DTCS: An Integrated Strategy for Enhancing Data Trustworthiness in Mobile Crowdsourcing

Jia Hu, Hui Lin*, Xuancheng Guo, Ji Yang

***Abstract* — Mobile Crowdsourcing Systems (MCS) are important sources of information for the positioning services in IoT such as gathering location information through employing citizens to participate in data collection. Although MCS have attracted significant research and development efforts, there are salient open issues and challenges in security and privacy for MCS, which is an essential factor for its success. This paper proposes an integrated strategy named DTCS to enhance data trustworthiness and defend against the internal threats for mobile crowdsourcing. The DTCS integrates effective methods including an evaluation scheme for the attribute relevancy and familiarity of participants, a trust relationship establishment method, a group division strategy based on attributes and metagraph, and a core-selecting based incentive mechanism. The simulation results show that the DTCS improves the performance of the crowdsourcing strategy compared to the state-of-the-art including the TSCM and PPPCM. The DTCS can effectively defend against internal conflicting behaviour attacks and collusion attacks to enhance data trustworthiness for mobile crowdsourcing.**

***Index Terms* — Internet-of-Things, Mobile crowdsourcing, Data trustworthiness, Mechanism design**

I. INTRODUCTION

INTERNET-of-Things (IoT), which use pervasive interconnected smart objects operating together to reach common goals, have become particularly popular with the rapid development of advanced low-cost sensors, wireless communications and networking technologies [1-3]. IoT technologies can effectively improve the intelligence of the positioning services, promote the interactions between the human and the environment, enhance the reliability, resilience, operational efficiency, and energy efficiency of smart city services [4-7].

Mobile crowdsourcing systems (MCS) are important sources of information for the positioning services in IoT such as gathering location related sensing data by employing ordinary citizens to participate in data collection [8, 9]. MCS has become very popular as the number of mobile devices equipped with sensors (including handsets, tablets, electronic devices, etc.) shows dramatic growth [10-12]. MCS relies on individual participants to collect data from their activities and surrounding environments, and then upload the data to the application server via any available networking facility. The application server will process all data reported by the participants, extract the information in which queriers are interested, and forward such information to the queriers [13]. MCS has been successfully adopted to enable many new IoT applications, ranging from highway congestion detection to social trend understanding and positioning services [14].

Although MCS has attracted significant research and development efforts, there are salient open issues and challenges in security and privacy for MCS, which is an essential factor for the success of the burgeoning MCS for positioning services [15-18]. Since MCS allows any voluntary participant to contribute data, the application server is exposed to erroneous or even malicious data. Moreover, malicious participants may deliberately contribute bad data. In order to avoid making decisions based on the analysis of uncertain and imprecise data, it is crucial to maintain a high level of data trustworthiness, which is defined by a number of factors including data origin, collection and processing methods, such as trusted infrastructure and facility [13, 19].

Meanwhile, security has often had a low priority for vendors of IoT devices and this has led to a situation where IoT is filled with security vulnerabilities in practice. Hence, mobile crowdsourcing based positioning services are often exposed to security attacks [3] targeting at data confidentiality, privacy and data trustworthiness. Data trustworthiness shows how much the data used are trusted, authentic and protected from unauthorized access and modification, which ensures data to be accurate, complete and up-do-date. There are many security challenges in data trustworthiness such as denial-of-service, credential stealing, remote code injection, data integrity attacks, internal attacks, and supply chain attacks [19, 20]. Consequently, the availability, confidentiality, and integrity of both the original data and the analytics data are threatened by these attacks, e.g., the degraded availability of the mobile crowdsourcing based positioning services, the compromised confidentiality of the data and analytics, and the violated integrity of the data and analytics.

As an effort to tackle the aforementioned challenges, this paper focuses on the aspect of data trustworthiness to enhance security and privacy-preserving for mobile crowdsourcing based positioning services in IoT through designing a new Data Trustworthiness enhanced Crowdsourcing Strategy (DTCS). The major contributions of this work include the following:

This work was supported by the National Natural Science Foundation of China (61363068, 61472083, 61402110, 61472121), Pilot Project of Fujian Province (2016Y0031), the Foundation of Science and Technology on Information Assurance Laboratory (KJ-14-109), and the GDAS Special Project of Science and Technology Development (2017GDASCX-0101).

Jia Hu is with College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK (e-mail: j.hu@exeter.ac.uk), and the College of Mathematics and Informatics, Fujian Normal University, China.

Hui Lin (*Corresponding author) and Xuancheng Guo are with the College of Mathematics and Informatics, Fujian Normal University, Fuzhou, 350117 China (e-mail: linhui@fjnu.edu.cn).

Ji Yang is with the Guangzhou Institute of Geography, Guangzhou, China. (email: yangji@gdas.ac.cn)

- We propose a data trustworthiness enhanced crowdsourcing strategy to defend against the internal threats for mobile crowdsourcing based positioning services.
- The DTCS innovatively integrates four methods to achieve its aim: an evaluation scheme for participants' attribute relevancy and familiarity, a trust relationship establishment method between persons and groups, a group division strategy bases on attributes and metagraph, and a core-selecting based incentive mechanism.
- Simulation experiments demonstrate that the DTCS improves the performance of the crowdsourcing strategy compared to the state-of-the-art including the TSCM [21] and PPPCM [8] strategies. The DTCS can effectively defend against internal conflicting behaviour attacks and collusion attacks to enhance data trustworthiness for MCS.

The remainder of this paper is organized as follows. Section II presents a brief review of related work, Section III describes the system and adversary models, Section IV introduces the implementation details of the DTCS, Section V presents the performance evaluation of the DTCS. Finally, Section VI concludes the paper.

II. Related Work

Data trustworthiness for IoT enabled MCS systems has become a research hotspot that attracts many interests [8,21-27]. For example, Kantarci et al. [21] proposed a reputation based Sensing-as-a-Service scheme to ensure data trustworthiness in crowdsourcing management for MCS systems. Cao et al. [23] proposed a Trust-based Data Usage architecture including trust-based data sharing system, data semantic and abstraction models, and a data transparency and accountability enhancing mechanism. Palaghias et al. [24] presented an opportunistic sensing system to reliably derive and quantify trust relationships for MCS systems by combining the extracted real-world social graph. Huang et al. [25] proposed a reputation system based on the Gompertz function to compute reputation scores of devices to measure the trustworthiness of the contributed sensing data for MCS systems. Wang et al. [26] proposed ARTSense, a framework to solve the problem of "trust without identity" in MCS network to achieves the anonymity and security requirements by combining the privacy-preserving provenance model, a data trust assessment scheme with an anonymous reputation management protocol. Zhang et al. [8] proposed a participant coordination framework, which includes a cooperative data aggregation, an incentive distribution method, and a punishment mechanism to both protect participant privacy and ensure the trustworthiness of the collected data. Both Li [22] and Liu [27] proposed privacy-preserving schemes that use the homomorphic encryption to protect the trustworthiness of the crowdsourced data for a mobile crowdsourcing based location system. Gong et al. [28] identified fundamental tradeoffs among utility, privacy, and efficiency in MCS and proposed a flexible optimization framework to collect reliable data and provide privacy protection. Zhang et al. [29] proposed a secure and dependable auction mechanisms for MCS to defend against dishonest bidders in the sensing process and to incentivize participants to provide trustworthy crowdsourced data.

In the existing research on data trustworthiness, many studies assumed that the authenticated participants are trustworthy, thus ignoring the internal security threats such as internal conflicting behavior attacks launched by an internal participant with a legal identity giving dishonest opinions to frame up good parties and/or boost trust values of malicious peers. Meanwhile, most existing data trustworthiness enhanced mechanisms for MCS in IoT didn't consider the collusion attack that represents the real-world nature of MCS in IoT. Consequently, it is an open problem and a challenging task to design a new strategy to prevent internal attacks to enhance the data trustworthiness for MCS.

III. System and Adversary Models

A. System Model

Different MCS applications may have different system models. To make it more general, in this paper we consider a typical MCS system architecture in IoT, which has three stages: sensing, learning and mining, disseminating [13, 30]. In the sensing stage, before the owner of a mobile device can participate in an MCS application, he/she first needs to download the corresponding application to become a participant. For a certain query, the application server informs all participants about their sensing tasks. In the learning and mining stage, there are two possible data collection models. In the first model, participants play an active role by deciding when to report data. In the second model, reporting occurs whenever the state of the mobile device satisfies the tasks' requirements. Therefore, the sensed data are uploaded to the application server through wireless networks. The application server then processes the sensed data to extract the desired information. In the disseminating stage, the results are formatted into suitable forms and made available to queriers. The participants are connected to the access point through the smartphone and senses the required data. The end users or queriers request data through tasks and then utilize the information acquired by participants. The MCS operator distributes tasks to participants who meet the requirements of applications.

B. Adversary Model

This paper focuses on the internal security threats [31, 32] that can affect data trustworthiness to mobile crowdsourcing based positioning service in IoT. The internal threats are launched by an inside attacker who is a legal and certified participant. The internal attacks may compromise certain participants and gain full control of them. Once participants are compromised, the attacker can gain access to all stored information, including public and private keys. The attacker could also reprogram the captured participants to behave in a malicious manner. Therefore, the traditional encryption and authentication techniques may no longer be effective. The specific internal attacks considered in this paper are below [3]:

- Conflicting behavior attack: The attackers can transmit partially trustful information (e.g., correct IP address) and partially incorrect information (e.g., fake positions). Attackers can also provide erroneous recommended opinions for their own benefits.
- Collusion attack: Attackers collude to provide false information and give misleading judgments.

IV. Data Trustworthiness enhanced Crowdsourcing Strategy (DTCS)

In this section, we elaborate on the proposed data trustworthiness enhanced crowdsourcing strategy (DTCS), which integrates the trust relationship evaluation [31-33] with the mechanism design [34, 35], metagraph theory [36, 37], user group division technologies [38] to improve the accuracy of the trust relationship evaluation, defend against internal attacks and enhance data trustworthiness for mobile crowdsourcing based positioning services in IoT. In the rest of the paper, the term "participant", "mobile device" and the term "user" are used interchangeably.

In DTCS, sensing data are classified into different categories based on the sensitivity level (SL) of data. In this work, the sensitivity level of sensing data is decided by the data owner, fixed and divided into five grades from 1 to 5. The higher is the sensitivity level of data, the greater is the need for the confidentiality and privacy protection. Also, we use metagraph [36, 37], a graphical data structure for representing a collection of directed set-to-set mappings, to divide all participants into different groups according to the participants' attribute relevancy and familiarity. Moreover, each participant will execute an incentive mechanism before it makes a behaviour decision. The details of the DTCS are described as follows.

Attribute and Metagraph based Participant Group Division Scheme (AMPGD)

In AMPGD, we firstly evaluate the attribute relevancy and familiarity (ARF) among participants, and then divide all the participants into different groups based on the ARF evaluation results.

We assume each participant has an attribute set **ATTR** = {attr$_1$, attr$_2$, …, attr$_k$}, the attribute set of a participant may include location, gender, age, major, hobby and so on. The attribute relevancy and familiarity of participant *j* toward participant *i* $ARF_{(i,j)}$ evaluation can be done as follows.

$$ARF_{(i,j)} = R_{(i:j)} * \tau * \left[\frac{1}{n} * \sum_{int=1}^{n} \frac{\left|\mathrm{ATTR}_i^{int} \cap \mathrm{ATTR}_j^{int}\right|}{\left|\mathrm{ATTR}'\right|} \right] \quad (1)$$
$$s.t.\ \left|\mathrm{ATTR}_i^{int} \cap \mathrm{ATTR}_j^{int}\right| > w$$

where *w* is the threshold of the attribute intersection's scale. ATTR' is the attribute set used in this interaction, ATTR_i^{int} and ATTR_j^{int} are the attribute set used in the *int*th interaction between participants *i* and *j*, respectively. *n* is the total number of the interactions between participants *i* and *j*. $R_{(i:j)}$ is the reputation of *j* toward *i* stored in the local reputation database of *i*. τ is the time factor that determine how much the interaction time affect $R_{(i:j)}$. We then formally define the τ as:

$$\tau = \tau_{i:j,T_n} * \beta_{T_n} \quad (2)$$

where β_{T_n} is the density of the historical interaction until time T_n and $\tau_{i:j,T_n}$ is the weight factor, which determines how much the distribution of the interactions affects the $R_{(i:j)}$ at time T_n. $\tau_{i:j,T_n}$ and β_{T_n} can be computed as follows.

$$\beta_{T_n} = 1 - e\wedge\left(-\frac{\sum_{sl=1}^{|\mathrm{SL}|} N_{sl}}{m*n}\right) \quad (3)$$

$$\tau_{i:j,T_n} = \sum_{l=1}^{n}\left(\frac{T_l}{m} * \frac{l}{n}\right) \quad (4)$$

where N_{sl} is the number of times the historical accessing behaviors or interactions are confirmed on the sensitivity level *sl*. *m* and *n* are the number of time slots and cycle *T* respectively, e.g., in this paper, *T* is equal to 10 seconds, *m* is 5, so one time slot equals 2 seconds.

Based on the attribute relevancy and familiarity evaluation results, all the participants will be divided into different groups by using the metagraph theory, and the different possible kinds of trust relations between persons and groups will be built as follows.

First, for any participant *p* the trust relationship between *p* and *p*' (TR(*p*, *p*')), and *p* and group *g* (TR(*p*, *g*)) will be computed as follows.

a) Trust relationship between *p* and *p*' when *p* has a direct interaction with *p*', $TR_{(p,p')}^{direct}$, can be computed as follow.

$$TR_{(p,p')}^{direct} = \frac{1}{|\mathrm{SL}|} * \sum_{sl=i}^{|\mathrm{SL}|} \left(\frac{SI^{sl}}{TI^{sl}} * \xi_{sl} \right) \quad (5)$$

$$\begin{cases} \xi = E(\gamma_t) \\ \gamma_t = \sum_{j=i}^{|\mathrm{SL}|} IA_j \Big/ \sum_{j=1}^{|\mathrm{SL}|} IA_j \end{cases}, \quad (t = 1 \ldots N_{slot}) \quad (6)$$

where *i* is the minimum sensitivity level requirement. SI^{sl}, TI^{sl} denote the number of successful and total interaction with sensitivity level *sl*, respectively. ξ is the weight factor that determine how much the sensitivity level *sl* of the interaction affect $TR_{(p,p')}^{direct}$. γ_t is the rate between the number of interaction with the sensitivity level higher than the current required sensitivity level *i* and the total number of interaction with all sensitivity levels. IA_j represents the number of times that the sensitivity level of historical interaction is confirmed as *j*, and N_{slot} denotes the number of the time slots.

b) The trust relationship between p and group g when p has a direct interaction with g, $TR^{direct}_{(p,g)}$, can be computed as follow.

$$\begin{cases} TR^{direct}_{(p,g)} = \lambda_1 * (\frac{1}{m_1} * \sum_{k=1,p' \in g}^{m_1} TR^{direct}_{(p,p')}) + \lambda_2 * (\frac{1}{m_2} * \sum_{k=1,p'' \in g}^{m_2} TR^{indirect}_{(p,p'')}) \\ \lambda_1 + \lambda_2 = 1 \\ m_1 + m_2 \le |g| \end{cases} \quad (7)$$

where $TR^{indirect}_{(p,p'')}$ is the indirect trust relationship between p and p' when p has not a direct link with p'. m_1 and m_2 are the number of participants in group g that have direct and indirect interaction with p respectively. λ_1 and λ_2 are the weight factors that determine how much the direct and indirect interaction affect the $TR^{direct}_{(p,g)}$.

c) Let $\text{DirR} = \{dir\text{-}rec_i \,|\, i = 1 \ldots n\}$ be the direct recommenders set. The direct recommenders who has the direct interaction with p' and has the direct trust relationship evaluation result about the p'. Indirect trust relationship between p and p' when p has not a direct interaction with p', $TR^{indirect}_{(p,p')}$, can be computed as follow.

$$TR^{indirect}_{(p,p')} = \frac{1}{n} * \sum_{j=1,p_j \in DirR}^{n} \left(\frac{sl_j}{sl_{max}} * TR^{direct}_{(p,p_j)} \right) \quad (8)$$

where sl_{max} is the maximal security level of the recommender in DirR.

d) Let $\text{DirRG} = \{dir\text{-}recg_i \,|\, i = 1 \ldots m\}$ be the direct recommender group set. The direct recommender group g_i who has the direct interaction with g and has the direct trust relationship evaluation result about g. Indirect trust relationship between p and group g when p has not a direct interaction with g, $TR^{indirect}_{(p,g)}$, can be computed as follow.

$$\begin{cases} TR^{indirect}_{(p,g)} = \frac{1}{m} * \sum_{j=1,g_j \in DirRG}^{m} \left(\frac{sl_j}{sl_{max}} * TR^{direct}_{(p,g_j)} \right) \\ sl_j = \frac{1}{|g_j|} \sum_{k=1,p_k \in g_j}^{|g_j|} sl_{p_k} \end{cases} \quad (9)$$

where sl_j is the average sensitivity level of group g_i. sl_{max} is the maximal security level of the recommender group in DirRG.

Secondly, the attribute and metagraph theory based group division are considered as follows.

a) A metagraph $S =< X, E >$ is built as a graphical construct specified by its generating set X (participant set and attribute set) and a set of edges E defined on the generating set (trust relationship set).

b) Generating set X represents participants and their attribute in their corresponding groups. Edges between two metagraph nodes (participants or groups) indicate the existence of trust relationship between them.

c) Each edge has a label $e = <V_e, W_e> \in E$, which is a couple of values $<t; c>$: the first component is the trust relationship value of metagraph node V_e (participants or groups V_e) toward node W_e (participants or groups W_e) while the second component is the quality of the trust relationship value assignment (i.e. a confidence value), both of these components are in the range [0, 1].

d) Each participant might possess different positions within a group, which is denoted as participant membership degree PMD. The higher a PMD in the group, the more likely the behavior of the participant will be based on the standards and norms of the group. Let $\bar{g}$ be the group, the PMD of a member $\bar{p}$ in $\bar{g}$ is defined as follows.

$$\begin{cases} PMD = \kappa_1 * \frac{1}{|\bar{g}|} * \sum_{p \in \bar{g}, p \ne \bar{p}} ARF(\bar{p}, p) + \kappa_2 * \frac{1}{|\bar{g}|} * \sum_{p \in \bar{g}, p \ne \bar{p}} TR(\bar{p}, p) \\ TR(\bar{p}, p) = \rho_1 * TR^{direct}_{(\bar{p},p)} + \rho_2 * TR^{indirect}_{(\bar{p},p)} \\ \kappa_1 + \kappa_2 = 1 \\ \rho_1 + \rho_2 = 1 \end{cases} \quad (10)$$

e) A participant p belongs to a group g if the following condition is satisfied.

$$\begin{cases} \frac{1}{|g|} * \sum_{\tilde{p} \in g, p \ne \tilde{p}} ARF(\tilde{p}, p) > \theta \\ \frac{1}{|g|} * \sum_{\tilde{p} \in g, p \ne \tilde{p}} TR(\tilde{p}, p) > \theta' \\ TR(\tilde{p}, p) = \rho_1 * TR^{direct}_{(\tilde{p},p)} + \rho_2 * TR^{indirect}_{(\tilde{p},p)} \\ \rho_1 + \rho_2 = 1 \end{cases} \quad (11)$$

where θ and θ' are the threshold of the attribute relevancy and familiarity, and trust relationship respectively.

f) A high trust relationship value means that the trustee has gained a good feedback, whereas a confidence value close to 1 indicates that the trustor estimates the correlated trust relationship value with precision.

As an example, consider the metagraph $S =< X, E >$ in Figure 1. The sets X is $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ and the set of edges is $E = \{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3, \tilde{e}_4\}$. In Fig. 2, the edge $\tilde{e}_1$ between groups G1 and G2 is labeled as < 0.7, 0.6>. It shows that there exists a trust relationship between group G1 and group G2 and the trust relationship value of group G1 to group G2 is 0.7, and it is estimated with precision 0.6.

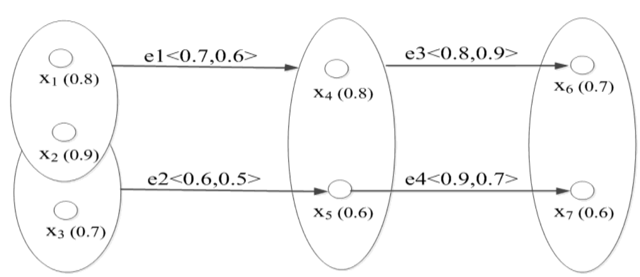


Fig. 1. An example of attribute and metagraph theory based group division.

Core-Selecting based Incentive Mechanism (CSIM)

This section presents the Core-Selecting based Incentive Mechanism (CSIM), which integrates the auction game into mobile crowdsourcing system to guarantee the reliability of the gathered crowdsourcing information through motivating all participants to provide true crowdsourcing information. The CSIM can also effectively defend against the internal collusion conflicting behavior and cheating attacks through implementing effective rewards and punishment mechanism.

Before introducing the CSIM, we present the mathematical descriptions as follows.

- Bidders: let $N = \{1, 2, \ldots, n\}$ denote the set of all bidders (crowdsourcing service participants);
- Auctioneers: the owner of crowdsourcing positioning service in IoT (crowdsourcing service initiator);
- Crowdsourcing services: let $M = \{1, 2, \ldots, m\}$ denote the set of crowdsourcing services, where $m >= 1$; S is the subset of services set, where $S \subseteq M$. Here we assume the auctioneer divides his own resources into m units and the bidders bid for some units' resources and pay after receiving the confirmation from the auctioneer.

1) Utility Function

The utility of the bidder i is defined as

$$u_i = \begin{cases} b_i(S) - p_i(S) & \text{if } i \text{ wins} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $b_i(S)$ is the revenue of the bidder i with service subset S, and $p_i(S)$ is the payment of bidder i when it wins the service subset S. The payment charged by the auctioneer to the winning bidder i can be computed as

$$p_i(S) = W\left(N/\{i\}\right) - \left(W\left(N\right) - b_i(S)\right) \tag{13}$$

where $W(N/\{i\})$ is the result of solving the Winner Determination Problem (WDP) in auction game [35, 36] using bids from all bidders except i, and the last term is the sum of the winning bids by all bidders except i. The WDP is defined as

$$W\left(N\right) = \max \sum_{i \in N} \sum_{S \subseteq M} b_i\left(S\right) x_i\left(S\right) \tag{14}$$

Subject to

$$\begin{aligned} &\sum_{S \subseteq M} x_i\left(S\right) \le 1 \quad \forall i \in N; \\ &x_i\left(k\right) + x_j\left(k\right) \le 1 \quad \forall i, j \in N, \forall k \in M; \\ &x_i\left(S\right), x_i\left(k\right) \in \{0,1\} \quad \forall i \in N, \forall S \subseteq M. \end{aligned} \tag{15}$$

where $x_i(S)$ is the indicator variable: $x_i(S) = 1$ if bidder i wins in an auction and 0 otherwise. The first constraints represent the use of XOR bids [35], to make an individual bidder's bids mutually exclusive.

2) The core of the auction

In the CSIM, the "core" is the set of allocations whose imputed payoffs are core imputations and we define the core function in the CSIM as follows, and formally motivate the use of core-selecting auctions.

$$Core(N,W) = \{u \ge 0 \mid \sum_{i \in N} u_i = W(N) \text{ and } \sum_{\substack{\forall S \subseteq M, \\ j \in S}} u_j \ge W(S)\} \tag{16}$$

An auction outcome is in core if no group of participants (including the auctioneer) are motivated to secede to settle for their own solution.

3) Effectiveness of CSIM

The proposed CSIM satisfies the following property that indicates its effectiveness.

Property: A core outcome is Pareto optimal if there is no other core outcome that can improve at least one bidder's utility without reducing any other bidder in a subset $S \subseteq N$. The property ensures that there exists no incentive for bidders to form the collusion.

Proof: Assume there exists a collusion $N' \subset N$. The bidder i that belongs to the collusion tries to earn profit $u_i^{'}$. Combining Eqs. (12) (14) (16), we have:

$$u_i^{'} = b_i(S') - p_i(S') = W(N' - \{i\}) - W(N') - 2p_i(S') \tag{17}$$

where the bidder tries to earn more profit and the sum of profit is fixed by (16), and then there is,

$$W(N' - \{i\}) - W(N') \le W(N - \{i\}) - W(N) \tag{18}$$

Based on the payment rules in (13), there is

$$p_i(S') \ge p_i(S) \tag{19}$$

Therefore, we have $u_i^{'} \le u_i$, which means that the bidder i has no profit gain by joining any collusion.

Data Trustworthiness enhanced Crowdsourcing Strategy

The process of the data trustworthiness enhanced crowdsourcing strategy (DTCS) is shown in Figure 2 and the details are described as follows, where the red lines represent the security and privacy demands, black lines represent the actions by initiators, and blue lines represent the actions by participants.

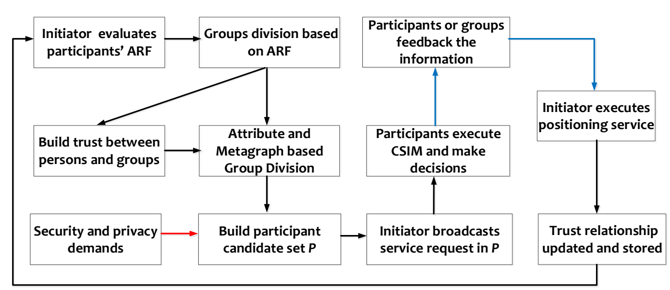


Fig. 2. The DTCS system structure.

First, crowdsourcing service initiator executes the attribute and metagraph based participant group division scheme (AMPGD) to evaluate its neighbor participants' ARF and divides all the neighbor participants into different groups based on the attribute relevancy and familiarity evaluation results.

Second, crowdsourcing service initiator builds the different possible kinds of trust relationship between persons and groups, and then selects the participant or participant group according to the security and privacy preserving demands of the crowdsourcing positioning service based on the established trust relationship.

Third, crowdsourcing service initiator broadcast the crowdsourcing positioning service request to the selected crowdsourcing service participants.

Fourth, each participant receiving the request executes the CSIM (the default behavior mode is cooperation, i.e., provide truth information), and decide which behavior it will take to respond the request.

Finally, after the service, the crowdsourcing service initiator re-evaluate and update the trust relationship with the crowdsourcing participant according to the provided information.

The details of the crowdsourcing service execution process in IoT are shown in the algorithm 1.

Algorithm 1: DTCS

```
1.  Begin
2.   The crowdsourcing service initiator evaluate neighbor participants' ARF;
3.   If the participants' ARF belong to different level then
4.       Initiator divides the neighbor participants into different groups;
5.   Else
6.       All the neighbor participants are in a same group.
7.   End if
8.   If there is more than one group then
9.     {
10.      Initiator builds the trust relationship between persons in a same group;
11.      Initiator builds the trust relationship between a person and a group;
12.      Initiator builds the trust relationship between persons in two groups;
13.      Initiator builds the trust relationship between two different groups;
14.    }
15.  Else
16.      Initiator builds the trust relationship between persons in the same group;
17.  End if
18.  If minimum (TR (p, p'))> Threshold then
19.      Put g into the participant set P;
20.  End if
21.  If TR (p, p') > Threshold then
22.      Put p' into the participant set P;
23.  End if
24.  If P is not empty, then
25.    {
26.      Initiator broadcast the crowdsourcing positioning service request in P;
27.      Initiator waits for the feedback;
28.    }
29.  End if
30.  Any participant receiving the request executes the CSIM and make a
       decision which behavior it will take.
31.  Participants or participant groups feedback the information to the initiator;
32.  Initiator converges the feedback information and executes the
       crowdsourcing positioning service;
33.  Initiator re-evaluate and update the trust relationship;
34.  End
```

V. Performance Evaluation

In this section, we developed a Java-based simulator to implement the proposed strategy DTCS and compare it with TSCM [21] and PPPCM [8] because they are the similar and latest related crowdsourcing strategies. The following performance metrics are evaluated when internal conflicting behaviour attacks and collusion attacks are present.

In the simulation tests, we evaluate three strategies in a 1000×1000 region (m^2) where 1000 participants are uniformly distributed as the crowd during a 30-min event. We assume that a certain number of participants is malicious, intending to provide disinformation. Moreover, good participant always sends correct sensing reports but an adversary does not necessarily always send false sensing reports.

The security parameters λ_1, λ_2, κ_1, κ_2, ρ_1, ρ_2 are 0.6, 0.4, 0.4, 0.6, 0.6, 0.4, which are empirical values obtained from multiple simulation experiments. λ_1 and λ_2 are the weight factors in (7) used to determine how much the direct and indirect interaction affect the $TR_{(p,g)}^{direct}$. κ_1 and κ_2 are the weight factors in (10) used to determine how much the attribute relevancy and familiarity of the participant and the trust relationship affect the participant membership degree (PMD). ρ_1 and ρ_2 are the weight factors in (10) used to determine how much the direct and indirect trust relationship between two participants affect the integrated trust relationship of them.

Because Utility rate of the crowdsourcing strategy (URCS), Disinformation ratio (DIR) and Trustworthy participant selection rate (TPSR) are three important and frequently used metrics to evaluate the feasibility and availability of the crowdsourcing strategy, they are chosen as the metrics in the performance evaluation when internal when internal conflicting behavior attacks and collusion attacks are present. These performance metrics are defined below.

➢ **URCS**: The Utility of the crowdsourcing strategy (i.e., the accuracy of decision and efficiency of the crowdsourcing strategy according to the crowdsourced information).

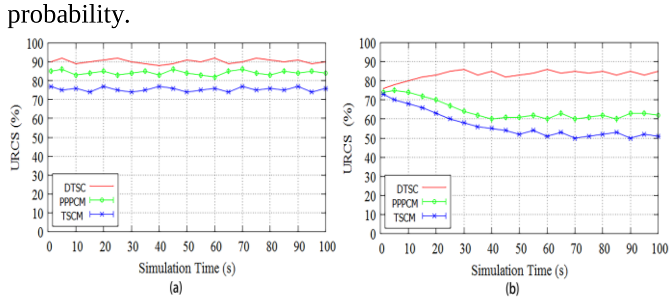
➢ **DIR**: The rate of the disinformation information to the total crowdsourced information.

➢ **TPSR**: The rate of the trustworthy cooperative crowdsourcing participants to the total number of selected crowdsourcing participants.

All experiments depicted in the following figures had been repeated at least 100 times (more for the random selection method), and the average values are taken as the final results.

1) Utility rate of the crowdsourcing strategy (URCS)

First, we investigate the utility rate of the DTCS, and compare it with those of the TSCM and PPPCM in an honest network and a hostile network when internal conflicting behavior attacks and collusion attacks are present, respectively. In the honest network, all the participants are good participants. While in the hostile network, the participants may be adversaries who give false information with a random probability.



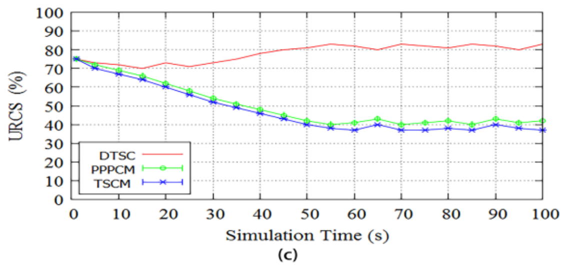


Fig. 3. Utility rate of the crowdsourcing strategies (a) in an honest network, (b) with conflicting behavior attacks, and (c) with collusion attacks.

The comparison result of the URCS of the three crowdsourcing strategies in an honest network is shown in Fig. 3 (a). The results show that in the honest network environment, all the three strategies have high decision accuracy because all the participants provide the truth information. Also, we can see that the URCS of the DTSC is higher than the other two strategies, the reason is that the DTSC divides all the participants into different groups by using the metagraph theory based on the attribute relevancy and familiarity evaluation results, therefore, more relevant and familiar participants will be selected as crowdsourcing participants that provide more accurate crowdsourced data, which efficiently enhances the accuracy of the crowdsourcing information. Moreover, the adoption of the user group division improves the efficiency of relevant and familiar participants' selection. The two advantages mentioned above make the URCS of the DTSC higher than the TSCM and PPPCM.

We also analyse the impact of the malicious attacks on the URCS of the three strategies. Comparing to the results in Fig. 3 (a), in Fig. 3 (b) and (c) where the conflicting behaviour attack and collusion attack are present, the URCS of DTSC decreases by 7% and 12%, the URCS of PPPCM decreases by 25% and 40%, and the URCS of TSCM decreases by 30% and 45%, respectively. In DTSC, the establishment of trust relations between persons makes possible the fine-grained reputation evaluation of participants and selection of trustworthy crowdsourcing participants. Meanwhile, the participant group division and the establishment of trust relations between groups effectively solve the reputation transferring and loss problem of participants during the participant movement. Furthermore, the core-selecting based incentive mechanism in DTSC provides better defending against the internal attacks than those of the PPPCM and TSCM through implementing effective rewards and punishment mechanism. The above-mentioned schemes make the URCS of DTSC the highest and slowest decreasing among the three strategies.

2) Disinformation ratio (DIR)

Next, we analyze the disinformation ratio (DIR) of the three strategies under two hostile network environments. In Fig. 4 (a) and (b), as expected, the DIR increases with the simulation rounds. It is observed that the DIR of the DTSC is the lowest among the three strategies. This is because that the integrated combination of trust relations establishment and participant group division improves the accuracy and efficiency of the participants' reputation evaluation and solves the participants' reputation transferring and loss problem, which enhances the reliability of the selected crowdsourcing participants and crowdsourced data trustworthiness and thus decreases the DIR of the DTSC. Moreover, the core-selecting based incentive mechanism proposed in DTSC motivates the selected participants to provide truthful information and decline to join any collusion attacks, which also improves the crowdsourced data trustworthiness and decreases the DIR of the DTSC. Although the other two strategies also adopt related technologies to improve the accuracy and reliability of the participants' selection and data trustworthiness, they do not consider the impact of the participants' movement on the accuracy and efficiency of the participants' reputation evaluation. Moreover, they do not take the collusion attacks into account and cannot defend against the internal collusion attacks. Therefore, their DIR is higher than that of the DTSC.

We also evaluate the DIR of the three strategies with different proportions of conflicting behavior attackers and collusion attackers, respectively. From the results shown in Fig. 4 (c) and (d), we can see that DIR is dramatically affected by the number of malicious participants and the DIR of all the three strategies increase as the proportion of malicious participants increases. However, the DIR of the DTSC is relative stable and lower than those of the PPPCM and TSCM. Neither PPPCM or TSCM can implement the more accurate reputation evaluation of participants and solve the reputation transferring and loss problem, therefore, they cannot effectively identify the mobile malicious participants and choose more relevant trustworthy participants, which makes their DIR decreases faster than the DTSC. Furthermore, neither PPPCM or TSCM can defend against the collusion attack, therefore, they will receive more false information and their DIR decreases faster than the DTSC.

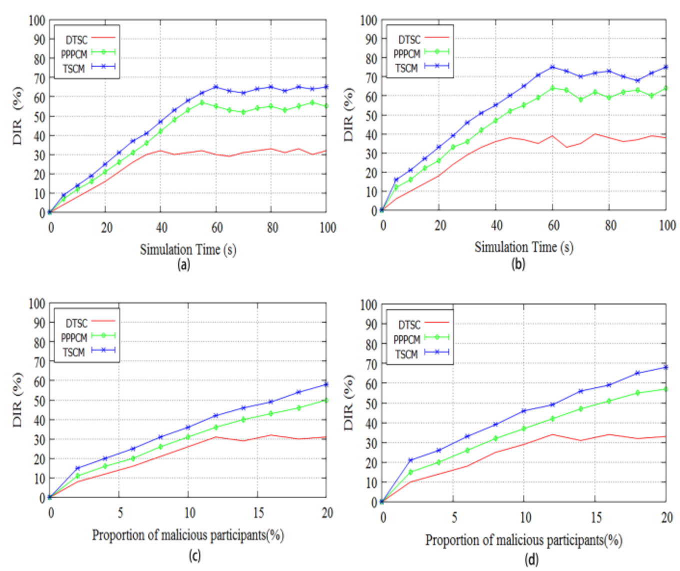


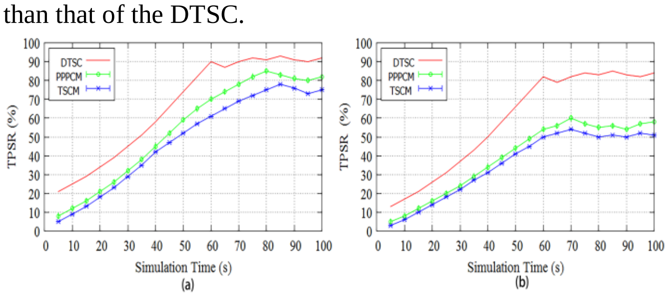
Fig. 4. Disinformation ratio of the crowdsourcing strategies (a) with conflicting behavior attacks, (b) with collusion attacks, and (c) with different proportions of conflicting behavior attackers, and (d) with different proportions of collusion attackers.

3) Trustworthy participant selection rate (TPSR)

Finally, we evaluate the trustworthy participant selection rate (TPSR) of the DTCS, and compare it with those of the TSCM and PPPCM. The comparison result of the TPSR of the three strategies in the honest network is shown in Fig. 5 (a). The results show that in the honest network environment, the TPSR of all the three strategies increases with the simulation time. In the honest network, all the participants will participate in collaboration actively and the reputation of the positive cooperative participants that provide more accurate and reliable information will increase more quickly, which enhances the participants' probability to be chosen greatly and thus improve the TPSR. In DTSC, the trust relations establishment and participant group division mechanisms make the participants' reputation evaluation more accurate and timely than that of the TSCM and PPPCM. Also, the establishment of the trust relations between participants and groups solves the participants' reputation transferring and loss problem effectively, therefore, the TPSR of the DTSC is higher than TSCM and PPPCM.

We also analyse the impact of the malicious attacks on the TPSR of the three strategies. From the results shown in Fig. 5 (b) and (c) where the conflicting behaviour attack and collusion attack are present, we can see that the TPSR of TSCM and PPPCM is affected by the malicious attacks more severely than the DTSC. Specifically, the TPSR of DTSC decreases by 8-10%, the TPSR of PPPCM decreases by 23-25%, and the TPSR of TSCM decreases by 30-33%. The reason is that neither PPPCM or TSCM can defend against the conflicting behaviour attack and collusion attack effectively and thus cannot evaluate and update the participants' reputation or identify the malicious participants accurately and timely, which makes the TPSR of TSCM and PPPCM lower than that of the DTSC.

At the same time, we evaluate the TPSR of the three strategies with different proportions of conflicting behaviour attackers and collusion attackers, respectively. In Fig. 5 (d) and (e), as expected, we see that the TPSR is dramatically affected by the number of malicious participants and the TPSR of all the three strategies decrease as the proportion of malicious participants increases. However, with the combination of the trust relations establishment, participant group division and core-selecting based incentive mechanism, DTSC can identify more malicious participants than TSCM and PPPCM, and thus can defend against the conflicting behaviour attacks and collusion attacks more effectively than the TSCM and PPPCM. Therefore, the TPSR of the PPPCM and TSCM decreases faster than that of the DTSC.



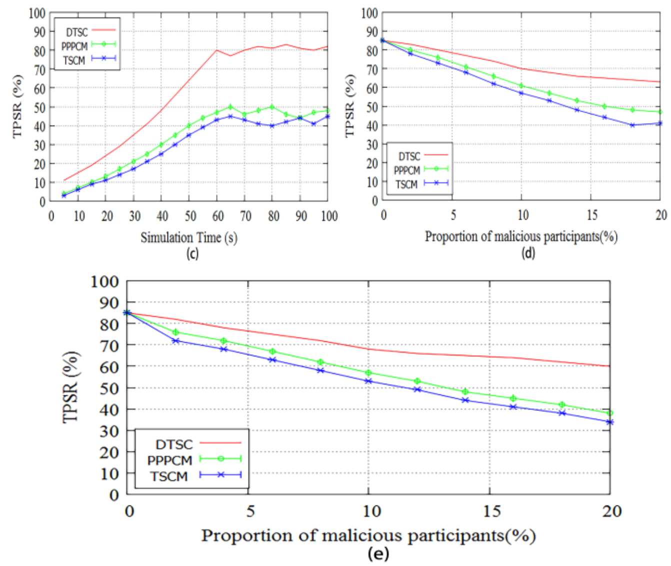


Fig. 5. Trustworthy participant selection rate (a) in an honest network, (b) with conflicting behavior attacks, (c) with collusion attacks, (d) with different proportions of conflicting behavior attackers, and (e) with different proportions of collusion attackers.

VI. Conclusions

This paper proposes an integrated strategy named DTCS to enhance data trustworthiness and defend against the internal threats for mobile crowdsourcing. The DTCS integrates four different methods including an evaluation scheme for the attribute relevancy and familiarity of participants, a trust relationship establishment method between persons and groups, a group division strategy bases on attributes and metagraph, and a core-selecting based incentive mechanism. The DTCS can effectively defend against internal conflicting behaviour attacks and collusion attacks to enhance data trustworthiness for mobile crowdsourcing. We have evaluated the performance metrics including the utility rate of the crowdsourcing strategy, disinformation ratio, and selection rate of trustworthy participant. Simulation experiments demonstrate that the DTCS improves the performance of the crowdsourcing strategy compared to the state-of-the-art including the TSCM and PPPCM. The DTCS can effectively defend against internal conflicting behaviour attacks and collusion attacks to enhance data trustworthiness for MCS. For the future work, we plan to introduce the encryption or signature based privacy preserving technology into the mobile crowdsourcing process to improve the data trustworthiness further.

References

[1] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 2013, 29(7), pp. 1645-1660.

[2] E. Borgia. The Internet of Things vision: Key features, applications and open issues. *Computer Communications*, 2014, 54, pp. 1–31.

[3] L. Chen, S. Thombre, K. Jarvinen, E. S. Lohan, A. K. Alen-Savikko, et.al. Robustness, Security and Privacy in Location-Based Services for Future IoT: A Survey. *IEEE Access*, 2017.

[4] L. Da Xu, W. He, S. Li. Internet of Things in industries: A survey. *IEEE Trans. Ind. Informat.*, 2014,10(4) pp. 2233–2243.

[5] Y. Cao, S. Yang, G. Min, X. Zhang, H. Song, Omprakash Kaiwartya and Nauman Aslam. "A Cost-Efficient Communication Framework or Battery Switch Based Electric Vehicle Charging". *IEEE Communications Magazine*,

vol. 5, no. 5, pp. 162-169, May, 2017.

[6] Y. Cao, Y. Miao, G. Min, T. Wang, Z. Zhao and H. Song. "Vehicular-Publish/Subscribe (V-P/S) Communication Enabled On-the-move EV Charging Management". *IEEE Communications Magazine*, vol. 54, no. 12, pp. 84-92, December, 2016.

[7] Y. Cao and N. Wang. "Towards Efficient Electric Vehicle Charging Using VANET-Based Information Dissemination". *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 2886-2901, April, 2017.

[8] B. Zhang, C. H. Liu, J. Lu, et.al. Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing. *Computer Networks*, 2016, 101, pp. 29-41.

[9] K. Wang, X. Qi, L. Shu, et al. Toward trustworthy crowdsourcing in the social internet of things. *IEEE Wireless Communications*, 2016, 23(5), pp. 30-36.

[10] X. Zhang, Z. Yang, W. Sun, et al. Incentives for Mobile Crowd Sensing: A Survey. *IEEE Communications Surveys & Tutorials*, 2016, 18(1), pp. 54-67.

[11] H. Ma, D. Zhao, P. Yuan. Opportunities in mobile crowd sensing. *IEEE Communications Magazine*, 2014, 52(8), pp.29-35.

[12] B. Guo, H. Chen, Z. Yu, et al. FlierMeet: a mobile crowdsensing system for cross-space public information reposting, tagging, and sharing. *IEEE Transactions on Mobile Computing*, 2015, 14(10), pp. 2020-2033.

[13] D. He, S. Chan, M. Guizani. User privacy and data trustworthiness in mobile crowd sensing. *IEEE Wireless Communications*, 2015, 22(1), pp. 28-34.

[14] H. Xiong, D. Zhang, L. Wang, et al. EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants. *ACM Transactions on Intelligent Systems and Technology*, 2015, 6(3), pp. 39.

[15] S. H. Chang, Z. R. Chen. Protecting Mobile Crowd Sensing against Sybil Attacks Using Cloud Based Trust Management System. *Mobile Information Systems*. 2016, pp. 1-10

[16] Y. Li, W. Dai, Z. Ming, et al. Privacy protection for preventing data over-collection in smart city. *IEEE Transactions on Computers*, 2016, 65(5), pp. 1339-1350.

[17] Q. Li, G. Cao, T. F. La Porta. Efficient and privacy-aware data aggregation in mobile sensing. *IEEE Transactions on dependable and secure computing*, 2014, 11(2), pp. 115-129.

[18] M. Pouryazdan, B. Kantarci, T. Soyata, et al. Anchor-Assisted and Vote-based Trustworthiness Assurance in Smart City Crowdsensing. *IEEE Access*, 2016, 4, pp. 529-541.

[19] S. Yin, O. Kaynak. Big Data for Modern Industry: Challenges and Trends Point of View. *Proceedings of the IEEE*, 2015, 103(2), pp. 143-146.

[20] J. Kepner, V. Gadepally, P. Michaleas, et al. Computing on masked data: a high performance method for improving big data veracity. *in Proc. IEEE High Performance Extreme Computing Conference (HPEC)*, 2014, pp.1-6.

[21] B. Kantarci, H. T. Mouftah. Trustworthy sensing for public safety in cloud-centric internet of things. *IEEE Internet-of-Things Journal*, 2014, 1(4), pp. 360-368.

[22] S. Li, H. Li, L. Sun. Privacy-preserving crowdsourced site survey in WiFi fingerprint-based localization. *EURASIP Journal on Wireless Communications and Networking*, 2016, 1, pp.123-132.

[23] Q. H. Cao, I. Khan, R. Farahbakhsh, et al. A Trust Model for Data Sharing in Smart Cities. in *Proc. IEEE International Conference on Communications (ICC 2016)*, 2016, pp.1-7.

[24] N. Palaghias, N. Loumis, S. Georgoulas, et al. Quantifying trust relationships based on real-world social interactions. in *Proc. IEEE International Conference on Communications (ICC 2016)*, 2016, pp. 1-7.

[25] K. L. Huang, S. S. Kanhere, W. Hu. Are you contributing trustworthy data?: the case for a reputation system in participatory sensing. in *Proc. of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, 2010, pp. 14-22.

[26] X. O. Wang, W. Cheng, P. Mohapatra, et al. Artsense: Anonymous reputation and trust in participatory sensing. in *Proc. INFOCOM*, 2013, pp. 2517-2525.

[27] B. Liu, L. Chen, X. Zhu, Y. Zhang, et.al. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. *In Proc. EDBT*, 2017, pp. 478-481.

[28] Y. Gong, L. Wei, Y. Guo, C. Zhang, and Y. Fang. Optimal Task Recommendation for Mobile Crowdsourcing with Privacy Control. *IEEE Internet-of-Things Journal*, 2017, 3(5), pp. 745-756.

[29] Y. Zhang, H. Zhang, S. Tang, et al. Designing Secure and Dependable Mobile Sensing Mechanisms With Revenue Guarantees. *IEEE Transactions on Information Forensics & Security*, 2016, 11(1), pp. 100-113.

[30] K. Yang, K. Zhang, J. Ren, X. Shen. Security and privacy in mobile crowdsourcing networks: challenges and opportunities. *IEEE Communications Magazine*, 2015, 53(8), pp.75-81.

[31] Hui Lin, Li Xu, Yi Mu, Wei Wu. A Reliable Recommendation and Privacy-Preserving Based Cross-Layer Reputation Mechanism for Mobile Cloud Computing. *Future Generation Computer Systems*, 2015, 52, pp.125–136.

[32] Hui Lin, Jia Hu, Jianfeng Ma, Li Xu, Li Yang. CRM: A New Dynamic Cross-Layer Reputation Computation Model in Wireless Networks. *The Computer Journal*. 2015, 58(4), pp. 656-667.

[33] Hui Lin, Li Xu, Xinyi Huang, Wei Wu, Yijie Huang. A Trustworthy Access Control Model for Mobile Cloud Computing Based on Reputation and Mechanism Design. *Ad Hoc Networks*, 2015, 35(C), pp. 51-64

[34] R. W. Day, S. Raghavan. Fair payments for efficient allocations in public sector combinatorial auctions. *Management Science*, 2007, 53(9), pp.1389-1406.

[35] Y. Zhu, B. Li, H. Fu, Z. Li. Core-selecting secondary spectrum auctions. *IEEE Journal on Selected Areas in Communications*, 2014,32(11), pp. 2268-2279.

[36] M. Ezhei, B. T. Ladani. Gtrust: A group based trust model. *International Journal of Information Security*, 2014, 5(2), pp. 155-169.

[37] A. Basu and R. W. Blanning, "Metagraphs and their applications", Springer, 2007, pp. 88-113.

[38] B. Kantarci, K. G. Carr, C. D. Pearsall. SONATA: Social Network Assisted Trustworthiness Assurance in Smart City Crowdsensing. *International Journal of Distributed Systems and Technologies (IJDST)*, 2016, 7(1), pp.59-78.