IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2019, accepted August 27, 2019, date of publication September 2, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938771

Ensuring Honest Data Collection Against Collusive CSDF Attack With Binary-Minmaxs Clustering Analysis in Mobile Crowd Sensing

JINGYU FENG, TAO LI, YUJIA ZHAI, SHAOQING LV, AND FENG ZHAO
Shaanxi Key Laboratory of Information Communication Network and Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

Corresponding author: Feng Zhao (peakzhao@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1836210 and Grant 61572460, in part by the National Science Foundation of Shaanxi Province under Grant 2019JM-442, and in part by the Science Foundation of Shaanxi Provincial Education Office under Grant 17JK0703.

ABSTRACT Mobile crowd sensing (MCS) is considered as a powerful paradigm which takes advantage of the pervasive sensor-embedded smartphones to collect data. However, MCS assumes all workers always are trusted, and thus offering opportunities for malicious workers to conduct the crowd sensing data falsification (CSDF) attack. To suppress such threat, recent efforts have been made to trust mechanism. Currently, some malicious workers can collude with each other to form a collusive clique, and thus not only increasing the power of CSDF attack but also avoiding the detection of trust mechanism. To ensure honest data collection in MCS, we must fight against such collusive CSDF attack. Noting that the duality of sensing data, we propose a defense scheme called BMCA from the design idea of binary-minmaxs clustering analysis to suppress collusive CSDF attack. In the BMCA scheme, the logic AND operation corresponding to the type of ''1'' and ''0'' historical sensing data is used to measure the similarity between any two workers. Based on this, we find the feature that collusive CSDF attackers usually hold high trust value and a low variance in their similarity vector. To detect collusive CSDF attackers, the min and max variance analysis is introduced to design a new binary-minmaxs clustering algorithm. Moreover, the BMCA scheme can perfect trust evaluation to prevent the trust value growth of collusive CSDF attackers. Simulation results show that the BMCA scheme can enhance the accuracy of trust evaluation, and thus successfully reducing the power of collusive CSDF attack against data collection in MCS.

INDEX TERMS Mobile crowd sensing, trust mechanism, clustering analysis, collusive attack.

I. INTRODUCTION

With the rapid development of mobile communication and wireless sensing, a large number of sensors are integrated into smart mobile terminals, which motivates mobile crowd sensing (MCS) as an emerging paradigm to collaboratively collect sensing data and extract knowledge in smart cities [1], [2]. MCS leverages the inherent mobility of mobile users (called workers), the sensors embedded in mobile phones and the existing communication infrastructures (Wi-Fi, 4G/5G networks) to collect and transfer urban sensing data [3]. Nowadays, MCS has been widely used in the intelligent traffic detection [4], environment detection [5], smart healthcare [6], social network [7], etc. These applications open the door for new innovative research and will significantly revolutionize our daily lives [8].

Although with the various MCS-enabled innovative applications, the new sensing paradigm also encounters new challenges as ''humans'' act as sensors [9]. Actually, the success of MCS requires the participation from a large number of workers [10]. Due to the nature of openness and mobility, MCS gives almost all mobile users the chance to participate in MCS activities as workers. However, the workers in MCS maybe unreliable and vary in terms of ability, honesty, depend-ability, loyalty and so on [11]. It may offer opportunities for malicious workers to corrupt the data collection by launching CSDF attack. Such CSDF attack pattern can be launched by two ways: individual or collusive. Compared

The associate editor coordinating the review of this manuscript and approving it for publication was Miguel López-BeníTez.

with collusive attack, individual attack is less harmful and can be suppressed. In the collusive pattern, the malicious workers who collude with each other to form a collusive clique can increase the power of CSDF attack and fake the sensing data intentionally. If there are the adequate malicious workers, a collusive clique can lead to a wrong decision in the data aggregation.

Accordingly, trust mechanism is recognized as an important part of the MCS platform in which the trust value of mobile workers is evaluated and used to suppress CSDF attack. With trust mechanism, trusted workers can not only perform honest behaviors in MCS, but also fulfill the requirements of a certain task with high quality. The malicious workers with low trust value will be rejected to in the MCS task assignment. Currently, various trust mechanism studies have been presented [12]–[15]. They evaluate whether a worker is trusted or not by his historical sensing behaviors and give the low weights to less trusted workers or even delete their sensing data when making a final decision in the data aggregation. Nevertheless, collusive CSDF attackers can improve their trust value with the help of each other, except for increasing the attack power. Therefore, they may avoid the detection of trust mechanism to corrupt the data collection in MCS.

In this paper, we analyze the characteristics of collusive CSDF attack and propose a defense scheme called BMCA from the design idea of binary-minmaxs clustering analysis to suppress such attack. The main contributions of this paper are as following:

- Abstract the MCS behaviors of each worker as a binary variable "1" or "0", in which "1" denotes the false behavior of a worker in a task by comparing their sensing data with the final result of the data aggregation. Noting that collusive CSDF attackers often launch "1" behaviors together in a certain task time and behave high similarity among themselves, the logic AND operation is introduced to measure the similarity, which can result in less complex and lightweight in mathematical computation.
- Introduce the min and max variance analysis to design a new binary-minmaxs clustering algorithm. Noting that collusive attackers often launch "1" behaviors together again, each of them may get a low variance in their similarity vector. While trusted workers often report honest sensing data individually, each of them may get a high variance in their similarity vector. Even though both collusive CSDF attackers and trusted workers have high trust value, this new clustering algorithm can differentiate collusive CSDF attackers and trusted workers effectively.
- Enhance the accuracy of trust evaluation. The special punishment factor to sudden false MCS behaviors is introduced to dynamically updating the trust value of workers at each MCS action. By sharply reducing the trust value of collusive CSDF attackers, the special punishment factor can prevent the trust value growth of collusive CSDF attackers. As a result, they can be detected by trust mechanism.

The organization of this paper is as follows: In section II, preliminaries related on MCS and trust mechanism are described. We analyze collusive CSDF attack and construct the BMCA scheme to suppress it in section III. Simulation analysis of the BMCA scheme is performed in section IV. Finally, we conclude this paper in section V.

II. PRELIMINARIES

A. MOBILE CROWD SENSING

Mobile crowd sensing (also called participatory sensing) is an emerging paradigm of IoT [16] in which citizens everywhere voluntarily use their computational devices to capture and share sensing data from their surrounding environments in order to monitor and analyze some phenomenon (e.g., weather, road traffic, pollution, etc.) [17]. As shown in Fig.1, there are three main parties in an MCS platform, namely MCS service provider (SP), end user (EU), and MCS worker [11].

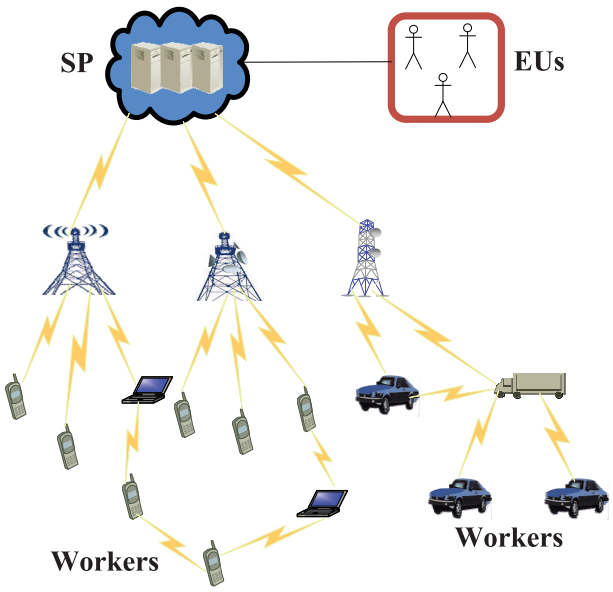


FIGURE 1. System architecture of MCS.

In the MCS platform, the SP could be an organizer that assigns a task for crowd sensing. An EU could be an individual or organization that lacks an ability to perform a certain data collection task. Workers are the mobile users who participate in MCS and perform the assigned tasks.

The life-cycle of MCS can be divided into four stages [3]: task creation, task assignment, data collection and data aggregation, which are briefly described as follows:

- **Task creation**: When EUs request services, the SP creates a sensing task to be given to workers with the corresponding mobile applications.
- **Task assignment**: The SP selects workers and assigns them with the specific sensing task.

- **Data collection**: Once the workers have received the assigned task, they can complete it and send their sensing data to the SP.
- **Data aggregation**: The SP aggregates the received sensing data to determine the final decision about the current task, and sends it to EUs.

B. TRUST MECHANISM

Trust mechanism is increasing influence on many application scenarios, including e-commerce [18], P2P file-sharing [19], cooperative spectrum sensing [20], online social communities [21], etc.

Trust mechanism also plays significant roles in MCS, such as 1) select trusted workers in the task assignment, 2) filter out false sensing data reported by attackers in the data collection, 3) prevent malicious workers from participating into MCS, and 4) assist the SP's rapid decision-making reliably in the data aggregation.

Representative trust mechanism schemes in MCS are as follows. To evaluate the trust value of workers, the reference [12] comprehensively considered a number of properties that affect the honest sensing data, such as link reliability, service quality, and region heat. In [13], a ranking-based is proposed scheme that introduces trust and worker ability into the evaluation of worker trust value. In [14], a dynamic-trust-based recruitment framework (DTRF) is proposed for MCS, in which real-time direct trust and lightweight feedback aggregation trust are combined to select the well-suited workers. In [15], a dynamical credibility assessment of privacy-preserving [22] strategy is designed, in which the sensing data are dynamically split into slices and the number of slices is based on the trust of encountered workers. Specially, worker trust is assessed in two dimensions including the quality of contribution trust and social trust, which indicates how likely a worker can fulfill its functionality and how trustworthy the relationship between a worker and other workers will be, respectively.

In trust mechanism, one of the most popular design is based on beta function. It counts the number of false and honest behaviors a worker has conducted in the data collection, and then evaluates the trust value with beta probability density function denoted by Beta(α, β) [23].

$$Beta(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{1}$$

where θ is the probability of MCS behaviors, $0 \leq \theta \leq 1$, $\alpha > 0$, $\beta > 0$.

Noting that the sensing data from workers can be abstracted as a binary variable (''1'' or ''0''), it is easy for workers to behave two types of sensing behaviors: false or honest. Based on the binary variable, a basic trust mechanism scheme called Baseline can be described, in which the trust value of each worker can be evaluated by two indexes: the number of false sensing (fal) and the number of honest sensing (hon).

Take the i-th worker (W_i) as an example, fal_i and hon_i denote the number of false sensing data and honest sensing data reported by W_i. In the Baseline, the trust value of W_i can be evaluated as

$$T_i = Beta(fal_i + 1, hon_i + 1) \tag{2}$$

Consider the condition $\Gamma(x) = (x-1)!$ when x is an integer [24]. Thus, the expectation value of the beta function is $E[Beta(\alpha, \beta)] = \alpha/(\alpha+\beta)$. In this case, T_i can be further evaluated as

$$T_i = \frac{1 + hon_i}{2 + fal_i + hon_i} \tag{3}$$

Let δ denotes the threshold of trust value. For $T_i \geq \delta$, W_i will be identified as an attacker, and vice verse. In order to guarantee the performance of MCS, δ should satisfy two requirements: 1) δ should be a rational value between 0 and 1 as $T_i \in [0, 1]$, 2) the value of δ can be adjusted to suppress malicious responses generated by malicious workers who report false sensing data.

However, how to suppress collusive CSDF attack has not been considered in current trust mechanism for MCS. In this paper, we argue that securing MCS with only trust mechanism is not enough. Malicious workers can attain high trust value with the help of each other. Specially, the malicious workers with high trust value can launch CSDF attack together in a collusive manner to corrupt the data collection.

III. COLLUSIVE CSDF ATTACK OVERVIEW

Due to the nature of openness and mobility, it is very easy for malicious workers to launch CSDF attack by reporting false sensing data, resulting in a wrong final decision of in the SP.

At first, malicious workers launch CSDF attack individually and respectively. The individual CSDF attack and defense have appeared in some existing works. In [25], the authors analyze the problem of aggregating noisy labels from crowd workers to infer the underlying true labels of binary tasks. To address this problem, they also design a reputation-based worker filtering algorithm that uses a combination of disagreement-based penalties and optimal semi-matchings to identify adversarial workers. In [26], the authors develop an optimal attack framework in which the attacker can not only maximize his attack utility but also disguise the introduced malicious workers as normal ones such that they cannot be detected easily. The strategy derived from the proposed optimal attack framework makes the malicious workers behave ''smarter''. If there is little hope to achieve the attack goal on some objects, they will tend to agree with the normal workers on those objects to gain higher weights, and in turn, can exert stronger impact on other objects [26]. However, it is difficult for a malicious worker to improve his trust value fast all by himself through this occasional reliable behavior. Overall, the power of individual CSDF attack is finite and can be suppressed by trust mechanism easily.

To avoid the detection of trust mechanism, some malicious workers attempt to collude with each other and submit false sensing data together at the same time. This attack pattern can be called collusive CSDF attack. Generally, the collusive

attack pattern can increase the power of malicious workers, which can be launched through three methods.

In the first method, a malicious worker can acquire multiple IDs to fake sensing data through the sybil attack [27]. There are many schemes to defense against the Sybil attack [28], [29]. Specially, if each IP is restricted to acquire an ID, this attack pattern can be addressed easily. In the second method, a malicious worker can control multiple computers by embedding trojan viruses. This attack pattern can suppressed by using a good antivirus software. In the third method, multiple malicious workers collaborate together to fake sensing data. In this attack pattern, each malicious worker only has an ID. Currently, this attack pattern is used as a popular collusive attack, especially in MCS to fake sensing data. Although this collusive attack pattern is considered in [30], how to improve the trust value of malicious workers is not discussed. Except for increasing the attack power, collusive CSDF attack can also be launched to avoid the detection of trust mechanism.

In this paper, we further analyze the characteristics of MCS and the collusive CSDF attack demand, and thus finding three types of threats that can be achieved by collusive CSDF attackers, including increase the attack power, improve trust value and disturb the data aggregation.

- **Increase the attack power**: As we know, ''more hands make light work'' and ''more people, more powerful''. Inspired by this, malicious workers can conspire with each other to form a collusive clique to increase the attack power, and thus faking sensing data intentionally to corrupt the data collection.
- **Improve trust value**: By collusion, collusive CSDF attackers can improve their trust value quickly. For instance, one of collusive CSDF attackers can disguise as an EU and preselect a data aggregation decision. Then, this attacker would tell it to his conspirers in advance as well as send a Query message to the SP. Their trust value can be improved quickly if their sensing data are consistent as the SP's decision in the data aggregation.
- **Manipulate the data aggregation**: With high trust value, collusive CSDF attackers can avoid the detection of trust mechanism, and thus reporting their false sensing data to manipulate the data aggregation successfully and mislead the SP to make a wrong decision.

To maintain the opportunity to manipulate the data aggregation, collusive CSDF attackers are extremely sensitive to their trust value. As shown in Fig.2, they begin collusive CSDF attack procedure under the constraint

$$\|T_i \leq \delta + \lambda\| \leq \frac{N_c}{2}$$

where N_c is the number of collusive CSDF attackers and W_i is one of them, $\|T_i \leq \delta + \lambda\|$ denotes the number of collusive CSDF attackers under the case $T_i \leq \delta + \lambda$. Here, $\lambda (0 \leq \lambda < 1 - \delta)$ is the initial trust warning line for collusive

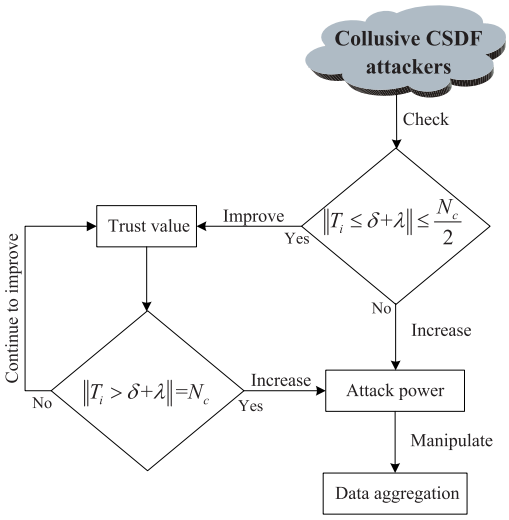


FIGURE 2. Collusive CSDF attack procedure.

CSDF attackers. It is too late to improve trust value when $\|T_i \leq \delta\| \leq \frac{N_c}{2}$. In this situation, the majority of collusive CSDF attackers may be marked as malicious. This attack pattern continues until $\|T_i > \delta\| = N_c$. Then, collusive CSDF attackers can increase their attack power to corrupt the data collection and manipulate the data aggregation successfully.

IV. PROPOSED DEFENSE SCHEME

In this section, we propose a defense scheme called BMCA from the design idea of binary-minmaxs clustering analysis to suppress collusive CSDF attack in MCS. Moreover, the BMCA scheme can be used to perfect trust evaluation.

A. DESIGN IDEA

To design the defense scheme of collusive CSDF attack, we analyze its attack threats, and thus finding three kinds of general features as follows.

- **Binary data**: Workers generally report honest or false sensing data in the data collection. Thus, the sensing behaviors of workers in the MCS environment can be abstracted as the binary data with the type of ''1'' or ''0'' sensing data.
- **Action together**: Collusive CSDF attackers always report false sensing data together no matter which threats they would launch.
- **High trust value**: Collusive CSDF attackers often have high trust value. With the help of each other, they can improve their trust value to avoid the detection trust mechanism.

Considering the ''Binary data'' and ''Action together'' of general features, we introduce the design idea of binary-minmaxs clustering analysis based on the logic AND operation to construct the defense scheme called BMCA to suppress collusive CSDF attack. In addition, the ''Action together'' of general feature can make collusive CSDF attackers behave a low variance in their similarity vector. In this

basis, we analyze the "High trust value" of general feature to design the algorithm of detecting collusive CSDF attackers.

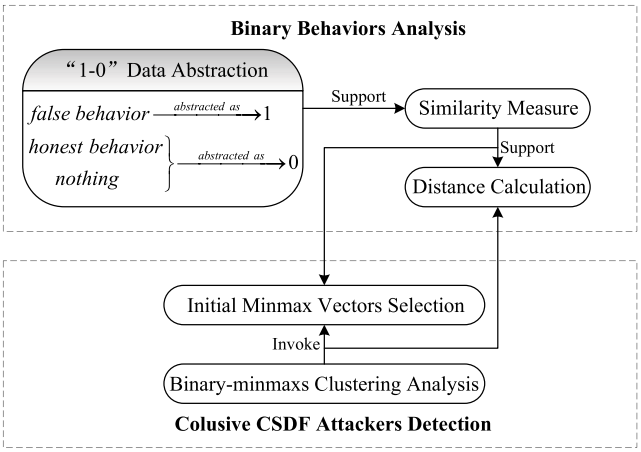


FIGURE 3. Architectural view of the BMCA scheme.

As shown in Fig.3, the BMCA scheme is conducted in two successive stages: binary behaviors analysis and collusive CSDF attackers detection. In the first stage, we design the logic AND distance calculation between any two workers in line with the type of "1" or "0" sensing data and measure their similarity. In the second stage, an algorithm with binary-minmaxs clustering analysis is designed to detect collusive CSDF attackers.

B. BINARY BEHAVIORS ANALYSIS

It is worth noting that the sensing data from each worker can be abstracted as a binary variable: false or honest. Based on this, the SP can abstract the MCS behaviors of workers as: "1" or "0". Specifically, the false behavior of a worker in a task by comparing their sensing data with the final decision of the SP can be abstracted as "1", whereas the honest behavior of a worker can be abstracted as "0".

The "1-0" database is designed as an extensible database, whose size corresponds to the task times of MCS actions. After each MCS action, the SP should add a row in the "1-0" database to record the abstracted binary behaviors related on the sensing data from workers. When the current MCS action is numbered as the k-th task time, the size of the "1-0" database is k. In the "1-0" database, n is the total number of workers in the MCS platform.

Take the i-th worker (W_i) as an example, $W_i(b_i)_k$ is recorded as $W_i(1)_k$ when W_i reported false sensing data, $W_i(b_i)_k \rightarrow W_i(0)_k$ when reported honest sensing data and $W_i(b_i)_k \rightarrow W_i(-)_k$ when reported nothing at the k-th task time.

In the collusive CSDF attack, we have known that malicious workers report false sensing data together to corrupt data collection. That is, they often launch "1" behaviors together in a certain task time and behave high similarity among themselves. Therefore, the logic AND operation (&) can be introduced to measure the similarity. Take W_i and W_j as an example of any two workers, if $W_i(b_i)_k \rightarrow W_i(1)_k$, $W_j(b_j)_k \rightarrow W_i(1)_k$, $1\&1 = 1$ under the logic AND operation. Otherwise, 1&0=0 or 0&1=0.

For the convenience of measuring the similarity between any two workers, the history binary behaviors of each worker should be extracted from the "1-0" database as a vector in the current MCS action. For W_i, his behaviors vector can be represented as $B_i = [W_i(b_i)_1, W_i(b_i)_2, \cdots, W_i(b_i)_k]$. If $W_i(b_i)_1 \rightarrow W_i(0)_1$, $W_i(b_i)_2 \rightarrow W_i(-)_2$, $W_i(b_i)_k \rightarrow W_i(1)_k$, B_i can be definitely described as $[0, -, \cdots, 1]$.

Obviously, the redundant data such as $W_i(-)_2$ or $W_j(-)_6$ are useless to measure the similarity between W_i and W_j. Since "1" is useful to analyze the collusive features of malicious workers, it can be set $W_i(-)_2 = W_i(0)_2$ and $W_j(-)_6 = W_j(0)_6$. As a consequence, both the honest behavior and nothing reported by a worker can be abstracted as "0".

For W_i and W_j, the logic AND operation between B_i and B_j can be described as

$$B_{ij} = B_i \& B_j \tag{4}$$

Then, the similarity between W_i and W_j can be measured as

$$sim_{ij} = \frac{|\ell_{ij}| + c_{ij}}{|B_{ij}| + c_{ij}}, \quad |1_{ij}| \leq |B_{ij}| \tag{5}$$

where $|\ell_{ij}|$ denotes the amount of "1" in B_{ij} and $|B_{ij}| = k - 1$ is corresponding to the amount of elements in B_{ij}.

Specially, c_{ij} is used to record the continuous false behaviors of malicious workers. For instance, if W_i and W_j suddenly report false sensing data simultaneously at the q-th task time, they may continues to report false sensing data at the $(q+1)$-th, $\cdots$, $(q+c_{ij})$-th task time. Such continuous false behaviors can increase the similarity between W_i and W_j. The continuous parameter c_{ij} can be recorded by Algorithm 1.

Algorithm 1 Record c_{ij} Value

Input: B_{ij}
Output: c_{ij}

```
Initialize c_ij = 0, q = 1
for each b_ij^q ∈ B_ij (1 ≤ q < |B_ij|) do
    if (b_ij^q == 1 && b_ij^{q+1} == 1) then
        c_ij++
        q++
    end if
end for
```

The continuous parameter c_{ij} can make W_i and W_j behave a higher similarity due to their continuous false behaviors. It can be proofed that

$$\frac{|\ell_{ij}| + c_{ij}}{|B_{ij}| + c_{ij}} \geq \frac{|\ell_{ij}|}{|B_{ij}|} \tag{6}$$

Proof:

$$\frac{|\ell_{ij}| + c_{ij}}{|B_{ij}| + c_{ij}} - \frac{|\ell_{ij}|}{|B_{ij}|} = \frac{c_{ij} * (|B_{ij}| - |\ell_{ij}|)}{|B_{ij}| * (|B_{ij}| + c_{ij})} \geq 0$$

TABLE 1. Description of the "1-0" database style.

Task times	W_ID (the abstracted binary behaviors related on sensing data)					
1	$W_1(b_1)_1$	$W_2(b_2)_1$	$\cdots$	$W_j(b_j)_1$	$\cdots$	$W_n(b_n)_1$
2	$W_1(b_1)_2$	$W_2(b_2)_2$	$\cdots$	$W_j(b_j)_2$	$\cdots$	$W_n(b_n)_2$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
k	$W_1(b_1)_k$	$W_2(b_2)_k$	$\cdots$	$W_j(b_j)_k$	$\cdots$	$W_n(b_n)_k$

Likewise, we can measure the similarity between W_i and other workers, and thus generating a similarity vector for W_i, which is expressed as

$$SIM_i = [sim_{i1}, \cdots, sim_{ij}, \cdots, sim_{in}] \tag{7}$$

For all workers, their similarity vectors of can compose a matrix $SIM_{n\times n}$.

$$SIM_{n\times n} = \begin{pmatrix} sim_{11} & \cdots & sim_{1n} \\ \vdots & \ddots & \vdots \\ sim_{n1} & \cdots & sim_{nn} \end{pmatrix} \tag{8}$$

With the similarity measure, the distance between W_i and W_j can be calculated as

$$d_{ij} = \frac{1}{n}\sum_{p=1}^{n}(sim_{ip} - sim_{jp}) \tag{9}$$

It can be found that collusive CSDF attack can make W_i and W_j behave high similarity but short distance between themselves.

C. COLLUSIVE CSDF ATTACKERS DETECTION

In the BMCA scheme, we design a binary-minmaxs clustering algorithm to differentiate collusive CSDF attackers and trusted workers. First of all, we need to select two samples as the initial minmax vectors of this algorithm by analyzing the variance of each similarity vector from workers.

Considering that both collusive CSDF attackers and trusted workers have high trust value, the variance analysis should be performed in $SIM_{h\times h}$ from the workers with high trust value (Ω) at the current MCS action, in which h is the number of the workers with high trust value. Obiviously, $h \leq n$. Compared with $SIM_{n\times n}$, it can prompt the detection efficiency to perform binary-minmaxs clustering algorithm in $SIM_{h\times h}$, since our main purpose is to detect collusive CSDF attackers who hold the characteristics of high trust value.

Since collusive attackers often launch "1" behaviors together, each of them may get a low variance in SIM_i. Let $var(\cdot)$ denote the variance function, the initial min vector (μ_1) is corresponding to collusive attackers, which can be derived from the SIM_i with the lowest variance.

$$\mu_1 = \min_{W_i\in\Omega}(var(SIM_i)) \tag{10}$$

Since trusted workers often report honest sensing data individually, each of them may get a high variance in SIM_i. The initial max vector (μ_2) is corresponding to trusted workers, which can be derived from the SIM_i with the highest variance.

$$\mu_2 = \max_{W_i\in\Omega}(var(SIM_i)) \tag{11}$$

With the initial minmax vectors $\{\mu_1, \mu_2\}$, the binary-minmaxs clustering analysis can be designed by Algorithm 2 to detect collusive CSDF attackers.

Algorithm 2 Binary-Minmaxs Clustering Analysis

Input: Ω, $SIM_{h\times h}$, $\{\mu_1, \mu_2\}$,

Output: the set of collusive CSDF attackers (Ψ_1) and trusted workers (Ψ_2)

repeat
Initialize $\Psi_1 = \Psi_2 = \emptyset$
for $i = 1, i \leq h, i++$ **do**
  Calculate the distance d_{is} between SIM_i and μ_s ($1 \leq s \leq 2$) with with equation (9)
  if $d_{i1} > d_{i2}$ **then**
    $\Psi_1 = \{W_i\} \cup \Psi_1$
  else
    $\Psi_2 = \{W_i\} \cup \Psi_2$
  end if
end for
Update the new min vector $\mu_1' = \min_{W_i\in\Psi_1}(var(SIM_i))$
Update the new min vector $\mu_2' = \max_{W_i\in\Psi_2}(var(SIM_i))$
for $s = 1, s \leq 2, s++$ **do**
  if $\mu_s' \neq \mu_s$ **then**
    Update $\mu_s' = \mu_s$
  else
    Keep the current minmax vectors unchanged
  end if
end for
until the current minmax vectors are not updated again

D. PERFECT TRUST EVALUATION

To suppress such threat, the special punishment to sudden false MCS behaviors is introduced to prevent the trust value growth of collusive CSDF attackers at each MCS action.

For instance, if W_i suddenly report false sensing data at his k-th MCS action, his trust value will be reduced by the special punishment factor, which can be evaluated as:

$$\varphi_i = (1-\theta) * log_{|hon_i - fal_i|}(k + c_i) \tag{12}$$

Once W_i continues to report false sensing data at the $(k+1)$-th, $\cdots$, $(k+c_i)$-th MCS action, his trust value will be continuously reduced by φ_i.

Therefore, the trust value of W_i at the k-th MCS action can be dynamically updated as:

$$DT_i = \begin{cases} T_i - \Delta_i * \varphi_i, & T_i > \varphi_i \\ 0, & T_i \leq \varphi_i \end{cases} \tag{13}$$

where $\Delta_i = 1$ means that the special punishment factor will militate when W_i suddenly reports false sensing data at his k-th MCS action. Or else, $\Delta_i = 0$.

Let Φ is the set of workers in the current MCS action and is Γ is the set of these workers' trust value. Algorithm 3 can be performed to perfect trust evaluation.

Algorithm 3 Perfect Trust Evaluation

Input: Φ, Γ, Ψ_1 and Ψ_2

Output: the updated Γ

1: **for** each $W_i \in \Phi$ **do**
2: **if** $W_i \in \Psi_1$ **then**
3: $fal_i = fal_i + 1$ and his sensing data are deleted
4: c_i++
5: **if** $c_i > 1$ **then**
6: Calculate φ_i with equation (12)
7: Dynamically update the trust value of W_i with equation (13)
8: **end if**
9: **else**
10: $hon_i = hon_i + 1$
11: **end if**
12: **end for**

V. SIMULATION ANALYSIS

A. SIMULATION SETUP

We perform simulations to validate the performance of the BMCA scheme and discuss the simulation results. The simulation elements are shown in table II.

TABLE 2. Description of simulation elements.

Parameters	Description	Default
N_w	Number of workers	60
cycle	Number of cycle simulation	100
round	Rounds of attack	50
P_c	Percentage of collusive CSDF attackers	10∼50%
λ	Initial trust warning line	0.1

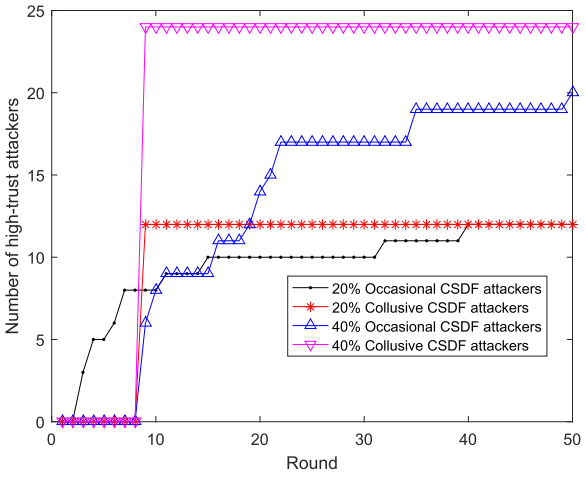


FIGURE 4. Collusive CSDF attack vs. occasional CSDF attack under forming high-trust attackers.

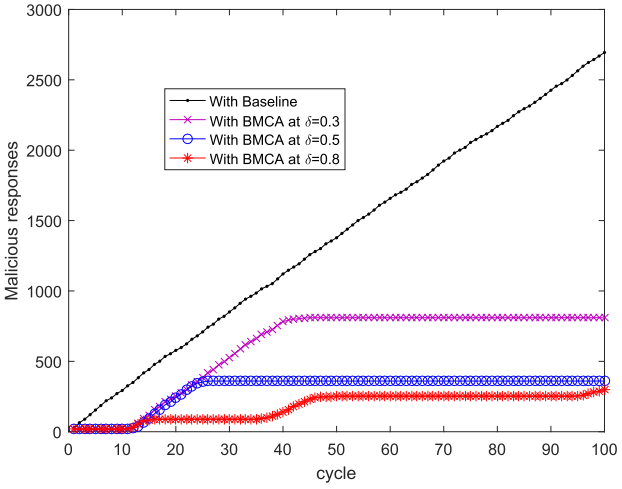
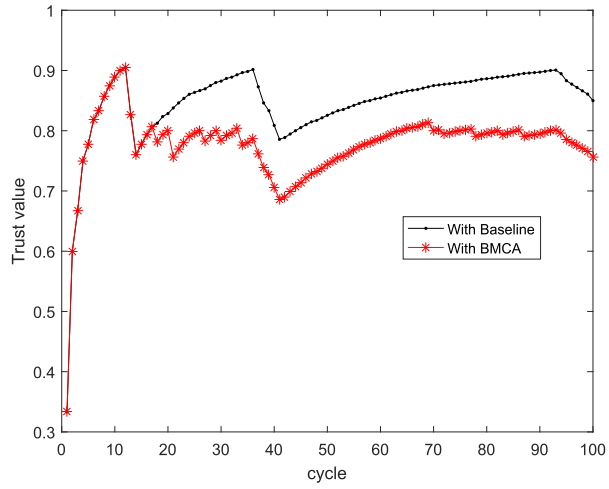
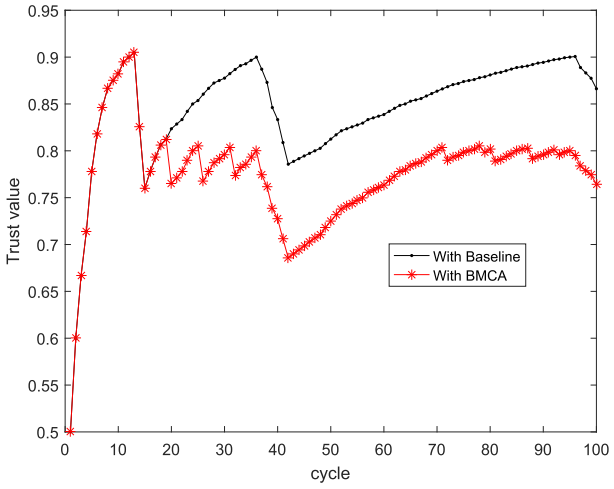


FIGURE 5. Suppressing malicious responses.

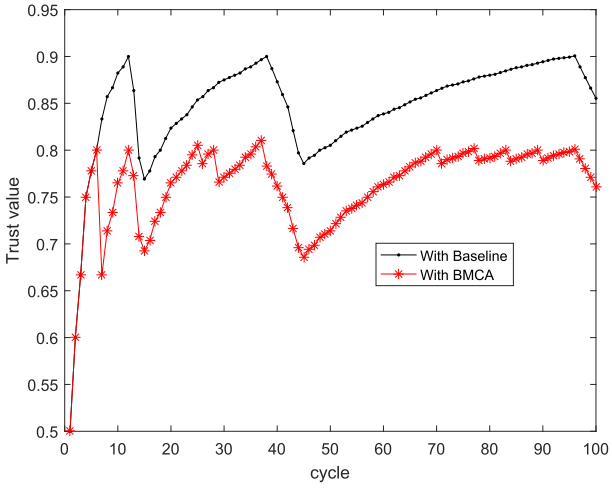
The simulations are performed by cycle-based fashion. At each cycle, workers are selected randomly to execute MCS actions with each other. After several cycles, a trusted network topology is gradually generated by trust mechanism. The SP then utilizes it to execute the following MCS actions, and update the trust value on the corresponding workers. In our simulations, the behavior pattern for trusted workers is modeled to always report honest sensing data, while the behavior pattern for collusive CSDF attackers is to alternatively report false or honest sensing data together depending on whether their trust value is greater than the threshold δ.



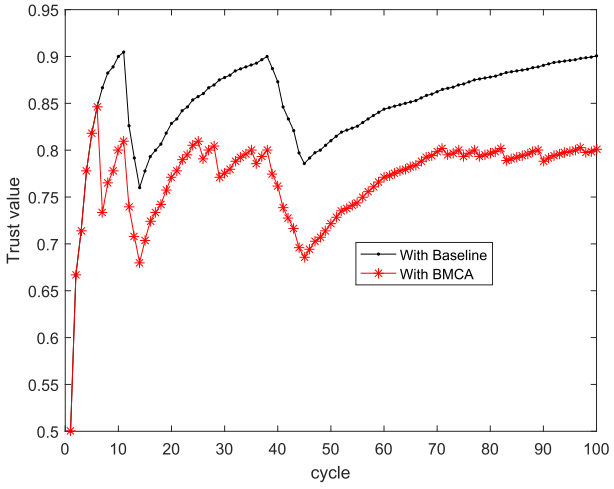
(a) collusive CSDF attacker 1



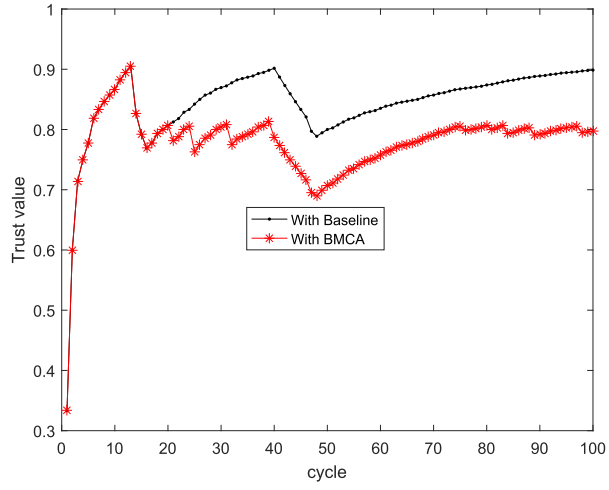
(b) Collusive CSDF attacker 2



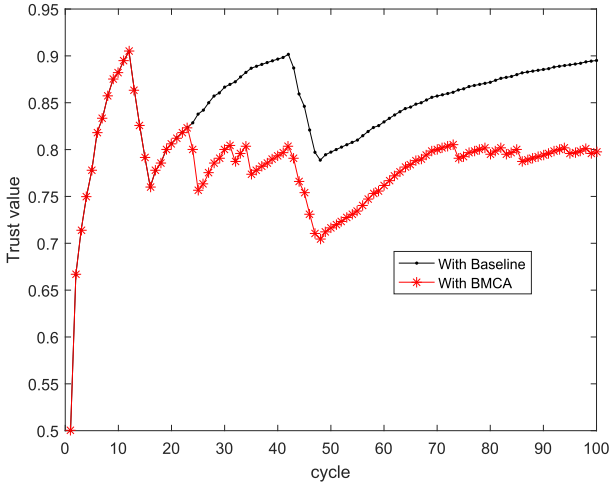
(c) Collusive CSDF attacker 3



(d) Collusive CSDF attacker 4



(e) Collusive CSDF attacker 5



(f) Collusive CSDF attacker 6

FIGURE 6. Variation of collusive CSDF attackers' trust value.

B. SIMULATION RESULTS

One of main goals of collusive CSDF attackers is to improve their trust value. As we know, an attacker such as W_i can be detected by $T_i < \delta$ in trust mechanism. To get more attack opportunities, W_i need to be disguised as a high-trust attacker, i.e. $T_i \geq \delta$. As shown in Fig.4, collusive CSDF attackers can form high-trust attackers with less rounds of attack. Generally speaking, occasional CSDF attackers improve their trust value by reporting honest sensing data when there is little hope to achieve the attack goal [26], whereas collusive CSDF attackers can help with each other. So, collusive CSDF attackers can become high-trust faster than occasional CSDF attackers.

To suppress collusive CSDF attack better under the defense of the BMCA scheme, it is necessary to select a rational value of δ. As $T_i \in [0, 1]$, δ can be considered from the three types of optional states [low, medium, high]. Then, we can perform the simulation of suppressing malicious responses to validate the effectiveness of the BMCA scheme under the three types of optional states of δ. In this simulation, 0.3, 0.5 and 0.8 denotes the low, medium and high state of δ respectively. As shown in Fig.5, the performance of the BMCA scheme at $\delta = 0.8$ is the best. Therefore, the rational value of δ should be selected as 0.8 in the simulations. We can also find that the BMCA scheme is better than Baseline in suppressing malicious responses, even though δ is selected as 0.8 for Baseline.

To analyze how collusive CSDF attack can affect the performance of trust mechanism, we firstly choose six collusive CSDF attackers randomly to observe the variation of their trust value in the Baseline and BMCA scheme. Fig.6 shows that collusive CSDF strategies make the attacker's trust value fluctuates along with cycles. The trust value usually outweighs δ in Baseline. Fortunately, the trust value of collusive CSDF attackers can be reduced by the BMCA scheme. This is because the special punishment to sudden false MCS behaviors is introduced to dynamically updating the trust value of collusive CSDF attackers at each MCS action.

We note that collusive CSDF attackers will deviate the real trust value by forming high-trust workers, and thus causing some network trust errors (*nte*). A higher errors indicate lower accuracy in the evaluation of trust value. With *nte*, we can analyze how collusive CSDF attack affects the performance of trust mechanism from the entire network. *nte* can be specified by:

$$nte = \frac{1}{N_w} \sum_{i=1}^{N_w} \sqrt{\frac{1}{T_i'}(T_i' - T_i)^2} \tag{14}$$

where T_i' and T_i are the actual and measured trust value of T_i, respectively.

In the simulation of *nte*, the actual trust value of a malicious worker is randomly assigned in the interval (0, 0.5]. Without loss of generality, we employ the averaged *nte* data of 100 cycles as the simulation results. As shown in Fig.7, the BMCA scheme can reduce *nte* better than Baseline. Even though the percentage of collusive CSDF attackers is 50%, the *nte* of the TFAA scheme merely achieves 0.28.

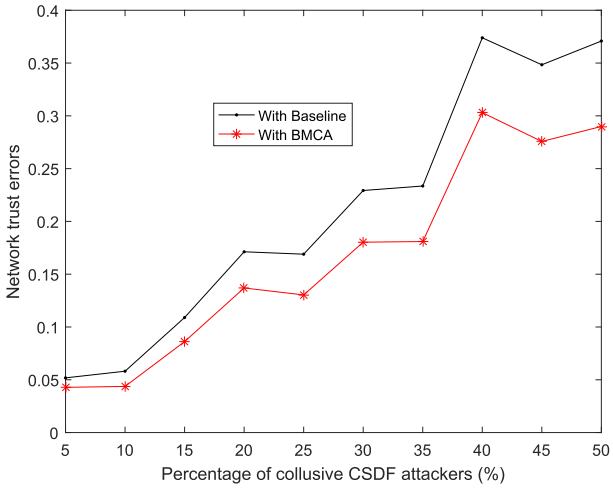


FIGURE 7. ***nte*** **with the guard of BMCA.**

Finally, we validate the performance of BMCA in terms of attack success ratio when collusive CSDF attackers fake sensing data with high trust value. Without loss of generality, we employ the averaged attack success ratio data of 50 rounds of attack as the simulation results. At each round of attack, several workers are selected randomly to perform an MCS action from trusted workers and collusive CSDF attackers. In an MCS action, if the collusive CSDF attackers with high trust value are more than the majority of workers, they would attack successfully.

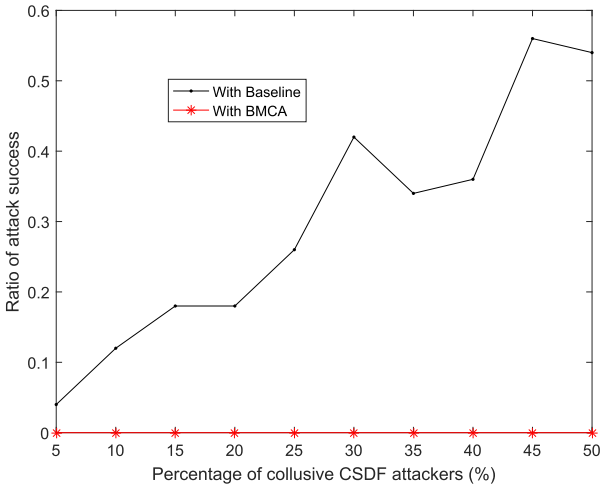


FIGURE 8. **Suppressing attack success ratio.**

As shown in Fig.8, the attack success ratio against the Baseline scheme amplifies with the percentage of collusive CSDF attackers. This is because collusive CSDF attackers' trust value usually outweigh δ in the Baseline scheme, so they can manipulate the decision result of the SP with false sensing data easily. Fortunately, collusive CSDF attackers are difficult to improve their trust value to outweigh δ in the BMCA

scheme. Consequently, collusive CSDF attackers are impossible to manipulate the decision result of the SP, since the collusive CSDF attackers with high trust value cannot become the majority of workers.

VI. CONCLUSION

In this paper, we analyze the threats of collusive CSDF attack and propose the BMCA scheme to suppress such attack. The BMCA scheme is conducted in two successive stages: binary behaviors analysis and collusive CSDF attackers detection, in which binary-minmaxs clustering analysis based on the logic AND operation is introduced to design the BMCA scheme due to the type of "1" or "0" historical sensing data. With the special punishment to sudden false MCS behaviors, the BMCA scheme can be used to prevent collusive CSDF attackers' trust value from growing in the trust evaluation. Simulation results show that our BMCA scheme can enhance the accuracy of trust evaluation and suppress collusive CSDF attack success ratio effectively. In addition, we must continue to fight against the potential threats against MCS. Except for collusive CSDF attack, malicious workers may launch their attack strategies in different tasks. For instance, they behave harmless in some uninterested tasks in order to attain high trust value and always behave harmful in their interested tasks. In the future work, we need to further study a defense scheme from the design idea of distinctive trust evaluation for different tasks to suppress such threat.

REFERENCES

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011.
[2] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 1, pp. 3–16, Jan./Mar. 2017.
[3] J. Wang, Y. Wang, D. Zhang, J. Goncalves, D. Ferreira, A. Visuri, and S. Ma, "Learning-assisted optimization in mobile crowd sensing: A survey," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 15–22, Jan. 2019.
[4] Z. Zhang, P. Zhang, D. Liu, and S. Sun, "SRSM-based adaptive relay selection for D2D communications," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2323–2332, Aug. 2018.
[5] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, and Z. Han, "The accuracy-privacy trade-off of mobile crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 132–139, Jun. 2017.
[6] C. de Capua, A. Meduri, and R. Morello, "A smart ECG measurement system based on web-service-oriented architecture for telemedicine applications," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 10, pp. 2530–2538, Oct. 2010.
[7] H. Amintoosi and S. S. Kanhere, "A reputation framework for social participatory sensing systems," *Mobile Netw. Appl.*, vol. 19, no. 1, pp. 88–100, 2014.
[8] C. Wu, T. Luo, F. Wu, and G. Chen, "EndorTrust: An endorsement-based reputation system for trustworthy and heterogeneous crowdsourcing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
[9] V. Kostakos, J. Rogstadius, and D. Ferreira, "Human sensors," in *Participatory Sensing, Opinions and Collective Awareness*. Cham, Switzerland: Springer, 2017, pp. 69–92.
[10] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, "Social-network-assisted worker recruitment in mobile crowd sensing," *IEEE Trans. Mobile Comput.*, vol. 18, no. 7, pp. 1661–1673, Jul. 2019.
[11] W. Feng, Z. Yan, H. Zhang, K. Zeng, Y. Xiao, and Y. T. Hou, "A survey on security, privacy, and trust in mobile crowdsourcing," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2971–2992, Aug. 2018.
[12] J. An, X. Gui, Z. Wang, J. Yang, and X. He, "A crowdsourcing assignment model based on mobile crowd sensing in the Internet of Things," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 358–369, Oct. 2015.
[13] H. Amintoosi and S. S. Kanhere, "A trust-based recruitment framework for multi-hop social participatory sensing," in *Proc. IEEE Int. Conf. Distrib. Comput. Sensor Syst.*, Cambridge, MA, USA, May 2013, pp. 266–273.
[14] Y. Gao, X. Li, J. Li, and Y. Gao, "DTRF: A dynamic-trust-based recruitment framework for Mobile Crowd Sensing system," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, May 2017, pp. 632–635.
[15] D. Wu, L. Fan, C. Zhang, H. Wang, and R. Wang, "Dynamical credibility assessment of privacy-preserving strategy for opportunistic mobile crowd sensing," *IEEE Access*, vol. 6, pp. 37430–37443, 2018.
[16] B. Kuang, A. Fu, S. Yu, G. Yang, M. Su, and Y. Zhang, "ESDRA: An efficient and secure distributed remote attestation scheme for IoT sWARMS," *IEEE Internet Things J.*, to be published.
[17] H. Mousa, S. B. Mokhtar, O. Hasan, O. Younes, M. Hadhoud, and L. Brunie, "Trust management and reputation systems in mobile participatory sensing applications: A survey," *Comput. Netw.*, vol. 90, no. 10, pp. 49–73, Oct. 2015.
[18] M. A. Morid and M. Shajari, "An enhanced e-commerce trust model for community based centralized systems," *Electron. Commerce Res.*, vol. 12, no. 4, pp. 409–427, Nov. 2012.
[19] X. Li, F. Zhou, and X. Yang, "Scalable feedback aggregating (SFA) overlay for large-scale P2P trust management," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 10, pp. 1944–1957, Oct. 2012.
[20] J. Feng, Y. Zhang, G. Lu, and W. Zheng, "Securing cooperative spectrum sensing against ISSDF attack using dynamic trust evaluation in cognitive radio networks," *Secur. Commun. Netw.*, vol. 8, no. 17, pp. 3157–3166, Nov. 2015.
[21] M. Li, Y. Xiang, B. Zhang, Z. Huang, and J. Zhang, "A trust evaluation scheme for complex links in a social network: A link strength perspective," *Appl. Intell.*, vol. 44, no. 4, pp. 969–987, Jun. 2016.
[22] A. Fu, S. Yu, Y. Zhang, H. Wang, and C. Huang, "NPP: A new privacy-aware public auditing scheme for cloud data sharing with group users," *IEEE Trans. Big Data*, to be published.
[23] A. Josang and R. Ismail, "The beta reputation system," in *Proc. 15th Bled Electron. Commerce Conf.*, 2002. pp. 1–14.
[24] Wikipedia. (Aug. 2016). *Gamma Function*. [Online]. Available: http://en.wikipedia.org/wiki/Gamma function
[25] S. Jagabathula, L. Subramanian, and A. Venkataraman, "Reputation-based worker filtering in crowdsourcing," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2014, pp. 2492–2500.
[26] C. Miao, Q. Li, H. Xiao, W. Jiang, M. Huai, and L. Su, "Towards data poisoning attacks in crowd sensing systems," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jun. 2018, pp. 111–120.
[27] J. R. Douceur, "The Sybil attack," in *Proc. Int. Workshop Peer-to-Peer Syst.*, Mar. 2002, pp. 251–260.
[28] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *Proc. IEEE Symp. Secur. Privacy*, May 2008, pp. 3–17.
[29] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "Sybilguard: Defending against sybil attacks via social networks," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, Sep. 2006, pp. 267–278.
[30] Z. Qin, Q. Li, and G. Hsieh, "Defending against cooperative attacks in cooperative spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2680–2687, Jun. 2013.

JINGYU FENG received the B.S. degree from the Lanzhou University of Technology, China, in 2006, and the Ph.D. degree from Xidian University, China, in 2011. He is currently a Vice Professor with the Xi'an University of Post and Telecommunication, China. His current research interests include the IoT security, trust management, and mobile crowd sensing.

TAO LI received the B.S. degree from the Wanfang Science and Technology Institute, Henan Technological University, China, in 2017. He is currently pursuing the master's degree with the Xi'an University of Posts and Telecommunications. His current research interests include trust management and mobile crowd sensing.

YUJIA ZHAI is currently pursuing the bachelor's degree with the Xi'an University of Posts and Telecommunications. Her current research interests include trust management and mobile crowd sensing.

SHAOQING LV received the B.S. degree from Tianjin Polytechnic University, China, in 2009, and the Ph.D. degree in information security from Xidian University, China, in 2016. He is currently a Lecturer with the Xi'an University of Posts and Telecommunications, China. His research interests include network security and online social networks analysis.

FENG ZHAO received the M.S. degree from the Xi'an University of Architecture and Technology, China, in 2005. He is currently a Senior Lecturer with the Xi'an University of Posts and Telecommunications, China. His current research interests include wireless security, trust management, and mobile crowd sensing.