node-To-Network Broadcasting detects colliding identities by flooding the location information all over the network. Further, Deterministic Multicast decreases the communication cost by broadcasting the information to a limited number of determined witness nodes. Alternatively, Randomized Multicast makes improvement by randomly nominating witness nodes. In the rear, LSM makes more progresses by designating witness nodes following a couple of route lines. As a result, the clone attack can, at best, be detected at the intersections of witness route lines. In this circumstance, witness routing becomes a fundamental design objective since the detection performance is highly constrained by the quantity and distribution of the witness route intersections. To achieve a better detection coverage, Zeng in RAWL [8] takes steps forward to come down in favor of random walking witness routes, no longer straight routes. To take a next step, [9] and [10] propose two more complex witness routing algorithms, LSCD and ERCD, while reducing memory cost and extending network lifetime. In spite of the complexity of routing, the potential of miss detection escalates with the increasing number of failing witness nodes.

The recent technology breakthrough provides a promising solution to address the above challenges by utilizing the non-repudiation of physical layer characteristics of devices. Physical layer security has been continuously expanding its application scope [11]–[13]. Physical layer channel state information (CSI) features good spatial differentiability and analytical unbreakability. It has been demonstrated by theory [14] and experiments [15], [16] that channel realizations are essentially diacritical when the transmitters are separated by more than half a wavelength. Accordingly, CSI based clone detection methods step on the stage by taking advantages of CSI uniqueness [17], [18]. To elaborate upon this idea, receivers extract useful information from noisy CSI traces and compare it with an appropriate reference CSI. If they are not similar enough, a clone attack alarm will be triggered. Correspondingly, locations are not claimed by transmitters or watchdogs and no witness routing is involved. The CSI based detection are better making its way against demerits of conservative solutions. Nonetheless, CSI is not absolute static which accordingly degrades detection accuracy. The comparison of the afore mentioned methods are presented in Table 1.

Proximity-based service (ProSe) is able to provide direct discovery and direct communication [20] in sensor systems. In general, the ProSe function discovers sensor nodes in proximity and authorizes them to communicate without base stations. The existing security mechanism in ProSe is accomplished at ProSe function which is mostly secret keys management [21]. Once legitimate nodes are captured, the attacker will get all the confidential information, including secret keys, and conduct clone attack. In this case, the ProSe function is not able to distinguish the clone nodes by secret keys which might leads to great loses. Among all the existing clone detection methods, the CSI based detection method features low resource consumption which corresponds to the aim of ProSe. However, as afore mentioned, the CSI is never absolute static due to the nature of wireless propagation. In order to make up for this shortcoming, a novel physical layer reputation based clone detection (PRCD) protocol for ProSe is proposed in this article. CSI from multiple packets is appropriately extracted and accumulated to generate reputations. Clone detection is processed periodically by reputation assessment. This physical layer reputation is immune to dishonesty, slander, failing witness and channel noise. In addition, different clone attacks are deliberated about and detected efficiently.

The rest of the paper is organized as follows. Firstly, the network model and adversary model are elaborated in Section II. Secondly, the PRCD protocol is elaborated in Section III. Then, the performance of PRCD protocol and the experiment results are analyzed in Section IV. At last, the conclusion is provided in Section V.

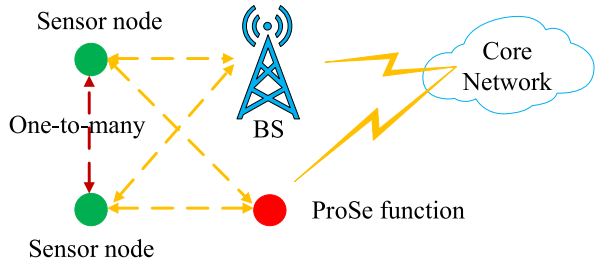


FIGURE 1. Proximity based service (ProSe) network model [20].

II. PRELIMINARIES

A. NETWORK MODEL

The ProSe network model is illustrated in Figure 1. The interfaces between sensor nodes, BS and ProSe function are wireless. The interface between ProSe sensor nodes can be one-to-many or point-to-point. In this article, it is assumed that the sensors in the network are motionless which results in a long coherence time. The ProSe function discovers the nodes in proximity and authorizes them to conduct direct communication. Considering multiple sensor nodes in proximity, the ProSe function divides them into clusters and assigns group IDs to them. One cluster-head node is selected in each cluster, as shown in Figure 2.

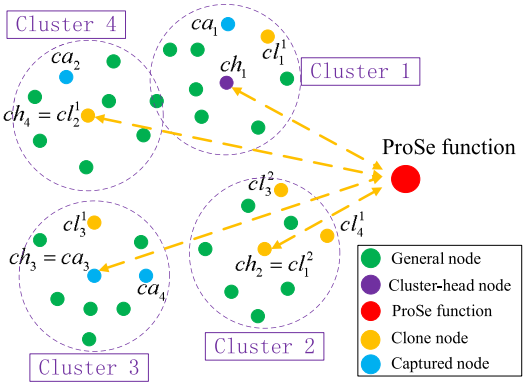


FIGURE 2. Clone attack in four clusters under the coverage of one ProSe function.

In this article, it is assumed that there are N sensor nodes in proximity, $N \in Z$, including four cluster-head nodes, ch_i, $i = 1, 2, 3, 4$. General sensor nodes $\{n_g | g = 1, 2, \cdots, N - 4\}$ are organized into local clusters. Signaling messages are transmitted from general sensor node to ProSe function by the relaying of cluster-head node.

B. ADVERSARY MODEL

We assume that the adversaries are only capable of capturing a limited number of nodes in the network and the ProSe function should always be secure and trusted. The nodes which are controlled by adversaries are referred to as captured nodes. The untouched nodes are regarding as general nodes. Adversaries reproduce replicas of captured nodes, namely clone nodes, and deploy them in the network. Adversaries attempt to conceal the existence of clone nodes and cover for each other. It is assumed that adversaries only allocate clone nodes within the same coverage of the ProSe function as their captured nodes.

TABLE 2. Captured nodes and clone nodes.

Captured Node	Clone Node
ca_1	$cl_1^1, ch_2 = cl_1^2$.
ca_2	cl_2^1.
ca_3	cl_3^1, cl_3^2.
ca_4	cl_4^1.

As shown in Figure 2, there are four clusters and the cluster-head nodes are ch_1, ch_2, ch_3, ch_4. There are six clone nodes duplicated from four captured nodes, as shown in Table 2. Note that the clone node cl_1^2 becomes cluster-head node ch_2.

III. PRCD PROTOCOL

The proposed PRCD protocol is designed to detect all kinds of clone attacks as previously mentioned. It consists of three stages: initiation, data transmission and clone detection. The initiation is conducted at the beginning of the system, followed by data transmission. The clone detection is performed periodically and the data transmission starts up again afterwards. The proposed protocol is illustrated in the flow chart Figure 3 and elaborated as follows.

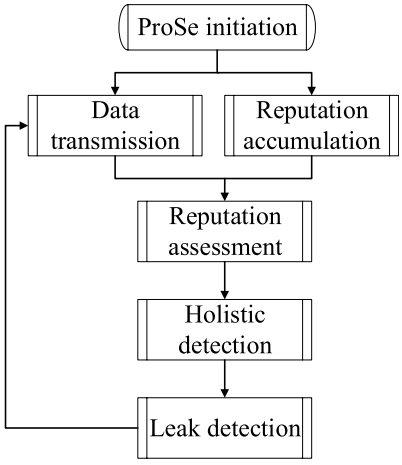


FIGURE 3. The flow chart of the PRCD protocol.

A. INITIATION STAGE

After ProSe direct discovery, the ProSe function nominates the cluster-head nodes. Each node becomes a cluster-head node with an average probability of p, $0 < p < 1$. To avoid hot nodes in the system, cluster-head nodes should be shifted, otherwise they might be memory overflowed or die away quickly. The cluster-head nodes broadcast their status and each general node chooses the nearest cluster-head

node to respond with its ID and pilot. The cluster-head nodes record the IDs and extract the physical layer CSI, i.e., the channel matrixes, as the initial channel matrixes $H_{g,0}$, $g = 1, 2, \cdots, N$.

B. DATA TRANSMISSION STAGE

After the clusters determined, as shown in Figure 2, general sensor nodes n_g begin to transmit messages $m_{g,t}$ to cluster-head nodes ch_i at timeslot t, as shown in formula (1), where $t = 1, 2, \cdots$ and $pilot_{g,t}$ is the pilot which is used to estimate channel information.

$$m_{g,t} \ \left\{ID(n_g), pilot_{g,t}, data_{g,t}\right\}. \tag{1}$$

The cluster-head node ch_i estimates the channel from $pilot_{g,t}$ to get the incoming CSI $H_{g,t}$ at timeslot t, as shown in formula (2)

$$H_{g,t} = \left[H_{g,t}(f_1), \cdots, H_{g,t}(f_r), \cdots, H_{g,t}(f_M)\right], \tag{2}$$

where $f_r = f_0 + ((r/M) - (1/2))\, W$, $r = 1, 2, \cdots, M$ and f_0 is the center frequency. W is the bandwidth. M is the number of the frequency over the bandwidth of the cluster.

There are several ways to get the channel differences, as shown in formula (3) - (7):

$$\Lambda_1(t) = \|H_{g,t} - H_{g,0}\|, \tag{3}$$

$$\Lambda_2(t) = \|H_{g,t} - H_{g,t-1}\|, \tag{4}$$

$$\Lambda_3(t) = \frac{1}{s}\sum_{d=0}^{s-1} \|H_{g,t-d} - H_{g,t-d-1}\|, \tag{5}$$

$$\Lambda_4(t) = \frac{\|H_{g,t} - H_{g,t-1}\|}{\|H_{g,0}\|}, \tag{6}$$

$$\Lambda_5(t) = \frac{\sum_{d=0}^{s-1} \|H_{g,t-d} - H_{g,t-d-1}\|}{s \cdot H_{g,0}}, \tag{7}$$

where $\Lambda_x(t) > 0$, $x = 1, 2, 3, 4, 5$, are the channel differences and $s = 1, 2, \cdots$. $\|A\|$ returns the 2-norm of the matrix A. Therefore, reputations correspond to different channel difference algorithms, as in formula (8):

$$R_i^x\left(n_g, t\right) = \frac{1}{\sum_{j=1}^{t} \Lambda_x(j)}, \tag{8}$$

where $R_i^x\left(n_g, t\right) > 0$. The initial reputation is assumed as $R_i^x\left(n_g, 0\right) = 0$. The closer the channel matrixes are, the bigger the probability of the legality of the transmitting nodes are, and the greater the reputation values are.

It is worth noting that the reputation of the node n_g is generated and recorded by the cluster-head node ch_i and even the node n_g itself does not know its own reputation value. On the other hand, the cluster-head node ch_i gets the reputation from the physical layer channel information which will never lie for any nodes. This ensures the reliability and non-repudiation of the reputation solution. The reputation list is shown in Table 3.

TABLE 3. Reputation list.

Node ID	Node Reputation
$ID\,(n_g)$	$R_i\,(n_g, t)$
⋮	⋮

C. CLONE DETECTION STAGE

After a period of τ, where $\tau > 0$ is less than the channel coherence time, the ProSe function triggers a clone detection procedure by broadcasting the clone detection request. There are three procedures to detect the clone attack:

1) A reputation assessment procedure to detect the clone nodes within the same cluster, as the node ca_1 and cl_1^1 in Figure 2.
2) A holistic detection procedure to detect clone attacks in two scenarios:
   a) The clone nodes and the captured nodes are deployed in different clusters like node ca_1 and cl_1^2.
   b) The cluster-head node is a captured node as the node ca_3 and cl_3^1. The clone node cl_3^2 and cl_4^1 will not be detected in this procedure because the cluster-head node ch_2 will cover for them.
3) A leak detection procedure to detect the rest of the clone nodes which are so smart to pass the former detections like the clone node cl_2^1, cl_3^2 and cl_4^1 in Figure 2.

1) REPUTATION ASSESSMENT PROCEDURE

Most of the clone attacks happen during the data transmission stage because the detection stage is much shorter than the transmission stage. The cluster-head node will forward the messages from the recorded nodes in the clustering procedure and will discard the messages from other nodes. As a result, in the data transmission stage, the adversaries try to deploy the clone nodes within the same cluster to make sure that their messages will not be discarded, as ca_1 and cl_1^1 in Figure 2. In reality, there may be several clone nodes in the same cluster with the captured node ca_1 though only one is taken as an example.

After receiving the clone detection request, the cluster-head nodes ch_i begin to assess reputations in their *Reputation List*. The null hypothesis $\mathcal{H}_0$ represents that the nodes can be trusted. The alternative hypothesis $\mathcal{H}_1$ represents that clone attacks happen:

$$R_i^x\left(n_g, t\right) \lessgtr_{\mathcal{H}_0}^{\mathcal{H}_1} \theta_x. \tag{9}$$

If there are nodes with different channels but the same ID, the reputation value $R_i^x\left(n_g, t\right)$ will be lower than the threshold θ_x. Thus the cluster-head node ch_i will find that clone attacks happened. It will delete these nodes and report the clone node IDs to the ProSe function. The ProSe function lists the clone IDs in the *Clone Node List*, as shown in Table 4.

Take Figure 2 as an example, the cluster-head node ch_1 will report the ID of the node ca_1 and cl_1^1 to the ProSe function and

TABLE 4. Clone node list.

Clone Node ID
$ID\,(ca_1, cl_1^1)$
$\vdots$

delete them in the cluster. Note that, the cluster-head node ch_2 will not report any clone attacks because itself is a clone node cl_1^2 and it will try to protect the clone node cl_3^2 and cl_4^1 by not reporting their IDs. The cluster-head node ch_3 will not report clone attack too, because the captured node ca_3 is the cluster-head ch_3 and it does not find any conflicts of the IDs in its cluster. The cluster-head node ch_4 also will not report clone attack because the cluster-head node ch_4 is a clone node cl_2^1 and it will never report the captured node ca_2 to hide itself and pretend to be the real ca_2. These clone attacks will be detected later.

The clone detection procedure is triggered every period of τ and the detection lasts a period of τ_d where $\tau_d > 0$ and $\tau_d \ll \tau$. Thus, the adversaries deploy the clone nodes in the data transmission stage with a probability of $\frac{\tau-\tau_d}{\tau}$ which is approaching to 1 and $\frac{\tau-\tau_d}{\tau} > 0$. That is, $\lceil l \cdot \frac{\tau-\tau_d}{\tau} \rceil$ clone nodes can be detected in the reputation assessment procedure where $l \in N$ is the number of the clone nodes. However, there are still a few of the clone nodes, $\lceil l \cdot \frac{\tau_d}{\tau} \rceil$ clone nodes, have been leaked away. These clone nodes were deployed at the beginning of a new round and scattered in different clusters. Some clone nodes even have chance to be selected as cluster-head nodes. To detect these clone nodes, we need to perform the holistic detection procedure and the leak detection procedure.

2) HOLISTIC DETECTION PROCEDURE

During the holistic detection procedure, the ProSe function broadcasts a holistic detection request. All the cluster-head nodes report all the node IDs in their clusters including the IDs of themselves as requested. The clone node IDs which have been sent in the reputation assessment procedure will not be sent again. The ProSe function compares the received node IDs with each other and with the IDs in the *Clone Node List*. If there are any IDs appear more than once, the ProSe function will add the IDs into the *Clone Node List*.

In Figure 2, the cluster-head ch_1 will not report the ID of the captured node ca_1 and the clone node cl_1^1 again, because they have been reported in the reputation assessment procedure. The cluster-head node ch_2 and ch_3 will report the IDs of themselves and the IDs of the clone node cl_3^1 and the captured node ca_4 which leads to the result that the ProSe function will receive the ID of the captured node ca_3 more than once and find that the node ch_2 has the same ID as the captured node ca_1 in the *Clone Node List*. The ProSe function will decide that the clone attack has happened and add the ID of the captured node ca_3 to the *Clone Node List*. However, the ProSe function will not find that the node ca_4 has been captured and there is another clone node cl_3^2 because the cluster-head node ch_2 did not report the IDs of the clone node cl_4^1 and cl_3^2. In fact, it does not matter whether the IDs of the clone node cl_3^2 and cl_1^2 are reported, because the captured node ca_3 and ca_1 are already listed in the *Clone Node List*. Note that the clone node cl_2^1 and cl_4^1 are so smart such that they have passed two detection procedures and still remain undetected. A leak detection procedure is needed to arrest these smart malicious nodes.

3) LEAK DETECTION PROCEDURE

The ProSe function broadcasts a leak detection request and re-cluster all the nodes. One cluster-head node is not allowed to serve for another term consecutively. The general nodes are required to report their IDs to the new cluster-head nodes. Then the new cluster-head nodes report the node IDs in their clusters to the ProSe function and the ProSe function checks the received IDs to find if there are any clone nodes left.

After the re-clustering, the new scenario is shown in Figure 4. There are five clusters within the sink ring which is one more cluster than in Figure 2, because the probability to be a cluster-head node for each node will change according to the energy remained. The average probability is p and the average number of clusters is $\lceil Np \rceil$, N is the number of the sensor nodes in proximity. As a result, the actual number of the clusters may be more or less than $\lceil Np \rceil$ sometimes.

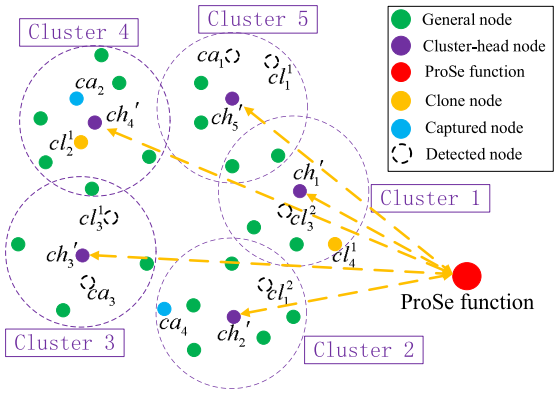


FIGURE 4. Clone attack scenario after re-clustering.

In Figure 4, the new cluster-head node ch_1', ch_2', ch_3', ch_4' and ch_5' send the node IDs in their clusters to the ProSe function. The ProSe function compares the received IDs with each other and with the IDs in the *Clone Node List*. Obviously, the new cluster-head node ch_1', ch_2' and ch_4' will no longer cover for the clone nodes. As a result, the clone node cl_2^1, cl_4^1 and their captured node ca_2 and ca_4 will be uncovered. Besides, the ProSe function will receive the ID of the clone node cl_3^2. There is no need for the ProSe function to add the ID of the clone node to the *Clone Node List*, because this ID is already in the list. The ProSe function will add the newly detected clone IDs of ca_2, cl_2^1, ca_4 and cl_4^1 to the *Clone Node List*.

TABLE 5. Detection summary.

Procedure	Detected Nodes
Reputation assessment	ca_1 and cl_1^1.
Holistic detection	ca_1 and cl_1^2, ca_3 and cl_3^1.
Leak detection	ca_2 and cl_2^1, ca_3 and cl_3^2, ca_4 and cl_4^1

The detection of all the clone nodes is summarized in Table 5. After the leak detection, the ProSe function broadcasts the *Clone Node List* and the cluster-head nodes delete the nodes with these IDs. The clone detection stage has been completed and a new data transmission stage begins.

IV. PERFORMANCE ANALYSIS AND NUMERICAL RESULTS

A. PERFORMANCE ANALYSIS

In what follows, we focus on the detection probability which depends on the node reputation $R_i\left(n_g, t\right)$ based on the difference of the channel $\Lambda_x(t)$. Our purpose of replacing the channel difference with the node reputation is to enlarge the differences between legal and illegal messages. We accumulate the channel differences and perform the clone detection procedure every τ time period. To simplify the proof, we take $\Lambda_1(t)$ as an example. The formula (8) can be written as:

$$R_i^1\left(n_g, t\right) = \frac{1}{\sum_{j=1}^{t} \|H_{g,j} - H_{g,0}\|}. \tag{10}$$

Theorem 1: If there exists clone nodes in the same cluster, the clone attack can be detected with a probability of $P_d = 1 - \varepsilon, \forall \varepsilon > 0$.

Proof: We take cluster 1 in Figure 2 as an example. The clone node cl_1^1 and the captured node ca_1 both send messages to the cluster-head node ch_1. In AWGN channel (Additive White Gaussian Noise channel), the channel difference between the captured node ca_1 and the cluster-head node ch_1 at timeslot t is denoted by $\Lambda_1^{ca}(t)$. The channel difference between the clone node cl_1^1 and the cluster-head node ch_1 at timeslot t is denoted by $\Lambda_1^{cl}(t)$. To simplify, we assume $\Lambda_1^{ca}(t)$ and $\Lambda_1^{cl}(t)$ follow the normal distribution independently, i.e., $\Lambda_1^{ca}(t) \sim N\left(\mu_1^{ca}, \sigma_{ca}^2\right)$, $\Lambda_1^{cl}(t) \sim N\left(\mu_1^{cl}, \sigma_{cl}^2\right)$, where μ_1^{ca} and μ_1^{cl}, σ_{ca}^2 and σ_{cl}^2 denote the expectations and variances of $\Lambda_1^{ca}(t)$ and $\Lambda_1^{cl}(t)$, respectively and $\mu_1^{ca} > \mu_1^{cl} > 0$. The cluster-head node has an equal chance to receive the messages from the captured node and the clone node. Thus we assume the cluster-head node has received T messages from the captured node and T messages from the clone node, where $T > 1$. We divide the reputation into two parts, R_{ca} and R_{cl}, where R_{ca} is accumulated from $\Lambda_1^{ca}(t)$ and R_{cl} is accumulated from $\Lambda_1^{cl}(t)$. Considering the property of the normal distribution, we obtain that $\frac{1}{R_{ca}}$ and $\frac{1}{R_{cl}}$ obey the normal distribution too, as shown in formula (11) and (12):

$$\frac{1}{R_{ca}} = \sum_{j=1}^{T} \Lambda_1^{ca}(j) \sim N\left(T\mu_1^{ca}, T\sigma_{ca}^2\right), \tag{11}$$

$$\frac{1}{R_{cl}} = \sum_{j=1}^{T} \Lambda_1^{cl}(j) \sim N\left(T\mu_1^{cl}, T\sigma_{cl}^2\right). \tag{12}$$

According to the property of the normal distribution, the difference between the reputations of the captured node and the clone node obey the normal distribution too, as in formula (13):

$$\frac{1}{R_{ca}} - \frac{1}{R_{cl}} \sim N\left(T\left(\mu_1^{ca} - \mu_1^{cl}\right), T^2\left(\sigma_{ca}^2 + \sigma_{cl}^2\right)\right). \tag{13}$$

The PRCD protocol is based on the difference between the reputations of the captured node and the clone node. If we are able to prove that the difference can be detected, we will be able to prove theorem 1. Therefore, we focus on the proof of $P(|R_{ca} - R_{cl}| > 0) = 1 - \varepsilon, \forall \varepsilon > 0$. Firstly, we transform the left part of the equation as the following:

$$\begin{aligned} &P\left(|R_{ca} - R_{cl}| > 0\right) \\ &= P\left(R_{ca} - R_{cl} \neq 0\right) \\ &= P\left(\frac{1}{R_{ca}} - \frac{1}{R_{cl}} \neq 0\right) \\ &= P\left(\sum_{j=1}^{T} \Lambda_1^{ca}(j) - \sum_{j=1}^{T} \Lambda_1^{cl}(j) \neq 0\right) \\ &= 1 - P\left(\sum_{j=1}^{T} \Lambda_1^{ca}(j) - \sum_{j=1}^{T} \Lambda_1^{cl}(j) = 0\right). \end{aligned} \tag{14}$$

From formula (11)-(13) and the property of the normal distribution, we can get the formula (15):

$$P\left(\sum_{j=1}^{T} \Lambda_1^{ca}(j) - \sum_{j=1}^{T} \Lambda_1^{cl}(j) = 0\right) = \varepsilon, \quad \forall \varepsilon > 0. \tag{15}$$

Bring formula (15) into formula (14), we get formula (16):

$$P\left(|R_{ca} - R_{cl}| > 0\right) = 1 - \varepsilon, \quad \forall \varepsilon > 0. \tag{16}$$

From formula (16), we can get that the probability of the existence of the difference between R_{ca} and R_{cl} is arbitrarily close to 1. Thus, the detection probability P_d is arbitrarily close to 1. That is $P_d = 1 - \varepsilon, \forall \varepsilon > 0$. Q.E.D. □

In fact, the boundary between the reputation accumulated from $\Lambda_1^{ca}(t)$ and the reputation accumulated from $\Lambda_1^{cl}(t)$, i.e., R_{ca} and R_{cl}, is clear even under low SNR scenarios.

Theorem 2: After receiving t messages, d_{PR} is the difference between the legitimate messages and the illegitimate messages in PRCD protocol and d_{CF} is the difference in CFCD protocol, where $E(d_{PR})/E(d_{CF}) = t$.

Proof: The difference between the legitimate messages and the illegitimate messages in CFCD protocol is

$d_{CF} = \Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)$. In PRCD protocol, the difference is $d_{PR} = \sum_{j=1}^{t} \Lambda_1^{ca}(j) - \sum_{j=1}^{t} \Lambda_1^{cl}(j)$.

As we discussed in theorem 1, $\Lambda_1^{ca}(t) \sim N\left(\mu_1^{ca}, \sigma_{ca}^2\right)$ and $\Lambda_1^{cl}(t) \sim N\left(\mu_1^{cl}, \sigma_{cl}^2\right)$. Thus,

$$\begin{aligned}
&\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t) \sim N\left(\mu_1^{ca} - \mu_1^{cl}, \sigma_{ca}^2 + \sigma_{cl}^2\right),\\
&\sum_{j=1}^{t}\Lambda_1^{ca}(j) - \sum_{j=1}^{t}\Lambda_1^{cl}(j) \sim N\left(t\left(\mu_1^{ca} - \mu_1^{cl}\right), t^2\left(\sigma_{ca}^2 + \sigma_{cl}^2\right)\right),\\
&E\left(\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)\right) = \mu_1^{ca} - \mu_1^{cl},\\
&E\left(\sum_{j=1}^{t}\Lambda_1^{ca}(j) - \sum_{j=1}^{t}\Lambda_1^{cl}(j)\right) = t\left(\mu_1^{ca} - \mu_1^{cl}\right),\\
&E\left(\sum_{j=1}^{t}\Lambda_1^{ca}(j) - \sum_{j=1}^{t}\Lambda_1^{cl}(j)\right) - E\left(\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)\right)\\
&\quad = t\left(\mu_1^{ca} - \mu_1^{cl}\right) - \left(\mu_1^{ca} - \mu_1^{cl}\right)\\
&\quad = (t-1)\left(\mu_1^{ca} - \mu_1^{cl}\right)\\
&\quad = (t-1)\,E\left(\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)\right).
\end{aligned} \tag{17}$$

From formula (17), we get

$$E\left(\sum_{j=1}^{t}\Lambda_1^{ca}(j) - \sum_{j=1}^{t}\Lambda_1^{cl}(j)\right) = tE\left(\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)\right).$$

Owing to $\mu_1^{ca} > \mu_1^{cl} > 0$ as we assumed before, we get formula (18):

$$\frac{E(d_{PR})}{E(d_{CF})} = \frac{E\left(\sum_{j=1}^{t}\Lambda_1^{ca}(j) - \sum_{j=1}^{t}\Lambda_1^{cl}(j)\right)}{E\left(\Lambda_1^{ca}(t) - \Lambda_1^{cl}(t)\right)} = t. \tag{18}$$

Thus, the difference in PRCD protocol between the legitimate messages and the illegitimate messages is as t times as the difference from one message. Q.E.D. □

Theorem 3: The memory cost of the PRCD protocol is $O(\lceil 1/p \rceil)$.

Proof: As we mentioned in the former section, there are N nodes in proximity and each node becomes a cluster-head node with an average probability of p. As a result, there will be $\lceil Np \rceil$ cluster-head nodes on average. The average number of nodes in the cluster will be $\lceil \frac{N}{Np} \rceil = \lceil \frac{1}{p} \rceil$. Each sensor nodes are prepared to be a cluster-head node, thus they should have the same memory capacity as the cluster-head node. Each cluster-head node needs to record the reputations of the nodes within the cluster. That is to say, the cluster-head node need to record $\lceil 1/p \rceil$ reputations. Thus, the memory cost is $O(\lceil 1/p \rceil)$. Q.E.D. □

Table 6 shows the performance comparison between the LSM [7], LSCD [9], CSI based detection [18] and the proposed PRCD protocol. N_w is the number of witness nodes.

TABLE 6. Performance comparison of different protocols.

	Lifetime	Computing Complexity	Memory Cost	Detection Probability
LSM	Short	Medium	$O\left(N_w\sqrt{N}\right)$	$(1-0.235)^{N_w}$
LSCD	Medium	High	N_w	≈ 1
CSI based	Long	Low	$O\left(\sqrt{N}\right)$	$O(SNR)$
PRCD	Long	Low	$O(\lceil 1/p \rceil)$	≈ 1

It is revealed in from the table that the CSI based protocol gathers the merits of LSM and LSCD. It features a long lifetime, low computing complexity and a reasonable memory cost while the detecting probability is negatively affected by the SNR. It is also revealed that the PRCD protocol is approaching to ideal. Its detection probability is approximate to ideal and the memory cost does not scale with the network. In addition, its lifetime is long and computing complexity is low.

Theorem 4: In the clone detection stage of PRCD protocol, the communication cost is $O(\lceil Np \rceil)$.

Proof: The communication cost is the number of packets which are sent in the network. In the reputation assessment procedure, cluster-head nodes need to send a clone list to the ProSe function and there are $\lceil Np \rceil$ cluster-head nodes. Thus, there are $\lceil Np \rceil$ packets which are sent from different cluster-head nodes. In the holistic detection procedure, the cluster-head nodes are required to send the ID lists to the ProSe function. Thus there are another $\lceil Np \rceil$ packets to be sent. In the leak detection procedure, the new cluster-head nodes are required to send the ID lists again. Thus, another $\lceil Np \rceil$ packets need to be sent from different cluster-head nodes. To sum up, there are $3 \cdot \lceil Np \rceil$ messages need to be sent in the clone detection stage. Thus the communication cost is $O(\lceil Np \rceil)$. Q.E.D. □

B. NUMERICAL RESULTS

To investigate the performance of the proposed PRCD protocol, we firstly simulated it in MATLAB under different SNRs. Parameters are listed in Table 7.

TABLE 7. Simulation parameters in MATLAB.

System Parameter	Value
Bandwidth W	1MHz
Center of frequency f_0	2GHz
Maximum Doppler frequency shift	10Hz
Number of rounds	1000
Number of messages in each round	100
Number of multi-paths	5

In section III-B, five reputation algorithms were introduced. Figure 5 presents the reputations acquired by different algorithms according to formula (8). From Figure 5, when $x = 1$, the deviation of the reputation value is comparatively smaller. Therefore, $\Lambda_1(t)$, in formula 3 is adopted to acquire

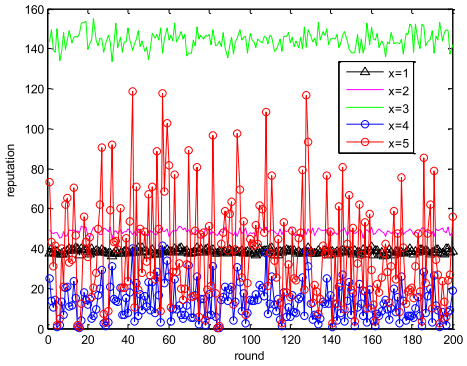


FIGURE 5. Reputations R^x acquired by five different algorithms, $x = 1, 2, 3, 4, 5$.

the channel difference and correspondingly, R^1 is adopted as the reputation algorithm in the rest of the article.

Detection rate and false alarm rate are two critical measurements. Detection rate indicates the probability of attack detection and false alarm rate indicates the probability of a legitimate message classified as illegitimate. Figure 6 depicts the pane of false alarm rate and detection rate for different thresholds. The proposed protocol is compared with the existing CSI based detection protocol, as in [18], under different SNRs. The corner (0, 1) is corresponding to an optimal threshold when the false alarm rate is 0 and detection rate is 1. It is clearly illustrated in Figure 6 that under each SNR from $0dB$ to $20dB$, there always exists an optimal threshold to reach the ideal performance of (0, 1). However, for the CSI based protocol in [18], the performance degrades significantly with the decrease of SNR.

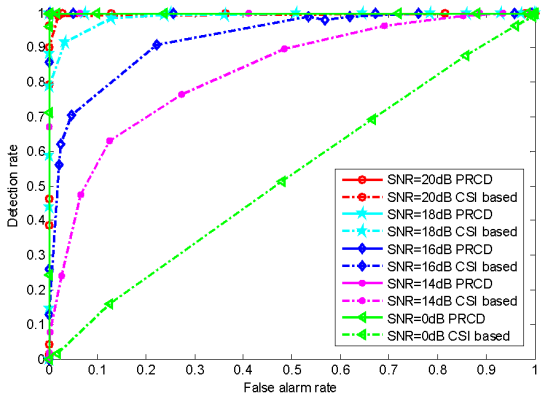


FIGURE 6. Comparison of the performances of PRCD protocol and the CSI based protocol in [18] under different SNRs.

Figure 7 shows the processes of the reputation accumulations. Even when the clone nodes send malicious messages with a probability of 0.1, there is an obvious difference between the reputation without clone attacks and the reputation under clone attacks.

The simulation has showed the advantages of the proposed protocol. The feasibility in reality is further investigated

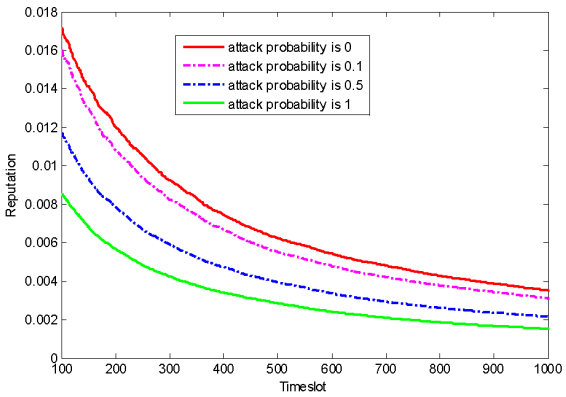


FIGURE 7. Accumulation process of reputation under different attack probabilities.

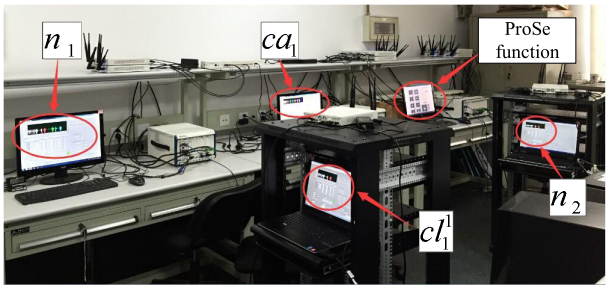


FIGURE 8. Simulation platform with USRPs.

with USRP (Universal Software Radio Peripheral) platform. As shown in Figure 8, five groups of USRPs are involved. The ProSe function and the captured node ca_1 possess 8 antennas. The sensor node n_1 is equipped with 4 antennas. The sensor node n_2 and the clone node cl_1^1 possess 2 antennas. The communication solution is based on MIMO-OFDM (Multiple Input and Multiple Output-Orthogonal Frequency Division Multiplexing) and ILS (Improved-scaled Least Squares) is adopted to estimate channels. Some parameters are shown in Table. 8.

TABLE 8. Simulation parameters in USRPs.

System Parameter	Value
Center of frequency f_0	3.5GHz
Bandwidth W	2MHz
Transmitting power	15dBm
Transmission gain	20dB
Digital modulation method	4QAM
Number of the subcarriers	128

In this article, we focus on improving the detection performance in worst case where the CSI from general nodes and clone nodes are mixed. Therefore, we choose some channel difference samples from the results, including 2000 channel differences Λ_1^{cl} from the clone node cl_1^1, 1000 Λ_1^{ca} from ca_1, 500 Λ_1^{ca} from n_1 and 500 Λ_1^{ca} from n_2. The channel differences are normalized and presented in Figure 9. We can

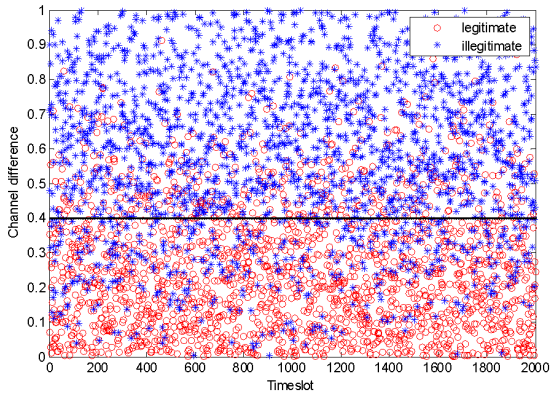


FIGURE 9. The channel differences from the legitimate messages and illegitimate messages. The threshold is 0.4.

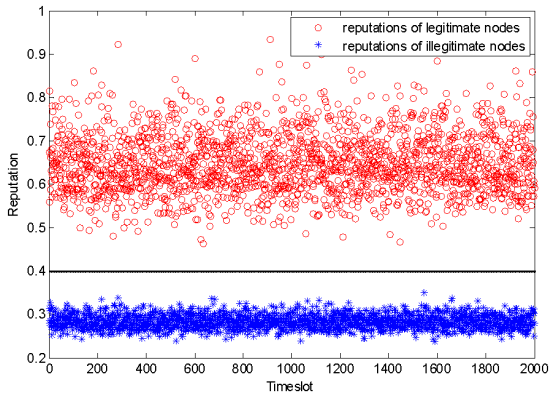


FIGURE 10. Normalized reputations of legitimate and illegitimate nodes.The threshold is 0.4.

see in Figure 9 that it is hard to find a boundary of the channel differences from legitimate and illegitimate messages. The threshold is around 0.4, but some channel differences of the legitimate messages are higher than the threshold owing to the channel randomness. This will leads to the false alarm.

After accumulating the channel differences, the boundary between the reputation accumulated from $\Lambda_1^{ca}(t)$ and the reputation accumulated from $\Lambda_1^{cl}(t)$, i.e. the boundary of R_{ca} and R_{cl}, is obvious as shown in Figure 10. We simulate the reputation accumulation procedure for 2000 rounds and 50 timeslots in each round. According to Figure 9 and Figure 10, it is much easier to acquire an optimal threshold to reach ideal performance in the proposed PRCD protocol than in the CSI based protocol in [18].

V. CONCLUSION

In this article, several clone detection protocols are compared. Among all the demerits, CSI based detection protocol seems promising for ProSe due to low computing complexity and honesty. However, CSI based detection protocol is sensitive to channel quality. Therefore, a novel physical layer reputation based clone detection (PRCD) protocol for ProSe is proposed to mitigate this demerit. The clone detection procedure in the proposed PRCD protocol composes of three steps: reputation assessment, holistic detection and leak detection. The clone nodes in the same cluster as their captured node, which is the most common scenario, are detected by reputation assessment. Other smart clone attacks are detected by holistic detection and leak detection. By this way, proximity-based services can provide stronger security services. The simulation and realization has proved the feasibility and advantages of the proposed protocol. The results show that the proposed protocol reaches ideal performance under different channel qualities and is resistant to slandering. Nonetheless, in this article, the sensor nodes are motionless and channels are quasi-static. It is necessary to further investigate the performance of the protocol in mobile scenarios in future.

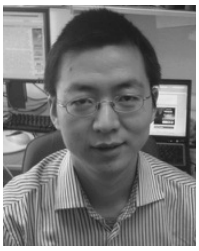
REFERENCES

[1] S. Buchegger and J.-Y. Le Boudec, "Self-policing mobile ad hoc networks by reputation systems," *IEEE Commun. Mag.*, vol. 43, no. 7, pp. 101–107, Jul. 2005.
[2] S. Bansal and M. Baker, "Observation-based cooperation enforcement in ad hoc networks," *CoRR*, vol. cs.NI/0307012, pp. 1–10, Jul. 2003. [Online]. Available: http://arxiv.org/abs/cs.NI/0307012
[3] P. Michiardi and R. Molva, "Core: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks," *Adv. Commun. Multimedia Secur.*, vol. 100, pp. 107–121, 2002, doi: 10.1007/978-0-387-35612-9_23.
[4] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Trans. Sensor Netw.*, vol. 4, no. 3, pp. 1501–1537, May 2008.
[5] D.-Q. Xiao, J.-Z. Feng, B. Yang, and H. Zhang, "Reputation formal model for wireless sensor network," *Comput. Sci.*, vol. 34, no. 6, pp. 84–87, 2007.
[6] G. Yang, S. Ying, and W. Yang, "Reputation model based on behavior of sensor nodes in WSN," *J. Commun.*, vol. 30, no. 12, pp. 18–26, 2009.
[7] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in *Proc. IEEE Symp. Secur. Privacy*, May 2005, pp. 49–63.
[8] Y. Zeng, J. Cao, S. Zhang, S. Guo, and L. Xie, "Random-walk based approach to detect clone attacks in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 5, pp. 677–691, Jun. 2010.
[9] M. Dong, K. Ota, L. T. Yang, A. Liu, and M. Guo, "LSCD: A low-storage clone detection protocol for cyber-physical systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 5, pp. 712–723, May 2016.
[10] Z. Zheng, A. Liu, L. X. Cai, Z. Chen, and X. Shen, "Energy and memory efficient clone detection in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1130–1143, May 2016.
[11] F. Pan, H. Wen, H. Song, T. Jie, and L. Wang, "5G security architecture and light weight security authentication," in *Proc. IEEE/CIC Int. Conf. Commun. China-Workshops (CIC/ICCC)*, Nov. 2015, pp. 94–98.
[12] L. Hu *et al.*, "Cooperative jamming for physical layer security enhancement in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 219–228, Feb. 2018.
[13] F. Xie *et al.*, "Optimized coherent integration-based radio frequency fingerprinting in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3967–3977, Oct. 2018.
[14] W. C. Jakes, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1974, pp. 635–642.
[15] X. He, H. Dai, W. Shen, P. Ning, and R. Dutta, "Toward proper guard zones for link signature," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2104–2117, Mar. 2016.
[16] S. Fang, Y. Liu, and P. Ning, "Mimicry attacks against wireless link signature and new defense using time-synched link signature," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1515–1527, Jul. 2016.
[17] H. Wen, J. Luo, and L. Zhou, "Lightweight and effective detection scheme for node clone attack in wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 1, no. 3, pp. 137–143, Sep. 2011.
[18] T. Ma, "Research on lightweight physical layer assist authentication technique in smart grid," M.S. thesis, Nat. Key Lab. Sci. Technol. Commun., Univ. Electron. Sci. Technol. China, Chengdu, China, Jul. 2015.

[19] C. Wang, Y. Lin, and Z. Zhang, "Research on physical layer security of cognitive radio network based on RF-DNA," in *Proc. IEEE Int. Conf. Softw. Qual., Rel. Secur. Companion (QRS-C)*, Jul. 2017, pp. 252–255.
[20] *Proximity-Based Services (ProSe); Stage 2*, document 3GPP ETSI TS 123 303 v15.1.0, Jul. 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123300_123399/123303/15.01.00_60/ts_123303v150100p.pdf
[21] *Proximity-Based Services (ProSe); Security Aspects*, document 3GPP ETSI TS 33.303 v12.1.0, Sep. 2014. [Online]. Available: http://www.doc88.com/p-9979705678314.html

FEI PAN (S'18) received the bachelor's degree in communication engineering from Northwest University, Xi'an, China, in 2011. She is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include wireless communications, physical layer security, and high performance wireless communications in industrial automation.

ZHIBO PANG (M'13–SM'15) received the B.Eng. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 2002, the M.B.A. degree in innovation and growth from the University of Turku, Turku, Finland, in 2012, and the Ph.D. degree in electronic and computer systems from the KTH-Royal Institute of Technology, Stockholm, Sweden, in 2013. He was a Co-Founder and a CTO of startups such as Ambigua Medito AB. He is currently a Principal Scientist in wireless communications with ABB Corporate Research, Västerås, Sweden, leading projects in digitalization solutions for smart buildings and homes, robotics and factories, healthcare and logistics, power electronics, and power systems. He is also an Affiliated Faculty with the KTH-Royal Institute of Technology. He was a recipient of the 2016 Inventor of the Year Award from ABB Corporate Research, Sweden. He is a Co-Chair of the Technical Committee on Industrial Informatics. He is an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, and a Guest Editor of PROCEEDINGS OF THE IEEE, the IEEE INTERNET OF THINGS JOURNAL, and the IEEE REVIEWS IN BIOMEDICAL ENGINEERING.



MING XIAO (S'02–M'07–SM'12) received the B.S. and M.S. degrees in engineering from the University of Electronic Science and Technology of China, Chengdu, in 1997 and 2002, respectively, and the Ph.D. degree from the Chalmers University of Technology, Göteborg, Sweden, in 2007. He is currently an Associate Professor in communications theory with the School of Electrical Engineering and Computer Science, KTH-Royal Institute of Technology, Stockholm, Sweden. Since 2012, he has been an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, a Senior Editor of the IEEE COMMUNICATIONS LETTERS, since 2015, and the IEEE WIRELESS COMMUNICATIONS LETTERS, from 2012 to 2016, and the Lead Editor of the IEEE JOURNAL ON SELECTED AREA IN COMMUNICATIONS and for the special issue on Millimeter Wave Communications for Future Mobile Networks.

HONG WEN (M'08–SM'13) received the M.Sc. degree from Sichuan University, Sichuan, Chengdu, in 1997, and the Ph.D. degree from Southwest Jiaotong University and the University of Waterloo, Waterloo, ON, Canada, in 2004 and 2018, respectively. From 2008 to 2009, she was a Visiting Scholar and a Post-Doctoral Fellow with the Electrical and Computer Engineering Department, University of Waterloo. She is currently a Professor with University of Electronic Science and Technology of China. Her current research interest includes communication systems and security.

RUN-FA LIAO was born in Chongqing, China. He is currently pursuing the Ph.D. degree in communication and information system with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, China. His current main interest includes wireless communication system security combined with intelligent algorithms.