Continuous Probabilistic Skyline Query for Secure Worker Selection in Mobile Crowdsensing

Xichen Zhang, Rongxing Lu, *Fellow, IEEE,* Jun Shao, Hui Zhu, *Senior Member, IEEE,* and Ali A. Ghorbani, *Senior Member, IEEE*

***Abstract*—Worker selection is always one of the most fundamental problems in Mobile Crowdsensing (MCS), since the reliability of workers' sensing data is hugely significant to the service quality. In the worker selection process, it is inevitable for the workers to share some of their sensitive information. Consequently, numerous studies are conducted on the problem of privacy-preserving worker selection in MCS platforms. However, most of the existing methods focus on static and short-term situations. As a result, they are inapplicable to the highly dynamic environments where the MCS tasks are long-term and the workers can continuously arrive at/leave the system. To solve these problems, in this paper, we propose a privacy-preserving worker selection scheme based on the probabilistic skyline over sliding windows. Specifically, the proposed scheme can select reliable workers for each current sliding window in terms of working experience, expiry time, and trustability. Besides, we design an ElGamal encryption-based scheme for securely outsourcing and comparing workers' personal information without revealing their privacy. Detailed security analysis shows that the workers' sensitive information, e.g., working experience and trustability, are not revealed to any authorized parties during the process of MCS under our security model. Furthermore, extensive experiments on both real-world and simulated datasets demonstrate that our proposed scheme outperforms the baseline method in two application scenarios, i.e., i) continuous worker arrival and ii) continuous worker departure.**

***Index Terms*—Mobile crowdsensing (MCS), continuous worker selection, probabilistic skyline, ElGamal encryption**

I. Introduction

With the explosion in the availability of smart devices brings a new paradigm of sensing network, called Mobile Crowdsensing (MCS), which has recently spurred lots of interests in both industries and academia [1]–[3]. By exploiting sensors embedded in mobile devices, MCS can be used in a large variety of real-world applications, such as location recommendation [4], air quality monitoring [5], traffic information sharing [6], and point-of-interest characterization [7]. As a representative example, *WAZE* is a GPS navigation software owned by *Google*, which can provide real-time services like traffic monitoring and route recommendation with users' smartphones and tablets [8]. With the pervasive sensor-embedded smart devices, the MCS platform can leverage large-scale and long-term crowdsensing services, which can hardly be finished by traditional sensing systems [9], [10].

In a typical MCS application, a crowd of mobile participants, namely workers, are selected by the service provider to collect and outsource their real-time sensing data. Evidently, worker selection is one of the most fundamental problems in MCS, and competent workers can significantly improve the quality and reliability of the sensing tasks. However, to achieve a satisfactory result for worker selection, it is inevitable for the participating workers to share some of their sensitive information, e.g., trustability, working experience, real-time location. In particular, a hostile MCS platform may exploit workers' trustability for unfair competition [11]; an attacker may infer workers' daily behaviors by analyzing their working experience of different MCS tasks. As a result, workers may be reluctant to take part in the MCS tasks due to the potential privacy disclosure [12]. To stimulate workers' interests and enthusiasm, an ideal MCS platform should protect workers' sensitive information while selecting proper workers efficiently and effectively.

Recently, numerous studies focused on privacy-preserving worker selection in MCS applications [6], [9]–[13]. However, instead of considering the dynamic situations, most of the studies treat MCS as a short-term and static process. In many real-world problems, a sensing task is required to be executed continuously over a long-time period. For example, a traffic monitoring platform may observe and detect car speeds for weeks; an environment evaluation system may keep track of the local air quality for months. Unlike a static system, such sensing problems need to select proper workers repeatedly in a long-term manner. Moreover, since the workers can continuously arrive at/leave the system, the stream of sensing data is always time-sensitive. The platform is generally more interested in the recent data than those in the far past. Nevertheless, most of the studies only take workers' non-temporal characteristics (e.g., asking price and real-time location) into account. The recency and time-sensitivity of the mobile workers have not been adequately studied in the past.

Aiming to address the issues above, we propose an efficient and privacy-preserving worker selection scheme for MCS applications. More specifically, a probabilistic skyline based approach is proposed for continuously selecting workers over sliding windows. Furthermore, an ElGamal encryption-based scheme is designed for securely outsourcing and comparing workers' sensitive information. Overall, the main contributions

X. Zhang, R. Lu, and A. Ghorbani are with the Canadian Institute for Cybersecurity, Faculty of Computer Science, University of New Brunswick, Fredericton, Canada E3B 5A3. e-mail: (xichen.zhang@unb.ca, rlu1@unb.ca, ghorbani@unb.ca).

J. Shao is with School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China 310018 e-mail: (chn.junshao@gmail.com).

H. Zhu is with School of Cyber Engineering, Xidian University, Xi'an, China 710126 e-mail: (zhuhui@xidian.edu.cn).

Manuscript received February xx, 2020.

of this work are three-fold as follows.

• First, we devote our attention to the problem of privacy-preserving and continuous worker selection in MCS services. By deploying probabilistic skyline queries over sliding windows, our approach can select qualified workers for each current window in terms of working experience, expiry time, and trustability. The MCS system can be kept reliable and sustainable in different application scenarios.

• Second, we design novel encryption schemes for securely outsourcing and comparing workers' sensitive attributes (i.e., working experience and trustability) without disclosing their real values. These schemes are later used for confidentially determining workers' skyline dominance relationships and calculating workers' probabilistic skyline values in worker selection.

• Third, we analyze the security of the proposed scheme and show that it can effectively preserve the privacy of workers' personal data. Besides, extensive performance evaluations are conducted on both real-world and simulated datasets, and the results demonstrate that our approach outperforms the baseline method in different application scenarios.

The remainder of this paper is organized as follows. In Section II, we introduce our system model, security model and design goals. In Section III, we describe some preliminaries. In Section IV, we present the proposed scheme in details. Then in Section V and Section VI, we respectively present the security analysis and performance evaluation, followed by the related works in Section VII. Finally, we recap the conclusions in Section VIII.

II. Models and Design Goals

In this section, we formalize our system model, security model, and identify our design goals.

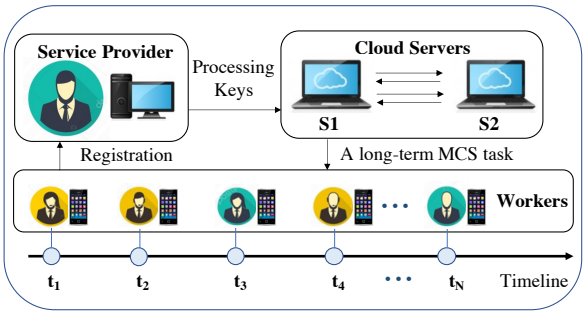


Fig. 1. The overview of the system model

A. System Model

In our system model, a long-term MCS task is denoted as $\mathcal{M}$, where workers can continuously arrive at/leave the MCS system. More specifically, we design an MCS worker selection scheme based on continuous probabilistic skyline computation. Our system model consists of four entities, namely a service provider $\mathcal{SP}$, two cloud servers $(\mathcal{S}_1, \mathcal{S}_2)$, and a set of participating workers $\mathcal{W} = \{w_1, w_2, \cdots\}$.

• Service Provider ($\mathcal{SP}$): $\mathcal{SP}$ is the service organizer and provider, and is responsible for bootstrapping the entire system. $\mathcal{SP}$ generates and distributes proper keys to different authorized entities so that a certain task can be completed cooperatively. Besides, $\mathcal{SP}$ assigns and computes a trustability score T_i to every registered worker w_i, which should be securely outsourced to $\mathcal{S}_1$ and $\mathcal{S}_2$ before a sensing task starts. After the task finishes, $\mathcal{SP}$ updates workers' T_i based on their sensing performance by the method in [9]. We assume that all parties fully trust $\mathcal{SP}$ in the system.

• Cloud Servers $(\mathcal{S}_1, \mathcal{S}_2)$: There are two cloud servers in our system model. After $\mathcal{M}$ starts, $\mathcal{S}_1$ and $\mathcal{S}_2$ will work together to select reliable workers over sliding windows. Concretely, when a new worker arrives, $\mathcal{S}_1$ and $\mathcal{S}_2$ need to update the subset of suitable workers for the current window based on their probabilistic skyline values.

• Workers $\mathcal{W} = \{w_1, w_2, \cdots\}$: Workers are the participants who wish to conduct $\mathcal{M}$. In our system model, each worker w_i is associated with a trustable score $T_i \in (0, 1)$, which can be considered as the trustable level of w_i. Higher T_i means w_i is more reliable for collecting and sharing high-quality sensing data. Once w_i plans to conduct $\mathcal{M}$, she/he is required to submit some real-time information to the cloud servers, e.g., arrival time $t_i^{\mathtt{arr}}$, the life span of the sensing data $t_i^{\mathtt{ls}}$, and working experience $\mathcal{E}_i$. w_i's expiry time $t_i^{\mathtt{exp}}$ can be calculated as $t_i^{\mathtt{exp}} = t_i^{\mathtt{arr}} + t_i^{\mathtt{ls}}$. After that, the cloud servers can select the suitable workers based on these information. In this work, $\mathcal{E}_i$ is defined as an integer within certain range (e.g., [1, 30]) to reveal how often w_i is allocated to conduct similar tasks in the past and the larger value is more preferred.

B. Security Model

In our security model, we consider $\mathcal{SP}$ is trustable, while $\mathcal{S}_1$ and $\mathcal{S}_2$ are honest-but-curious, which means both of them strictly follow the protocol procedure, yet may be curious to learn additional personal information in the process of worker selection. In addition, there is no collusion between $\mathcal{S}_1$ and $\mathcal{S}_2$. For the workers, we assume that they are strategic and selfish for maximizing their profits. Specifically, our model selects workers in terms of working experience, expiry time, and trustability. Therefore, if a worker w_i provides dishonest personal information (e.g., a larger $\mathcal{E}_i$ or $t_i^{\mathtt{ls}}$), she/he may have a higher probability to be selected. However, w_i' trustability T_i will be evaluated and updated based on his/her sensing performance after each task finishes [9]. So the value of T_i will be decreased largely if w_i submits false information, and the sensing results do not match the provided information.

Moreover, each worker ID can only be registered once, so malicious attackers cannot participate in $\mathcal{M}$ with multiple IDs. As a result, for long-term considerations, in order to be selected again in the future, the workers need to provide their personal information as correct as possible. It is worth noting that there may exist outside attackers who want to exploit the vulnerability of the platform and try to monitor and modify the sensing data, but they are beyond the scope of this work, and will be discussed in our future work.

TABLE I
THE SUMMARY OF NOTATIONS

Notation	Definition
System Notations	
$\mathcal{SP}$	Service provider
$(\mathcal{S}_1, \mathcal{S}_2)$	Two cloud severs
$\mathcal{M}$	A long-term MCS Task
Worker Notations	
$\mathcal{W} = \{w_1, w_2, \cdots\}$	A set of registered workers
$\mathcal{W}_c = \{w_1, \cdots, w_{win}\}$	The N_{win} workers in current window
w_i	The i^{th} worker
N_i	# of workers who do not dominate w_i
T_i	Trustability of worker w_i
$\overline{T}_i$	$1 - T_i$
$t_i^{\mathtt{arr}}$	Arrival time of w_i
$t_i^{\mathtt{ls}}$	Life span of w_i's sensing data
$t_i^{\mathtt{exp}}$	Expiry time of w_i's sensing data
$\Pr^{\mathtt{sky}}(w_i)$	Probabilistic skyline of worker w_i
Worker Selection Notations	
N_{win}	The size of each sliding window
$S_{\mathcal{E}}$	Sum of probabilistic work experience
$t_i^{\mathtt{rexp}}$	Relative departure time of w_i
$S_{\mathcal{E}}^{\mathtt{r}}$	$S_{\mathcal{E}} - \mathcal{E}_i$ when w_i leaves the system

C. Design Goals

This work aims to securely select a subset of reliable and trustable workers over sliding windows for conducting the MCS tasks. Specifically, the following two objectives should be satisfied.

• *Privacy preservation:* Our study needs to consider the leakage of workers' privacy since lots of sensitive information can be involved and inferred in the designed model. For example, a worker w_i's working experience $\mathcal{E}_i$ may indicate his/her daily routine and behavior, which should be highly protected. Moreover, w_i's trustability T_i can be used by MCS opponents for unfair competition. So, neither $\mathcal{S}_1$ nor $\mathcal{S}_2$ can get access to the real values of $\mathcal{E}_i$ and T_i during the process of worker selection.

• *Efficiency:* In order to preserve workers' privacy, certain secure comparison protocols should be designed for continuous skyline computation. Processing the encrypted data between cloud servers will bring extra computational cost, which should be minimized in our proposed scheme. More specifically, numerous sensing data can speedily arrive in our system model. In order to find the most qualified workers for executing the MCS task in a real-time fashion, the worker selection scheme should be performed efficiently without considering network delays.

III. PRELIMINARIES

In this section, we briefly introduce the background about probabilistic skyline computation over sliding windows. The frequently used notations are listed in Table I.

A. Skyline

Skyline computation is a well-known approach for multi-dimensional decision analysis [14] [15]. Given a dataset $\mathcal{D} = \{\gamma_1, \gamma_2, \gamma_3, ..., \gamma_D\}$ that contains D objects. Each object is in d dimensional space, where $\gamma_i = (\gamma[1], \gamma[2], \gamma[3], ..., \gamma[d])$ for $i \in [1, D]$. For simiplicity, we assume for each dimension, larger values are more preferred. Let γ_a and γ_b denote two different objects in $\mathcal{D}$ where $a, b \in [1, D]$ and $a \neq b$. We define γ_a dominates γ_b, denoted as $\gamma_a \prec \gamma_b$, if for all $k \in [1, d]$, $\gamma_a[k] \geq \gamma_b[k]$, and there exists at least one k such that $\gamma_a[k] > \gamma_b[k]$. The skyline points of $\mathcal{D}$ are all the objects that are not dominated by others in $\mathcal{D}$. If $\gamma_a \not\prec \gamma_b$, it means either $\gamma_b \prec \gamma_a$ or they do not dominate each other. Notably, $\gamma_a \not\prec \gamma_a$ (i.e., γ_a does not dominate itself) because we have $\gamma_a[k] = \gamma_a[k]$ for $k \in [1, d]$.

B. Probabilistic Skyline over Sliding Windows

Probabilistic skyline can compute the likelihood that one object to be selected as the skyline points [16] [17]. However, in many real-time monitoring platforms, objects appear sequentially. The arrival time and active time-span of the objects are essential aspects for deciding the dominance relationship between different objects. [18] studied the problem of operating probabilistic skyline over sliding windows, which can be described as follows.

Let $\mathcal{W}_c = \{w_1, w_2, ..., w_{N_{win}}\}$ denotes the subset of current N_{win} workers. Each worker w_i is associated with a trustability score $T_i \in (0, 1)$ for $i \in [1, N_{win}]$. In addition, we define $\overline{T}_i = 1 - T_i$ as the complement of workers' trustability and $\overline{T}_i \in (0, 1)$ as well. According to [18], we use $\overline{\Pr}(w_i)$ to denote the probability that w_i is not dominated by any other workers in $\mathcal{W}_c$, so $\overline{\Pr}(w_i)$ can be calculated as:

$$\overline{\Pr}(w_i) = \prod_{\substack{w_k \in \mathcal{W}_c, \\ i \neq k, w_k \prec w_i}} (1 - T_k) \cdot 1^{N_i} = \prod_{\substack{w_k \in \mathcal{W}_c, \\ i \neq k, w_k \prec w_i}} \overline{T}_k \cdot 1^{N_i}. \quad (1)$$

Let $\mathcal{W}_{c/i}$ denote the set of workers in $\mathcal{W}_c$ (not include w_i) who do not dominate w_i, i.e., $\mathcal{W}_{c/i} = \{w_k | w_k \in \mathcal{W}_c, w_k \prec w_i, k \neq i\}$. In Eq. (1), N_i is the cardinality of $\mathcal{W}_{c/i}$, i.e., $N_i = |\mathcal{W}_{c/i}|$.

Let $\Pr^{\mathtt{sky}}(w_i)$ denote w_i's probabilistic skyline value that indicates the possibility that w_i appears in the skyline of $\mathcal{W}_c$. $\Pr^{\mathtt{sky}}(w_i)$ is calculated based on workers' trustability, which is shown in the following equation

$$\Pr^{\mathtt{sky}}(w_i) = T_i \cdot \overline{\Pr}(w_i) = T_i \cdot \prod_{\substack{w_k \in \mathcal{W}_c, \\ i \neq k, w_k \prec w_i}} \overline{T}_k \cdot 1^{N_i}. \quad (2)$$

To better present how probabilistic skyline over sliding windows can be used for worker selection, a motivating example is given as follows.

Example 1: A city government plans to recruit numbers of workers to monitor local air quality for one month. The skyline computation depends on worker's working experience and the expiry time of their sensing data. In addition, each worker is associated with a trustable score which is derived from historical reporting records, and can be used to indicate the trustability of this worker. Due to the large-scale and fast arrival of the sensing data, the platform needs to select workers efficiently over sliding windows. For simplicity, we assume that the platform only focuses on the most 5 recent sensing data points. In Table II, each worker w_i' personal information is listed for $i \in [1, 6]$, including worker $\mathtt{ID}_i$,

TABLE II
AN EXAMPLE OF AIR QUALITY MOINTORING SYSTEM

ID	$t^{\mathtt{arr}}$	$t^{\mathtt{ls}}$	$t^{\mathtt{exp}}$	$\mathcal{E}$	T	$\overline{T}$
w_1	9:01 am	11 min	9:12 am	21	0.8500	0.1500
w_2	9:02 am	15 min	9:17 am	15	0.8500	0.1500
w_3	9:03 am	10 min	9:13 am	14	0.9500	0.0500
w_4	9:05 am	12 min	9:17 am	26	0.5500	0.4500
w_5	9:08 am	10 min	9:18 am	27	0.4500	0.5500
w_6	9:11 am	16 min	9:27 am	18	0.6000	0.4000

arrival time t_i^{arr}, the life span of the sensing data t_i^{ls}, the expiry time of the sensing data t_i^{exp} (i.e., $t_i^{exp} = t_i^{arr} + t_i^{ls}$), working experience $\mathcal{E}_i$, trustability T_i and the complement of trustability $\overline{T}_i$. Specifically, at time 9:10 am, w_1 to w_5 are the qualified candidates, since they are the only and most 5 recent workers, and none of their sensing data expires. However, when w_6 arrives at 9:11 am, this platform needs to decide which 5 workers can be kept in the system among the 6 candidates. Here, probabilistic skyline is computed based on Eq. (2) and is used to rank the qualification of workers. For instance, if we want to compute $Pr^{sky}(w_3)$, we first know that w_3 is dominated by w_2, w_4, w_5 and w_6 (see in Fig. 2), then $\overline{Pr}(w_3) = (1 - T_2) \cdot (1 - T_4) \cdot (1 - T_5) \cdot (1 - T_6) = \overline{T}_2 \cdot \overline{T}_4 \cdot \overline{T}_5 \cdot \overline{T}_6 = 0.0149$, and finally $Pr^{sky}(w_3) = T_3 \cdot \overline{Pr}(w_3) = 0.9500 * 0.0149 = 0.0141$. After we computing all workers' probabilistic skyline values, the worker with the minimum value will be removed from the system, while the rest of the workers will be kept for conducting $\mathcal{M}$.

From this example, we can see that probabilistic skyline computation over sliding windows is a practical and useful approach for worker selection. It can consider the trade-off between multiple criteria (e.g., working experience, recency of the sensing data, longer sojourn time) and discount the dominating criteria with too low trustability. Therefore, the workers with higher probabilistic skyline values are more reliable and trustable for fulfilling the task.

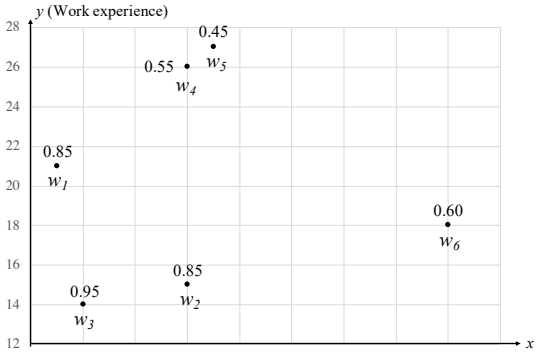


Fig. 2. The illustration of the example in Table II of calculating probabilistic skyline over sliding window based on working experience and life span

IV. THE PROPOSED SCHEME

In this section, we present the probabilistic skyline based scheme for continuous and secure worker selection in MCS applications. The proposed scheme mainly consists of the following five phases, namely, System Preparation (`SysPre`), Outsourcing Workers' Information (`OutInf`), Comparing Workers' Working Experience (`CmpExp`), Comparing Workers' Probabilistic Skyline Values (`CmpSky`), and Worker Selection (`WrkSel`). First, in `SysPre` phase, $\mathcal{SP}$ performs some preparations for the sensing task $\mathcal{M}$, including registering and authenticating workers, generating and distributing proper keys to the correct entities, outsourcing workers' trustability to the cloud servers, and releasing the details of the sensing task $\mathcal{M}$ to both $(\mathcal{S}_1, \mathcal{S}_2)$ and $\mathcal{W}$. Once a registered worker w_i wants to participate in $\mathcal{M}$, he/she needs to outsource his/her working experience T_i, arrival time $t_i^{\mathtt{arr}}$, and the life span of the sensing data $t_i^{\mathtt{ls}}$ to the cloud servers in `OutInf` phase. Upon receiving workers' information, the cloud servers needs to compare the relation of their working experience and determine their skyline dominance relationship in `CompExp` phase. Consequently, the cloud servers can compare workers' probabilistic skyline values in `CmpSky` phase and finally select the suitable workers for conducting the sensing task $\mathcal{M}$ in `WrkSel` phase. The details of the five phases are introduced as follows.

A. The `SysPre` Phase

In this phase, $\mathcal{SP}$ will run the following procedures for preparing and initializing the system before the sensing task $\mathcal{M}$ starts.

• *Registration:* Once a worker w_i with identity $\mathtt{ID}_i$ wants to register him/herself to the system, $\mathcal{SP}$ first validates the authenticity of $\mathtt{ID}_i$ and check whether this $\mathtt{ID}_i$ has been previously registered or not. If the $\mathtt{ID}_i$ is authentic and valid (i.e., it has not been resigstered before), then $\mathcal{SP}$ uses a cryptographic hash function to compute a pseudo-id $\mathtt{PID}_i$ based on $\mathtt{ID}_i$ [19]. Moreover, w_i's trustability T_i is initialized as 0.5. For $\mathcal{S}_1$ and $\mathcal{S}_2$, $\mathcal{SP}$ needs to authenticate their identities as well before $\mathcal{M}$ starts. Once w_i starts sensing the data, the pseudo-id $\mathtt{PID}_i$ needs to be authenticated again to guarantee that only registered workers can participate in $\mathcal{M}$. After $\mathcal{M}$ finishes, $\mathcal{SP}$ updates each T_i in an offline manner based on worker w_i's sensing performance [9]. The updated T_i will be used for worker selection in the next sensing task.

• *Key Generation and Distribution:* $\mathcal{SP}$ generates a key pair $(pk_{s_1}, sk_{s_1}) = ((p, q, g, y), x)$ of ElGamal encryption. Specifically, p and q are large primes and $q|(p-1)$. Let $\mathbb{Z}_p^* = \{1, 2, ..., p-1\}$, and $\mathbb{Z}_q^* = \{1, 2, ..., q-1\}$, g is a generator of a subgroup with order q in $\mathbb{Z}_p^*$, $y = g^x \bmod p$ where x is a random number from $\mathbb{Z}_q^*$. Moreover, $\mathcal{SP}$ selects another generator $\tilde{g}$ such that $\tilde{g} \neq g$, and creates a bloom filter BF based on Algorithm 1. After that, pk_{s_1} is published as system parameter to all the entities, and $\mathcal{SP}$ securely sends (sk_{s_1}, BF) and $\tilde{g}$ to $\mathcal{S}_1$ and $\mathcal{W}$, respectively.

• *Outsourcing Workers' Trustability to $\mathcal{S}_1$ and $\mathcal{S}_2$:* As we mentioned before, each worker w_i's T_i and $\overline{T}_i$ are owned by $\mathcal{SP}$, and both are essential for worker selection. In this work, we design a privacy-preserving scheme based on ElGamal encryption, which enables $\mathcal{SP}$ to securely outsource T_i and $\overline{T}_i$ to the cloud servers. The details of the proposed scheme are described as follows.

Algorithm 1: Generation of a bloom filter

Input : A N-bit length array $A[N]$ where all the bits are initialized to 0, k independent hash functions $\mathcal{H} = \{H_1, H_2, ..., H_k\}$ where $H_n : \{0,1\}^* \rightarrow \{0, 1, ..., N-1\}$ for $n \in [1, k]$, and $\mathcal{E}_{max}$ which indicates the maximum value of working experience (e.g., $\mathcal{E}_{max} = 30$).

Output: The bloom filter BF that contains all the elements of $o = \tilde{g}^{i-j} \bmod p$ for $i, j \in [1, \mathcal{E}_{max}]$ and $i > j$

```
for i = 1 to E_max do
    for j = 1 to E_max do
        if i > j then
            o = g̃^(i−j) mod p
            for l = 1 to k do
                set A[H_l(o)] = 1
return The bloom filter BF
```

Step 1. T_i and $\overline{T}_i$ are both in the range of (0, 1) where the modulo operations can not be adopted (e.g., $T_i = 0.6755$, and $\overline{T}_i = 0.3245$). So at the beginning, $\mathcal{SP}$ multiplies both T_i and $\overline{T}_i$ by a large integer θ (e.g., $\theta = 10^4$). After that, we can get $T_i = T_i \cdot \theta = 6755$ and $\overline{T}_i = \overline{T}_i \cdot \theta = 3245$, such that the modulo operations can be applied in the further steps. It is worth noting that after expansion, $T_i + \overline{T}_i = 10^4$ rather than 1. So the Eq. (2) will be modified as:

$$\Pr^{\mathtt{sky}}(w_i) = T_i \cdot \overline{\Pr}(w_i) = T_i \cdot \prod_{\substack{w_k \in \mathcal{W}_c, \\ i \neq k, w_k \prec w_i}} \overline{T}_k \cdot 10^{4N_i}. \quad (3)$$

Step 2. $\mathcal{SP}$ selects two random numbers $r_i, \tilde{r}_i \in \mathbb{Z}_q$, and then encrypts T_i and $\overline{T}_i$ as follows:

$$\begin{aligned} T'_{i1} &= T_i \cdot y^{r_i} \bmod p, \quad and \quad T'_{i2} = g^{r_i} \bmod p \\ \overline{T}'_{i1} &= \overline{T}_i \cdot y^{\tilde{r}_i} \bmod p, \quad and \quad \overline{T}'_{i2} = g^{\tilde{r}_i} \bmod p \end{aligned} \quad (4)$$

Step 3. $\mathcal{SP}$ securely distributes $(T'_{i1}, \overline{T}'_{i1})$ and $(T'_{i2}, \overline{T}'_{i2})$ to $\mathcal{S}_2$ and $\mathcal{S}_1$, respectively.

• *Launching the Sensing Task $\mathcal{M}$*: After outsourcing workers' trustability, $\mathcal{SP}$ releases the detailed information of $\mathcal{M}$ to $\mathcal{S}_1, \mathcal{S}_2$, and $\mathcal{W}$, e.g., the task content, location, and starting/ending time. Moreover, we assume that only a small number of workers are useful and can be kept in each sliding window. So, $\mathcal{SP}$ defines N_{win} as the size of each sliding window and announces N_{win} to $\mathcal{S}_1$ and $\mathcal{S}_2$ as well. In `WrkSel` phase, at most N_{win} workers can be selected to fulfill $\mathcal{M}$ in each window. Usually, N_{win} is a small integer (e.g., 5 or 10), which can be determined based on previous task experience.

B. The `OutInf` Phase

After $\mathcal{M}$ starts, any worker w_i who wants to participate in $\mathcal{M}$ needs to send his/her working experience $\mathcal{E}_i$, arrival time $t_i^{\mathtt{arr}}$, life span of the sensing data $t_i^{\mathtt{ls}}$ to the cloud servers. As mentioned before, $\mathcal{E}_i$ is sensitive and should be protected from being disclosed. The following details describe how w_i outsources $\mathcal{E}_i$, $t_i^{\mathtt{arr}}$, and $t_i^{\mathtt{ls}}$ to $\mathcal{S}_1$ and $\mathcal{S}_2$.

Step 1. At first, w_i generates a random number $\bar{r}_i \in \mathbb{Z}_q$, and computes $\mathcal{E}'_{i1}$ and $\mathcal{E}'_{i2}$ as follows:

$$\mathcal{E}'_{i1} = \tilde{g}^{\mathcal{E}_i} \cdot y^{\bar{r}_i} \bmod p, \ and \ \ \mathcal{E}'_{i2} = g^{\bar{r}_i} \bmod p \quad (5)$$

Step 2. Then, w_i calculates expiry time $t_i^{\mathtt{exp}}$ by $t_i^{\mathtt{exp}} = t_i^{\mathtt{arr}} + t_i^{\mathtt{ls}}$, where $t_i^{\mathtt{arr}}$ indicates his/her arrival time, and $t_i^{\mathtt{ls}}$ indicates the how long he/she plans to stay in the system.

Step 3. Finally, w_i securely distributes $(\mathcal{E}'_{i1}, t_i^{\mathtt{exp}})$ and $\mathcal{E}'_{i2}$ to $\mathcal{S}_2$ and $\mathcal{S}_1$, respectively.

C. The `CmpExp` Phase

In this phase, for w_i and w_j, $\mathcal{S}_2$ and $\mathcal{S}_1$ need to compare the relation of their working experience and determine their skyline dominance relationship.

• *Comparing workers' working experience*: Both $\mathcal{E}_i$ and $\mathcal{E}_j$ are sensitive information and should be protected from being disclosed. The following steps are conducted by $\mathcal{S}_1$ and $\mathcal{S}_2$ for comparing $\mathcal{E}_i$ and $\mathcal{E}_j$ without revealing their real values.

Step 1. Given $\mathcal{E}'_{i1}$ and $\mathcal{E}'_{j1}$ respectively from w_i and w_j, $\mathcal{S}_2$ calculates C_1 by Eq. (6) and then sends C_1 to $\mathcal{S}_1$.

$$C_1 = \frac{\mathcal{E}'_{i1}}{\mathcal{E}'_{j1}} = \frac{\tilde{g}^{\mathcal{E}_i} \cdot y^{\bar{r}_i}}{\tilde{g}^{\mathcal{E}_j} \cdot y^{\bar{r}_j}} = \tilde{g}^{(\mathcal{E}_i - \mathcal{E}_j)} \cdot y^{(\bar{r}_i - \bar{r}_j)} \bmod p \quad (6)$$

Step 2. Upon receiving C_1, $\mathcal{S}_1$ calculates C_2 as follows:

$$C_2 = \frac{C_1 \cdot \mathcal{E}'^{x}_{j2}}{\mathcal{E}'^{x}_{i2}} = \frac{\tilde{g}^{(\mathcal{E}_i - \mathcal{E}_j)} \cdot y^{(\bar{r}_i - \bar{r}_j)}}{y^{(\bar{r}_i - \bar{r}_j)}} = \tilde{g}^{(\mathcal{E}_i - \mathcal{E}_j)} \bmod p \quad (7)$$

Step 3. After obtaining C_2, $\mathcal{S}_1$ can determine the relation between $\mathcal{E}_i$ and $\mathcal{E}_j$ as follows: If $C_2 = 1$, then $\mathcal{E}_i = \mathcal{E}_j$. Otherwise, $\mathcal{S}_1$ needs to check whether C_2 is in the BF or not: if yes, then $\mathcal{E}_i > \mathcal{E}_j$, if no, then $\mathcal{E}_i < \mathcal{E}_j$. After that, $\mathcal{S}_1$ securely sends the comparison result to $\mathcal{S}_2$.

The correctness of *Step 3* is as follows. If $C_2 = 1$, it is easy to have $\mathcal{E}_i - \mathcal{E}_j = 0$ and $\mathcal{E}_i = \mathcal{E}_j$. Otherwise, under the condition of $\mathcal{E}_i > \mathcal{E}_j$, we know that all the possible values of C_2 have already been added into the BF based on Algorithm 1. Therefore, by checking whether C_2 is in BF or not, $\mathcal{S}_1$ can easily know whether $\mathcal{E}_i > \mathcal{E}_j$ or $\mathcal{E}_i < \mathcal{E}_j$. Notably, since $\tilde{g}$ is only securely shared to $\mathcal{W}$ in the `SysPre` phase, $\mathcal{S}_1$ can only determine the relation between $\mathcal{E}_i$ and $\mathcal{E}_j$. Without knowing $\tilde{g}$, $\mathcal{S}_1$ has no idea on the real values of $\mathcal{E}_i$ and $\mathcal{E}_j$.

• *Comparing workers' skyline dominance relationship*: The skyline dominance relationship between each pair of workers is captured in terms of working experience and expiry time, where for these two attributes, the larger values are more preferred. Specifically for workers w_i and w_j, i) $\mathcal{S}_1$ can securely compare the relation between $\mathcal{E}_i$ and $\mathcal{E}_j$ and send the relation to $\mathcal{S}_2$, and ii) $t_i^{\mathtt{exp}}$ and $t_j^{\mathtt{exp}}$ can be readily compared by $\mathcal{S}_2$ in their plaintexts. Therefore, the skyline dominance relationship between w_i and w_j can be easily determined by $\mathcal{S}_2$ (see in Algorithm 2).

D. The `CmpSky` Phase

In this phase, $\mathcal{S}_1$ and $\mathcal{S}_2$ can compare the relation between $\Pr^{sky}(w_i)$ and $\Pr^{sky}(w_j)$ for w_i and w_j. During the comparison, their trustability should not be disclosed to other entities.

Let $\Pr^{\mathtt{sky}}_{max}$ denote the possible largest value for $\Pr^{sky}(w_i)$. Given $T_i, \overline{T}_k \in (0, 10^4)$, $\Pr^{\mathtt{sky}}_{max} = 10^4 \cdot \prod_1^{N_{win}} 10^4 = 10^{4(N_{win}+1)}$, where N_{win} is the window size. In our system model, the largest value for N_{win} is 20, then $\Pr^{\mathtt{sky}}_{max} = 10^{84}$

Algorithm 2: Workers' skyline dominance relationship comparison *WorkerDom*(w_i, w_j)

Input : $t_i^{\exp}, t_j^{\exp}, BF, C_2$
Output: The skyline dominance relationship between w_i and w_j
if C_2 *in* BF **then**
  if $t_i^{exp} \geq t_j^{exp}$ **then return** $w_i \prec w_j$;
  if $t_i^{exp} < t_j^{exp}$ **then return** $w_i \not\prec w_j$ *and* $w_j \not\prec w_i$;
if C_2 *not in* BF **then**
  if $t_i^{exp} \leq t_j^{exp}$ **then return** $w_j \prec w_i$;
  if $t_i^{exp} > t_j^{exp}$ **then return** $w_i \not\prec w_j$ *and* $w_j \not\prec w_i$;
if $C_2 = 1$ **then**
  if $t_i^{exp} > t_j^{exp}$ **then return** $w_i \prec w_j$;
  if $t_i^{exp} = t_j^{exp}$ **then return** $w_i \not\prec w_j$ *and* $w_j \not\prec w_i$;
  if $t_i^{exp} < t_j^{exp}$ **then return** $w_j \prec w_i$;

and the big length $|\mathrm{Pr}^{\mathrm{sky}}_{max}| = 280$. Hence, $\mathrm{Pr}^{\mathrm{sky}}(w_i)$ is a large value $\in \{0,1\}^{280}$, and the method of building a bloom filter for comparing small integers in phase `CmpExp` is computationally expensive. As a result, we design a novel ElGamal-based comparison approach for large values, which is introduced as follows.

Step 1. $\mathcal{S}_1$ generates the skyline dominance relationships for all pairs of workers in the current window, and then securely sends the relationships to $\mathcal{S}_2$.

Step 2. $\mathcal{S}_2$ first selects a random number $\alpha \in \mathbb{Z}_p$. Then, based on Eq. (8), $\mathcal{S}_2$ calculates C_3 and C_4 for w_i and w_j, respectively. Finally $\mathcal{S}_2$ sends $(C_3,\ C_4)$ to $\mathcal{S}_1$.

$$\begin{aligned} C_3 &= \alpha \cdot T'_{i1} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} \overline{T}'_{k1} \cdot 10^{4N_i} \bmod p, \\ C_4 &= \alpha \cdot T'_{j1} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq j, w_k \prec w_j}} \overline{T}'_{k1} \cdot 10^{4N_j} \bmod p, \end{aligned} \tag{8}$$

where N_i and N_j represent the number of workers who do not dominate w_i and w_j, respectively.

Step 3. Upon receiving C_3, $\mathcal{S}_1$ checks the dominance relationships between w_i and other workers and find all the $\overline{T}'_{k2}$ for w_k where $w_k \prec w_i$. After that, $\mathcal{S}_1$ calculates C_5 for w_i as:

$$\begin{aligned} C_5 &= \frac{C_3}{(T'_{i2} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} \overline{T}'_{k2})^x} \bmod p \\ &= \frac{\alpha \cdot T'_{i1} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} \overline{T}'_{k1} \cdot 10^{4N_i}}{(g^{r_i} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} g^{\tilde{r}_k})^x} \bmod p \\ &= \frac{\alpha \cdot T_i \cdot y^{r_i} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} \overline{T}_k \cdot y^{\tilde{r}_k} \cdot 10^{4N_i}}{(y^{r_i} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} y^{\tilde{r}_k})} \bmod p \\ &= \alpha \cdot (T_i \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq i, w_k \prec w_i}} \overline{T}_k \cdot 10^{4N_i}) \bmod p \\ &= \alpha \cdot \mathrm{Pr}^{\mathrm{sky}}(w_i) \bmod p \end{aligned} \tag{9}$$

Similarly, $\mathcal{S}_1$ calculates C_6 for worker w_j based on C_4 as:

$$\begin{aligned} C_6 &= \frac{C_4}{(T'_{j2} \cdot \prod_{\substack{w_k \in \mathcal{W}_c \\ k \neq j, w_k \prec w_j}} \overline{T}'_{k2})^x} \bmod p \\ &= \alpha \cdot \mathrm{Pr}^{\mathrm{sky}}(w_j) \bmod p \end{aligned} \tag{10}$$

After that, $\mathcal{S}_1$ generates a random number $\beta \in \mathbb{Z}_q$, calculates C_7 by Eq. (11), and then sends C_7 to $\mathcal{S}_2$.

$$C_7 = \alpha \cdot \beta \cdot (C_5 - C_6) = \alpha \cdot \beta \cdot (\mathrm{Pr}^{\mathrm{sky}}(w_i) - \mathrm{Pr}^{\mathrm{sky}}(w_j)) \bmod p \tag{11}$$

Step 4. After obtaining C_7, $\mathcal{S}_2$ calculates C_8 as follows.

$$C_8 = \frac{C_7}{\alpha} = \beta \cdot (\mathrm{Pr}^{\mathrm{sky}}(w_i) - \mathrm{Pr}^{\mathrm{sky}}(w_j)) \bmod p \tag{12}$$

Finally, $\mathcal{S}_2$ can determine the relation between $\mathrm{Pr}^{\mathrm{sky}}(w_i)$ and $\mathrm{Pr}^{\mathrm{sky}}(w_j)$ based on the value and bit length of C_8. Specifically, if $C_8 = 0$, then it means $\mathrm{Pr}^{\mathrm{sky}}(w_i) = \mathrm{Pr}^{\mathrm{sky}}(w_i)$. Otherwise, if $|C_8| < \frac{|p|}{2}$ then $\mathrm{Pr}^{\mathrm{sky}}(w_i) > \mathrm{Pr}^{\mathrm{sky}}(w_j)$; if $|C_8| > \frac{|p|}{2}$ then $\mathrm{Pr}^{\mathrm{sky}}(w_i) < \mathrm{Pr}^{\mathrm{sky}}(w_j)$.

The correctness of the comparison is as follows. First, given a non-negative integer β, it is easy to know that when $C_8 = \beta \cdot (\mathrm{Pr}^{\mathrm{sky}}(w_i) - \mathrm{Pr}^{\mathrm{sky}}(w_j)) \bmod p = 0$, then $\mathrm{Pr}^{\mathrm{sky}}(w_i) = \mathrm{Pr}^{\mathrm{sky}}(w_i)$. Second, as both $\mathrm{Pr}^{\mathrm{sky}}(w_i)$ and $\mathrm{Pr}^{\mathrm{sky}}(w_j) \in (0, \mathrm{Pr}^{sky}_{max}]$, we know that if $\mathrm{Pr}^{\mathrm{sky}}(w_i) > \mathrm{Pr}^{\mathrm{sky}}(w_j)$, then $C_8 \in (0, \beta \cdot \mathrm{Pr}^{\mathrm{sky}}_{max})$. Therefore, $|C_8| < |\beta| \cdot |\mathrm{Pr}^{\mathrm{sky}}_{max}| = 160 + 280 = 440 < \frac{|p|}{2} = 512$. Last, if $\mathrm{Pr}^{\mathrm{sky}}(w_i) < \mathrm{Pr}^{\mathrm{sky}}(w_j)$, then $\beta \cdot (\mathrm{Pr}^{\mathrm{sky}}(w_i) - \mathrm{Pr}^{\mathrm{sky}}(w_j)) < 0$ and $C_8 = (p + \beta \cdot (\mathrm{Pr}^{\mathrm{sky}}(w_i) - \mathrm{Pr}^{\mathrm{sky}}(w_j))) \bmod p$. Here, $|C_8| \sim |p|$, hence $|C_8| > \frac{|p|}{2}$.

E. The `WrkSel` Phase

After $\mathcal{M}$ starts, workers will continuously arrive at the system. Particularly, $\mathcal{W}_c = \{w_1, w_2, ..., w_{N_{win}}\}$ denotes the N_{win} workers in current sliding window, w_n denotes the newly arrived worker, (w_i, w_j) denotes any pair of workers in $\mathcal{W}_c$, $\mathrm{Pr}^{\mathrm{sky}}(w_i)$ and $\mathrm{Pr}^{\mathrm{sky}}(w_i)_n$ denote w_i's probabilistic skyline values before and after w_n arrives, $\mathcal{T}[i][j]$ denotes the value at the ith row and the jth column in $\mathcal{T}$. The following steps are performed by $\mathcal{S}_1$ and $\mathcal{S}_2$ for worker selection.

Step 1. Before w_n arrives the platform, $\mathcal{S}_1$ and $\mathcal{S}_2$ will build a global worker selection table $\mathcal{T}$ based on Algorithm 3 (an example is shown in Table III). There are three values (i.e., 0, 1, -1) in $\mathcal{T}$, which represents $\mathrm{Pr}^{\mathrm{sky}}(w_i) = \mathrm{Pr}^{\mathrm{sky}}(w_j), \mathrm{Pr}^{\mathrm{sky}}(w_i) > \mathrm{Pr}^{\mathrm{sky}}(w_j)$ and $\mathrm{Pr}^{\mathrm{sky}}(w_i) < \mathrm{Pr}^{\mathrm{sky}}(w_j)$, respectively.

Step 2. After w_n arrives, the platform needs to update the relations of probabilistic skyline among workers in $\mathcal{W}_c$ (i.e., to update the values of $\mathcal{T}[i][j]$). A naive way is to update the relations for all pairs of workers. However, it is inefficient by involving many repetitive calculations. For the workers who are not dominated by w_n, their probabilistic skyline values do not change and the comparisons between them should be avoided. To address this issue, we propose an efficient approach for updating $\mathcal{T}[i][j]$ based on the following cases.

Case 1. *Both w_i and w_j are dominated by w_n, (i.e., $w_n \prec w_i$ and $w_n \prec w_j$), denoted as $w_n \prec (w_i, w_j)$.*

Algorithm 3: Generation of global worker selection table $\mathcal{T}$

Input : $\mathcal{W}_c = \{w_1, w_2, ..., w_{N_{win}}\}$
Output: A two-dimensional array $\mathcal{T}$ that indicates the realtion of workers' probabilistic skyline values

Initialize $\mathcal{T}$ to an empty 2-dimensional array with the size of $(N_{win}+1) \times (N_{win}+1)$;
for $i = 1$ *to* N_{win} **do**
  for $j = 1$ *to* N_{win} **do**
    if $i = j$ **then** $\mathcal{T}[i][j] = 0$;
    // indicating $\Pr^{\text{sky}}(w_i) = \Pr^{\text{sky}}(w_j)$
    else
      $\mathcal{S}_1$ and $\mathcal{S}_2$ calculate C_8 based on phase CmpSky
      if $C_8 = 0$ **then** $\mathcal{T}[i][j] = 0$;
      // indicating $\Pr^{\text{sky}}(w_i) = \Pr^{\text{sky}}(w_j)$
      if $C_8 < \frac{|p|}{2}$ **then** $\mathcal{T}[i][j] = 1$ & $\mathcal{T}[j][i] = -1$;
      // indicating $\Pr^{\text{sky}}(w_i) > \Pr^{\text{sky}}(w_j)$
      if $C_8 > \frac{|p|}{2}$ **then** $\mathcal{T}[i][j] = -1$ & $\mathcal{T}[j][i] = 1$;
      // indicating $\Pr^{\text{sky}}(w_i) < \Pr^{\text{sky}}(w_j)$
return *The global worker selectiono table* $\mathcal{T}$

Theorem 1. *If $w_n \prec (w_i, w_j)$, the relation between $Pr^{sky}(w_i)$ and $Pr^{sky}(w_j)$ does not change, so does the value of $\mathcal{T}[i][j]$.*

Proof. Without loss of generality, we assume that $\Pr^{\text{sky}}(w_i) > \Pr^{\text{sky}}(w_j)$. Given $w_n \prec (w_i, w_j)$, it is easy to obtain the following equations based on Eq. (2):

$$\Pr^{\text{sky}}(w_i)_n = \Pr^{\text{sky}}(w_i) \cdot \overline{T}_n, \quad \Pr^{\text{sky}}(w_j)_n = \Pr^{\text{sky}}(w_j) \cdot \overline{T}_n \tag{13}$$

Obviously after w_n arrives, $\Pr^{\text{sky}}(w_i)_n > \Pr^{\text{sky}}(w_j)_n$ due to $\Pr^{\text{sky}}(w_i) > \Pr^{\text{sky}}(w_j)$. Therefore, the relation between $\Pr^{\text{sky}}(w_i)$ and $\Pr^{\text{sky}}(w_i)$ keeps the same, and the value of $\mathcal{T}[i][j]$ does not need to update. □

Case 2. *In this case, w_n does not dominate either w_i or w_j (i.e., $w_n \not\prec w_i$ and $w_n \not\prec w_j$), denoted as $w_n \not\prec (w_i, w_j)$.*

Theorem 2. *If $w_n \not\prec (w_i, w_j)$, the values of $Pr^{sky}(w_i)$ and $Pr^{sky}(w_i)$ do not change, so does the value of $\mathcal{T}[i][j]$.*

Proof. If $w_n \not\prec w_i$, then the subset of workers who dominate w_i stays the same, and the value of the following equation $\Pr^{\text{sky}}(w_i) = T_i \cdot \overline{\Pr}(w_i) = T_i \cdot \prod_{\substack{w_k \in \mathcal{W}_c, \\ i \neq k, w_k \prec w_i}} \overline{T}_k \cdot \prod_{n=0}^{N_i} 10^4 \bmod p$ does not change. The same conclusion can be made for w_j. Therefore, the values of $\Pr^{\text{sky}}(w_i)$ and $\Pr^{\text{sky}}(w_i)$ do not change, so does the value of $\mathcal{T}[i][j]$. □

Case 3. *In this case, $w_n \prec w_i$, $w_n \not\prec w_j$ and $\mathcal{T}[i][j] = -1$ (i.e., $Pr^{sky}(w_i) < Pr^{sky}(w_j)$), denoted as $w_n \otimes (w_i, w_j)$.*

Theorem 3. *If $w_n \otimes (w_i, w_j)$, then the relation between $Pr^{sky}(w_i)$ and $Pr^{sky}(w_i)$ does not change, so does the value of $\mathcal{T}[i][j]$.*

Proof. Based on the definition of probabilistic skyline, if $w_n \otimes (w_i, w_j)$, $\Pr^{\text{sky}}(w_i)_n$ and $\Pr^{\text{sky}}(w_j)_n$ can be calculated as:

$$\Pr^{\text{sky}}(w_i)_n = \Pr^{\text{sky}}(w_i) \cdot \overline{T}_n, \quad \Pr^{\text{sky}}(w_j)_n = \Pr^{\text{sky}}(w_j) \cdot 10^4 \tag{14}$$

In this case, $\Pr^{\text{sky}}(w_i) < \Pr^{\text{sky}}(w_j)$, and also $\overline{T}_n < 10^4$, then it is easy to have $\Pr^{\text{sky}}(w_i)_n < \Pr^{\text{sky}}(w_j)_n$. Therefore, the relation between $\Pr^{\text{sky}}(w_i)$ and $\Pr^{\text{sky}}(w_i)$ does not change, so does the value of $\mathcal{T}[i][j]$. □

TABLE III
AN EXAMPLE OF GLOBAL WORKER SELECTION TABLE $\mathcal{T}$

	w_1	w_2	w_3	w_4	w_5	w_n
w_1	0	-1	-1	-1	-1	-1
w_2	1	0	1	-1	-1	-1
w_3	1	-1	0	-1	-1	-1
w_4	1	1	1	0	-1	1
w_5	1	1	1	1	0	1
w_n	1	1	1	-1	-1	0

Case 4. *In this case, $w_n \prec w_i$, $w_n \not\prec w_j$ and $\mathcal{T}[i][j] = 1$ (i.e., $Pr^{sky}(w_i) > Pr^{sky}(w_j)$), which is denoted as $w_n \oplus (w_i, w_j)$.*

Theorem 4. *If $w_n \oplus (w_i, w_j)$, then the relation between $Pr^{sky}(w_i)_n$ and $Pr^{sky}(w_j)_n$ cannot be determined without calculation. So $\mathcal{S}_1$ and $\mathcal{S}_2$ need to compare their relation based on the method in Section IV-D, and update $\mathcal{T}[i][j]$ accordingly.*

Proof. Similar to Case 3, given $w_n \oplus (w_i, w_j)$, $\Pr^{\text{sky}}(w_i)_n$ and $\Pr^{\text{sky}}(w_j)_n$ can still be calculated by Eq. (14). However, only given $\Pr^{\text{sky}}(w_i) > \Pr^{\text{sky}}(w_j)$ and $\overline{T}_n < 10^4$, we cannot determine the relation of $\Pr^{\text{sky}}(w_i)_n$ and $\Pr^{\text{sky}}(w_j)_n$ without updating their real values. □

In summary, after w_n arrives the platform, $\mathcal{S}_1$ and $\mathcal{S}_2$ can update the relation between $\Pr^{\text{sky}}(w_i)_n$ and $\Pr^{\text{sky}}(w_j)_n$ (i.e., the values of $\mathcal{T}[i][j]$) according to the cases mentioned above.

Step 3. Next, $\mathcal{S}_1$ and $\mathcal{S}_2$ compare the relations of probabilistic skyline values between w_n and all the workers in $\mathcal{W}_c$, and then fill in $\mathcal{T}$. For example, we assume $\mathcal{W}_c = \{w_1, w_2, w_3, w_4, w_5\}$. If the ranking of probabilistic skyline values for all the workers is $w_5 > w_4 > w_n > w_2 > w_3 > w_1$, then after this step, $\mathcal{T}$ is shown in Table III. Finally, $\mathcal{S}_2$ can select the top-N_{win} workers based on the values in $\mathcal{T}$. Specifically, $\mathcal{S}_2$ checks all the rows in $\mathcal{T}$, and finds the row which does not contain 1 (i.e., the first row in Table III). This means that the worker in this row has the lowest probabilistic skyline value among all the workers (i.e., w_1 in this example). At last, w_1 will be removed from the platform, and the rest of the workers are kept for fulfilling $\mathcal{M}$.

V. SECURITY ANALYSIS

In this section, we will analyze the security properties of the proposed scheme. Notably, following the design goals illustrated in Section II-C, the analysis will focus on how our scheme is privacy-preserving in protecting workers' personal information, i.e., working experience and trustability, from being disclosed in the process of worker selection. Specifically, let $\mathcal{W}_c = \{w_1, w_2, .., w_{N_{win}}\}$ denote the N_{win} workers in the current window and for $\forall w_i \in \mathcal{W}_c$, we have the following two theorems.

Theorem 5. *The worker w_i's working experience $\mathcal{E}_i$ cannot be revealed during the whole process under the assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$ are honest-but-curious.*

Proof. We give the proof in four parts according to the corresponding roles in our scheme, i.e., $\mathcal{S}_1$, $\mathcal{S}_2$, other workers and outsiders.

- Under the assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$ are honest-but-curious, $\mathcal{S}_1$ can only obtain $\mathcal{E}'_{i2}$ and many $(C_{1,(i,j)}, C_{2,(i,j)})$'s in our scheme, where $i, j \in [1, N_{win}]$ and $i \neq j$. From the following equations

$$C_{1,(i,j)} = \tilde{g}^{\mathcal{E}_i - \mathcal{E}_j} \cdot g^{\bar{r}_i - \bar{r}_j} \bmod p,$$
$$C_{2,(i,j)} = \tilde{g}^{\mathcal{E}_i - \mathcal{E}_j} \bmod p, \quad \mathcal{E}'_{i2} = g^{\bar{r}_i} \bmod p,$$

we can see that $C_{2,(i,j)}$ comes from $C_{1,(i,j)}$, and the information related to $\mathcal{E}_i$ in $C_{2,(i,j)}$ is not less than that in $C_{1,(i,j)}$. Furthermore, $C_{2,(i,j)}$ is independent from $\mathcal{E}'_{i2}$ that contains nothing related to $\mathcal{E}_i$. Hence, we only need to analyze whether $\mathcal{S}_1$ can obtain $\mathcal{E}_i$ from $C_{2,(i,j)}$'s.
It is easy to see that $\mathcal{S}_1$ can obtain at most the following $N_{win} - 1$ independent equations regarding $C_{2,(i,j)}$'s.

$$\begin{cases} C_{2,(i,1)} &= \tilde{g}^{\mathcal{E}_i - \mathcal{E}_1} \bmod p \\ &\vdots \\ C_{2,(i,i-1)} &= \tilde{g}^{\mathcal{E}_i - \mathcal{E}_{i-1}} \bmod p \\ C_{2,(i,i+1)} &= \tilde{g}^{\mathcal{E}_i - \mathcal{E}_{i+1}} \bmod p \\ &\vdots \\ C_{2,(i,N_{win})} &= \tilde{g}^{\mathcal{E}_i - \mathcal{E}_{N_{win}}} \bmod p \end{cases}$$

Without loss of generality, we assume that all the worker experience values are different. Under the assumption that $\mathcal{S}_1$ is honest-but-curious, $\mathcal{S}_1$ faces $N_{win} + 1$ unknown values in the above equations. Those $N_{win} + 1$ unknown values are $(\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_{N_{win}}, \tilde{g})$. Hence, even if $\mathcal{E}_i$ has a small range, $\mathcal{S}_1$ cannot obtain any information about $\mathcal{E}_i$ from the above equations. In other words, $\mathcal{E}_i$ is kept secret from $\mathcal{S}_1$ in our scheme.
- Under the assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$ are honest-but-curious, $\mathcal{S}_2$ can only obtain $\mathcal{E}'_{i1}$ that contains information about $\mathcal{E}_i$, where $\mathcal{E}'_{i1} = \tilde{g}^{\mathcal{E}_i} \cdot y^{\bar{r}_i} \bmod p$. However, $\mathcal{S}_2$ cannot obtain $\mathcal{E}_i$ only from $\mathcal{E}'_{i1}$ according to the correctness and security of ElGamal encryption. Hence, $\mathcal{E}_i$ is kept secret from $\mathcal{S}_2$ in our scheme.
- Regarding worker w_j $(i \neq j)$, he/she only knows whether him/herself is chosen or not in our scheme. w_j does not even know which worker he/she is compared with. Hence, w_j has no idea about $\mathcal{E}_i$ in our scheme.
- Regarding the ones outside of our scheme, it is clear that $\mathcal{E}_i$ is kept secret from them, since they obtain less information than $\mathcal{S}_1$, $\mathcal{S}_2$ or workers in our scheme.

□

Theorem 6. *The worker w_i's trustability T_i cannot be revealed during the whole process under the assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$ are honest-but-curious.*

Proof. Given $T_i + \overline{T}_i = 10^4$, in order to protect T_i from being disclosed to other entities, our scheme also needs to guarantee that $\overline{T}_i$ is privacy-preserving in the whole process of worker selection. Specifically, the following parts are provided to prove the confidentiality of T_i to $\mathcal{S}_1$, $\mathcal{S}_2$, other workers, and outsiders.

- In the proposed scheme, $\mathcal{S}_1$ keeps both T'_{i2} and $\overline{T}'_{i2}$. According to the correctness and security of the ElGamal encryption, $\mathcal{S}_1$ can compute T_i or $\overline{T}_i$ as long as $\mathcal{S}_1$ can also obtain T'_{i1} or $\overline{T}'_{i1}$. Therefore, in order to prove this theorem, we need to demonstrate that $\mathcal{S}_1$ cannot obtain T_i, $\overline{T}_i$, T'_{i1}, and $\overline{T}'_{i1}$, respectively. The proofs are presented as follows.
  - *T_i will not be revealed to $\mathcal{S}_1$.* In our proposed scheme, the views of $\mathcal{S}_1$ that related to T_i are many $(C_{5,(i,j)}, C_{7,(i,j)})$'s, such that

$$C_{5,(i,j)} = \alpha_{(i,j)} \cdot T_i \cdot \overline{\mathbb{T}}_i \bmod p,$$
$$C_{7,(i,j)} = \alpha_{(i,j)} \cdot \beta_{(i,j)} \cdot (T_i \cdot \overline{\mathbb{T}}_i - T_j \cdot \overline{\mathbb{T}}_j) \bmod p,$$

where $\overline{\mathbb{T}}_i$ is the continuous multiplication of $\overline{T}_k$'s and $10^{4N_i} \bmod p$ if there exists $w_k \in \mathcal{W}_c$ and $w_k \prec w_i$. If no such worker exists (w_i is a skyline worker), then $\overline{\mathbb{T}}_i = 10^{4N_i} \bmod p$, i.e.,

$$\overline{\mathbb{T}}_i = \begin{cases} \prod \overline{T}_k \cdot 10^{4N_i} \bmod p & \exists w_k \in \mathcal{W}_c, w_k \prec w_i, \\ 10^{4N_i} \bmod p & \text{otherwise.} \end{cases}$$

Based on Eq. (11) in Section IV-D, we know that $C_{7,(i,j)}$ is a linear combination of $C_{5,(i,j)}$ and $C_{6,(i,j)}$, where $C_{6,(i,j)} = \alpha_{(i,j)} \cdot \beta_{(i,j)} \cdot T_j \cdot \overline{\mathbb{T}}_j \cdot 10^{4N_j} \bmod p$. Since $C_{6,(i,j)}$ contains nothing about T_i, so the information related to T_i in $C_{7,(i,j)}$ is not more than that in $C_{5,(i,j)}$. Therefore, we only need to analyze whether $\mathcal{S}_1$ can obtain T_i from $C_{5,(i,j)}$'s or not.
It is easy to know that $\mathcal{S}_1$ can get the following $N_{win} - 1$ equations for $C_{5,(i,j)}$'s in the process of worker selection.

$$\begin{cases} C_{5,(i,1)} &= \alpha_{(i,1)} \cdot T_i \cdot \overline{\mathbb{T}}_i \bmod p \\ &\vdots \\ C_{5,(i,i-1)} &= \alpha_{(i,i-1)} \cdot T_i \cdot \overline{\mathbb{T}}_i \bmod p \\ C_{5,(i,i+1)} &= \alpha_{(i,i+1)} \cdot T_i \cdot \overline{\mathbb{T}}_i \bmod p \\ &\vdots \\ C_{5,(i,N_{win})} &= \alpha_{(i,N_{win})} \cdot T_i \cdot \overline{\mathbb{T}}_i \bmod p \end{cases}$$

If we consider $\mathbb{T}_i$ as a whole variable, then depending on whether $\mathbb{T}_i = 10^{4N_i} \bmod p$ or not, $\mathcal{S}_1$ faces at least N_{win} and at most $N_{win} + 1$ unknown values in the above $N_{win} - 1$ equations. The definite N_{win} unknown values are $(\alpha_{(i,1)}, \alpha_{(i,2)}, ..., \alpha_{(i,i-1)}, \alpha_{(i,i+1)}, ...\alpha_{(i,n)}, T_i)$. Therefore, $\mathcal{S}_1$ cannot directly solve the above equations, i.e., $\mathcal{S}_1$ cannot obtain any information about T_i.
  - *$\overline{T}_i$ will not be revealed to $\mathcal{S}_1$.* Let $\mathcal{W}_c^i$ denote a subset of workers in $\mathcal{W}_c$ (i.e., $\mathcal{W}_c^i \in \mathcal{W}_c$) such that every worker in $\mathcal{W}_c^i$ is dominated by w_i, so for $\forall w_{k'} \in \mathcal{W}_c^i$, we have $w_i \prec w_{k'}$. If no such worker exists, then $\mathcal{W}_c^i = \emptyset$.
  It is worth noting that if $\mathcal{W}_c^i = \emptyset$, then $\mathcal{S}_1$ can get nothing related to $\overline{T}_i$. Without lossing of generality, we assume that $\mathcal{W}_c^i \neq \emptyset$, and there is at least one worker $w_{k'}$ in $\mathcal{W}_c^i$. Accordingly, in our proposed

scheme, the views of $\mathcal{S}_1$ related to $\overline{T}_i$ are many $(C_{5,(k',j)}, C_{7,(k',j)})$'s for $k' \neq j$, such that

$$C_{5,(k',j)} = \alpha_{(k',j)} \cdot T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} \bmod p,$$
$$C_{7,(k',j)} = \alpha_{(k',j)} \cdot \beta_{(k',j)} \cdot (T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} - T_j \cdot \overline{\mathbb{T}}_j) \bmod p,$$

where

$$\overline{\mathbb{T}}_{k'/i} = \begin{cases} \prod \overline{T}_k \cdot 10^{4N_{k'}} \bmod p & \exists w_k \in \mathcal{W}_c, \\ & w_k \prec w_{k'}, k \neq i, \\ 10^{4N_{k'}} \bmod p & \text{otherwise.} \end{cases}$$

Similarly, the information that related to $\overline{T}_i$ in $C_{7,(k',j)}$ is not more than that in $C_{5,(k',j)}$, so in the next, we only need to analyze whether $\mathcal{S}_1$ can obtain $\overline{T}_i$ from $C_{5,(k',j)}$'s or not.

It is easy to know that $\mathcal{S}_1$ can get at most the following $N_{win}-1$ equations regarding $C_{5,(k',j)}$'s in our proposed scheme.

$$\begin{cases} C_{5,(k',1)} = \alpha_{(k',1)} \cdot T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} \bmod p \\ \quad \vdots \\ C_{5,(k',k'-1)} = \alpha_{(k',k'-1)} \cdot T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} \bmod p \\ C_{5,(k',k'+1)} = \alpha_{(k',k'+1)} \cdot T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} \bmod p \\ \quad \vdots \\ C_{5,(k',N_{win})} = \alpha_{(k',N_{win})} \cdot T_{k'} \cdot \overline{T}_i \cdot \overline{\mathbb{T}}_{k'/i} \bmod p \end{cases}$$

Again, let consider $\overline{\mathbb{T}}_{k'/i}$ as a whole variable. It is easy to see that $\mathcal{S}_1$ faces at least $N_{win}+1$ and at most $N_{win}+2$ unknown values, depending on whether $\mathbb{T}_{k'/i} = 10^{4N_{k'}} \bmod p$ or not. The $N_{win}+1$ definite unknown values are $(\alpha_{(k',1)}, ..., \alpha_{(k',k'-1)}, \alpha_{(k',k'+1)}, ..., \alpha_{(k',N_{win})}, T_{k'}, \overline{T}_i)$. Accordingly, $\mathcal{S}_1$ cannot obtain any information about $\overline{T}_i$ from the above equations.

– T'_{i1} *will not be revealed to* $\mathcal{S}_1$. In the proposed scheme, the views of $\mathcal{S}_1$ that related to T'_{i1} are many $C_{3,(i,j)}$'s, such that

$$C_{3,(i,j)} = \alpha_{(i,j)} \cdot T'_{i1} \cdot \overline{\mathbb{T}}'_{i1} \bmod p,$$

where $\overline{\mathbb{T}}'_{i1}$ is the continuous multiplication of $\overline{T}'_{k1}$'s and $10^{4N_i} \bmod p$ if there exists $w_k \in \mathcal{W}_c$ such that $w_k \prec w_i$. If no such worker exists (w_i is a skyline worker), then $\overline{\mathbb{T}}'_{i1} = 10^{4N_i} \bmod p$, i.e.,

$$\overline{\mathbb{T}}'_{i1} = \begin{cases} \prod \overline{T}'_{k1} \cdot 10^{4N_i} \bmod p & \exists w_k \in \mathcal{W}_c, w_k \prec w_i, \\ 10^{4N_i} \bmod p & \text{otherwise.} \end{cases}$$

In the proposed scheme, $\mathcal{S}_1$ can get at most the following $N_{win}-1$ equations regarding $C_{3,(i,j)}$'s in our proposed scheme.

$$\begin{cases} C_{3,(i,1)} = \alpha_{(i,1)} \cdot T'_{i1} \cdot \overline{\mathbb{T}}'_{i1} \bmod p \\ \quad \vdots \\ C_{3,(i,i-1)} = \alpha_{(i,i-1)} \cdot T'_{i1} \cdot \overline{\mathbb{T}}'_{i1} \bmod p \\ C_{3,(i,i+1)} = \alpha_{(i,i+1)} \cdot T'_{i1} \cdot \overline{\mathbb{T}}'_{i1} \bmod p \\ \quad \vdots \\ C_{3,(i,N_{win})} = \alpha_{(i,N_{win})} \cdot T'_{i1} \cdot \overline{\mathbb{T}}'_{i1} \bmod p \end{cases}$$

Similarly, $\mathcal{S}_1$ will face at least N_{win} and at most $N_{win}+1$ unknown values in the above $N_{win}-1$ equations. The definite N_{win} unknown values are $(\alpha_{(i,1)}, ..., \alpha_{(i,i-1)}, \alpha_{(i,i+1)}, ..., \alpha_{(i,N_{win})}, T'_{i1})$. Therefore, $\mathcal{S}_1$ cannot compute T'_{i1}. Without T'_{i1}, based on the correctness and security of ElGalmal encryption, $\mathcal{S}_1$ cannot obtain T_i in the designed scheme.

– $\overline{T}'_{i1}$ *will not be revealed to* $\mathcal{S}_1$. In the proposed scheme, the view of $\mathcal{S}_1$ that related to $\overline{T}'_{i1}$ are many $C_{3,(k',j)}$, such that

$$C_{3,(k',j)} = \alpha_{(k',j)} \cdot T'_{k'1} \cdot \overline{T}'_{i1} \cdot \overline{\mathbb{T}}'_{k'1/i} \bmod p,$$

where $\overline{\mathbb{T}}'_{k'1/i}$ is the continuous multiplication of $\overline{T}'_{k1}$'s and $10^{4N_{k'}} \bmod p$ if there exists $w_k \in \mathcal{W}_c$ $(k \neq i)$ such that $w_k \prec w'_k$. If no such worker exists (i.e., $w_{k'}$ is only dominated by w_i), then $\overline{\mathbb{T}}'_{k'1/i} = 10^{4N_{k'}} \bmod p$, i.e.,

$$\overline{\mathbb{T}}'_{k'1/i} = \begin{cases} \prod \overline{T}'_{k1} \cdot 10^{4N_i} \bmod p & \exists w_k \in \mathcal{W}_c, \\ & w_k \prec w_{k'}, k \neq i, \\ 10^{4N_{k'}} \bmod p & \text{otherwise.} \end{cases}$$

Again, $\mathcal{S}_1$ can get the following $N_{win}-1$ equations regarding $C_{3,(k',j)}$'s at most in our scheme.

$$\begin{cases} C_{3,(k',1)} = \alpha_{(k',1)} \cdot T'_{k'1} \cdot \overline{T}'_{i1} \cdot \overline{\mathbb{T}}'_{k'1/i} \bmod p \\ \quad \vdots \\ C_{3,(k',k'-1)} = \alpha_{(k',k'-1)} \cdot T'_{k'1} \cdot \overline{T}'_{i1} \cdot \overline{\mathbb{T}}'_{k'1/i} \bmod p \\ C_{3,(k',k'+1)} = \alpha_{(k',k'+1)} \cdot T'_{k'1} \cdot \overline{T}'_{i1} \cdot \overline{\mathbb{T}}'_{k'1/i} \bmod p \\ \quad \vdots \\ C_{3,(k',N_{win})} = \alpha_{(k',N_{win})} \cdot T'_{k'1} \cdot \overline{T}'_{i1} \cdot \overline{\mathbb{T}}'_{k'1/i} \bmod p \end{cases}$$

Let consider $\mathbb{T}'_{k'1/i}$ as a whole variable, then it is easy to know that $\mathcal{S}_1$ faces at least $N_{win}+1$ unknown variables from the above $N_{win}-1$ equations. The definite $N_{win}+1$ variables are $(\alpha_{(k',1)}, ..., \alpha_{(k',k'-1)}, \alpha_{(k',k'+1)}, ..., \alpha_{(k',N_{win})}, T'_{k'1}, \overline{T}'_{i1})$. Similar with the previous discussions, $\mathcal{S}_1$ has no idea about $\overline{T}'_{i1}$, so $\mathcal{S}_1$ cannot compute $\overline{T}_i$ based on ElGamal encryption.

In summary, we have proved that $\mathcal{S}_1$ can obtain neither T_i, $\overline{T}_i$, T'_{i1}, nor $\overline{T}'_{i1}$ in the proposed scheme. Therefore, the real value of T_i for any worker $w_i \in \mathcal{W}_c$ will not be revealed to $\mathcal{S}_1$ in the process of worker selection.

- In the proposed scheme, $\mathcal{S}_2$ keeps T'_{i1}, $\overline{T}'_{i1}$, and many $(C_{3,(i,j)}, C_{7,(i,j)})$'s. Based on the correctness and security of ElGamal encryption, $\mathcal{S}_2$ cannot compute T_i or $\overline{T}_i$ with only knowing T'_{i1} or $\overline{T}'_{i1}$. In addition, $C_{3,(i,j)}$ is computed by T'_{i1} and $\overline{T}'_{k1}$'s for $w_k \prec w_i$. So the information related to T_i or $\overline{T}_i$ in $C_{3,(i,j)}$ is no more than that in T'_{i1} and $\overline{T}'_{i1}$. Therefore, $\mathcal{S}_2$ cannot obtain T_i or $\overline{T}_i$ from either T'_{i1}, $\overline{T}'_{i1}$, or $C_{3,(i,j)}$. In the next, we will demonstrate why $\mathcal{S}_2$ cannot get access to T_i and $\overline{T}_i$ from $C_{7,(i,j)}$'s.

It is easy to see that $\mathcal{S}_2$ can obtain the following $N_{win}-1$ equations regarding about $C_{7,(i,j)}$.

$$\begin{cases} C_{7,(i,1)} = \alpha_{(i,1)} \cdot \beta_{(i,1)} \cdot (T_i \cdot \overline{\mathbb{T}}_i - T_j \cdot \overline{\mathbb{T}}_j) \bmod p \\ \quad \vdots \\ C_{7,(i,i-1)} = \alpha_{(i,i-1)} \cdot \beta_{(i,i-1)} \cdot (T_i \cdot \overline{\mathbb{T}}_i \\ \qquad - T_j \cdot \overline{\mathbb{T}}_j) \bmod p \\ C_{7,(i,i+1)} = \alpha_{(i,i+1)} \cdot \beta_{(i,i+1)} \cdot (T_i \cdot \overline{\mathbb{T}}_i \\ \qquad - T_j \cdot \overline{\mathbb{T}}_j) \bmod p \\ \quad \vdots \\ C_{7,(i,N_{win})} = \alpha_{(i,N_{win})} \cdot \beta_{(i,N_{win})} \cdot (T_i \cdot \overline{\mathbb{T}}_i \\ \qquad - T_j \cdot \overline{\mathbb{T}}_j) \bmod p \end{cases}$$

Finally, $\mathcal{S}_2$ will face at least $N_{win}+1$ unknown values from the above $N_{win}-1$ equations, i.e., $(\beta_{(i,1)}, ..., \beta_{(i,i-1)}, \beta_{(i,i+1)}, ..., \beta_{(i,N_{win})}, T_i, T_j)$. Therefore, $\mathcal{S}_2$ cannot obtain T_i from $C_{7,(i,j)}$'s. In summary, T_i is privacy-preserving for $\mathcal{S}_2$ in the whole process of worker selection.

- Same as Theorem 5, it is easy to prove that neither other workers nor outsiders can obtain the information about T_i in our designed scheme.

□

VI. Performance Evaluation

In this section, we first study the performance of our proposed scheme in terms of storage overhead and computational overhead. Then, we provide a detailed description of our experimental configuration and comparison results.

A. Theoretical Analysis

• Storage Overheads: Assume there are totally N workers registered in the MCS system, then in the `SysPre` phase, for each worker w_i, $\mathcal{S}_1$ needs to keep $T'_{i2}, \overline{T}'_{i2}$; $\mathcal{S}_2$ needs to store $T'_{i1}, \overline{T}'_{i1}$. After $\mathcal{M}$ starts, participating workers need to outsource $\mathcal{E}'_{i2}$ and $\mathcal{E}'_{i1}$ to $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. In summary, for both $\mathcal{S}_1$ and $\mathcal{S}_2$, the storage overheads in the process of worker selection can be computed as $\sum_1^N 2|p| + \sum_1^{N_{win}+1} |p|$, where $|p|$ (e.g., 1024) is the bit-length of large prime p generated in Section IV-A and N_{win} is window size.

• Computational Overheads: The computational costs for `CmpExp` phase can be calculated as $\frac{(N_{win}+1)(N_{win}+2)}{2} \cdot t_{sky}$, where t_{sky} is the time cost for determining the skyline dominance relationship between each pair of workers. For a newly formed window with $N_{win}+1$ workers, the computational costs for `WrkSel` phase can be calculated as $N_{win} \cdot t_{w_i,w_j}$, where t_{w_i,w_j} is the time cost for comparing $\Pr^{\mathrm{sky}}(w_i)$ and $\Pr^{\mathrm{sky}}(w_j)$ in phase `CmpSky`. Therefore, the overall computational costs for worker selection is $\frac{(N_{win}+1)(N_{win}+2)}{2} \cdot t_{sky} + N_{win} \cdot t_{w_i,w_j}$.

B. Dataset, Experimental Settings, and the Baseline Method

• Real-world Dataset: The real-world dataset used in our experiment is Reality Mining Data [20]. This dataset tracks the mobile communication information of ninety-four people in the Massachusetts Institute of Technology from September 2004 to June 2005. After removing all the logs with none communication time duration, we finally extract 111,586 smartphone logs. For each log record, the time of picking up the phone represents the start time, the time duration of the communication means the life-span of the worker's sensing data. Moreover, for each worker w_i, we randomly assign $\mathcal{E}_i \in [1, 30]$ as working experience and $T_i \in (0, 1)$ as trustability, where both $\mathcal{E}_i$ and T_i follow uniform distribution.

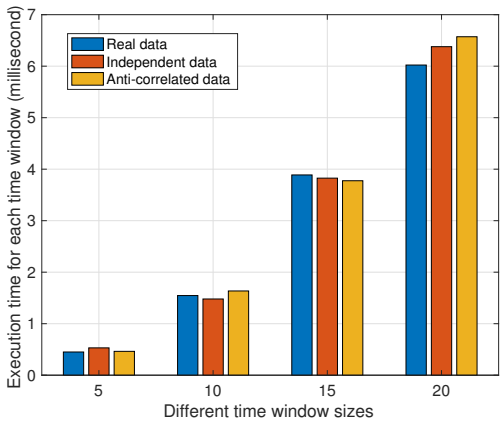


Fig. 3. The execution time (s) of the proposed worker selection scheme for different datasets with window size N_{win} varies from 5 to 20 (the communication costs between two cloud servers $\mathcal{S}_1$ and $\mathcal{S}_2$ are not considered).

• Synthetic Datasets: In the experiment, we also generate two different kinds of synthetic datasets (independent and anti-correlated) based on [14], both their sizes are 100k. For the independent data, each worker w_i's $\mathcal{E}_i \in [1, 30]$, $t_i^{\mathrm{ls}} \in [1s, 100s]$ and $T_i \in (0, 1)$ are generated independently using uniform distributions. For the anti-correlated data, we first uniformly generate a random probability $T_i \in (0, 1)$ as workers' trustability. Then, we assume that if a worker has higher $\mathcal{E}_i$ for a certain task $\mathcal{M}$, she/he tends to take shorter time to finish $\mathcal{M}$, i.e., smaller t_i^{ls}. Hence, $\mathcal{E}_i$ and t_i^{ls} are generated using the method in [14] such that they are anti-correlated with each other. For both simulated datasets, we randomly assign an order for workers' arrival in a data stream.

• Experimental settings and the baseline method: First, we need to decide the key sizes of the proposed encryption methods, i.e., the bit length of p and q. As we mentioned before, our proposed schemes are original from ElGamal encryption, whose chosen-plaintext attack security is based on the decisional Diffie-Hellman (DDH) assumption [21]. According to [22]–[24], when the bit lengths of p and q are set as 1024 and 160, the DDH assumtion holds. As a result, we set $|p| = 1024$ and $|q| = 160$ in this paper.

Next, we need to decide the experimental parameters for the MCS tasks. The window size N_{win} varies from 5, 10, 15, to 20, which means that our worker selection scheme will lanuch when $N_{win}+1$ (i.e., 6, 11, 16, and 21) workers are in the system. We also proposed a baseline method for comparison purposes. More specifically, when a new worker arrives in the system, the baseline method discards the worker who has the earliest expiry time t_i^{exp} and keeps the rest of the

workers for conducting the task. We perform the experiments with Python programming language on an Intel(R) Core(TM) i7-6700 CPU @3.60GHz Windows 64-bit Operating System with 32 GB RAM. We repeated each experiment 10000 times, and the average results are reported.

C. Experimental Results

• The computational cost for worker selection in each sliding window: It is noteworthy that our proposed scheme requires constant interactions between $\mathcal{S}_1$ and $\mathcal{S}_2$. However, the communication time cost between them is hard to be simulated. So we ignored the communication cost in our experiment. From Fig. 3, we observe that i) for each sliding window, the running time for worker selection can be finished fast (from 0.5*ms* to 6.5*ms*), which validates the efficiency and feasibility of our selection scheme. ii) The running time for worker selection is proportional to N_{win}, which equals to $\frac{(N_{win}+1)(N_{win}+2)}{2} \cdot t_{sky} + N_{win} \cdot t_{w_i,w_j}$. As a result, more workers in the current sliding window, the longer time it will take for selecting workers. iii) The computational cost is similar among different datasets. In the experiment, we simulate worker's $\mathcal{E}_i$ and T_i in each dataset. Essentially, the computational costs for securely comparing working experience in phase `CmpExp` and probabilistic skyline values in phase `CmpSky` account for a substantial part of the overall time costs in worker selection, which leads to the result that all the datasets have a similar running time when N_{win} is confirmed.

• Two scenarios for performance evaluation: For both the baseline and our proposed scheme, the following two real-world scenarios are deployed on all the datasets, i.e., *scenario 1: workers continuously arrive at the system* and *scenario 2: workers continuously leave the system.*

In *scenario 1*, with the successive arrival of the workers, the proposed worker selection scheme is maintained constantly to choose the top-N_{win} workers among the $N_{win}+1$ candidates. The sum of probabilistic worker experience $S_{\mathcal{E}} = \sum_{i=1}^{N_{win}}(T_i \times \mathcal{E}_i)$ is used as the evaluation metric in this scenario. $S_{\mathcal{E}}$ can be used to represent workers' overall quality in each sliding window, so it is a good indicator for the reliability of the proposed scheme. $S_{\mathcal{E}}$ is calculated for the selected N_{win} workers, and will be updated after a new worker arrives.

In *scenario 2*, for each current window, we assume that no worker arrives at the system anymore, and each worker leaves the MCS platform after they reach their expiry time. More specifically, for each worker w_i in the system, the following two variables are calculated: i) $t_i^{\mathtt{rexp}} = t_i^{\mathtt{exp}} - t_{min}^{\mathtt{exp}}$ which means w_i's relative departure time, where $t_{min}^{\mathtt{exp}}$ is the earliest expiry time in current window, and ii) $S_{\mathcal{E}}^{\mathtt{r}} = S_{\mathcal{E}} - \mathcal{E}_i$ which means the sum of probabilistic worker experience after w_i leaves the system. Consequently, *scenario 2* provides a useful circumstance for us to identify the overall sustainability and reliability of the system in terms of working experience and persevering duration when no worker arrives.

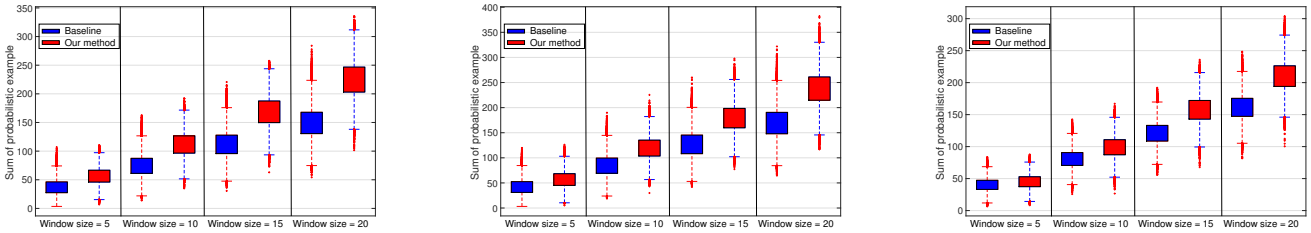
• Experimental results for *scenario 1*: Fig. 4 shows the experimental results for *scenario 1*. First of all, it is obvious that our proposed scheme performs better than the baseline method on selecting workers with more working experience. Specifically, $S_{\mathcal{E}}$ of our approach is larger than the baseline method for almost all the cases. This result indicates that given the same N_{win}, our scheme can keep each sliding window more reliable by maintaining workers with more experience. There even exists some cases such that the $S_{\mathcal{E}}$ of our proposed method when $N_{win} = 15$ is larger than baseline method when $N_{win} = 20$ (e.g., Fig. 4 (a) and (b)). Second, we can see that there is an exceptional example in Fig. 4 (c), where two methods are nearly indistinguishable for anti-correlated dataset when $N_{win} = 5$. Under the anti-correlated situation, a worker with more working experience tends to have a shorter life-span of the sensing data, and vice versa. This property leads to a result that when N_{win} is small (e.g. equals 5), the workers are more likely not to be dominated by each other. Thus, their $\Pr^{\mathtt{sky}}(w_i)$ tends to not be affected by the newly arrived worker (e.g., the case 2 in the `WrkSel` phase). So the workers may just be selected only by their expiry time. As a result, our proposed scheme shows a similar statistical pattern with the baseline. Third, as N_{win} increases, the differences of $S_{\mathcal{E}}$ between our scheme and baseline increase simultaneously. Consequently, it is important to define a sufficiently large N_{win} for better distinguishing our scheme with baseline in real-world MCS applications.

• Experimental results for *scenario 2*: Fig. 5 to Fig. 7 show the comparison results of *scenario 2* with N_{win} varies from 5 to 20. Intuitively, our proposed worker selection scheme is superior to the baseline method for all the cases in terms of $t_i^{\mathtt{rexp}}$ and $S_{\mathcal{E}}^{\mathtt{r}}$. This indicates that given the same N_{win}, our approach can provide better and longer services than baseline. Specifically, once a worker reaches the expiry time and leaves the platform, the rest of the workers in our system have more working experience than those in baseline method. In addition, the workers in our proposed selection scheme can stay 30s $\sim$ 200s longer than baseline workers under different N_{win}. Therefore, when no worker arrives anymore, the workers selected by our scheme can maintain the platform better and longer, which is significant for certain MCS services such as traffic jam/accident monitoring. This observation further validates our proposed scheme's reliability and sustainability in a special case, where the platform's durability plays a fundamental role in service quality.

VII. Related Work

Recently, there have been numerous endeavors that aim to study the problem of worker selection in MCS platforms. In this section, we briefly review some of the typical works in this area.

Jin et al. [11] proposed a privacy-preserving framework for MCS worker selection based on a novel incentive mechanism. Specifically, their framework can compensate workers' costs for both sensing leakage and privacy leakage while selecting workers who are more likely to provide reliable sensing data. Moreover, they integrated their framework with data perturbation and aggregation technologies, ensuring highly accurate aggregated results and guarantees workers' privacy. Zhang et al. [9] proposed a probabilistic skyline based worker

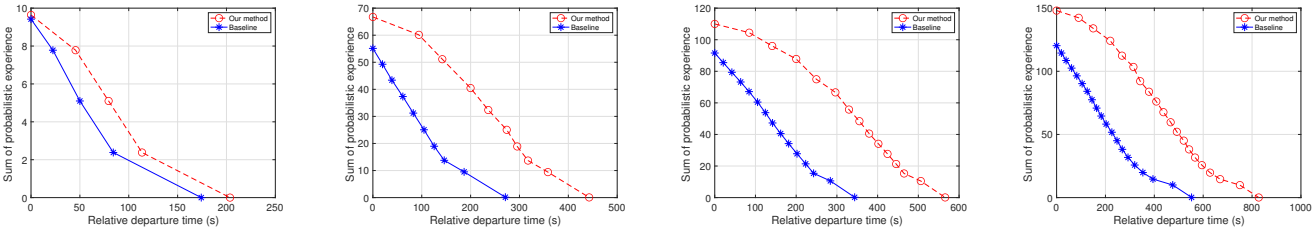


(a) Real dataset: Comparison of workers' quality between our method and baseline based on $S_{\mathcal{E}}$ with window size N_{win} varies from 5 to 20.

(b) Independent dataset: Comparison of workers' quality between our method and baseline based on $S_{\mathcal{E}}$ with window size N_{win} varies from 5 to 20.

(c) Anti-correlated dataset: Comparison of workers' quality between our method and baseline based on $S_{\mathcal{E}}$ with window size N_{win} varies from 5 to 20.

Fig. 4. Comparison and evaluation of workers' quality between our method and baseline method for different datasets with varied window size N_{win} from 5, 10, 15 to 20.



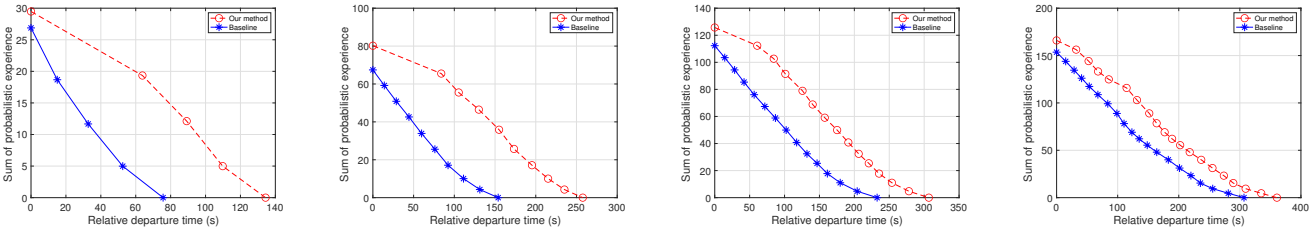
(a) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 5$.

(b) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 10$.

(c) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 15$.

(d) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 20$.

Fig. 5. Real dataset: The sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with varied window size $N_{win} = 5$, 10, 15 and 20, i.e., for a given window size N_{win}, $S^{\mathrm{r}}_{\mathcal{E}} = S_{\mathcal{E}} - \mathcal{E}_i$ is measured after a worker w_i leaves the system at t_i^{rexp}.



(a) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 5$.

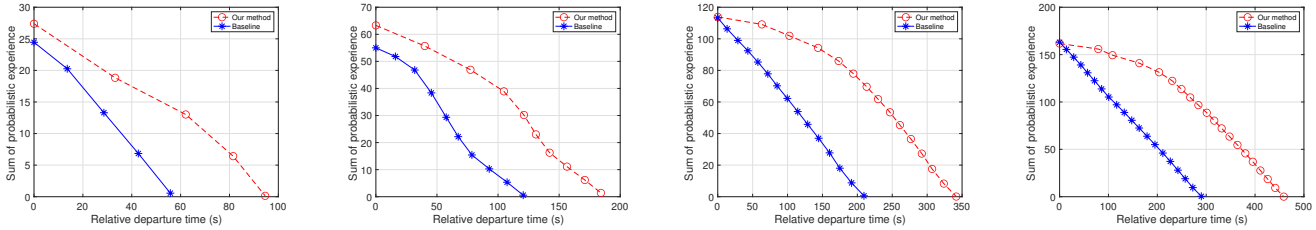
(b) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 10$.

(c) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 15$.

(d) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 20$.

Fig. 6. Independent dataset: The sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with varied window size $N_{win} = 5$, 10, 15 and 20, i.e., for a given window size N_{win}, $S^{\mathrm{r}}_{\mathcal{E}} = S_{\mathcal{E}} - \mathcal{E}_i$ is measured after a worker w_i leaves the system at t_i^{rexp}.

selection method, which solved a fundamental problem of calculating workers' trustability based on their historical reviews. In addition, they designed a non-interactive encrypted integer comparison protocol to compare the skyline dominance relationship between workers securely. Skyline operator is used for selecting workers based on ask price and task relativity. This is the first work using a skyline-based method for worker selection in MCS platforms, which can consider the trade-off between multiple aspects and select workers that are not dominated by others. Gong et al. [25] devised truthful crowdsensing mechanisms for incentivizing strategic workers to truthfully reveal their private quality and truthfully make efforts as desired by the requester. In their work, the authors assumed that a strategic worker with low quality may pretend to have a high quality to receive a high reward from the requester. Under the proposed mechanisms, they showed that the requester can assign the task only to the best workers that has the smallest virtual valuation. However, in our work, by introducing a performance feedback mechanism for calculating trustability, the workers need to submit the real-time information and sensing data as accurate as possible. Ren et al. [26] introduced a socially aware reputation management scheme for selecting well-suited participants and allocating the task rewards in MCS services. Concretely, social attributes, task delays, and reputation are focused under a fixed task budget. Moreover, a rewarding scheme is devised to measure the quality of the sensing reports and allocate reliable workers to certain MCS tasks under the consideration of the trustworthiness and cost performance of task participants. Ni et al. [27] described a privacy-preserving MCS framework for location-



(a) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 5$.

(b) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 10$.

(c) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 15$.

(d) Sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with $N_{win} = 20$.

Fig. 7. Anti-correlated dataset: The sum of probabilistic worker experience $S^{\mathrm{r}}_{\mathcal{E}}$ vs. relative departure time t_i^{rexp} with varied window size $N_{win} = 5$, 10, 15 and 20, i.e., for a given window size N_{win}, $S^{\mathrm{r}}_{\mathcal{E}} = S_{\mathcal{E}} - \mathcal{E}_i$ is measured after a worker w_i leaves the system at t_i^{rexp}.

based applications, which can balance the trade-off between privacy preservation and task allocation. A matrix-based location matching mechanism is used by the service provider to achieve location-based task allocation without disclosing workers' sensing location. Proxy re-encryption technology is developed to enhance privacy preservation during the MCS services. Jin et al. [28] proposed a differentially private incentive mechanism that preserves the privacy of each worker's bid, which is based on the single-minded reverse combinatorial auction. However, in this work, we select workers based on working experience, expiry time, and trustability, without considering workers' asking price information. Liu et al. [29] studied the problem of multi-task allocation in MCS services. For particular, two typical multi-task MCS environments are considered in their work, e.g., FPMT (few participants with more tasks) and MPFT (more participants with few tasks). Unique mechanisms are devised for different scenarios with different optimization goals. Wang et al. [13] proposed a location privacy-preserving worker selection framework with geo-obfuscation for protecting workers' locations during task allocation. More specifically, by obfuscating their real locations under the guarantee of differential privacy, workers can protect their location privacy regardless of adversaries' prior knowledge. In addition, a mixed-integer non-linear programming problem is defined for minimizing workers' expected travel distance and optimizing worker assignments. Wu et al. [30] proposed a context-aware multi-armed bandit incentive method for selecting high-quality workers in MCS systems. In their work, workers' service quality is evaluated by their context and cost. Then, by accurately assessing workers' quality information, a modified Thompson sampling approach is utilized for selecting reliable workers based on reinforcement learning. Sun et al. [31] studied the problem of truth discovery in crowdsourced question answering system based on a contract-based privacy-preserving incentive mechanism, whereas the scope of this work is reliable and continuous worker selection in MCS.

In summary, most of the previous works treat the MCS task as a one-time service. They allocate tasks to suitable workers by measuring their personal qualification, location, and trustability in particular sensing environment (e.g., multi-task allocation), without considering a dynamic situation such that workers may continuously arrive or leave the system. Furthermore, little systematic and data-driven evidence has been published for a typical real-world case in which the MCS task should be constantly maintained for a long time. Unlike the above, our work focuses on privacy-preserving and continuous worker selection for MCS platforms. Our proposed scheme can dynamically select reliable workers while guarantee privacy preservation of their sensitive information.

VIII. Conclusion

In this work, we have proposed a privacy-preserving worker selection scheme for MCS based on probabilistic skyline query over sliding windows. The proposed scheme can continuously select reliable workers in terms of working experience, expiry time, and trustability without revealing sensitive information. More specifically, a probabilistic skyline approach is designed for a dynamic situation where workers may constantly arrive at/leave the platform. For protecting workers' privacy, we have designed an ElGamal-based encryption approach for securely outsourcing sensitive information and comparing workers' personal information. Security analysis demonstrates that the proposed scheme is privacy-preserving. Extensive experiments have been conducted on both real-world and simulated datasets for two scenarios: i) continuous worker arrival and ii) continuous worker departure. The comparison results validate the efficiency and effectiveness of our proposed worker selection scheme. For future work, the group skyline technique will be studied to enhance the performance of the current system.

Acknowledgements

The authors generously acknowledge the funding from the National Science and Engineering Research Council of Canada (NSERC) through the discovery grant and Canada Research Chair to Dr. Ghorbani, NSERC (no. Rgpin 04009) to Dr. Lu. This research was also supported by NSF of Zhejiang Province (grant no. LZ18F020003), NSFC (grant no. U1709217), and NSFC (grant no. 61672411).

References

[1] D. Zhang, L. Wang, H. Xiong, and B. Guo, "4w1h in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 42–48, 2014.

[2] X. Zhang and A. A. Ghorbani, "Human factors in cybersecurity: Issues and challenges in big data," in *Security, Privacy, and Forensics Issues in Big Data*. IGI Global, 2020, pp. 66–96.

[3] X. Zhang, R. Lu, J. Shao, H. Zhu, and A. A. Ghorbani, "Achieve secure and efficient skyline computation for worker selection in mobile crowdsensing," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2019, pp. 1–6.

[4] S. Tiwari and S. Kaushik, "Information enrichment for tourist spot recommender system using location aware crowdsourcing," in *2014 IEEE 15th International Conference on Mobile Data Management*, vol. 2. IEEE, 2014, pp. 11–14.

[5] A. Overeem, J. R. Robinson, H. Leijnse, G.-J. Steeneveld, B. P. Horn, and R. Uijlenhoet, "Crowdsourcing urban air temperatures from smartphone battery temperatures," *Geophysical Research Letters*, vol. 40, no. 15, pp. 4081–4085, 2013.

[6] C. Zhang, L. Zhu, C. Xu, X. Du, and M. Guizani, "A privacy-preserving traffic monitoring scheme via vehicular crowdsourcing," *Sensors*, vol. 19, no. 6, p. 1274, 2019.

[7] Y. Cui, L. Deng, Y. Zhao, B. Yao, V. W. Zheng, and K. Zheng, "Hidden poi ranking with spatial crowdsourcing," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 814–824.

[8] Wikipedia, "WAZE, a GPS navigation software app," https://en.wikipedia.org/wiki/Waze, accessed: 2019-09-08.

[9] X. Zhang, R. Lu, J. Shao, H. Zhu, and A. Ghorbani, "Secure and efficient probabilistic skyline computation for worker selection in mcs," *IEEE Internet of Things Journal*, 2020, to appear.

[10] Z. Wang, J. Hu, R. Lv, J. Wei, Q. Wang, D. Yang, and H. Qi, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, 2018.

[11] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2016, pp. 341–350.

[12] J. Lin, D. Yang, M. Li, J. Xu, and G. Xue, "Frameworks for privacy-preserving mobile crowdsensing incentive mechanisms," *IEEE Transactions on Mobile Computing*, vol. 17, no. 8, pp. 1851–1864, 2017.

[13] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma, "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 627–636.

[14] S. Börzsönyi, D. Kossmann, and K. Stocker, "The skyline operator," in *Proceedings of the 17th International Conference on Data Engineering, April 2-6, 2001, Heidelberg, Germany*. IEEE Computer Society, 2001, pp. 421–430.

[15] Y. Zheng, R. Lu, B. Li, J. Shao, H. Yang, and K.-K. R. Choo, "Efficient privacy-preserving data merging and skyline computation over multi-source encrypted data," *Information Sciences*, vol. 498, pp. 91–105, 2019.

[16] J. Pei, B. Jiang, X. Lin, and Y. Yuan, "Probabilistic skylines on uncertain data," in *Proceedings of the 33rd international conference on Very large data bases*. VLDB Endowment, 2007, pp. 15–26.

[17] J. Liu, H. Zhang, L. Xiong, H. Li, and J. Luo, "Finding probabilistic k-skyline sets on uncertain data," in *Proceedings of the 24th acm international on conference on information and knowledge management*. ACM, 2015, pp. 1511–1520.

[18] W. Zhang, X. Lin, Y. Zhang, W. Wang, and J. X. Yu, "Probabilistic skyline operator over sliding windows," in *2009 IEEE 25th International Conference on Data Engineering*. IEEE, 2009, pp. 1060–1071.

[19] R. Lu, X. Lin, H. Zhu, P.-H. Ho, and X. Shen, "Ecpp: Efficient conditional privacy preservation protocol for secure vehicular communications," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 2008, pp. 1229–1237.

[20] A. Pentland, N. Eagle, and D. Lazer, "Inferring social network structure using mobile phone data," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

[21] D. Boneh, "The decision diffie-hellman problem," in *International Algorithmic Number Theory Symposium*. Springer, 1998, pp. 48–63.

[22] A. K. Lenstra and E. R. Verheul, "Selecting cryptographic key sizes," *Journal of cryptology*, vol. 14, no. 4, pp. 255–293, 2001.

[23] N. Ferguson and B. Schneier, *Practical cryptography*. Wiley New York, 2006, vol. 141.

[24] C. Paar and J. Pelzl, *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media, 2009.

[25] X. Gong and N. B. Shroff, "Truthful mobile crowdsensing for strategic users with private data quality," *IEEE/ACM Transactions on Networking*, vol. 27, no. 5, pp. 1959–1972, 2019.

[26] J. Ren, Y. Zhang, K. Zhang, and X. S. Shen, "Sacrm: Social aware crowdsourcing with reputation management in mobile sensing," *Computer Communications*, vol. 65, pp. 55–65, 2015.

[27] J. Ni, K. Zhang, X. Lin, Q. Xia, and X. S. Shen, "Privacy-preserving mobile crowdsensing for located-based applications," in *ICC'17*. IEEE, 2017, pp. 1–6.

[28] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2016, pp. 344–353.

[29] Y. Liu, B. Guo, Y. Wang, W. Wu, Z. Yu, and D. Zhang, "Taskme: Multi-task allocation in mobile crowd sensing," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 403–414.

[30] Y. Wu, F. Li, L. Ma, Y. Xie, T. Li, and Y. Wang, "A context-aware multiarmed bandit incentive mechanism for mobile crowd sensing systems," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7648–7658, 2019.

[31] P. Sun, Z. Wang, Y. Feng, L. Wu, Y. Li, H. Qi, and Z. Wang, "Towards personalized privacy-preserving incentive for truth discovery in crowdsourced binary-choice question answering," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 1133–1142.

Xichen Zhang received the B.E. degree from Changsha University of Science and Technology in 2010. He received his M.S. degree in Computer Science at the Canadian Institute for Cybersecurity (CIC), Faculty of Computer Science (FCS), University of New Brunswick (UNB) in 2018. After that, he worked as a research assistant in CIC. He is currently working toward the Ph.D. degree with FCS, UNB. His research interests are data mining in cybersecurity, privacy enhancing technologies, and IoT-Big Data security and privacy.

Rongxing Lu (S'09-M'11-SM'15-F'21) is currently an associate professor at the Faculty of Computer Science (FCS), University of New Brunswick (UNB), Canada. Before that, he worked as an assistant professor at the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore from April 2013 to August 2016. Rongxing Lu worked as a Postdoctoral Fellow at the University of Waterloo from May 2012 to April 2013. He was awarded the most prestigious "Governor General's Gold Medal", when he received his PhD degree from the Department of Electrical & Computer Engineering, University of Waterloo, Canada, in 2012; and won the 8th IEEE Communications Society (ComSoc) Asia Pacific (AP) Outstanding Young Researcher Award, in 2013. He is presently a senior member of IEEE Communications Society. His research interests include applied cryptography, privacy enhancing technologies, and IoT-Big Data security and privacy. He has published extensively in his areas of expertise, and was the recipient of 8 best (student) paper awards from some reputable journals and conferences. Currently, Dr. Lu currently serves as the Vice-Chair (Conferences) of IEEE ComSoc CIS-TC (Communications and Information Security Technical Committee). Dr. Lu is the Winner of 2016-17 Excellence in Teaching Award, FCS, UNB.

Jun Shao received the Ph.D. degree from the Department of Computer Science and Engineering at Shanghai Jiao Tong University, Shanghai, China in 2008. He was a postdoc in the School of Information Sciences and Technology at Pennsylvania State University, USA from 2008 to 2010. He is currently a professor of the School of Computer Science and Information Engineering at Zhejiang Gongshang University, Hangzhou, China. His research interests include network security and applied cryptography.

Hui Zhu (M'13-SM'19) received the B.Sc. degree from Xidian University, Xian, China, in 2003, the M.Sc. degree from Wuhan University, Wuhan, China, in 2005, and the Ph.D. degree from Xidian University, in 2009. He was a Research Fellow with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, in 2013. Since 2016, he has been a Professor with the School of Cyber Engineering, Xidian University. His current research interests include applied cryptography, data security, and privacy.

Ali A. Ghorbani has held a variety of academic positions for the past 39 years and is currently a Professor of Computer Science, Tier 1 Canada Research Chair in Cybersecurity, and Director of the Canadian Institute for Cybersecurity, which he established in 2016. He served as the Dean of the Faculty of Computer Science at the University of New Brunswick from 2008 to 2017. He is also the founding Director of the laboratory for intelligence and adaptive systems research. He has spent over 29 years of his 39-year academic career, carrying out fundamental and applied research in machine learning, cybersecurity, and Critical Infrastructure Protection. Dr. Ghorbani is the co-inventor on three awarded and one filed patent in the fields of Cybersecurity and Web Intelligence and has published over 280 peer-reviewed articles during his career. He has supervised over 190 research associates, postdoctoral fellows and students during his career. His book, Intrusion Detection and Prevention Systems: Concepts and Techniques, was published by Springer in October 2010. He developed several technologies adopted by high-tech companies and co-founded three startups, Sentrant Security, EyesOver Technologies, and Cydarien Security in 2013, 2015, and 2019. He is the co-founder of the Privacy, Security, Trust (PST) Network in Canada and its annual international conference and served as the co-Editor-In-Chief of Computational Intelligence: An International Journal from 2007 to 2017. Dr. Ghorbani is the recipient of the 2017 Startup Canada Senior Entrepreneur Award, and Canadian Immigrant Magazines RBC top 25 Canadian immigrants of 2019.