© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Mo, Wen, Anfeng Liu, Neal N. Xiong, and Houbing Song. “MPS: A Truth Discovery Service Scheme by Using History Data to Maximize Profit for Mobile Crowd Sensing.” IEEE Transactions on Services Computing, 2024, 1–14. https://doi.org/10.1109/TSC.2024.3433541.

https://doi.org/10.1109/TSC.2024.3433541

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it’s important to you. Thank you.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

MPS: A Truth Discovery Service Scheme by Using History Data to Maximize Profit for Mobile Crowd Sensing

Wen Mo, Anfeng Liu*, Neal N. Xiong*, *Senior Member, IEEE*, Houbing Song, *Fellow, IEEE*

***Abstract*—Mobile Crowd Sensing (MCS) has emerged as a novel paradigm in massive data collection, which leverages many individual mobile devices (called workers) to collect data. MCS platform utilizes the collected data to construct various services for service requesters, thus obtaining profit based on the data values contributed by workers. However, untrustworthy data would greatly reduce the data value, leading to a decline in platform profit, so it is crucial for the platform to recruit high-trust workers and collect truthful data, thereby providing high-quality service and obtaining high profit. To address this problem, we propose a Maximize Profit Scheme, called MPS, for MCS platforms, which consider that the data value declines as data trust decreases and discounts over time. MPS scheme is the first work that systematically addresses the impact of untruthful data on the platform profit, which is not well addressed in previous research. First, we utilize historical data of trusted workers as truthful data to identify the truth of data, which is a low-cost method. Then, a trust-discounting and time-discounting value model is proposed, which is more practical than previous methods. Based on the proposed value model, we propose a novel worker recruitment strategy combined with a trust-related and time-dependent reward threshold, which prioritizes workers with high trust and low latency, thereby promoting the data value of workers and maximizing the platform's profit. By comparing the MPS with existing schemes, the experimental results show that our MPS can achieve better performance in terms of total profit.**

***Index Terms*—Mobile Crowd Sensing, maximize profit, truth discovery services, data truth discovery, trust and time-discounting.**

I. Introduction

MOBILE Crowd Sensing (MCS) recruits ubiquitous mobile devices and leverages their spatial and temporal coverage to collect and process data, performing complex data collection tasks and enabling a variety of novel services or applications that are provided to users [1], [2], [3]. A MCS model usually consists of three parts: the platform, the service requesters, i.e. users, and mobile devices, i.e. workers [4], [5], [6]. The platform is usually located in the cloud and is responsible for publishing tasks and recruiting mobile workers to collect sensory data [7]. Generally, smartphones, mobile vehicles, laptops or wearables equipped with sensing devices act as workers, capable of collecting and sensing data [4], [5]. Based on the data collected, the platform constructs the desired service or application for users, and users pay fee to platform for purchasing the service [4], [8], [9]. A wide variety of services and applications has emerged in recent years [10], including AirVisual for air quality detection [11], Waze for navigation [12], NoiseTube for noise level detection [13], Sensorly [14] for network coverage mapping, Geograph [15] for street view collection, etc.

Maximizing profit of the platform is one of the key issues in MCS [4], [16], [17]. In MCS, the platform recruit workers to collect data [4]. As workers collect data, they consume resources such as time, bandwidth [18] and device battery [19]. Therefore, the platform gives rewards to compensate for the loss of workers, thereby motivating workers to actively participate in data collection [20], [21]. If the service fee paid by the user is higher than the total budget of the platform, the platform can obtain positive profit, otherwise the platform will subside or even die. The key factor for the platform to attract users to purchase services is the high quality of the services. The premise of high-quality service is that the service should be authentic. Otherwise, even if the service has less latency or higher coverage, false service is still worthless to users, and the service purchase fees paid by users to the platform are very minimal. Therefore, it is necessary to identify the truth of various services constructed by the platform, that is, *truth discovery service*. Since the services are constructed based on data collected by workers, the key to truth discovery service is to identify the authenticity of the data, i.e., *data truth discovery*, thus enabling the platform to accurately collect high-quality data from a large amount of data.

To improve data quality, some researchers have proposed many incentive mechanisms, which intend to incentivize workers to report higher quality data by increasing rewards [21], [22], but it is difficult to identify the authenticity of the data. The premise of high-quality data is that the data should be truthful [10], [18], [23]. However, there are some untrustworthy workers or even malicious workers in the actual network. Malicious workers are unwilling to spend the cost to collect truth data, but report false data to cheat the reward, and even submit malicious data to attack the platform. As shown in Fig. 1, false data submitted by malicious data will lead to the services have low credibility, low value or even invalid

Manuscript created xxx. This work was supported in part by National Natural Science Foundation of China under Grant 62072475. (Corresponding author: Anfeng Liu, Neal N. Xiong).

W. Mo and A. Liu are with the School of Computer Science and Engineering, Central South University, Changsha 410083 China. E-mails: mowen_@csu.edu.cn; afengliu@mail.csu.edu.cn.

Neal N. Xiong is with the Department of Computer Science and Mathematics, Sul Ross State University, Alpine, TX 79830 USA, E-mail: xiongnaixue@gmail.com; neal.xiong@sulross.edu.

Houbing Song is with the Department of Information Systems, University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250 USA, e-mail: h.song@ieee.org.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

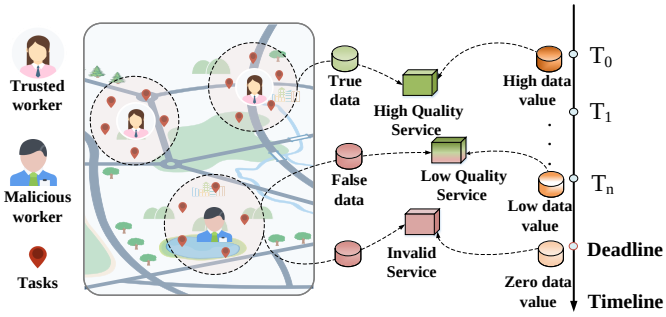


Fig. 1. The issue of malicious data and time impacting service quality.

[18]. Therefore, to ensure the healthy development of the platform, one is to identify the truth of the data [18], [23], [24], thus enabling the platform to recruit high-trust workers. The other is to accurately evaluate the data value, thus enabling the platform to collect high-value data to maximize profit [9]. However, these problems have not been well resolved in previous studies.

For data truth discovery issues, most of the existing research obtain the estimated true value in the following ways [18], [24], [25]. For the same task, the platform assigns n workers to collect data [23], [25], and takes the average value [23] or weighted sum [25] or median [23] of the n data as the true value. The main drawback is that if there are too many malicious workers, the estimated true value is still malicious data, allowing malicious workers to fool or control the platform at will. An effective way to verify the truth of the data submitted by workers is to compare the data submitted by workers with the real data [24], [26]. But this type of method needs to obtain the real value of the data, which is challenging. In addition, even if real data can be obtained, such as sending dedicated devices like drones, the cost of the MCS system is too high to fail [24], [27]. Therefore, how to conduct low-cost truth discovery is a challenging issue.

Another problem is how to calculate the value of data [9], [28]. There are some ways to measure the data value assuming the system can recognize the quality of the data. But in the actual network, it is very difficult to judge whether the data is truthful [18], [21]. So it is even more difficult to identify the multi-dimensional quality of the data, which leads to the method of calculating data value by data quality is hard to implement in practice. Many researchers have realized that many factors affect the data value, such as data submission time. In Fig. 1, the earlier the data is collected, the higher its value. As time passes, the data value is constantly attenuated. Zheng et al. [9] design a participant selection mechanism with time-discounting values. However, this method does not take into account the negative impact of false data on the data value. The property of trust-discounting value introduces new challenges in data value assessment.

In most studies, it is believed that the collected data can contribute positive value to the services contruction [4]. However, in an actual MCS, malicious data with low trust will sharply drop the platform's profit. More seriously, if the malicious workers reach a certain large proportion, the profit may be negative. Therefore, when calculating the value of data, it is necessary to consider the negative effect caused by malicious data. In addition, data value is related to time and trust, among which the data trust is the most influential factor. How to accurately measure the value of data is an urgent problem to be solved.

To tackle the above challenges, we first propose a low-cost data truth discovery method to identify the trust of the data, which is the most influential factor in the data value, thus overcoming the shortcoming that MCS cannot really select the credible data in previous studies. Then a Maximize Profit Scheme (MPS) is proposed, which considers the impact of time and trust on the value of data, truly reflecting the platform's profit. In summary, the contribution of our work are as follows:

- We consider that the data value declines as trust decreases and discounts as time goes by. As far as we know, our scheme is the first work that systematically addresses the impact of untruthful data on the profit of the MCS system. We propose a trust-discounting and time-discounting value model, which is more practical than previous methods.
- We use historical data that trusted workers submit after the task deadline as truthful data to verify the data submitted by unknown workers for the same task, which can effectively identify the truth of the data and the trust of workers, thereby preventing the platform from collecting untrustworthy data. Since data value of historical data itself is low, the platform only needs to pay rewards for lower historical data, thus reducing the trust verification cost of the platform. We give the maximum proportion of identifying workers when the platform's profit reaches the maximum.
- Considering the time-discounting and trust-discounting properties of workers, a novel worker recruitment algorithm is proposed, which introduces the trust-bid ratio and efficiency of workers to prioritize workers with high trust and low latency. We also propose a reward threshold optimization algorithm that combines trust and time to calculate the rewards received by workers, which promotes workers to provide high-value data, thereby maximizing platform profit.
- The experimental results of this paper prove that the proposed MPS can not only effectively identify the credibility of workers in a wide range, but also increase the platform revenue at a low cost. By comparing proposed MPS with several existing solutions, we show that our design can achieve better performance in terms of the total profit. Compared with the TDM scheme, the MPS increases the total platform revenue by 31.37%-81.11%, and compared with the Random scheme, MPS increases the total revenue by 10.31%-74.81%.

The rest of this paper is organized as follows. The related works are reviewed in Section II. The system model and problem statement are presented in Section III. Section IV introduces our novel MPS which includes data truth discovery service, trusted and timely workers selection scheme. Then, Section V provides the performance analysis. Finally, the

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

conclusion and future work are given in Section VI.

II. Related Work

A. Workers Selection and Incentive Scheme

Workers recruitment is one of the most important research goals in the MCS [29], [30], [31], [32]. There are different worker selection strategies according to different research goals [33], [34], [35]. In some studies, the platform requires the number of workers selected to reach a specified number [4]. The most commonly used method is the game method, which is mainly divided into user-centric and platform-centric methods. In the user-centric methods, workers submit their bids to the platform, and then the platform selects game winners based on the workers' bids [27]. If workers report high prices while other workers' bid are low, workers with low prices are more likely to be selected by the platform, but platform's revenue is relatively small [27]. The other is the platform-centric methods [27], [36]. The platform gives a price for workers to report data. If the price is relatively high, many workers are willing to participate in data collection. It is easy for the platform to receive enough data [27] but the cost is relatively high. If the price drops to a certain level, the workers receive very little compensation and are no longer willing to collect data. Therefore, there is a game process between the platform and workers, and finally the game reaches a balance.

The platform not only needs a certain number of workers, but also requires workers to provide high-quality data [37], [38]. The quality of data has different definitions in different studies [39], [40], [41]. Tham etc. al [39] used the Quality of Contributed Service (QCS) as the evaluation index. QCS is defined as the data contribution to the services constructed by the platform. If the data plays a large role in the platform's services, then its QCS is a large value [39]. However, in many studies, QCS is a relatively abstract concept and difficult to calculate accurately in practice [39]. Song et al. [40] uses Quality-of-Data (QoD) to measure the quality of data. They divided the sensing region into many grids. If a worker collects data for a grid with no data or very little data, QoD of the worker will be larger, thus covering more monitoring areas. The work of Ren et al. [16] is similar. The platform pays lower rewards for workers in the city center to save costs and increases the reward for workers in remote areas to motivate them to actively collect data, so as to increase the coverage of the data and services quality. Some researchers have proposed some strategies to select data based on the data value. Zheng et al. [9] proposed that the value of data is time-discounting value, i.e., the value of each mobile worker's contribution decreases as time goes by. But their study assumed that the data submitted by workers was truthful. However, data value primarily depends on the truth of the data. Only the data is credible or worker is trustworthy, data is valuable. In our paper, we propose a data value evaluation model which combines trust and time to recruit high-credibility and low-latency data, thus maximizing platform profit.

B. Data Truth Discovery in MCS

Truth discovery is also becoming more and more important with the development of MCS recently [18], [24], [25]. Most studies believe that the data reported by workers is true. However, with the deepening of research, researchers find that MCS is different from previous studies. In a general platform, the devices are owned by the platform itself and the perceived data is controlled by the platform, so the data obtained is credible. However, in MCS, mobile devices usually use a participatory perception approach to collect data [4], [18]. Some dishonest workers may submit false data to save data-aware costs and defraud rewards [24], [25].

Researchers have proposed many data truth discovery approaches. The most commonly used methods are as follows. (1) Average method: taking the average value of n data as the truth data. (2) Median method: taking the median of n data as the truth data. (3) Weighted average method [25]. Sun et al. [25] believed that n data obey normal distribution. Data closer to the center value is closer to the truth data, and is assigned a larger weight, while the data that is far from the normal center is assigned a smaller weight. Then the estimated truth data is obtained after weighting. (4) Vote method [23]. Experts vote on n data, and the data with the most votes is the "truth data" [23]. The above methods need to hire n workers to collect data, resulting in the cost of the system increasing exponentially. Second, these methods can only obtains estimated truth data. Moreover, these methods are vulnerable to attacks.

Some researchers have utilized emerging technologies such as drones for data truth discovery [24], [29], [35]. In the research of Huang et al. [24], the platform dispatches drones and workers to collect data at same data collection points. The data collected by drones is truthful data, then the data collected by the workers is compared with data of the drone. If the data is consistent, the data reported by workers is trusted, thus increasing the trust of workers, and vice versa [24], [35]. Through this process, some highly trusted workers can be obtained. The truth data reported by trusted workers can also be used as a standard to verify the data of other workers with unknown trust. Thus, more workers' trust can be obtained. However, this method is only suitable for scenarios where drones can collect data, which limits its application scope and spends high cost. Therefore, finding a universally applicable and lower-cost truth discovery method is an important issue to be solved in this paper.

III. System Model and Problem Statement

A. System Model

We consider a typical mobile crowdsensing system, as shown in Fig. 2, which consists of a service provider (usually the crowdsensing platform) service requester (users), and mobile devices (workers). First, users request services such as environmental monitoring or noise pollution monitoring in Region of Interest (RoI), which is divided into m discrete Points of Interest (PoIs). Then platform recruits a group of workers to collect sensing data of PoIs, and then provides services to requesters based on the sensing data. The major notations used in the paper are listed in Table I.

Definition 1 (Task): The platform aims to collect data for m PoIs, with data collection from one PoI as a task. Thus, the

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

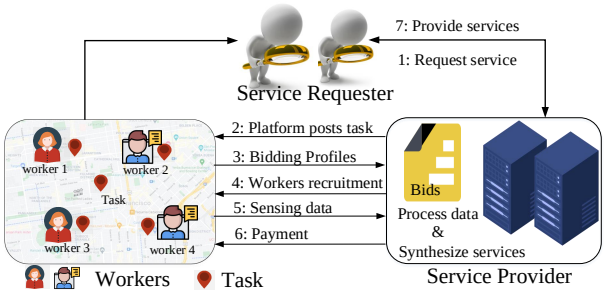


Fig. 2. Worker recruitment workflow in the Mobile Crowd Sensing.

platform issues m PoI-related sensing tasks [9], [42], [43]. The set of sensing tasks is denoted as $\mathcal{M} = \{1, 2, \ldots, m\}$, $r \in \mathcal{M}$. We consider time-sensitive tasks, all of which have a deadline T_e. $T_{e,r}$ is the deadline of the task r. Assuming the start time of the task is T_s, then the data is valid which is submitted within the range $[T_s, T_e]$. Without Loss of Generality, we assume $T_s = 0$ and $T_{e,r} = T$, $\forall r \in \mathcal{M}$.

Definition 2 (Workers): There are n workers in RoI, denoted by $\mathcal{N} = \{w_1, w_2, \ldots, w_m\}$, $w_i \in \mathcal{N}$. w_i can accomplish a set of tasks, denoted by $\mathcal{M}_i = \{r_1, r_2, \ldots, r_i\}$. $|\mathcal{M}_i| \in [r_{min}, r_{max}]$. Each worker collects a data when performing a task. Before the platform recruits workers, each worker i submits its bid $\theta_i = (a_i, c_i, \mathcal{M}_i)$ to the platform [30], [43]. a_i is the arrival time of worker i, which is in the range $[0, T_{max}]$. The time T for a round of data collection process is divided into l stages, denoted as $\{1, 2, ..., \lfloor \log_2 T \rfloor + 1\}$ (introduced in section IV-D) [9], [40]. c_i is the reported cost that the worker performs a task in $\mathcal{M}_i$. The cost c_i does not exceed the range $[c_{min}, c_{max}]$. Considering the individual rationality of workers, c_i is the minimum reward that w_i receives from the platform. We assume that the time it takes each worker to collect data can be ignored since it is much smaller than that of the worker arriving at the PoI in magnitudes [9]. Hence, the decay in data value comes from worker arrival time.

Definition 3 (Data Value): We use v_i^r to represent the data value contributied by worker i of executing task r. The platform has a price for each task, which is the maximum value of the data, denoted by v_{max}. We consider time-sensitive services and the platform expects to collect data as early as possible to get a higher profit and provide the better services to the requester. However, there is a delay between the arrival time of workers and the start time of tasks, leading to the delay of the collected data. As time passes, the data value would decrease. Besides time, the data value would be affected by the trust of the workers collecting the data. The higher the trust of the workers, the higher the quality of the collected data, and thus, the higher the data value.

Definition 4 (Marginal Value): We denote that worker i can perform a set of tasks $\mathcal{M}_i$. The data value $V_i(\mathcal{S})$ contributed by worker i to the platform is the sum of the data values v_i^r of all tasks in $\mathcal{M}_i$. However, due to the potential overlap between tasks that workers can complete, one worker's collected data for a specific task has already been submitted by an existing recruited worker, resulting in a small marginal contribution, which we define as the marginal value. In such cases, the data value participating in the overlapping task is 0. As a result, the total data value provided by a worker would decrease. Assuming that $\mathcal{S}$ represents the set of workers recruited by the platform, and i is a newly recruited worker by the platform, then the marginal value of worker i is defined as follows:

$$V_i(\mathcal{S}) = V(\mathcal{S} \cup i) - V(\mathcal{S}). \tag{1}$$

The marginal value is the sum of the data value of each valid PoIs. Valid covered PoIs are obtained by removing the already covered points from the set of PoIs covered by worker i. Assuming that the set of PoIs covered by the recruited workers is represented as $\mathcal{M}_{\mathcal{S}} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{i^+}\}, i^+ \in \mathcal{S}$, and the set of PoIs covered by the newly arrived worker i is represented as $\mathcal{M}_i$, then the set of valid PoIs, after removing the overlapping points is given by $\mathcal{M}_i = \{\mathcal{M}_i \backslash \{x | x \in \mathcal{M}_{\mathcal{S}} \cap \mathcal{M}_i\}$. Therefore, the marginal value of worker i is calculated as follows:

$$V_i = \sum_{r \in \mathcal{M}_i} v_i^r, \mathcal{M}_i = \{\mathcal{M}_i \backslash \{x | x \in \mathcal{M}_{\mathcal{S}} \cap \mathcal{M}_i\}. \tag{2}$$

TABLE I
DESCRIPTION OF MAJOR NOTATIONS

Notations	Description
$\mathcal{M}$, m	task set, number of tasks
$\mathcal{N}$, n	worker set, number of workers
$\mathcal{M}_i$, $\mathcal{M}_{\mathcal{S}}$	task set completed by worker i, task set completed by the recruited workers set $\mathcal{S}$
a_i, b_i, c_i	arrival time, departure time, reported cost of worker i
T, t_i	deadline of each task, data submission time of worker i
$\mathcal{B}$, v_{max}	budget of the platform, maximum value of the task
v_i^r	value of data collected by worker i for task r
$v_i(t, \mathcal{T})$	trust-time function of data value
V_i	marginal value of worker i
$\mathcal{S}$	set of workers recruited by platform
p_i	reward of worker i for submitting timely data
$\mathcal{C}_{\mathcal{T}}$, $\mathcal{C}_{total}$	total inspection cost, total cost
$\mathcal{Q}_p$	platform profit
α, β	trust decay factor, time decay factor
χ, ϖ	trust recognition rate, error threshold
λ, φ	parameter of adjusting the trust decay magnitude, weight decay factor
$\mathcal{V}_i^r$, $\mathcal{H}_j^r$	data collected by worker i for task r, historical data collected by trusted worker j for task r
BWs	set of trusted workers
$\mathcal{T}_{\mathcal{D}}^i$, $\mathcal{T}_{\mathcal{A}}^i$	direct trust and comprehensive trust value of worker i
$\mathbb{S}^{\mathcal{T}}$, $\mathbb{S}_{\mathcal{A}}$	set of workers with known trust value, set of workers whose trust value is greater than the threshold
$\mathbb{S}^{\kappa}$, $\mathcal{S}^r$	set of workers arriving in stage κ, set of workers that complete task r
η_{κ}	reward factor of stage κ
$\overline{\mathcal{T}_{\kappa}}$	average trust value of workers in set $\mathbb{S}^{\kappa}$
ph_i	reward of w_i for submitting historical data

B. Theat Model

In our model, the threats to MCS stem from low-quality data attacks reported by malicious workers, which include Malicious Data Attacks (MDA) and False Data Attacks (FDA) [28], [30]. In FDA, workers are unwilling to invest time or effort into collecting data but submit false data at a low cost to gain rewards from the platform. In MDA, an attacker controls workers to report malicious data, aiming to manipulate the system or degrade the quality of services.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

C. Problem Statement

We denote the set of selected workers as $\mathcal{S}$. The profit $\mathcal{Q}_p$ of the platform is the sum of the marginal data value of all workers recruited, as in Eq. (3).

$$\mathcal{Q}_p = \sum_{i \in \mathcal{S}} V_i = \sum_{i \in \mathcal{S}} \sum_{r \in \mathcal{M}_i} v_i^r. \tag{3}$$

We comprehensively considers the positive effect of worker trust on data value and the reverse effect of data submission time on data value. That is, the data value increases with the increase of worker $i's$ trust $\mathcal{T}_i$, and decreases with the delay of data submission time t_i. We give the specific time-trust function $v_i(t, \mathcal{T})$ of data value in section IV-C. In the following sections, v_i^r and $v_i(t, \mathcal{T})$ are used to represent data value, and we will no longer distinguish between them.

The platform has a total budget $\mathcal{B}$ which is used to reward workers and verifying the truth of the data. The former is called the reward cost and the latter is called the inspection cost $\mathcal{C}_\mathcal{T}$. The total reward cost of the platform is $\sum_{i \in \mathcal{S}} p_i$. $\sum_{i \in \mathcal{S}} p_i + \mathcal{C}_\mathcal{T} \leq \mathcal{B}$. If the total cost of the platform reaches the budget, the platform stops recruiting workers. The platform expects to maximize the profit, i.e., maximize the total data value provided by the recruited workers, under the budget constraint. The research objective is expressed as Eq. (4).

$$\begin{aligned} \max \mathcal{Q}_p &= max \sum_{i \in \mathcal{S}} \sum_{r \in \mathcal{M}_i} v_i^r \\ s.t. &\sum_{i \in \mathcal{S}} p_i + \mathcal{C}_\mathcal{T} \leq \mathcal{B}. \end{aligned} \tag{4}$$

IV. Our Proposed MPS

A. Research Motivation

The data value decays as the trust of workers decreases. For a worker's trust value $\mathcal{T}_i$, we use a numeric value in the range of $[0, 1]$ to denote it, from very untrustworthy 0, neutral 0.5 to very trustworthy 1 [44]. The lower the worker's trust level, the higher the probability that the submitted data is not trustworthy, and therefore the lower the data quality and the lower the data value used for platform building services. When $\mathcal{T}_i = 1$, the data value reaches the maximum value, $v_i = v_{max}$. When $\mathcal{T}_i$ decreases from 1 to 0, the data value also gradually decreases. When $\mathcal{T}_i = 0$, it means that the data is completely false and we consider $v_i^r = 0$.

The data value decays over time. For time-sensitive applications, the earlier the data is submitted, the higher the data value. Similar to Ref. [9], when a worker w_i moves to RoI at time a_i and submits data at time t, the contributed data value to the platform equals v_i multiplying a time discounting factor β^{-t}, i.e., $v_i(t) = v_i \times \beta^{-t}$, $\beta > 1$. We assum that workers can submit data once they are selected, so t is the arrival time of workers, i.e., $t = a_i$. When $t = 0, v_i(t) = v_{max}$. As t increases from 0 to deadline T, the data value gradually decines. When $t = T, v_i(t) = v_{min}$. When $t > T$, the data is invalid and the data value decreases rapidly to 0.

The trust recognition rate χ of the system is defined as the ratio of the number of workers whose trust have been recognized to the total number of workers. Initially, as the trust recognition rate increases, the platform can select highly trusted workers, so the profit continues to increase, as does the cost and the number of workers recruited. However, since the budget of platform is limited, when the recognition rate continues to increase, the inspection cost becomes too high, fewer workers can be recruited, so and profit begins to decline. In the Fig. 3, the platform profit first increases and then decreases as the trust recognition rate increases. There is an optimal trust identification rate χ_{op} that allows the platform to select as many highly trusted workers as possible under a limited cost to maximize the platform profit.

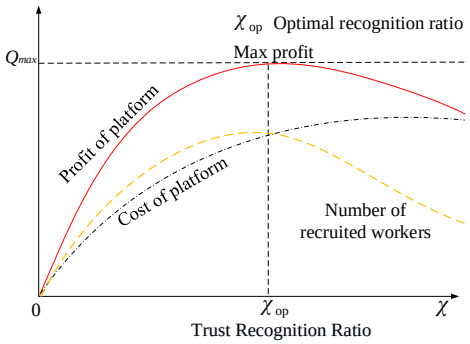


Fig. 3. Platform profit and trust recognition ratio.

In our proposed MPS, we first propose a low-cost trust identification method in section IV-B, which utilizes historical data as the truthful data, thereby obtaining ground truth data of tasks at low cost. By comparing workers' reported data with ground truth data, a trust evaluation method is proposed to calculate workers' trust degree, thereby obtaining highly trustworthy workers. Then in section IV-C, we establish the mathematical relationship between data value and worker's trust, time to accurately calculate the data value of workers. In section IV-E, we propose a worker recruitment strategy to recruit high-value workers by computing the trust-bid ratio and the value-cost ratio of workers. On the one hand, MPS can reduce the cost of truth discovery, and on the other hand, MPS can increase the data value by recruiting highly reliable workers, thereby maximizing the profits of the platform.

B. Truth Discovery Service Based on Data Truth Discovery

This paper adopts a data truth discovery method to identifying the truth of servics, which is a low-cost method to obtain the trust of workers. Data truth discovery can identify trusted data, allowing the platform to collect highly credible data to construct high-quality services. On the other hand, it can also identify false services constructed with malicious data, thereby discarding untrustworthy services. Since each task has a deadline T, if the worker arrives after the deadline of the task, the worker's data has become invalid for the platform and the data value is very low. We call the data perceived by the worker after the deadline as *historical data*. Here, we utilize historical data to conduct trust detection.

Since trusted workers have a high probability of submitting data with a small difference from the true value, the data collected by trusted workers can be considered truthful [2], [24], [28]. When the platform recruits unknown workers to

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

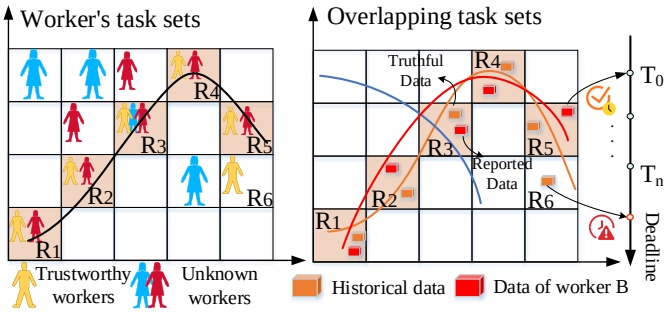


Fig. 4. Truth discovery using historical data.

perform tasks, from the workers' bids, platform knows that there are trusted workers and unknown workers who can perform the same tasks. But trusted workers arrive after the deadline, the platform still recruits these workers to submit historical data of all tasks it can accomplish. After receiving historical data, the platform compares it with the data reported by workers to identify the authenticity of the data of the same task submitted by unknown workers. Historical data is not used to constitute services, so the platform only needs to give lower rewards to obtain evidence of trust verification, thereby reducing the cost of trust evaluation. There are a very small number of high-trust workers such as Unmanned Aerial Vehicles (UAVs) known to the platform before truth discovery, which is called Baseline Workers (BWs) [22], [28], [35]. BWs are highly trustworthy and the data collected by BWs is used as ground truth data.

As shown in Fig. 4, each grid represents a task. A worker's movement trajectory passing through the grid indicates that the worker can perform this task. Worker A is trustworthy and worker B is unknown. The set of tasks that worker A can perform is $\mathcal{M}_a = \{r_1, r_2, r_3, r_4, r_5, r_6\}$, and the set of tasks that worker B can perform is $\mathcal{M}_b = \{r_1, r_2, r_3, r_4, r_5\}$. From Fig. 4, the overlapping task set of worker A and B is $\mathcal{M}_b$. Mobile worker B submit the data of task $r_1 - r_5$, and the platform evaluates the authenticity of the data and computes the trust of worker B using the historical data of worker A. Assuming the set of submitted data by an unknown worker w_i is denoted as $\mathcal{V}_i = \{\mathcal{V}_i^1, \mathcal{V}_i^2, \ldots, \mathcal{V}_i^k\}$. The bid of trusted worker w_j shows that the task set overlaps with the task set of w_i but w_j arrives later than deadline. Then w_j is recruited to submit the historical data $\mathcal{H}_j = \{\mathcal{H}_j^1, \ldots, \mathcal{H}_j^k, \ldots, \mathcal{H}_j^r\}$ to the platform. Based on the difference between trusted data and reported data, the direct trust degree $\mathcal{T}_\mathcal{D}^i$ of w_i after completing the kth subtask is calculated using Eq. (5):

$$\mathcal{T}_\mathcal{D}^{i,k} = \begin{cases} e^{-\frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi}}, & \text{if } |\mathcal{V}_i^k - \mathcal{H}_i^k| \leq \varpi. \\ e^{-\lambda \times \frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi}}, & \text{if } |\mathcal{V}_i^k - \mathcal{H}_i^k| > \varpi. \end{cases} \tag{5}$$

Where ϖ is the the threshold of the difference between real-time data and historical data and λ is a constant, which is used to adjust the trust decay magnitude. If $|\mathcal{V}_i^k - \mathcal{H}_i^k| \leq \varpi$, the greater the difference, the greater the decrease in trust; when the difference exceeds the threshold, we consider the data deviation is very large, and the trust degree drops rapidly.

The comprehensive trust degree of w_i after the latest r is $\mathcal{T}_\mathcal{A}^i = \{\mathcal{T}_\mathcal{D}^{i,1}, \mathcal{T}_\mathcal{D}^{i,2}, \ldots, \mathcal{T}_\mathcal{D}^{i,r}\}$, where $\mathcal{T}_\mathcal{D}^{i,r}$ is the direct trust

Algorithm 1 Trust recognition algorithm for workers

Input: φ, h, ϖ, λ,
BWs: set of baseline workers, $\mathcal{M}_B$: set of tasks of BWs,
$\mathcal{S}^\mathcal{T}$: the set of workers to be verifited,
$\mathcal{V}_i$: the set of data submitted by w_i;
Output: $\mathcal{T}_\mathcal{A}$: the set of workers' trust values of $\mathcal{S}^\mathcal{T}$;

$\mathcal{T}_\mathcal{A} \leftarrow \emptyset$;
for worker $w_i \in \mathbb{S}^\mathcal{T}$ **do**
  for data $\mathcal{V}_i^k$ submitted by w_i for r_k **do**
    if the baseline workers BW_i participate in the task r_k and submit the data $\mathcal{H}_i^k$ for task r_k **then**
      Platform compare $\mathcal{V}_i^k$ and $\mathcal{H}_i^k$;
      Using (5) to compute the direct trust $\mathcal{T}_\mathcal{D}^i$;
    else
      Trusted workers w_j have also performed task r_k but did not submit data in time;
      Platform requests w_j to submit historical data $\mathcal{H}_i^k$ and gives w_j rewards;
      Using (5) to compute the direct trust $\mathcal{T}_\mathcal{D}^i$;
    end if
    Using (6) to calculate comprehensive trust $\mathcal{T}_\mathcal{A}^i$;
    $\mathcal{T}_\mathcal{A} \leftarrow \mathcal{T}_\mathcal{A} \cup \{\mathcal{T}_\mathcal{A}^i\}$;
  end for
end for
return $\mathcal{T}_\mathcal{A}$;

obtained from task r closest to the current time. The closer to the current time, the more weight is given to direct trust value, which is calculated by a weight decay function $h(x)$, as in Eq. 7. Then the comprehensive trust degree $\mathcal{T}_\mathcal{D}^i$ of w_i after h tasks is as Eq. (6). φ is a constant and $h(x)$ is a weight decay function that is used to reduce the weight of earlier tasks, i.e., $h(x) > h(x+1)$. The algorithm of trust verification is shown in Algorithm 1.

$$\mathcal{T}_\mathcal{A}^i = \frac{\sum_{x=1}^{r} h(x)\mathcal{T}_\mathcal{D}^i}{r}. \tag{6}$$

$$h(x) = \begin{cases} 1, & \text{when } x = r. \\ h(x+1) \times \varphi, & \text{when } 1 \leq x < r. \end{cases} \tag{7}$$

Theorem 1: The trust of workers $\mathcal{T}_\mathcal{A}^i$ is in the range of [0, 1].

Proof: In order to prove that $\mathcal{T}_\mathcal{A}^i$ is in the range of [0, 1], we need to show that $\mathcal{T}_\mathcal{D}^{i,k}$ and $h(x)$ are both in the range of [0,1]. For the $\mathcal{T}_\mathcal{D}^{i,k}$, the data of the worker compared to the baseline data has two results. (a) When $|\mathcal{V}_i^k - \mathcal{H}_i^k| \leq \varpi$, $0 \leqslant \frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi} \leqslant 1$, $\frac{1}{e} \leqslant e^{-\frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi}} \leqslant 1$, thus $0 < \mathcal{T}_\mathcal{D}^{i,k} \leq 1$; (b) when $|\mathcal{V}_i^k - \mathcal{H}_i^k| > \varpi$, $\lambda \times \frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi} > 1$, $0 < e^{-\lambda \times \frac{|\mathcal{V}_i^k - \mathcal{H}_i^k|}{\varpi}} < \frac{1}{e}$, thus $0 < \mathcal{T}_\mathcal{D}^{i,k} \leq \frac{1}{e}$. In summary, the direct trust value $\mathcal{T}_\mathcal{D}^{i,k}$ is in the range of [0,1].

Next, we prove that $0 < h(x) \leqslant 1$. When $x = r$, $h(x) = 1$. When $1 \leq x < r$, $h(x)$ is equal to the weight of the next task multiplied by a constant between 0 and 1. And the maximum weight of the next task is 1, thus $0 < h(x) \leqslant 1$ is proved. $\mathcal{T}_\mathcal{D}^{i,k}$ and $h(x)$ are both in the range of [0,1], so $0 < h(x) \times \mathcal{T}_\mathcal{D}^{i,k} \leq$

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

Algorithm 2 Data Value and Rewards Calculation

Input: T: deadline of a task, $\mathcal{S}$: the set of selected workers, $\mathcal{T}_\mathcal{A}$: the set of workers' trust value, $\mathcal{S}^\mathcal{T}$: the set of verifited workers, $\{\theta_i, \theta_j, \ldots, \theta_n\}$: bid of workers;
Output: η_κ, p_i.

$\mathcal{S}_k^w \leftarrow \emptyset$;
for a worker $j \in \mathcal{S}$ **do**
  if $T_{\kappa-1} \leq a_j \leq T_\kappa$ **then**
    $\mathcal{S}_k^w = \mathcal{S}_k^w \cup \{j\}$;
  end if
  for a worker $i \in \mathcal{S}_k^w$ **do**
    $\bar{\mathcal{T}_\kappa} = \frac{\sum_{i \in \mathcal{S}_k^w} \mathcal{T}_\mathcal{A}^i}{n_\kappa}$;
    $\rho_{max} = max\{\frac{v_1}{b_1}, \frac{v_2}{b_2}, \ldots, \frac{v_\kappa}{b_\kappa}\}$;
    $\rho_{min} = min\{\frac{v_1}{b_1}, \frac{v_2}{b_2}, \ldots, \frac{v_\kappa}{b_\kappa}\}$;
    $\eta_\kappa = \frac{1/\bar{\mathcal{T}_\kappa}}{l-\kappa} \times \frac{\rho_{max}}{\rho_{min}} \times \frac{\sum_{i \in \mathcal{S}_k^w} v_i(t,\mathcal{T})}{\sum_{i \in \mathcal{S}_k^w} b_i}$;
    $p_i = \frac{v_i(t,\mathcal{T})}{\eta_\kappa}$;
  end for
end for
return p_i;

1. The average of the sum of the r times products is also between 0 and 1, i.e., $0 < \frac{\sum_{k=1}^{k=r} h(x) \times \mathcal{T}_\mathcal{D}^{i,k}}{r} \leqslant r$. Therefore, based on the above derivation we can prove that the trust of workers $\mathcal{T}_\mathcal{A}^i$ is in the range of [0,1]. ■

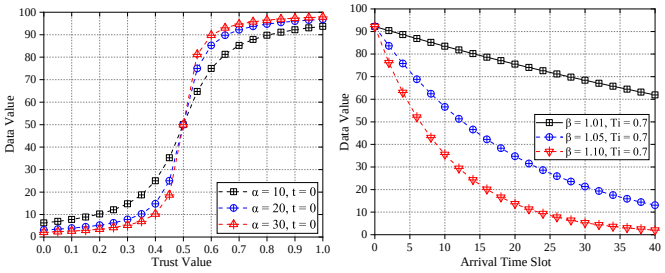
After workers report data to the platform, the platform adopts the truth discovery method based on historical data to verify the trust of workers. Before the first round of worker recruitment starts, the trust of all workers is unknown, after the first round of recruitment, some workers report data to the platform, the platform selects some workers in the set $\mathcal{S}$ as the verified workers and adds them to the set $\mathcal{S}^\mathcal{T}$. We consider the trust to be constant once verified. In the second round of worker recruitment, the platform increases the trust recognition rate and expands the validation for unknown workers excluding already validated workers.

C. Data Value

The data submitted by workers to the platform contributes a certain data value to the trusted service construction. In this section, we introduce the mathematical model for calculating the value of data. The value of worker i's data decays over time as the data is submitted. In our model, we assume that the time each worker takes to perform a task can be ignored since it is far less than that of the worker arriving a PoI in magnitudes [9], [45], so the time of data submission is the arrival time, i.e., $t_i = a_i$. The later the arrival time of the worker, the more the data value decays. The data value decay function over time is expressed by Eq. (8).

$$v_i(t) = v_{max} \times \beta^{-t_i} = v_{max} \times \beta^{-a_i}. \tag{8}$$

Where β is the value decay factor. The larger the value of β, the faster the value decay. On this basis, we integrates the effects of time and trust degree on data value. When $0 < \mathcal{T}_i < 0.5$, the data value grows slowly because the trust value is low,



(a) (b)

Fig. 5. Impact of trust value and arrival time on data value. (a) Trust value and data value. (b) Arrival time and data value.

and when $0.5 < \mathcal{T}_i < 1$, the data value grows faster and tends to the maximum value. As $\mathcal{T}_i$ increases from 0 to 1, the value of data shows a trend of first growing slowly and then rapidly. Based on Eq. (8), we propose a trust-time function of data value $v_i(t, \mathcal{T})$ for workers, as in Eq. (9).

$$v_i(t, \mathcal{T}) = v_{max}\beta^{-t_i} \times (\frac{1}{\pi} \arctan(\alpha \times (\mathcal{T}_\mathcal{A}^i - 0.5)) + \frac{1}{2}). \tag{9}$$

α is the trust decay factor, which determines the data value that a worker gains (or loses). Let $t_i = 0, v_{max} = 100$, Fig. 5a shows the data value increases with trust for α = 10, 20, and 30, respectively. The larger the value of α, the more the data value grows. Let $\mathcal{T}_i = 0.7$, Fig. 5b shows the data value decays with the increase of arrival time for $\beta = 1.01, 1.05, 1.1$. In the beginning, the value decays fast and then becomes slow. The larger the value of β, the more the data value decays. As can be seen from Fig. 5a, the function arctan(·) of Eq. (9) is a monotonically increasing function , and when $T_\mathcal{A}^i \to 0$, $v_i(t, \mathcal{T}) \to 0$; $\mathcal{T}_\mathcal{A}^i \to 1$, $v_i(t, \mathcal{T}) \to v_{max}$, which satisfies the demand that a worker with the high trust has a higher the data value. In the first round of worker recruitment, whether a worker is selected depends on the arrival time of the worker due to $\mathcal{T}_\mathcal{A}^i = 0.5, \forall i \in \mathcal{N}$, which is fair for each worker at the beginning of worker selection. In particular, if the platform has residual costs to recruit workers whose trust is unknown, we regard its trust level as 0.5, i.e., $v_i(t, \mathcal{T}) = \frac{1}{2} v_{max}\beta^{-t_i}$. If the worker arrives after the task deadline, we consider the data value to be 0, i.e., when $t > T$, $v_i(t, \mathcal{T})$ = 0.

D. Inspection Cost and Reward Cost of the Platform

The total budget $\mathcal{B}$ of the platform is used to reward workers and identify the trust of workers, which are called reward costs and inspection costs respectively. Next, we give the calculation method of two types of costs.

(1) The inspection cost $\mathcal{C}_\mathcal{T}$ of the platform is the reward paid to trusted workers for collecting historical data. In the previous method, when w_i uploads real-time data of task r, the platform should recruit trustworthy worker to collect truthful data of task r, and then identify data trust by comparing the real data. This method is not only slow to gain trust, but also leads to extremely high verification costs, while the platform pays the remuneration much less for historical data than for real-time data and can obtain real evidence of data verification, thereby greatly reducing the inspection cost of the platform.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

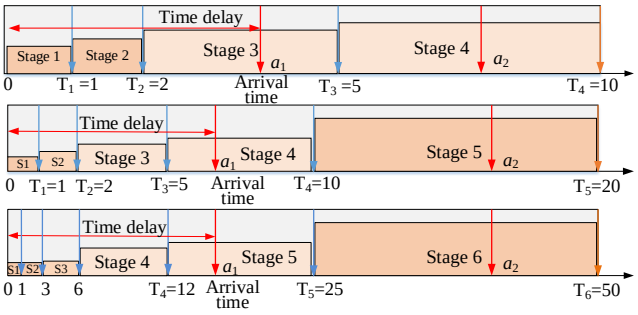


Fig. 6. Illustration of a multiple-stage arriving process when $T = 10, 20, 50$.

When a worker i with unknown trust submits the data of task r in time, the platform requests trusted workers to submit historical data of task r and give payment ph_i^r, which is calculated by Eqs. (10) and (11).

$$ph_i^r = \frac{v_{max}}{\bar{\rho}_r} \times \frac{\rho_{ave}}{\rho_{max}}. \tag{10}$$

$$\bar{\rho}_r = \frac{\sum_{i \in \mathcal{S}^r} v_{i,r}(t, \mathcal{T})}{\sum_{i \in \mathcal{S}^r} b_i}. \tag{11}$$

$\mathcal{S}^r$ is the set of workers that can perform the task r, and $\bar{\rho}_r$ is the average efficiency of workers in $\mathcal{S}^r$, calculated by Eq. (11). ρ_{max} is the maximum efficiency of workers in $\mathcal{S}^r$ and ρ_{ave} is the average of the efficiency of all workers in $\mathcal{S}^r$. Assume that the set of tasks that a unknown worker i overlaps with trusted worker is $\mathcal{R}_i$. Trusted worker submits historical data for task $r \in \mathcal{R}_i$ and the platform rewards ph_i^r, so the total cost for identifying w_i is the sum of the rewards paid to all historical data in task $\mathcal{R}_i$, i.e., $\mathcal{C}_{\mathcal{T}}^i = \sum_{r \in \mathcal{R}_i} ph_i^r$. When the trust recognition rate of the system reaches χ, the platform identifies multiple workers, which is in the set $S^{\mathcal{T}}$, so the total inspection cost of the platform is as Eq. (12).

$$\mathcal{C}_{\mathcal{T}}^{\chi} = \sum_{i \in \mathcal{S}^{\mathcal{T}}} \sum_{r \in \mathcal{R}_i} ph_i^r = \sum_{i \in \mathcal{S}^{\mathcal{T}}} \sum_{r \in \mathcal{R}_i} \frac{v_{max}\rho_{ave}}{\rho_{max}} \cdot \frac{\sum_{i \in \mathcal{S}^r} b_i}{\sum_{i \in \mathcal{S}^r} v_{i,r}(t, \mathcal{T})}. \tag{12}$$

(2) The total reward cost of the platform is the reward paid to all recruited workers. The rewards of workers are proportional to the value of their data. Workers with high data value will be paid more. The reward cost of recruiting worker i is p_i, calculated by Eq. (13).

$$p_i = \frac{v_i(t, \mathcal{T})}{\eta_\kappa}, i \in \mathcal{S}, a_i \in [T_{\kappa-1}, T_\kappa]. \tag{13}$$

$$\eta_\kappa = \frac{1/\bar{\mathcal{T}}_\kappa}{l - \kappa} \times \frac{\rho_{max}}{\rho_{ave}} \times \frac{\sum_{i \in \mathcal{S}_k^w} v_i(t, \mathcal{T})}{\sum_{i \in \mathcal{S}_k^w} b_i}. \tag{14}$$

$$\rho_{ave} = \frac{\sum_{i \in \mathcal{S}_k^w} \rho_i}{|\mathcal{S}_k^w|}, \quad \bar{\mathcal{T}}_\kappa = \frac{\sum_{i \in \mathcal{S}_k^w} \mathcal{T}_{\mathcal{A}}^i}{|\mathcal{S}_k^w|}. \tag{15}$$

Note that if $\frac{v_i(t,\mathcal{T})}{\eta_\kappa} \leq b_i$, we consider the reward to be equal to the worker's cost b_i, i.e., $p_i = b_i$. Thus, the reward received by each worker is not less than the data collection cost that he or she pays, and the utility of worker $i \in \mathcal{S}$, $u_i = p_i - b_i \geq 0$, thus guaranteeing individual rationality of the mechanism.

Algorithm 3 Workers Selection Algorithm

Input: $\mathcal{S}^{\mathcal{T}}$, $\mathcal{B}$, T, Ω, $\{\theta_i, \theta_j, \ldots, \theta_n\}$: bid of workers;
Output: The set of selected workers $\mathcal{S}$;

$\mathcal{S} \leftarrow \emptyset$, $\mathcal{S}_{\mathcal{A}} \leftarrow \emptyset$, $\mathcal{S}^w \leftarrow \emptyset$;
for each worker $i \in \mathcal{S}^{\mathcal{T}}$ **do**
  if $\mathcal{T}_{\mathcal{A}}^i > \Omega$ and $a_i < T$ **then**
    $\mathcal{S}_{\mathcal{A}} = \mathcal{S}_{\mathcal{A}} \cup \{i\}$;
  end if
  while $\mathcal{S}_{\mathcal{A}} \neq \emptyset$ **do**
    $i \leftarrow argmax \frac{\mathcal{T}_{\mathcal{A}}^i}{a_i b_i}$;
    if $\sum_{j \in \mathcal{S}} p_j + \mathcal{C}_{\mathcal{T}} + p_i \leq \mathcal{B}$ **then**
      $\mathcal{S} = \mathcal{S} \cup \{i\}$;
    end if
    $\mathcal{S}_{\mathcal{A}} = \mathcal{S}_{\mathcal{A}} \backslash \{i\}$;
  end while
end for
if $\sum_{i \in \mathcal{S}} p_i + \mathcal{C}_{\mathcal{T}} \leq \mathcal{B}$ **then**
  for each worker $i^* \in \mathcal{N} \backslash \mathcal{S}^{\mathcal{T}}$ **do**
    if $a_i^* < \mathcal{T}$ and $\frac{v_{i^*}(t,\mathcal{T})}{b_{i^*}} > 1$ **then**
      $\mathcal{S}^w = \mathcal{S}^w \cup \{j\}$;
    end if
  end for
  while $\mathcal{S}^w \neq \emptyset$ **do**
    $j^* \leftarrow argmax \frac{v_k(t,\mathcal{T})}{b_k}$;
    if $\sum_{i \in \mathcal{S}} p_i + \mathcal{C}_{\mathcal{T}} + p_{j^*} \leq \mathcal{B}$ **then**
      $\mathcal{S} = \mathcal{S} \cup \{j^*\}$;
    end if
    $\mathcal{S}^w = \mathcal{S}^w \backslash \{j^*\}$;
  end while
end if
return $\mathcal{S}$;

In Eq. (13), η_κ is the reward factor of stage κ. The worker recruitment time T is divided into l stages, $l = \lfloor \log_2 T \rfloor + 1$, i.e., $\kappa \in \{1, 2, ..., \lfloor \log_2 T \rfloor, \lfloor \log_2 T \rfloor + 1\}$. The time range of the κth stage is $[T_{\kappa-1}, T_\kappa]$. $T_\kappa = \lfloor 2^{\kappa+1} \times T / 2^{\lfloor \log_2 T \rfloor + 1} \rfloor$ [9], [43], [44]. Fig. 6 shows the stage division when $T = 10, 20, 50$. $\mathcal{S}_k^w$ is the set of workers with the arrival time in stage κ among all workers $\mathcal{N}$. $\rho_\kappa = \frac{\sum_{i \in \mathcal{S}_k^w} v_i(t,\mathcal{T})}{\sum_{i \in \mathcal{S}_k^w} b_i}$, which is the average efficiency of workers in stage κ. ρ_{max} is the maximum efficiency of workers in $\mathcal{S}_k^w$ and ρ_{ave} is the average of the efficiency of workers in $\mathcal{S}_k^w$. $\bar{\mathcal{T}}_\kappa$ is the average trust of all workers in $\mathcal{S}_k^w$. Eq. (13) shows that the higher the average trust value and the earlier the arrival time, the smaller the reward factor, so the larger the reward, which satisfies the demand that a participant's high data value will increase his/her reward. Data value and cost calculation method is as in Algorithm 2.

E. Workers Selection Strategy

When recruiting a new mobile worker, the platform calculates whether the total cost exceeds the budget $\mathcal{B}$. If the remaining cost is not enough to pay the reward to the newly recruited worker, the platform ends the worker recruitment. Workers are divided into workers with known trust and workers with unknown trust. The platform prioritizes those workers

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

with high trust, and then uses the remaining cost to select workers with unknown trust. Below we describe the process of selecting different types of workers for the platform.

1) Strategies for selecting workers with known trust: The platform first selects the worker i ($i \in \mathcal{S}^{\mathcal{T}}$) with $\mathcal{T}_{\mathcal{A}}^{i} > \Omega$ and adds worker i the candidate set $\mathcal{S}_A$. Ω is called the trust threshold. As can be seen from Fig. 5, when $\mathcal{T}_{\mathcal{A}}^{i} = 0.5$, $v_i = \frac{1}{2}v_{max}$, which is called the critical value. When $\mathcal{T}_{\mathcal{A}}^{i} > 0.5$, the value of the data is greater than the critical value and growing rapidly, so we take $\Omega = 0.5$. The workers in the candidate set should satisfy the following conditions:

$$\mathcal{S}_{\mathcal{A}} = \mathcal{S}_{\mathcal{A}} \cup \{i\},\ s.t. \mathcal{T}_{\mathcal{A}}^{i} > 0.5 \text{ and } a_i < T. \tag{16}$$

Workers with higher trust and early arrival time can provide higher data value, and the cost declared by workers is the lowest reward cost, so the lower cost, the lower reward cost paid by the platform. Therefore, these workers are preferred among candidates to obtain the maximum data value. We define the *Trust- Bid Ratio* (TBR) of workers as $\frac{\mathcal{T}_{\mathcal{A}}^{i}}{a_i b_i}$. The platform selects workers in descending order according to the TBR. Workers are sorted according to TBR, as in Eq. (17).

The platform prioritizes the worker i with the highest TBR. When worker i is recuited, if the total cost does not exceed the budget, the worker is the winner and is added to the set $\mathcal{S}$, and vice versa to end the worker recruitment.

$$\frac{\mathcal{T}_{\mathcal{A}}^{i}}{a_i b_i} > \frac{\mathcal{T}_{\mathcal{A}}^{j}}{a_j b_j} > \cdots > \frac{\mathcal{T}_{\mathcal{A}}^{n}}{a_n b_n}, i, j, \ldots, n \in \mathcal{S}_{\mathcal{A}}. \tag{17}$$

2) Strategies for worker selection with unknown trust: In the initial stage, the trust recognition rate is low, and the number of workers identified with trust is small, so when the platform has finished selecting the appropriate workers with known trust, there may be remaining cost to be able to select workers with unknown trust. For such workers, the platform tends to recruit workers with high data value and low cost. We define the *efficiency* ρ of a worker as the ratio of data value $v_i(t, \mathcal{T})$ to cost b_i, i.e.,

$$\rho = \frac{v_i(t, \mathcal{T})}{b_i}. \tag{18}$$

The platform judges whether a worker i ($i \in \mathcal{N} \backslash \mathcal{S}^{\mathcal{T}}$) satisfies the constraints: $a_i < \mathcal{T}$, $\rho_i > 1$. $\rho_i > 1$ is to ensure that the platform's profit (i.e., $v_i(t, \mathcal{T}) - b_i$) from worker i to be greater than 0. $\mathcal{N}$ is the set of all workers and $\mathcal{S}^{\mathcal{T}}$ is the set of workers with verified trustworthiness. Workers that meet the above constraints are added in the set $\mathcal{S}^w$, and then the efficiency of the workers in $\mathcal{S}^w$ is sorted in descending order, as in Eq. (19). The platform selects workers from high to low efficiency. Each time a worker is selected, platform calculates whether the total cost paid exceeds the budget. If not, w_i can be selected, and then select the next workers until the total cost reaches the budget and the workers recruitment is stopped. The workers selection is as in Algorithm 3.

$$\frac{v_{i^+}(t, \mathcal{T})}{b_{i^+}} > \frac{v_{j^+}(t, \mathcal{T})}{b_{j^+}} > \cdots > \frac{v_{n^+}(t, \mathcal{T})}{b_{n^+}}, i^+, \ldots, n^+ \in \mathcal{S}_w. \tag{19}$$

TABLE II
EXPERIMENTAL PARAMETERS SETTINGS

Parameters	Default value
Number of tasks m	100
Number of workers n	50
Deadline of each task T	20
Maximum value of the task v_{max}	100
Budget of the platform $\mathcal{B}$	5000
Cost range of workers c_i	[20, 50]
Number of tasks of worker i	[1, 5]
Time-discounting factor β	1.01
Trust-discounting factor α	20
Data error threshold ϖ	2
Trust decay rate λ	1.2

TABLE III
REWARD FACTORS WHEN PROFIT IS MAXIMIZED

Parameters	η_1	η_2	η_3	η_4	η_5
ρ_{max}	4.401	4.662	3.989	4.662	3.488
ρ_{ave}	2.589	3.157	2.796	1.468	2.330
ρ_{max}/ρ_{ave}	1.699	1.477	1.426	2.615	1.497
ρ_κ	2.441	2.948	2.577	1.439	2.200
$\mathcal{T}_a$	0.702	0.793	0.770	0.542	0.737
η_κ	1.181	1.372	1.591	3.469	4.470

V. PERFORMANCE EVALUATION

A. Experimental Settings

To evaluate the performance of our profit maximization mechanism, we implement a simulation experiment with the following experimental setup. The parameters used in the experiment are listed in Table II. The total number of tasks $m = 100$ and the total number of workers $n = 50$. The deadline of each task $T = 20$, which is divided into $\lfloor \log_2 T \rfloor + 1 = 5$ stages. The number of tasks m_i that each worker i can perform does not exceed [1, 5], and then m_i tasks are randomly selected from $\mathcal{M}$ to generate the set of tasks for worker i. Total budget of the platform is $\mathcal{B} = 5000$ and sensing cost c_i of each worker for each task satisfies $20 \leqslant c_i \leqslant 50$.

We simulate collecting temperature data from different region as data collection tasks, and the data collected by the workers is in [10, 30]. The error threshold between the reported data and truth data is set as $\varpi = 2$. We use following metrics to evaluate the effectiveness of proposed MPS.

(a) *Total Profit*, is the sum of the marginal data values of the recruited workers. Our goal is to maximize the total profit of the platform and obtain the optimal recognition rate corresponding to the maximum profit.

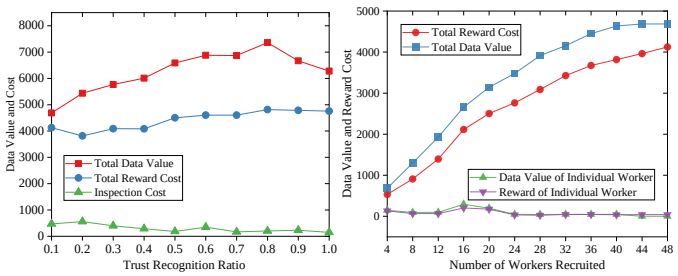
(b) *Trusted workers' Ratio (TR)*, is defined as the ratio of the number of trusted workers to the number of all workers.

(c) *Task Implementation Ratio (TIR)* and *Worker Selection Ratio (SR)*. TIR is the ratio of the number of tasks that can be completed by the recruited workers to the total number of tasks, while SR is the ratio of the number of recruited workers to the total number of workers.

(d) *Budget Utilization Ratio (BUR)*, is defined as the ratio of the total cost of recruiting workers to the total budget.

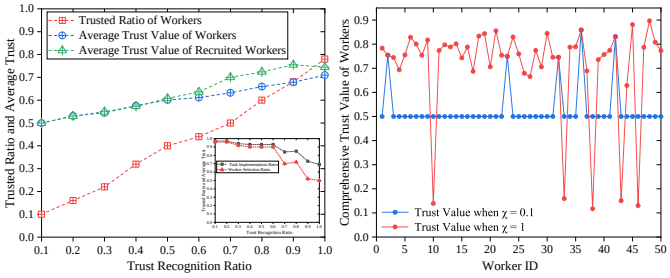
Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541



(a) (b)

Fig. 7. Impact of trust recognition rate on platform profit. (a) Total profit and total cost. (b) Data value and cost for individual workers.



(a) (b)

Fig. 8. Impact of trust recognition rate on different indicators. (a) TR, SR, TIR and average trust value. (b) Worker's trust value.

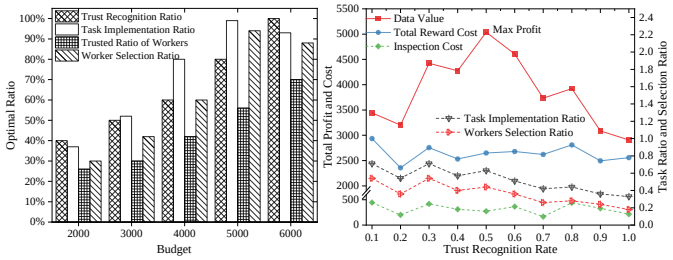
B. Impact of Trust Recognition Ratio on Total Value

Fig. 7 illustrates the variation of data value and total cost under different trust recognition ratios when $\mathcal{B} = 5000, n = 50, m = 100$. In Fig. 7, as the trust recognition rate increases from 0.1 to 1, the total data value of the platform first increases slowly and then decreases, and is higher than the total cost. When $\chi = 0.8$, the platform's profit reaches its maximum value, and when $\chi > 0.8$, the profit starts to decrease. Table III shows the reward factor when profit is maximized. The total cost of the platform in each round is close to the total budget, indicating that the platform has a very high budget utilization ratio. Compared to the total reward cost, the total inspection cost of the platform is very low. Fig. 7b shows the data value and cost for individual workers when $\chi = 0.1$. The inspection cost of a single worker is also very small compared to the rewards received by the worker, which indicates that our proposed data truth discovery method is a low-cost method.

Fig. 8 shows the impact of trust recognition rate on different indicators such as TR and TIR. In Fig. 8a, as the trust recognition rate increases, the platform is able to recognize more trustworthy workers, so the value of TR gradually increases to 0.8, and the average trust value of recruited workers and the trust level of all workers in the system gradually increases from 0.5 to more than 0.7. As shown in Fig. 8a, when the value of χ is small, the reward cost used for worker recruitment is sufficient and the trust value of the worker is unknown, the reward cost is also low and more number of workers can be recruited, so the TIR and SR values are maintained at 0.9. But as the value of χ increases, the platform is able to recruit more high trustworthy workers but also pays a higher reward cost, and fewer workers are recruited, and there is a clear trend of decreasing TIR and SR values. Fig. 8b demonstrates the trust values of all workers in the MCS when $\chi = 0.1$ and $\chi = 1$. When $\chi = 0.1$, 90% of the workers' trust is unknown which is 0.5. When all the workers' trust is identified, the workers' trust is concentrated above 0.7, and the trusted workers' ratio of MCS system is higher than 70%, which shows that our trust recognition scheme is effective.

C. Optimal Trust Recognition Ratio

In this section, we evaluate the optimal trust recognition ratios under different budgets. As shown in Fig. 9a, by increasing the budget, the platform has more cost to spend on trust verification, so the trusted ratio of the system continues



(a) (b)

Fig. 9. Optimal trust recognition Ratio under different budgets. (a) Recognition Ratio, TR, TIR, SR under different budgets. (b) Optimal trust recognition rate when $\mathcal{B} = 3000$.

to increase, and the optimal trust recognition ratio gradually increases from 40%. When the budget is large enough such as $\mathcal{B} = 6000$, the platform has the ability to verify the trust values of all workers with $\chi_{op} = 1$ and is able to reach the maximum profit. The changes in TIR and SR values are consistent with an overall increasing trend with the increase of the budget. Fig. 9b shows the optimal recognition rate when $\mathcal{B} = 3000$. Due to the low budget, the platform can only obtain the trust of some of the workers. The total profit shows a trend of increasing and then decreasing, and the profit reaches the maximum when $\chi = 0.5$. As the recognition rate increases, the profit gradually decreases as the inspection costs increases. The task completion rate and worker selection rate change simultaneously, showing a slow decline trend overall.

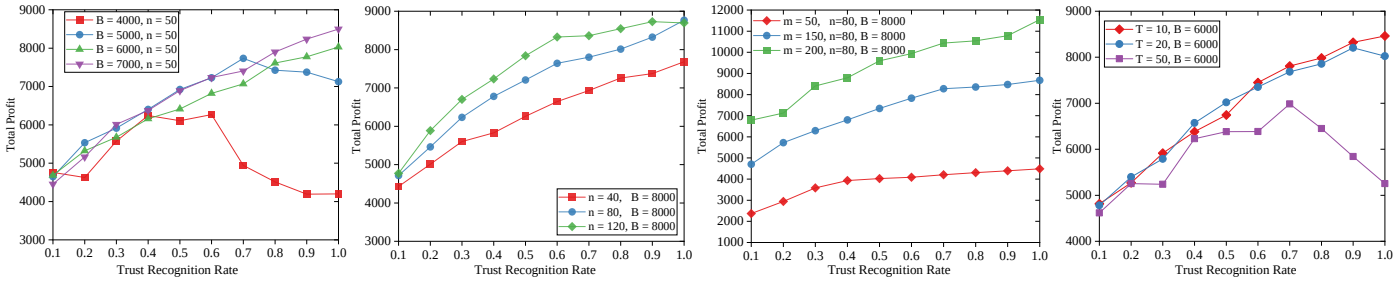
D. The Effects of System Parameters on Total Profit

In this section, we evaluate the impact of trust recognition ratio on the platform's total profit under different system parameters, such as budget, number of workers or number of tasks, etc. As shown in Fig. 10a, with a fixed number of workers ($n = 50$), we set the budget is increased from 4000 to 7000. When the budget is small such as $\mathcal{B} = 4000$ and $\mathcal{B} = 5000$, the platform's profit increases and then decreases, while as $\mathcal{B}$ increases, the profit rises continuously as the recognition rate increases, which is due to the fact that the platform has enough cost to recognize the trustworthiness of more workers.

If the number of workers increases from 40 to 120 with a limited budget that is set to $\mathcal{B} = 8000$, it can be seen from Fig. 10b that the higher the number of workers, the higher the profit. This means that given a sufficient budget, the platform is able to detect the trustworthiness of all workers, and the

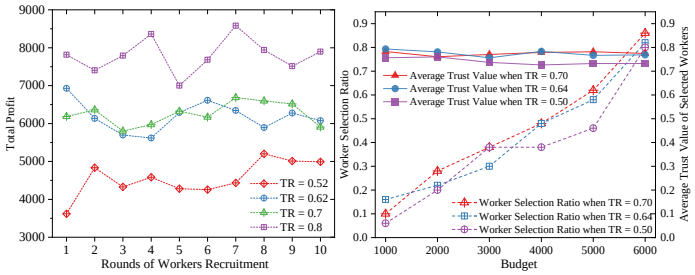
Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541



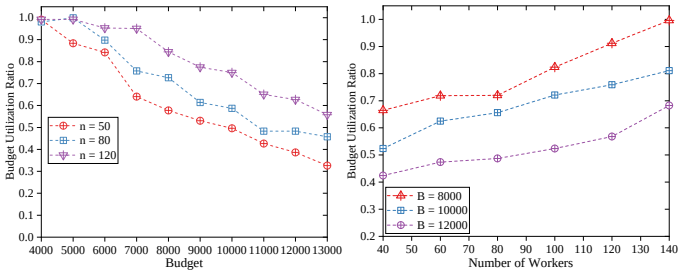
(a) (b) (c) (d)

Fig. 10. Impact of trust recognition ratio on profit under different system parameters. (a) Different budgets. (b) Different number of workers. (c) Different number of tasks. (d) Different deadlines.



(a) (b)

Fig. 11. Impact of system trustworthiness ratio on profit and selected Ratio. (a) Profit under different system Trustworthiness Ratio. (b) Selection ratio under different system trustworthiness ratio.



(a) (b)

Fig. 12. Impact of varying budget and number of workers on Budget Utilization Ratio. (a) The impact of budgets. (b) The impact of number of workers.

number of highly trustworthy workers selected by the platform is also increasing, so the profit will continue to rise.

We next consider the evaluation results about the effect of number of tasks and task deadline T on the total value. As in Fig. 10c, we set $n = 80, \mathcal{B} = 8000$, and the total number of tasks increases from 50 to 200. We can see a significant increase in profit, which is due to the fact that the number of tasks increases, the overlap of tasks that can be performed between workers becomes less, and the marginal value of the workers becomes larger, which increases the revenue of the platform. Fig. 10d shows the platform's revenue under different task deadlines. The shorter the task deadline, the platform is able to recruit workers submitting data earlier, so the less the data value decays. As T decreases, the total value will be higher than the that with a larger task deadline. In Fig. 10d, the platform's profit first rises and then declines when $T = 50$. This is due to the fact that the platform allows later arriving workers to participate in the data collection, which provide low data value, so as the recognition rate increases, the platform's revenue instead starts to decline.

E. Impact of System Trust Ratio on Profit

In this section, we discuss the impact of the trust ratio (TR) of the MCS system on the value of the data and SR. We set the four distribution ratio of trusted workers in the networks, as shown in Fig. 11a. The more trusted workers distributed in the system, the more trusted workers the platform can recruit, and thus the higher the data value. When TR = 0.5, the number of trusted workers is small and the total profit of the platform is the lowest. As TR increases, the profit rises significantly. When TR = 0.7, the platform's profit is the highest.

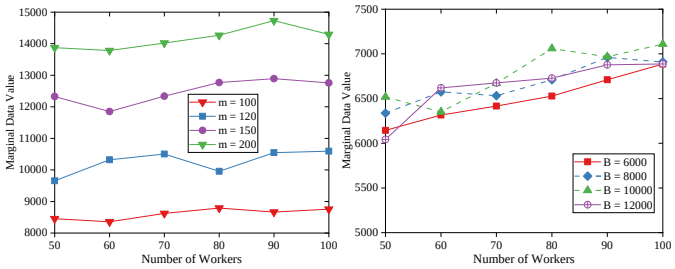
In Fig. 11b, the average trust values of workers selected under different TR values are relatively stable. When TR = 0.7, the average trust value is maintained at 0.8, and as the TR value decreases, the average trust value has a small decrease but is able to stay above 0.7. This shows that even though the TR decreases, the worker selection strategy still selects highly trusted workers and screens out the lowly trusted workers, and thus the average trust value can maintain a high value. The above results show that our scheme is able to accurately distinguish between trusted and malicious workers, and can also select high-quality workers to increase the profit of the platform. For the worker selection ratio, the larger the TR value, the more highly trusted workers can be selected, hence the larger the value of SR. When the TR value is fixed, the worker selection ratio increases as the budget increases because the platform is able to reward more workers to participate in the task. As the TR value decreases, the number of trusted workers in the system is then reduced, and there are fewer candidate workers that meet the requirements for worker recruitment, so the SR value is lower.

F. Budget Utilization Ratio

Fig. 12 shows the variation of budget utilization ratio when the cost changes from 4000 to 13000 and the number of workers changes from 40 to 140. As shown in Fig. 12a, when the number of workers is fixed, the BUR decreases as the cost increases. This is because that the data value provided by a fixed number of workers is limited, which indicates adding more budget may not improve the platform's revenue. When the budget is small, the costs are fully spent on recruiting

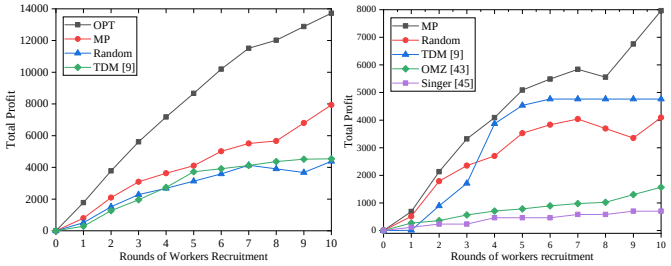
Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541



(a) (b)

Fig. 13. Impact of varying budget and number of tasks on marginal data value. (a) The impact of number of tasks. (b) The impact of budgets.



(a) (b)

Fig. 14. Profit comparison of MPS and the other schemes [9], [43], [46]. (a) Profit comparison of Manuscript with TDM [9]. (b) Impact of reward factor on profit compared with OMZ [43], Singer [46].

workers and the budget utilization rate is above 0.8, while as the budget increases, more redundant costs are incurred, resulting in low utilization rate. In Fig. 12b, under the fixed budget, the BUR rises with the number of workers. This is because the platform is able to recruit more credible and low-latency workers, but also needs to pay more reward costs, so the platform has a higher budget utilization rate.

We also evaluate the impact of varying budget and number of tasks on marginal data value. In Fig. 13a, when the number of tasks is fixed, the platform revenue may not improve significantly as the number of workers increases. Fig. 13b also shows that with a constant number of tasks $m = 100$, if the budget of the platform is increased, the marginal value of the data is close and there are no major improvements. This is because considering the marginal data value of the workers, the platform does not double count the value of data that has already been submitted by other workers. When the number of tasks is small, the tasks performed by a new worker overlap significantly with the tasks of recruited workers, so the marginal value contributed by this worker is very small, and thus expanding the number of workers or budgets may not improve the platform's profit. But if the number of tasks is increased, the overlap between the workers is reduced. so there is a significant rise in platform's profit in Fig. 13a.

G. Platform Profit Comparison

In order to verify the effectiveness of the proposed MPS, we compare the profit of MPS with that of the following schemes.

(1) Time-Discounting Mechanism (TDM) [9]. TDM proposes a time-discounting value model to measure the value of the collected data, i.e., $v_i(t_i) = v_i \times \beta^{-t}$. The total value is $V_{TDM} = \sum_{w_i \in \mathcal{S}} v_i(t_i) = \sum_{w_i \in \mathcal{S}} v_i \times \beta^{-t}$. $\mathcal{S}$ is the recruited workers set. TDM does not identify the credibility of the data and and total budget is spent on recruiting workers.

(2) OMZ [43] and Singer Mechanism [46]. Workers selection in the above two schemes is similar to MPS. The biggest difference is the calculation of the reward factor for each stage. In OMZ, the reward factor is the average value in the data value-cost ratio of the selected workers, while in the Singer scheme, the reward factor is the maximum value in the data value-cost ratio of the selected workers.

(3) Random mechanism. The platform randomly selects workers until the platform's budget is exhausted.

(4) Optimal solution (OPT). In OPT, each worker is highly trusted and the latency of the data is 0.

We compare the platform profit of MPS with OPT, TDM [9], and Random schemes, as shown in Fig. 14a. With the increase of worker recruitment rounds, the profit of the four schemes is gradually increasing and the total profit of the OPT is much higher than that of the other three schemes. Compared with the Random scheme, because the MPS selects the most efficient workers each time instead of randomly selecting workers with known trust, while the Random scheme may select workers with low trust and large delay, MPS can increase the revenue by 31.37% -81.11%. Compared with the TDM scheme, since TDM lacks the truth discovery, the TDM scheme is at a disadvantage in choosing high-trust workers, leading that the data value is greatly reduced. The TDM mechanism introduces the parameter $(\frac{U}{L})^{l-\kappa}$ to refract the time decay effect, while our MPS introduces the parameter $\frac{1/\bar{\mathcal{T}}_\kappa}{l-\kappa}$ based on the TDM to capture the comprehensive effect of trust and time decay on the reward factor. Experimental results show that the optimized MPS can increase the platform revenue by 10.31%-74.81%.

To verify the effect of the reward factor η_κ on system performance, we compare the platform profit of MPS with OMZ [43] and Singer scheme [46]. The above two schemes increase the reward factor, which will exclude some workers with low latency and high trust to be selected. It can be seen from Fig. 14b that the reward factor has a great influence on the platform's revenue. Although our scheme increases the inspection cost and introduces a time decay factor, the reward factor is optimized through the time parameter $\frac{1}{l-\kappa}$ and the trust parameter $\frac{1}{\bar{\mathcal{T}}_\kappa}$. Experimental results show that this optimization can make the platform's profit significantly higher than the OMZ and Singer schemes.

H. Engineering Applicatons

As shown in Fig. 15, our proposed MPS can be applied to various time-sensitive engineering applications in mobile computing. For example, in urban traffic monitor applications, mobile vehicles running on the road and pedestrians on the roadside serve as workers. Mobile workers distributed in different locations perceive road flow, traffic accidents, road closures data in real-time and report honestly to MCS platform. Platform calculates congestion on multiple alternative routes for users to choose smoother roads, improving traffic efficiency and safety. But since these workers are not specialized devices for monitoring traffic conditions, there are malicious workers out there. MPS is able to identify these malicious workers

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

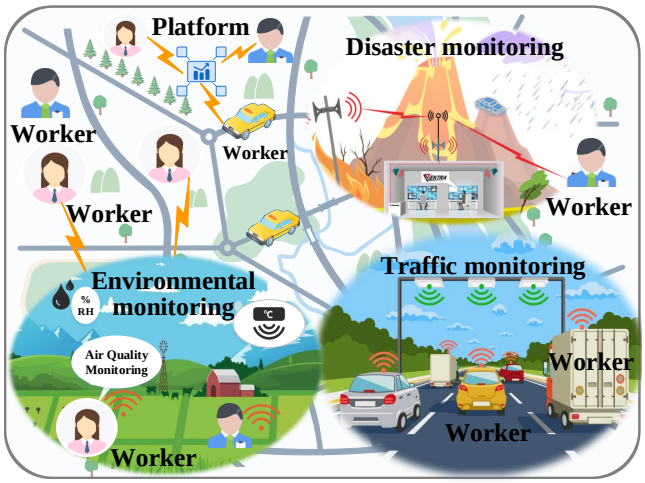


Fig. 15. Practical engineering applications of MPS.

and recruit high-value workers to improve the accuracy of the information provided by the monitoring system. In addition, in environmental monitoring, workers distributed in different locations in the city such as smartphones and laptops collect dynamically changing environmental data, such as air quality, temperature and humidity, etc. Leveraging mobile computing of workers, the monitoring system can detect environmental changes in time and respond quickly to environmental emergencies. However, these devices may not be able to work properly due to aging or external exposure. Therefore, our system model can be applied by dispatching some high-quality devices to collect data at certain location and test whether the devices are functioning normally.

VI. Conclusion and Future Work

In the MCS, the platform recruits mobile workers to complete data collection, task calculation, and service construction. However, untrustworthy data would lead to poor services and platform's low profit. Identifying the data truth is an important issue to ensure the high quality of services provided by the platform. This paper first proposes a low-cost service truth discovery method, which utilizes the historical data of mobile trusted workers as truthful data to compute the data truth. Secondly, our proposed MPS comprehensively considers that the data value decreases with the decrease of the worker's trust and decreases over time. Then the platform calculates the data value of mobile workers and prioritizes recruiting workers with high trust and less delay to maximize the platform's profit. With the trust recognition rate increases, the platform can identify more workers' trust and select high-trust workers, but this increase the inspection costs, so there is a balance between the platform's profit and the trust recognition rate. We give the optimal trust recognition rate when platform profit is maximized. The experimental results show that MPS shows good performance in terms of platform profit and is significantly better than the existing schemes. For the further work in this paper, we intend to consider the time decay of trust and obtain the validity period of the trust of workers, recruiting credible workers within the validity period, and thus maintaining the long-term maximum profit of the platform.

References

[1] Y. Cheng, X. Wang, P. Zhou, X. Zhang, and W. Wu, "Freshness-aware incentive mechanism for mobile crowdsensing with budget constraint," *IEEE Transactions on Services Computing*, vol. 16, no. 6, pp. 4248–4260, 2023.

[2] H. Wang, A. Liu, N. N. Xiong, S. Zhang, and T. Wang, "Tvd-ra: A truthful data value discovery based reverse auction incentive system for mobile crowd sensing," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 5826–5839, 2024.

[3] Z. Wang, J. Li, J. Hu, J. Ren, Q. Wang, Z. Li, and Y. Li, "Towards privacy-driven truthful incentives for mobile crowdsensing under untrusted platform," *IEEE Transactions on Mobile Computing*, vol. 22, no. 2, pp. 1198–1212, 2023.

[4] M. Huang, Z. Li, F. Xiao, S. Long, and A. Liu, "Trust mechanism-based multi-tier computing system for service-oriented edge-cloud networks," *IEEE Transactions on Dependable and Secure Computing*, 2023, early access, doi: 10.1109/TDSC.2023.3285927.

[5] Y. Zhou, F. Tong, and S. He, "Bi-objective incentive mechanism for mobile crowdsensing with budget/cost constraint," *IEEE Transactions on Mobile Computing*, vol. 23, no. 1, pp. 223–237, 2024.

[6] X. Zhu, Y. Luo, A. Liu, W. Tang, and M. Z. A. Bhuiyan, "A deep learning-based mobile crowdsensing scheme by predicting vehicle mobility," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4648–4659, 2021.

[7] Y. Liu, Z. Zeng, X. Liu, X. Zhu, and M. Z. A. Bhuiyan, "A novel load balancing and low response delay framework for edge-cloud network based on SDN," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5922–5933, 2020.

[8] Z. Liu, Q. Z. Sheng, Z. Zhang, X. Xu, D. Chu, J. Yu, and S. Wang, "Accurate and reliable service recommendation based on bilateral perception in multi-access edge computing," *IEEE Transactions on Services Computing*, vol. 16, no. 2, pp. 886–899, 2023.

[9] Z. Zheng, S. Yang, J. Xie, F. Wu, X. Gao, and G. Chen, "On designing strategy-proof budget feasible online mechanisms for mobile crowdsensing with time-discounting values," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 2088–2102, 2022.

[10] L. Ma, X. Liu, Q. Pei, and Y. Xiang, "Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 786–799, 2019.

[11] "Air quality app." [Online]. Available: https://www.iqair.com/us/

[12] N. Hoseinzadeh, Y. Liu, L. D. Han, C. Brakewood, and A. Mohammadnazar, "Quality of location-based crowdsourced speed data on surface streets: A case study of waze and bluetooth speed data in sevierville, tn," *Computers, Environment and Urban Systems*, vol. 83, p. 101518, 2020.

[13] J. Ni, K. Zhang, Q. Xia, X. Lin, and X. S. Shen, "Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1317–1331, 2020.

[14] "Sensorly." [Online]. Available: http://www.sensorly.com

[15] "Geograph: Geographically representative photographs of the whole great britain and ireland." [Online]. Available: http://www.geograph.org.uk/

[16] Y. Ren, Y. Liu, N. Zhang, A. Liu, N. N. Xiong, and Z. Cai, "Minimum-cost mobile crowdsourcing with qos guarantee using matrix completion technique," *Pervasive and Mobile Computing*, vol. 49, pp. 23–44, 2018.

[17] X. Yang, Z. Zeng, A. Liu, N. N. Xiong, T. Wang, and S. Zhang, "A decentralized trust inference approach with intelligence to improve data collection quality for mobile crowd sensing," *Information Sciences*, vol. 644, p. 119286, 2023.

[18] C. Huang, H. Yu, R. A. Berry, and J. Huang, "Using truth detection to incentivize workers in mobile crowdsourcing," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 2257–2270, 2022.

[19] Z. Lu, Y. Wang, X. Tong, C. Mu, Y. Chen, and Y. Li, "Data-driven many-objective crowd worker selection for mobile crowdsourcing in industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 531–540, 2023.

[20] Y. Zhao, K. Zheng, Y. Li, J. Xia, B. Yang, T. B. Pedersen, R. Mao, C. S. Jensen, and X. Zhou, "Profit optimization in spatial crowdsourcing: Effectiveness and efficiency," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8386–8401, 2023.

[21] M. Xiao, B. An, J. Wang, G. Gao, S. Zhang, and J. Wu, "Cmab-based reverse auction for unknown worker recruitment in mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 21, no. 10, pp. 3502–3518, 2021.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Services Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSC.2024.3433541

[22] X. Fu, A. Liu, N. N. Xiong, T. Wang, and S. Zhang, "Atwr-smr: An area-constrained truthful-worker recruitment-based sensing map recovery scheme for sparse mcs in extreme-environment internet of things," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 3711–3724, 2024.

[23] L. Jiang, X. Niu, J. Xu, D. Yang, and L. Xu, "Incentive mechanism design for truth discovery in crowdsourcing with copiers," *IEEE Transactions on Services Computing*, vol. 15, no. 5, pp. 2838–2853, 2021.

[24] S. Huang, A. Liu, S. Zhang, T. Wang, and N. N. Xiong, "Bd-vte: A novel baseline data based verifiable trust evaluation scheme for smart network systems," *IEEE transactions on network science and engineering*, vol. 8, no. 3, pp. 2087–2105, 2021.

[25] P. Sun, Z. Wang, L. Wu, Y. Feng, X. Pang, H. Qi, and Z. Wang, "Towards personalized privacy-preserving incentive for truth discovery in mobile crowdsensing systems," *IEEE Transactions on Mobile Computing*, vol. 21, no. 1, pp. 352–365, 2020.

[26] X. Liu and J. Liu, "A truthful double auction mechanism for multi-resource allocation in crowd sensing systems," *IEEE Transactions on Services Computing*, vol. 15, no. 5, pp. 2579–2590, 2022.

[27] Y. Zhan, Y. Xia, J. Zhang, and Y. Wang, "Incentive mechanism design in mobile opportunistic data collection with time sensitivity," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 246–256, 2017.

[28] J. Tang, K. Fan, W. Xie, F. Han, Z. Qu, A. Liu, N. N. Xiong, S. Zhang, and T. Wang, "Btv-cmab: A bi-directional trust verification-based combinatorial multiarmed bandit scheme for mobile crowdsourcing," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 1925–1938, 2024.

[29] H. Gao, J. Feng, Y. Xiao, B. Zhang, and W. Wang, "A uav-assisted multi-task allocation method for mobile crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 22, no. 7, pp. 3790–3804, 2023.

[30] J. Yu, Z. Li, N. Xiong, S. Zhang, A. Liu, and A. V. Vasilakos, "A reliability and truth-aware based online digital data auction mechanism for cybersecurity in mcs," *Future Generation Computer Systems*, vol. 141, pp. 526–541, 2023.

[31] S. Gao, X. Chen, J. Zhu, X. Dong, and J. Ma, "Trustworker: A trustworthy and privacy-preserving worker selection scheme for blockchain-based crowdsensing," *IEEE Transactions on Services Computing*, vol. 15, no. 6, pp. 3577–3590, 2022.

[32] A. Alagha, S. Singh, H. Otrok, and R. Mizouni, "Influence- and interest-based worker recruitment in crowdsourcing using online social networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1924–1936, 2023.

[33] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, and G. Chen, "On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 832–847, 2017.

[34] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial iots," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[35] M. Huang, A. Liu, N. N. Xiong, and J. Wu, "A uav-assisted ubiquitous trust communication system in 5g and beyond networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 11, pp. 3444–3458, 2021.

[36] Z. Cai, Z. Duan, and W. Li, "Exploiting multi-dimensional task diversity in distributed auctions for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 8, pp. 2576–2591, 2021.

[37] C. Zhao, S. Yang, and J. A. McCann, "On the data quality in privacy-preserving mobile crowdsensing systems with untruthful reporting," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 647–661, 2019.

[38] X. Ding, J. Guo, G. Sun, and D. Li, "Optimizing worker selection in collaborative mobile crowdsourcing," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 7172–7185, 2024.

[39] C. K. Tham and T. Luo, "Quality of contributed service and market equilibrium for participatory sensing," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 829–842, 2014.

[40] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, "Qoi-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4618–4632, 2014.

[41] G. Ji, Z. Yao, B. Zhang, and C. Li, "Quality-driven online task-bundling-based incentive mechanism for mobile crowdsensing," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7876–7889, 2022.

[42] Y. Qian, Y. Ma, J. Chen, D. Wu, D. Tian, and K. Hwang, "Optimal location privacy preserving and service quality guaranteed task allocation in vehicle-based crowdsensing networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4367–4375, 2021.

[43] D. Zhao, X. Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, 2014, pp. 1213–1221.

[44] H. Gao, C. H. Liu, J. Tang, D. Yang, P. Hui, and W. Wang, "Online quality-aware incentive mechanism for mobile crowd sensing with extra bonus," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2589–2603, 2018.

[45] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2982–2997, 2018.

[46] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 1157–1166.



Wen Mo received master's degree from Central South University in 2023. Currently, she is pursuing her Ph.D. in Computer Science and Technology at the School of Computer Science and Engineering, Central South University, Changsha, China. Her research interests include trust computing, crowd sensing networks, and Internet of Things.



Anfeng Liu received the M.Sc. and Ph.D. degrees in computer science from Central South University, China, in 2002 and 2005, respectively. He is currently a Professor with the School of Information Science and Engineering, Central South University. He has published four books and over 100 international journal and conference papers, among which there are more than 30 ESI highly-cited papers. His major research interests include wireless sensor networks, the Internet of Things, information security, edge computing, and crowdsourcing..



Neal N. Xiong (Senior Member, IEEE) received the first Ph.D. degree in sensor system engineering from Wuhan University, Wuhan, China, in 2007, and the second Ph.D. degree in dependable communication networks from Japan Advanced Institute of Science and Technology, Nomi, Japan, in 2008. He is current an Associate Professor (four year credits) with the Department of Computer Science and Mathematics, Sul Ross State University, Alpine, TX, USA. He has published over 200 international journal articles and over 100 international conference papers. Some of his works were published in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE or ACM Transactions, ACM SIGCOMM Workshop, IEEE INFOCOM, ICDCS, and IPDPS. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory.



Houbing Song (Fellow, IEEE) received the M.S. degree in civil engineering from The University of Texas at El Paso, El Paso, TX, USA, in December 2006, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in August 2012. He has been serving as an Associate Technical Editor for IEEE COMMUNICATIONS MAGAZINE since 2017, an Associate Editor for IEEE INTERNET OF THINGS JOURNAL since 2020, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS since 2021, and IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS since 2020, and the Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is an ACM Distinguished Member and an ACM Distinguished Speaker. He is a Highly Cited Researcher identified by Clarivate™ (2021, 2022) and a Top 1000 Computer Scientist identified by Research.com. He received the Research.com Rising Star of Science Award in 2022 (World Ranking: 82; USA Ranking: 16).

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on August 02,2024 at 18:54:35 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.