Anchor-Assisted and Vote-Based Trustworthiness Assurance in Smart City Crowdsensing

Maryam Pouryazdan
Burak Kantarci
Tolga Soyata
Houbing Song

Follow this and additional works at: https://researchrepository.wvu.edu/faculty_publications

Part of the Engineering Commons
Anchor-Assisted and Vote-Based Trustworthiness Assurance in Smart City Crowdsensing

MARYAM POURYAZDAN1, (Student Member, IEEE), BURAK KANTARCI1, (Senior Member, IEEE), TOLGA SOYATA2, (Member, IEEE), AND HOUBING SONG3, (Senior Member, IEEE)
1Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699, USA
2Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA
3Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV 25136, USA
Corresponding author: B. Kantarci (bkantarc@clarkson.edu)

This work was supported by the U.S. National Science Foundation through the Division of Computer and Network Systems under Grant CNS-1239423 and Grant CNS-1464273.

ABSTRACT
Smart city sensing calls for crowdsensing via mobile devices that are equipped with various built-in sensors. As incentivizing users to participate in distributed sensing is still an open research issue, the trustworthiness of crowdsensed data is expected to be a grand challenge if this cloud-inspired recruitment of sensing services is to be adopted. Recent research proposes reputation-based user recruitment models for crowdsensing; however, there is no standard way of identifying adversaries in smart city crowdsensing. This paper adopts previously proposed vote-based approaches, and presents a thorough performance study of vote-based trustworthiness with trusted entities that are basically a subset of the participating smartphone users. Those entities are called trustworthy anchors of the crowdsensing system. Thus, an anchor user is fully trustworthy and is fully capable of voting for the trustworthiness of other users, who participate in sensing of the same set of phenomena. Besides the anchors, the reputations of regular users are determined based on vote-based (distributed) reputation. We present a detailed performance study of the anchor-based trustworthiness assurance in smart city crowdsensing through simulations, and compare it with the purely vote-based trustworthiness approach without anchors, and a reputation-unaware crowdsensing approach, where user reputations are discarded. Through simulation findings, we aim at providing specifications regarding the impact of anchor and adversary populations on crowdsensing and user utilities under various environmental settings. We show that significant improvement can be achieved in terms of usefulness and trustworthiness of the crowdsensed data if the size of the anchor population is set properly.

INDEX TERMS
Crowdsensing, reputation, sensing as a service, smart cities, smart city sensing, smartphone sensing, truthfulness, trustworthiness.

I. INTRODUCTION
The advances in cellular communications, as well as consumer electronics have led smartphones to serve ubiquitous computing and distributed sensing in smart city applications [1], [2]. Figure 1 illustrates various building blocks of a smart city. While smart transportation, smart energy, smart urban services, smart water and smart homes are major application areas, smart citizens close the loop by participating in sensing, actuating and decision making processes [3]. This emerging community sensing paradigm is called mobile crowdsensing, where individuals with sensing and computing devices collectively share data and extract information to measure and map phenomena of common interest [4]. As mentioned by the forum in [5], having citizens as sensors in a smart city, makes groups of individuals collaboratively work towards the same goal with strong interaction links even though this does not always require strong social links between them. Thus, individuals form communities; collaborating communities form social networks where interaction is in the form of community-based participatory sensing [6].

Today’s smartphones are equipped with various built-in sensors that can be accessed by crowdsourcing applications for various purposes such as urban traffic monitoring [7], weather monitoring or noise level [8], public safety [9], [10] and emergency preparedness [11]. Integration of smartphones, built-in sensors, cloud computing and big data analytics into the smart city architecture accelerates the smart city services [12].

To the best of our knowledge, almost all crowdsensing applications are either voluntary-based [13] or they do not
incorporate dynamic recruitment of users [2], [14], [15]. It is worthwhile noting that there are proposals that analyze the performance of rewarding mechanisms [16]–[19], while some of them stand out with an emphasis on truthfulness [20], [21]. Besides, social behavior and social trust dimensions have also been considered to be incorporated into crowdsensing via social credits [22]. Related work also reports that several factors such as time and location affect users’ decision to participate in crowdsensing even for public safety purposes [23].

Although all smart city applications call for highly accurate sensing data, trustworthiness of the crowdsensed data is of paramount importance in particular cases as follows. Smart urban services require continuity and robustness as they include public safety services such as crowd monitoring, crowd management under extreme situations such as disaster preparedness and recovery [24], [25]. Furthermore, smart connected vehicles require trustworthiness in information sharing [26] as infrastructure-less data sharing is prone to maliciously altered information dissemination. Electric vehicles are becoming an inseparable part of smart cities, and energy-aware route selection is critical when cars run on batteries. As studied by the related work [27], energy-minimizing routes can be crowdsensed for electric vehicles where trustworthiness of crowdsensed data is highly desired.

Despite the differences among the concepts, all studies agree on the fact that users need to be awarded based on the usefulness of the data that they provide. The trade-off between rewarding users and maximizing the value of crowdsensed data is generally addressed by game theoretical or auction-based approaches. As pointed out in previous work, in the presence of adversaries, the crowdsourcer platform may get dis-informed [28]. In [9], reputation-based smartphone user recruitment methods have been proposed. According to the proposal, recruitment probability of a user is a function of its reputation/trustworthiness. Hereafter, we use trustworthiness to denote the average reputation of a user over time. We also use a crowdsensing node and a smartphone user interchangeably. As shown in several studies, malicious users may intermittently attack the crowdsourcer in order to avoid being identified as an adversary. Thus, a malicious user may keep track of its reputation based on the truthfulness of the data, aiming to build reputation up to a certain value. Once the upper threshold is reached, the user starts to send false data, and this is where its reputation starts to decrease. Given that centralized and statistical reputation-awareness can lead to degradation in the utility of the crowdsourcer, distributed and vote-based mechanisms can be considered. In [29], a distributed voting solution was proposed which adopts the vote-based Sybil detection approach in social networks [30] such that users of a particular interest group vote for the trustworthiness of a user that joins the group and transfer their vote capacities to the newly joining user. Here, an interest group denotes a group of smartphones that have at least
In this section, we review the research in the field of Smart City Crowdsensing and the challenges that arise from this concept, as well as metrics that quantify the accuracy and usefulness (i.e., utility) of the sensed data.

### A. Smart City Crowdsensing

Smart city crowdsensing has recently been an emerging topic as smartphones can provide accelerometers, gyroscopes and GPS data as a service in order to monitor crowd density and recognize human activities [24], [32]. A mobile crowdsensing architecture was presented in [33] by adopting service-oriented architecture design principles for resource optimization in smart city crowdsensing. In [2], besides defining a crowdsensing system in a smart city, the authors outline the requirements for a crowdsensing system design as minimum disruption on user devices, fast processing and prompt feedback, content sharing, extensibility, security and complete data management workflow. In [34], the authors present a prototype crowdsensing application for public transport in a smart city. The crowdsensed data is aggregated by using the Extensible Messaging and Presence Protocol (XMPP), an open source communication protocol based on publish/subscribe model, which enables communicating sensed tasks via Extensible Markup Language (XML). Another transportation application in smart city crowdsensing was presented in [35] where cyclists provide vibration data in a participatory manner in order for the cloud platform to determine road conditions in an urban area. In [36], crowdsensing motion pictures via wearable cameras in a smart city were studied. In the corresponding study, the authors particularly focused on optimal transcoding of the videos on wearable cameras, and to this end, an adaptive transcoding algorithm was proposed to improve latency, energy consumption in wearables and peak signal to noise ratio in crowdsensed video transmission. In [37], a service oriented framework has been presented to assess the quality of data crowdsensed through citizens during an emergency. Authors in [38]–[40] elaborate on an application where multiple users perform face recognition using multiple smartphones and multiple cloud servers.

### B. Incentivizing Users

As previously mentioned, incentivizing users is a crucial challenge for effective crowdsensing. The sensor data can be considered to be aggregated at a cloud platform, and the cloud platform serves as a middleware between crowdsourcer and the smartphone users who participate in crowdsensing. Indeed, it is viable to delegate these tasks to a cloud platform as the cloud platform provides flexibility in resource allocation, service provisioning and improved security [41]. To the best of our knowledge, the authors in [42] proposed
user-centric and platform-centric incentives for the first time for a generic application of crowdsourcing using smartphones.

C. PLATFORM UTILITY OF CROWDSENSING

Platform-centric incentives aim to maximize the benefit (i.e., utility) of the crowdsourcer (i.e., cloud platform) and the crowd (i.e., smartphone users) by recruiting a set of smartphone users from the crowd and assigning them sensing tasks along with sensing schedules. User-centric incentives rely on users’ own sensing plans and pre-announced sensing costs, and the platform runs a reverse auction to select a subset of users based on their sensing costs (i.e., bids in the auction) and the value of the sensing tasks in the terrain.

1) AUCTION THEORETIC CROWDSENSING

MSensing auction was proposed to maximize both platform and user utility which was explained and analyzed in detail in [20]. Let \( \omega(b_i) \) denote a binary function that evaluates to TRUE (i.e., 1) if the user \( i \) wins an auction by bidding \( b_i \). In this auction, the critical value, \( b^c_i \), is defined as the highest winning bid among all users, thereby implying \( b_i < b^c_i \) during the auction. The authors use the monotone property of \( b_i \) to allow only the user that bids truthfully to win by using Eq. 1

\[
\omega(b'_i) = \begin{cases} 
1, & (b'_i \leq b_i \land \omega(b_i) = 1) \lor (b'_i = b^c_i) \\
0, & \text{else}
\end{cases}
\] (1)

where \( b'_i \) quantifies a bid, \( b'_i \leq b_i \), that could also allow the user to win. Furthermore, we know that \( b_i > b^c_i \) would not guarantee a winning bid for user \( i \). Therefore, the additional margin between \( b_i \) and \( b^c_i \) allows the crowdsourcer to increase its platform utility, i.e., the cost of crowdsourcing.

2) GAME THEORETIC CROWDSENSING

In [16], the authors presented the concept of discretized crowdsensing where users participate in crowdsensing event in discrete timeslots. A game theoretic approach has been formulated based on the Perfect Bayesian Equilibrium which aims to maximize crowdsourcer platform utility and define an effective strategy to adjust user utility. In [43], the authors propose an approximation mechanism to address the trade-off between platform utility and user utility where users compete for the announced tasks based on the availability of their time slots.

D. TRUSTWORTHINESS OF THE CROWDSENSED DATA

Trustworthiness of the crowdsensed data is directly related to the accuracy of smartphone sensors and reputation of their users who have been recruited to sense the tasks of the corresponding data [44]. The fundamental assumption among the schemes above is that the auction participants are trustworthy and the only malicious activity of the users in the crowd can be joining the auction with higher bids to increase their income. Related work proposes to adopt reputation systems to enhance the trustworthiness of the crowdsensed data. Users that report false sensing data to the crowdsensing platform can either be due to a sensor malfunction or intentional disinformation.

1) REPUTATION BASED USER RECRUITMENT

Regardless of the cause, the crowdsensing system recruits users that have a higher reputation by rewarding the ones that provide useful data [9]. In [45], the authors present the concept of S, U and E reports to assess user credibility. S reports denote sensing activity whereas E reports denote reports from review requests and U reports are the reports from authorities. In [29], the authors adopt a Sybil detection technique that was proposed for social networks [46] and tailor it to the detection of adversaries in a mobile crowdsensing system.

2) RECRUITING USERS USING SOCIAL NETWORKS

The proposed scheme in [29] called Social Network-Aided Trustworthiness Assurance (SONATA) formulates vote-based trustworthiness of crowdsensed data through incentivized users. Users are linked in the form of a social network community as long as they have common sensing tasks. Thus, a community is represented as a fully connected graph, and interconnection of these communities forms a social network. The trustworthiness of a user is voted by the existing members of the community upon joining of the user to the corresponding community. It is worthwhile mentioning that joining a community denotes having a new common sensing task with other community members.

3) VOTE BASED CROWDSENSING

Each member of the community has a vote capacity and a reputation as proposed in [46]. \( \mathcal{N}_t \) is defined as the trustworthiness of user \( j \) at time \( t \), having a vote capacity \( \omega_j \) and an actual vote \( \chi_t^j \) for user \( i \), the trustworthiness of newly joining user (user \( i \) is calculated as the weighted sum of votes of the community members normalized by the total reputable vote capacity as formulated in Eq. 2.

\[
\mathcal{N}_t = \frac{\sum_{j \in \mathcal{C}_i} \omega_j \chi_t^j \mathcal{N}_t}{\sum_{j \in \mathcal{C}_i} \omega_j \mathcal{N}_t}
\] (2)

The advantage of maintaining vote-based trustworthiness in a crowdsensing system is that historical data does not need to be stored; reputation calculation is run in a distributed manner, allowing the platform only to deal with running a reverse auction procedure in order to select the smartphone users (i.e., crowd) and assess the usefulness of the data they provide. Before we proceed with the details of the vote-based crowdsensing with anchor nodes, it is worthwhile to present the notation used in the paper. Table 1 summarizes the most important details about our notation.
TABLE 1. The notation used in the formulation.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{A}$</td>
<td>Anchor nodes in the terrain</td>
</tr>
<tr>
<td>$\mathcal{U}$</td>
<td>Non-anchor nodes in the terrain</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>Set of participating nodes; $\mathcal{P} = \mathcal{A} \cup \mathcal{U}$</td>
</tr>
<tr>
<td>$W$</td>
<td>Set of winners in the auction; $W \subseteq \mathcal{P}$</td>
</tr>
<tr>
<td>$\tilde{W}$</td>
<td>Set of the social network of mobile devices</td>
</tr>
<tr>
<td>$</td>
<td>\tilde{W}</td>
</tr>
<tr>
<td>$(i)$</td>
<td>User $i$. Also termed “node $i$” or “mobile device $i$”</td>
</tr>
</tbody>
</table>

III. ANCHORS IN VOTE-BASED CROWDSENSING

As reported in [29], the success of vote-based trustworthiness in crowdsensing systems is closely related to the i) ratio of malicious users in the crowd, ii) sensing task load on the crowdsensing system, and iii) initial reputations of the users. In [3], possible improvement to SONATA was proposed by assigning “anchor” roles to some of the nodes. An anchor node has 100% trustworthiness during a pre-determined period of time regardless of the accuracy of its sensor readings. Thus, wrong sensor data reported by an anchor node is presumably due to the malfunctioning of its built-in sensors and it is not considered as malicious activity. Furthermore, anchor nodes have full vote capacity when they are needed to vote for a newly joining node in their communities. On the other hand, it is worthwhile noting that an anchor node cannot evaluate the trustworthiness of any other node better than a non-anchor node is able to do.

A. TRUSTWORTHINESS MANAGEMENT

Figure 2 depicts a minimalist overview of vote-based mobile crowdsensing with anchor nodes in a smart city application. Each node is equipped with a mobile interface for crowdsensing. The interface can access a node’s built-in sensors and determine the most recent readings of a node’s connections. Each node casts a vote for a newly joining node and publishes its sensor data through this interface, which is basically a mobile application.

$\mathcal{R}_{\text{inst}}(i)(t)$ is the instantaneous trustworthiness of node $i$ at time $t$ and is formulated in Eq. 3 below:

$$
\mathcal{R}_{\text{inst}}(i)(t) = \begin{cases} 
\delta \cdot \mathcal{R}_{\text{inst}}(i)(t) + (1 - \delta) \cdot \mathcal{R}_{\text{i}}(t^-), & \{i\} \in \mathcal{U} \\
\mathcal{R}_{\text{inst}}(i)(t^-) = \mathcal{R}_{\text{i}}(t^-) = 1, & \{i\} \in \mathcal{A}
\end{cases}
$$

$\mathcal{R}_{\text{inst}}(i)(t)$ is computed as the weighted sum of the votes of all nodes in its community ($\sum \omega_j \cdot \mathcal{R}_{j}(t^-)$), normalized by total vote capacity of the voting nodes ($\sum \omega_j \cdot \mathcal{R}_{j}(t^-)$). The weight of the vote of a node $j$ is the product of its trustworthiness and vote capacity ($\omega_j \cdot \mathcal{R}_{j}(t^-)$).

1) DYNAMIC CHANGES IN TRUSTWORTHINESS

As a node’s communities change dynamically due to dynamically arriving sensing task requests, its trustworthiness needs to be adjusted dynamically. As formulated in Eq. 4, overall trustworthiness of a node at the end of time period $t$ ($\mathcal{R}_{\text{i}}(t)$), is a weighted sum of its instantaneous trustworthiness at $t$ ($\mathcal{R}_{\text{inst}}(i)(t)$), and its overall trustworthiness immediately before time $t$ ($\mathcal{R}_{\text{i}}(t^-)$). It is worthwhile noting that if the node is assigned an anchor role (node $i \in \mathcal{A}$), its trustworthiness is always 1.0 as seen in the equations.

$\mathcal{R}_{\text{i}}(t)$ is computed as the weighted sum of the votes of all nodes in its community ($\sum \omega_j \cdot \mathcal{R}_{j}(t^-)$), normalized by total vote capacity of the voting nodes ($\sum \omega_j \cdot \mathcal{R}_{j}(t^-)$). The weight of the vote of a node $j$ is the product of its trustworthiness and vote capacity ($\omega_j \cdot \mathcal{R}_{j}(t^-)$).

2) DYNAMIC CHANGES IN VOTE CAPACITY

Vote capacity of a node also changes dynamically as the node inherits the vote capacity of the nodes with which it has just been linked. As seen in Eq. 5, instantaneous vote capacity of newly joining user is calculated as the average vote capacity...
of its connections.

\[
\omega_i^{\text{inst}}(t) = \begin{cases} 
\sum_{j : |T_i \cap T_j| \neq \emptyset} \frac{\omega_j}{|W|} , & [i] \in U \\
1, & [i] \in A
\end{cases}
\] (5)

The actual vote capacity of the node is calculated based on running average method as shown in Eq. 6. Note the 100% vote capacity implied for an anchor node in Eq. 5.

\[
\omega_i(t) = \begin{cases} 
\gamma \cdot \omega_i^{\text{inst}}(t) + (1 - \gamma) \cdot \omega_i(t^-), & [i] \in U \\
\omega_i^{\text{inst}}(t) = \omega_i^{\text{inst}}(t^-) = 1, & [i] \in A
\end{cases}
\] (6)

where \( \gamma \) is the transition coefficient for trustworthiness, which plays a similar role to \( \delta \) to quantify the transition speed.

3) INTRODUCING ANCHOR NODES

As seen in Fig. 2, \( U_2 \) and \( U_4 \) are assigned anchor roles in a crowdsensing application. At \( (t = t_1) \), \( U_0 \) joins the community that consists of the following nodes: \( \{ U_2, U_4, U_6, U_7 \} \). At \( (t = t_1) \), \( U_0 \) has an initial trustworthiness value, \( \Re_0(t_0) \), and it is combined with the voted trustworthiness value to obtain its trustworthiness at \( t = t_1 \). Voted trustworthiness of the node is contributed by the votes of \( U_4, U_5, U_6, \) and \( U_7 \). At \( (t = t_2) \), the new sensing task is scheduled within a sensing range covering \( \{ U_0, U_1, U_2, U_3, U_4 \} \). Given that \( U_0 \) did not have any common tasks with the nodes in this new community, \( U_1, U_2, U_3, U_4 \) vote the trustworthiness of \( U_0 \). Finally, weighted sum of \( U_0 \)'s trustworthiness at \( (t = t_1) \) and its voted trustworthiness are stored as its trustworthiness at \( (t = t_2) \).

B. USER RECRUITMENT VIA REVERSE AUCTION

As mentioned before, users can be recruited by user-centric or platform-centric incentives [20]. As we aim at running the data acquisition in a distributed manner, we adopt the reputation-aware user-centric reverse auction in [9] and [47]. This reverse auction consists of the following two steps: 1) winner selection and 2) rewarding. Reverse auction-based recruitment of users was initially proposed in [42] for the first time. Sensing tasks arrive in bursts and the nodes that fall in the range of a sensing task participate in the first step of the auction by reporting its sensing cost, particularly its bid \( (b_i) \) in the auction.

1) WINNER SELECTION

From the standpoint of the platform and the end user, each task has a predefined value \( (\vartheta_i) \). Therefore, \( W \) and \( T_W \) being the set of selected nodes out of the participates set and the set of tasks sensed by the users in \( W \), respectively; the value of the set \( W \) is the sum of the values of the tasks in \( T_W \) as formulated in Eq. 7.

\[
\vartheta(W) = \sum_{i \in T_W} \vartheta_i
\] (7)
Winner selection out of the participants set involves a sequential search process that aims at selecting the participants that lead to positive marginal contribution to platform utility. By positive marginal contribution, we denote the difference between the marginal value of the corresponding node over the selected nodes set and its sensing cost. Equation 8 formulates the marginal value of node \( i \) (\( \vartheta_i(W) \)) as the difference between the values of the set \( W \), before and after recruiting node \( i \).

\[
\vartheta_i(W) = \vartheta(W) - \vartheta(W \cup \{i\})
\]

\[
\Rightarrow \vartheta^R_i(W) = \vartheta^R_i(W \cup \{i\}) - \vartheta^R_i(W)
\] (8)

The left-hand side of Eq. 8 formulates the marginal value whereas the right-hand side presents calculation of reputation-based marginal value of node \( i \) over the set \( W \). As seen in Eq. 8, reputation-based marginal value is the difference between reputation-based values of the set before and after recruiting node \( i \). Here, one can ask what reputable value of a set (\( \vartheta_i^R(W) \)) stands for. As formulated in Eq. 9, total value of the tasks sensed by the nodes in the set, \( W \), is scaled by the average trustworthiness of the nodes in the set where \( \Gamma_i \) denotes the set of users that sense task \( r \) in a participatory manner.

\[
\vartheta^R_i(W) = \sum_{r \in \mathcal{R}_W} \sum_{j \in \Gamma_i} \left( \frac{\vartheta_j \cdot \vartheta^R_i}{|\Gamma_i|} \right)
\] (9)

As for the winner selection, the platform adds nodes until the difference between the marginal value of node \( i \) over the set, \( W \), and its sensing cost is non-positive, i.e., \( \vartheta_i(W) - b_i > 0 \Rightarrow \vartheta^R_i(W) - b_i/\vartheta^R_i > 0 \). Similar to the previous work [9], [47], sensing costs of the nodes are scaled by their trustworthiness.

2) REWARDING OF THE WINNERS

Any recruited node is guaranteed to be rewarded no less than its sensing cost. The rewarding mechanism in [9] is adopted in this work which was built on the reputation-unaware rewarding in [42]. For each winner node, \( w \), the platform runs a secondary search to find a maximum hypothetical sensing cost/bid that would still make node \( w \) preferred over every other non-rewarded node, \( w_v \). At the beginning, each user is assumed to be rewarded zero. For each winner, user \( w \), a temporary set \( \mathcal{P}' \) is constructed which is equal to the set of participants excluding user \( w \) as shown in Eq. 10.

\[
\mathcal{P}' = \mathcal{P} \setminus \{w\}
\] (10)

At this point, a temporary winners set \( \Delta \) is constructed out of \( \mathcal{P} \). The rewarding module searches for maximum hypothetical value for the corresponding user’s bid which would enable winning of user \( w \) instead of user \( w_v \in \mathcal{P}' \). User \( w_v \) stands for a user in \( \mathcal{P} \) whose reputable marginal value is greater than its quantified bid. Therefore, the rewarding algorithm keeps adding those users to set \( \Delta \) starting from the user that maximizes the platform utility, i.e., the difference between the value and quantified bid as shown in Eq. 11.

\[
w_v = \arg\max_{v \in \mathcal{P} \setminus \Delta} \left( \vartheta_v(\Delta) - \frac{b_v}{\vartheta_v} \right), \quad \{v\} \in \mathcal{U}
\]

\[
w_v = \arg\max_{v \in \mathcal{P} \setminus \Delta} \left( \vartheta_v(\Delta) - b_v \right), \quad \{v\} \in \mathcal{A}
\] (11)

When a new user, user \( w_v \), is added to the temporary user set \( \Delta \), the maximum reward that can be made to user \( w \) given the bid value of user \( w_v \) is searched. Equation 12 formulates this process.

\[
\rho_w = \max \left( \vartheta_w, \min \left( \vartheta_v(\Delta) - (\vartheta_w(\Delta) - b_w), \vartheta_w(\Delta) \right) \right)
\] (12)

Once the temporary set \( \Delta \) is constructed, the nodes are sorted with respect to their marginal contribution as proposed in [20] and [42], however reputation-based values are used here to ensure trustworthiness as shown in Eq. 13 and user \( w \) is rewarded the maximum of final value of \( \rho_w \) and its marginal value over the set \( \Delta \).

\[
\left( \vartheta^R_w - \frac{b_w}{\vartheta^R_w} \right) > \left( \vartheta^R_{w+1} - \frac{b_{w+1}}{\vartheta^R_{w+1}} \right)
\] (13)

The algorithm uses the quantified bids only for constructing the temporary set \( \Delta \) as seen in Eq. 11, whereas the actual bids are used for payment calculation as seen in Eq. 12. Furthermore, the actual values of the set \( \Delta \) is used instead of its reputation-based value. The idea behind this distinction is to avoid cuts in non-malicious user rewards due to possibly sensing the same tasks with other malicious users. In [9], the rewarding module is proposed to operate in two modes, and the mode presented above is called the non-aggressive mode. In case, the rewarding module always uses the quantified bids and reputation-based values in Eqs. 11 and 12, cuts should be expected in user rewards for the sake of increased platform utility, and the corresponding mode is called the aggressive rewarding mode.

In order to address the trade-off between aggressive and non-aggressive rewarding modes, an adaptive rewarding mode was proposed in [9] which is also adopted and adapted to the anchor-based crowdsourcing here. Thus, as formulated in Eq. 14, if the trustworthiness of user \( w \) has increased since its last recruitment and the user is a non-anchor user, its reputation-based marginal value over the set \( W \) is defined as the difference between the reputation-based values of the set \( W \) after and before the selection of user \( w \). On the other hand, if the user is assigned an anchor role or its trustworthiness has been degraded, the actual marginal value of the set \( W \) is considered to be the reputation-based value of the set \( W \). This approach is used in both winner selection and rewarding phases of the user recruitment.

\[
\vartheta^R_w(W) = \begin{cases} 
\vartheta^R(W \cup \{w\}) - \vartheta^R(W), & \vartheta_w \leq \vartheta^R_w \wedge \{w\} \in \mathcal{U} \\
\vartheta^R(W), & \vartheta_w \geq \vartheta^R_w \vee \{w\} \in \mathcal{A}
\end{cases}
\] (14)

where \( \vartheta^R_w \) and \( \vartheta_w \) are the trustworthiness values of node \( w \) before and after the recruitment, respectively.
In addition, it is worthwhile noting that recruitment of anchor nodes is exactly the same as that of non-anchor nodes. However, reputation-based contributions and/or bids of anchor users are identical to their reputation-unaware formulations as an anchor node is presumed to be 100% trustworthy until the end of the monitoring period.

IV. PERFORMANCE EVALUATION

We evaluate anchor-based trustworthiness assurance in smart city crowdsensing through simulations, and develop performance specifications.

A. SIMULATION SETTINGS

We implement an object-oriented simulation environment in Java and conduct simulations based on the settings reported in Table 2. The sensing tasks arrive at a terrain that covers a 1000 m × 1000 m geographic area and 1000 smartphone users are uniformly distributed as the nodes of the crowd. As we aim at developing performance specifications, we set the ratio of malicious smartphone users (i.e., malicious nodes) to 3% and 5% of the entire crowd population, covering two different scenarios. Furthermore, we set the anchor percentage in the crowd (i.e., anchor nodes) to 3% and 5% of the entire crowd population; therefore, these combinations allow us to investigate four total scenarios, as shown in Table 3.

### TABLE 2. Simulation settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrain size</td>
<td>1000 m × 1000 m</td>
</tr>
<tr>
<td>Sensing Range</td>
<td>30 m</td>
</tr>
<tr>
<td>Number of users</td>
<td>1000</td>
</tr>
<tr>
<td>Task arrival rate (λ)</td>
<td>(20, 40, 60, 80) / min</td>
</tr>
<tr>
<td>Initial reputation</td>
<td>0.7</td>
</tr>
<tr>
<td>Malicious user probability</td>
<td>{0.03, 0.05}</td>
</tr>
<tr>
<td>Anchor node percentage</td>
<td>{0.03, 0.05}</td>
</tr>
<tr>
<td>Task value</td>
<td>{1, 2, 3, 4, 5}</td>
</tr>
<tr>
<td>Bid value</td>
<td>{1, 2, ..., 10}</td>
</tr>
<tr>
<td>Simulation duration (τduration)</td>
<td>30 min.</td>
</tr>
<tr>
<td>Sensing inaccuracy probability</td>
<td>[0.02, 0.03]</td>
</tr>
<tr>
<td>Adversary recognition probability</td>
<td>0.2</td>
</tr>
<tr>
<td>δ trust. transition coefficient</td>
<td>0.5</td>
</tr>
</tbody>
</table>

### TABLE 3. Simulated scenarios based on malicious user probability and anchor node percentage.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Malicious user probability</th>
<th>Anchor node percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario I</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Scenario II</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>Scenario III</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Scenario IV</td>
<td>5%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 3 presents the four scenarios that are considered in performance evaluation of the proposed distributed crowdsensing system. The crowdsensing platform assumes that the users are trustworthy by 70% at the beginning of the monitoring event. The event duration is set to 30 minutes where sensing tasks arrive following a Poisson distribution with a mean [20, 40, 60, 80] tasks/min.

As illustrated in Fig. 2, sensor readings are published through a sensor publishing layer where historical readings are available to the nodes that are in the same community. The value of a vote can be either 1 or −1. The fundamental assumption in our study is that altered sensor readings behave like the Sybil attacks in social networks [46]. Hence one fundamental assumption in this study is that there is a 20% chance that the recommendation of a truthful node casts its votes as −1 for a malicious node. The value of a task is uniformly distributed in the range {1, 2, 3, 4, 5}, whereas the sensing cost of a mobile device ranges in {1, 2, ..., 10}. Furthermore, built-in sensors can provide accurate readings with a probability between 97% and 98% [48]. We consider three performance metrics as follows:

1) **Crowdsensing platform utility (U_p)** represents the benefit of the platform as a function of the value of total sensing tasks and the total rewards made to the mobile users [9], [42].

2) **Manipulation probability (M)** can be defined as the ratio of the sensing tasks for which at least one malicious node has been rewarded throughout the monitoring event. However, we hereby manipulation denote the total rewards made to the malicious users during the monitoring event.

3) **Average user utility per unit time (U_user)** is formulated as the benefit of a smartphone user as a function of the rewards and the sensing cost of the user, averaged by the total number of winners in each auction. To be clearer, we formulate this metric in Eq. 15 below:

\[
U_{user} = \left( \sum_{i} \left( \frac{\left( \sum_{t} p_{i}^{\tau} - \sum_{t} c_{i}^{\tau} \right)}{|W_{\tau}|} \right) \right) / \tau_{duration}
\]

(15)

where \( p_{i}^{\tau} \) and \( c_{i}^{\tau} \) are the rewards received by mobile device user \( i \) at the \( \tau \)th auction and the sensing cost of smartphone user \( i \) at the \( \tau \)th auction, respectively. In the equation, \( W_{\tau} \) stands for the number of winners in the \( \tau \)th period which lasts as long as \( \tau_{duration} \).

As reported by related work, it is reasonable to assume that malicious users intermittently manipulate their sensor readings instead of continuously altering them. As smartphones are equipped with improved computing capability compared to previous generation mobile devices, an adversary keeps track of its historical sensing data, and based on that it retrieves some statistical trustworthiness value for itself. The statistical trustworthiness translates into the percentage of unaltered sensor readings of the smartphone user among all readings that have been transmitted to the platform so far. Thus, a malicious smartphone user aims at building...
Therefore, it can be concluded that under the remaining utility cannot be maximized under either of the load levels. 

is as large as 7% of the entire crowd population, platform both cases, when the anchor and/or malicious user population arrival rates, i.e., 60 tasks/min and 80 tasks/min. As seen in the entire crowd population under two different sensing task probability in the crowd and the anchor user percentage in study the relation between platform utility, malicious user The first set of results are presented in Fig. 3 where we simulated platform utility (FIGURE 3. Simulated platform utility ($U_p$) for different anchor node percentage and malicious user probabilities; two different task arrival rates are evaluated.

B. SIMULATION RESULTS

The first set of results are presented in Fig. 3 where we study the relation between platform utility, malicious user probability in the crowd and the anchor user percentage in the entire crowd population under two different sensing task arrival rates, i.e., 60 tasks/min and 80 tasks/min. As seen in both cases, when the anchor and/or malicious user population is as large as 7% of the entire crowd population, platform utility cannot be maximized under either of the load levels. Therefore, it can be concluded that under the remaining settings, having 3–5% of the crowd population as anchor nodes is feasible under heavier sensing task loads (i.e., 80 tasks/min) whereas for lighter loads, in order to ensure high platform utility, it is not feasible to have less anchor nodes than the size of the malicious user population. Therefore, we run our simulations under the four scenarios presented in Table 3.

In Fig. 4, we test the presented framework under the four scenarios each of which corresponds to a different anchor-adversary percentage combination in the crowd population. As seen in Fig. 4a and Fig. 4d, if the anchor population is limited to 3% of the entire population it can be clearly beneficial under heavier sensing task arrival rates and if and only if the malicious user population is less than 5% of the entire crowd population. The reason behind this phenomenon is as follows. Having malicious population larger than the anchor population will not help increase platform utility via anchors. Instead, simple voting will work better as the platform will be intending to recruit the anchors due to their high trustworthiness regardless of the accuracy of their votes and recommendations. The impact of unconsidered regular behavioral assessment capacity of anchor users become more severe under lightly arriving sensing tasks. On the other hand, when the anchor and adversary population are equal to each other (Fig. 4a), having trusted nodes help improve platform utility when vote-based trustworthiness is adopted by the crowdsensing incentives.

When anchor population is bigger, it is expected to improve platform utility as observed in Fig. 4b. However, too many anchors will dominate selection process under heavier sensing task arrival rates when the anchor population is larger than the malicious user population as seen in the figure. This suppression can be eliminated by reducing the size of the anchor population. As seen in Fig. 4c, anchor and adversary populations are set to 5% of the entire crowd population, and the benefit of anchor-assisted and vote-based crowdsensing is clearly observed under all load levels.

In the next step, we evaluated the impact of anchors in the rewards made to the malicious users while they are aiming at disinformation at the crowdsensing platform. As previous work presents, vote-based trustworthiness assurance (i.e., SONATA [29]) significantly reduces disinformation probability when compared to the conventional MSensing-based crowdsensing where reputation awareness is left to future work [42]. Only scenario-I and Scenario-III lead to negligible disinformation under SONATA as seen in Fig. 5. Deploying anchor nodes in the crowdsensing application can eliminate rewarding malicious users; hence it overcomes disinformation at the crowdsensing platform.

Fig. 6 illustrates the average user utility in the crowdsensing system under study. As reputation-based system scales the sensing costs (i.e., auction bids) by the reputation of the corresponding devices while compensating the mobile device users [9], [47], cuts are expected in the payments to
FIGURE 4. Simulated platform utility (\(U_p\)) vs. sensing task arrival rate for four different scenarios enumerated in Table 3, each representing a different (malicious user probability, anchor node percentage) tuple: (a) Scenario I [3%, 3%], (b) Scenario II [3%, 5%], (c) Scenario III [5%, 5%], (d) Scenario IV [5%, 3%].

FIGURE 5. Average reward made to malicious users vs. sensing task arrival rate for two different scenarios enumerated in Table 3, each representing a different (malicious user probability, anchor node percentage) tuple: (a) Scenario I [3%, 3%], (b) Scenario III [5%, 5%].

the mobile device users. It is worthwhile mentioning that every winner is compensated no less than its sensing cost. Besides, having said that improvement in platform utility of reputation-unaware crowdsensing is promising, reduction in user rewards is acceptable as the value of information provided to the end users is not counted as a part of the reward in the crowdsensing business model. Moreover, deployment of anchors does not lead to degradation in user utility regardless of the size of the anchor and/or adversary population.
FIGURE 6. Average user utility ($U_{\text{user}}$) vs. sensing task arrival rate for four different scenarios enumerated in Table 3, each representing a different (malicious user probability, anchor node percentage) tuple: (a) Scenario I (3%, 3%), (b) Scenario II (3%, 5%), (c) Scenario III (5%, 5%), (d) Scenario IV (5%, 3%).

V. CONCLUSION

Sensing-as-a-Service ($S^2$aaS) has appeared as a promising solution with the advent of cloud computing and Internet of Things (IoT) paradigms accelerating smart city applications. Crowdsensing systems enable forming connected communities of mobile devices that provide their built-in sensors as a service for sensing several phenomena. Trustworthiness has been pointed out as an important challenge in crowdsensing systems since adversaries may lead to disinformation at the service requester site through manipulation of sensor readings. In this paper, we have studied vote-based crowdsensing frameworks focusing on anchor-based trustworthiness assurance by adopting a previously proposed crowdsensing scheme which uses only votes of the participating smartphone users [29]. We have incorporated vote-based social trustworthiness assessments via a dynamic social network structure. Furthermore, we have assigned anchor roles to a subset of participating users denoting their 100% trustworthiness remaining unchanged during a pre-determined period of time. According to the anchor-assisted and vote-based trustworthiness assurance in a crowdsensing system, trustworthiness of a smartphone user (i.e., crowdsensing node) is obtained through votes of the users in the same community whereas anchor nodes have 100% reputation and full vote capacity throughout the monitored event.

We have studied design specifications for fully distributed and vote-based smart city crowdsensing via simulations. Through simulations, we have shown that deployment of anchor nodes in smart city crowdsensing improves crowd-sourcer utility by up to 20% when compared to purely vote-based trustworthiness. We have also shown that by the deployment of anchor nodes in a smart city crowdsensing scenario, crowd (user) utility is not reduced when compared to purely vote-based trustworthiness assurance without anchors. Moreover, disinformation probability at the crowd-sourcer platform can be eliminated if anchor-based trustworthiness assurance is integrated with the decentralized component of the trustworthiness assurance module of smart city crowdsensing. Based on our research findings, we conclude that the size of the anchor population has to be selected properly with respect to the detected adversary population; hence we advocate the emergency of adaptive anchoring solutions for smartphone users in smart city crowdsensing applications.

REFERENCES


MARYAM POURYAZDAN (S’16) received the B.Sc. degree in computer engineering from the University of Kerman-Iran, in 2011, and the M.Sc. degree in computer engineering from UTM-Malaysia, in 2013, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY. Her research is on trustworthiness assurance in crowd sensing via mobile social networks.

BURAK KANTARCI (S’05–M’09–SM’12) received the M.Sc. and Ph.D. degrees in computer engineering from Istanbul Technical University, in 2005 and 2009, respectively. He was a Post-Doctoral Researcher with the School of Electrical Engineering and Computer Science, University of Ottawa (UOttawa). During his Ph.D. study, he studied as a Visiting Scholar at UOttawa, where he completed the major content of his thesis. He is an Assistant Professor with the Department of Electrical and Computer Engineering, Clarkson University, and the Founding Director of the Next-Generation Communications and Computing Networks (NEXTCON) Research Lab. He has co-authored over eighty papers in established journals and conferences, and contributed to 11 book chapters. He is a member of ACM. He received the Siemens Excellence Award for his studies in optical burst switching in 2005. He is the Co-Editor of the book entitled Communication Infrastructures for Cloud Computing. He has served as the Technical Program Co-Chair of seven international conferences/symposia/workshops. He is an Editor of the IEEE Communications Surveys and Tutorials. He also serves as the Secretary of the IEEE ComSoc Communication Systems Integration and Modelling Technical Committee.

TOLGA SOYATA (M’08) received the B.S. degree in electrical and communications engineering from Istanbul Technical University, in 1988, the M.S. degree in electrical and computer engineering from Johns Hopkins University, in 1992, and the Ph.D. degree in electrical and computer engineering from the University of Rochester, in 1999. He joined the ECE Department, University of Rochester, in 2008, where he is currently an Assistant Professor-Research. He manages the GPU Research Center and GPU Teaching Center programs for the University of Rochester, and Xilinx University Program and MOSIS Educational Program for the UR ECE Department. He teaches courses in VLSI ASIC design, GPU parallel programming, FPGA-based advanced digital design, and advanced medical data analytics (data science practicum). His current research interests include cyber physical systems, digital health, and GPU-based high-performance computing.

HOUBING SONG (M’12–SM’14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in 2012, and the M.S. degree in civil engineering from the University of Texas, El Paso, TX, in 2006. In 2012, he joined the Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV, where he is currently an Assistant Professor and the Founding Director of the West Virginia Center of Excellence for Cyber-Physical Systems, sponsored by the West Virginia Higher Education Policy Commission, and the Security and Optimization for Networked Globe Laboratory. He was an Engineering Research Associate with the Texas A&M Transportation Institute in 2007. His research has been supported by the West Virginia Higher Education Policy Commission. His research interests lie in the areas of cyber-physical systems, Internet of Things, big data analytics, communications and networking, connected vehicle, and smart and connected health.

Dr. Song is a member of ACM. He is an Associate Editor for several international journals, including the IEEE ACCESS, KSII Transactions on Internet and Information Systems, and SpringerPlus, and a Guest Editor of several special issues. He was the General Chair of five international workshops, including the first IEEE International Workshop on Security and Privacy for Internet of Things and Cyber-Physical Systems in London, U.K., the first IEEE International Workshop on Big Data Analytics for Smart and Connected Health in Washington, DC, and the first/second/third IEEE ICCC International Workshop on Internet of Things (2013/2014/2015) in Xi’an/Shanghai/Shenzhen, China. He also served as the Technical Program Committee Chair of the fourth IEEE International Workshop on Cloud Computing Systems, Networks, and Applications in San Diego, USA. He has served on the Technical Program Committee for numerous international conferences, including ICC, GLOBECOM, INFOCOM, and WCNC. He has published more than 80 academic papers in peer-reviewed international journals and conferences.