Journal of Network and Computer Applications 215 (2023) 103634

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Joint privacy and data quality aware reward in opportunistic Mobile Crowdsensing systems

Luca Bedogni [a],[*], Federico Montori [b]

[a] University of Modena and Reggio Emilia, Italy
[b] University of Bologna, Italy

ARTICLE INFO

Keywords:
IoT
Mobile Crowdsensing
Reward in Mobile Crowdsensing
Privacy

ABSTRACT

Mobile Crowdsensing (MCS) is a paradigm involving a crowd of participants, called workers, into sensor data gathering campaigns through their personal devices. Some campaigns require workers to contribute with small amounts of geolocalized data at a constant rate, while being not directly aware of the global conditions of the system. In the scope of this reduced awareness, it is crucial to consider the privacy preservation of single workers at design time, as the disclosure of their exact location may lead to severe privacy issues. In this paper we design a privacy by design MCS framework that leverages variable rewards for workers willing to submit their location with an higher precision than others. Privacy is ensured through a negotiation phase that estimates the reward of the workers for different levels of location precision. This way, it helps them decide autonomously the spatial granularity of their data in order to preserve their privacy, yet obtaining a reward for their data. We design a metric based on k-anonymity to evaluate the level of privacy achieved, and validate the proposed framework over a real dataset. Our results show the efficacy of the framework as well as interesting effects caused by the topology of the environment.

1. Introduction

Mobile Crowdsensing (MCS) is often considered as a viable technology to gather data of interest in an area, leveraging personal devices of a crowd of users. In this scenario the participants to the MCS campaign are called *workers*, and are committed to provide data to the entity running the campaign, called *crowdsourcer*, in exchange of a reward (Capponi et al., 2019). MCS campaigns often require sensor data and are subject to the well-known problems of recruiting and rewarding workers efficiently, as well as to maximize the data quality while dealing with highly variable and uncertain environments (Montori et al., 2018). Furthermore, there are many different categories of MCS platforms; in this work we focus on Opportunistic MCS, meaning that data is reported opportunistically without any human interaction as workers move around. This typically involves workers to install an application either on their smartphone or on a specialized device. Opportunistic MCS is different compared to Participatory MCS; in the latter the tasks issued by the crowdsourcer can be heterogeneous, and involve a necessary human participation on the workers' side. Most of the times, data in MCS is geolocalized, so workers also implicitly provide their position whilst sending the data of interest. This behavior raises a significant issue in MCS, since sending the personal position of workers may also lead to the reconstruction by an attacker of the history of locations visited by each worker. This may potentially highlight routine behaviors of individual workers, manifesting a privacy issue which needs to be addressed, although the perception of the issue by workers may vary depending on their awareness (Kim et al., 2022). Moreover MCS workers are typically rewarded by the crowdsourcer, either for simply taking part in it or for each measurement they send. While in Participatory MCS there are different proposals on this aspect, in Opportunistic MCS it is more challenging, since reporting data and deciding the reward to provide to the worker needs to be automatically addressed, without any explicit human interaction. Another challenge is related to the noise and to the amount of information needed in any part of the city by the crowdsourcer. In fact, having identical measurements close in time and space may be less valuable than having a wider coverage of the whole area of interest. This is also the motivation for crowdsourcers to provide different rewards based on the amount of data sent, the time of the day, or the location of the workers. In order to protect the geographical privacy of single workers, a widespread approach is to alter the precise location of measurements to obfuscate the real data – the so-called location cloaking (Pournajaf et al., 2014a) – to reduce possible issues if such data is accessed by a malicious entity. However, this is applicable if

[*] Corresponding author.
E-mail addresses: luca.bedogni@unimore.it (L. Bedogni), federico.montori2@unibo.it (F. Montori).

https://doi.org/10.1016/j.jnca.2023.103634
Received 14 June 2022; Received in revised form 15 February 2023; Accepted 30 March 2023
Available online 8 April 2023

workers are partially aware of the conditions of other workers, due to fog layers, overlay networks or trusted entities. In real use cases these might not be available, therefore, purely optimization-based solutions such as in Wang et al. (2016) or Wang et al. (2017) might not be viable. For these reasons it is important to design a rewarding system which has privacy by design principles, so that workers are protected while yielding a meaningful data collection to the crowdsourcer. This must reward workers willing to share more precise measurements compared to workers which provide coarser data, although not being aware of the presence of others.

In this paper we propose a Privacy-by-Design (PbD) MCS system which tackles the aforementioned challenges: users' privacy, data quality and a variable reward. We do so by designing a framework which advertises the rewards to each of the workers through a negotiation phase, and allows workers to determine the desired privacy level. The rewards are estimated on top of the willingness of workers to provide more precise data. At the same time, workers are willing to receive the highest possible reward, while also disclosing little information about their real position. There is an evident trade off between the measurement location precision and the reward which the remote server may offer. Each worker always starts by sending a measurements with the coarser possible precision, then possibly increasing it upon exchanging information with the server. More specifically, upon receiving each measurement with a specific uncertainty in the real position, the server looks for measurements already collected and stored, and foresees an increased reward for the worker, in case he/she is willing to provide the same measurement with an increased precision. This is then sent back to the worker, who has the possibility to decide whether to increase the precision or abort the process, keeping the reward collected with the last measurement sent. This enables the central entity to adopt algorithms for variable rewards, which can take into account what the crowdsourcer values the most, and reward users providing such data appropriately. In other words, workers may refine with subsequent messages their measurements, in case the server is willing to offer a higher reward in exchange for such increased precision.

The rest of this paper is structured as follows: Section 2 discusses related work on this topic; Section 3 details the model we have designed and the communication which takes place between the workers and the remote server; Section 4 outlines the negotiation phase that leads workers to select the location precision; Section 5 discusses the privacy metrics which we considered to evaluate our proposal; Section 6 analyses the performance of our system with a real dataset, and Section 7 concludes this work and presents future works on this topic.

2. Related work

MCS is currently a significantly studied topic in literature. As it is a wide area of research, there are many contributions which target the challenges that this scenario presents.

Concerning the privacy of users, many works have explored how to send, store and retrieve data in MCS campaigns while maintaining workers' privacy Cheng et al. (2022a). Recent survey papers (Kim et al., 2022) Zhao et al. (2022b) underline how the privacy problem is one of – if not the – most important challenges in MCS and presents some of the newest works in the area. In Ni et al. (2017) the authors explore the privacy problem in MCS by building a location matrix through which the server can get precise information about data measurements, but users do not have to disclose their exact location. The focus of Bou Abdo et al. (2016) is on exploring different privacy metrics for MCS, and among the contributions of it the authors identify that a privacy preserving MCS should protect location privacy, identity privacy and identity-query privacy. Moreover, Montori and Bedogni (2020) focuses on achieving an increased privacy for users participating in the MCS campaign, by obfuscating IDs with hashes, hence breaking the correlation between chunks of different measurements. Nevertheless, in some scenarios this may not be acceptable, as the crowdsourcer needs to identify measurements reported by the same user. In Wu and Luo (2020) the authors present the concepts of Activity Point Exposure and Activity Transitions Exposure, providing three main privacy preservation techniques which are anonymization, obfuscation and encryption. Recently also privacy by design mechanism have been proposed, which preserve the workers' privacy thanks to the exchange of specific messages that do not disclose the identity of the worker Montori and Bedogni (2023). More recently, in Cheng et al. (2022b), authors envision a privacy-preserving MCS system for vehicular scenarios, using k-anonymity and a reputation-location matrix computed by both the mobile worker and the data requester. Through the Hadamard product, the suitability of the worker for the task is calculated without the need for explicit disclosure of location data. As a concluding remark, we observe that the vast majority of location privacy-aware MCS scenarios are mostly founded on the concepts of pure k-anonymity or differential privacy, such as in Wang et al. (2016), Pournajaf et al. (2014b), Sun et al. (2019) or Yan et al. (2019). All such works are adding some sort of trusted layer to the architecture, which instructs the workers on the environmental conditions. This allows to build algorithms that optimize location cloaking against the location precision, however, these trusted layers are not viable in many of the actual opportunistic deployments, which are better represented by the architecture proposed in this paper. Clearly, these systems should also cope with additional constraints, such as the energy efficiency (Marjanović et al., 2016), the scalability of them (Mota et al., 2018) and the data quality, which may be noisy due to user generated data (Cheng et al., 2017). Another very recent set of solutions comes from the data aggregation, using for instance homomorphic encryption schemes to aggregate encrypted data without the need to know their content. The solution in Zhao et al. (2022a) combines this concept with Federated Learning techniques to securely aggregate individual participants' models.

Regarding rewarding mechanisms, the main challenge lies in determining the appropriate reward depending on a number of factors such as the data quality, the data freshness or the type of data sensed. For participatory crowdsensing, Wang et al. (2018) proposes an optimal dynamic programming solution and also a greedy based solution. In Hu et al. (2020) the authors propose a rewarding system based on blockchain, where the incentive is computed through a three step Stackelberg game. Finally Klopfenstein et al. (2019) proposes an anonymous rewarding system, through an aggregation service which forwards requests to the rewarding platform. Finally, regarding the data quality, the main challenge in MCS is to understand whether crowdsensed data has to be trusted or not, concerning the noise which may result from opportunistic measurements (Arkian et al., 2017). In Luo et al. (2019) the authors tackle the problem by proposing a cross validation of data by other participants, to double check the data reported to find possible issues. In Liu et al. (2017) the authors focus on the same problem, although analyzing it by determining the context in which the sensing takes place, thus foreseeing the possibility that such sensing operations may result in noisy data. An et al. (2020) focuses instead on selecting participants which provide higher quality data to the campaign, by also employing a blockchain mechanism. In fact, participant selection is a key task in participatory crowdsensing (Azzam et al., 2016; Alagha et al., 2021). Finally Zhao et al. (2021) proposes PACE, and focuses both on the data quality and on the privacy of users, to reward users while also preserving their privacy. A work that is close to the one proposed in this paper is Jin et al. (2019), which aims to build an auction market between workers and the campaign owner. However, the study is focused on participatory MCS, in which the campaign owner publishes a set of tasks on which workers can bid depending on their privacy preferences, and are rewarded based on the precision of the location through which they report their measurement.

3. Framework

The proposed system operates onto a PbD pull-based MCS scenario, where mobile workers spontaneously and opportunistically send data to a central data storage without any communication mediator, i.e. no fog layer is required.

3.1. System model

More in detail, let us define the set of mobile workers as $\mathbf{W} = \{W_1, \ldots, W_n\}$. Each mobile worker is then defined as a list of events, which correspond to location and times: $W_i = \{\langle t_0^{\{i\}}, l_0^{\{i\}}\rangle, \langle t_1^{\{i\}}, l_1^{\{i\}}\rangle, \ldots\}$. We assume that time is sliced into discrete time slots with a defined duration Δt in seconds. As for location, in this paper we do not use punctual location encoding based on coordinates, such as GPS; instead, we leverage the hierarchical areal encoding offered by MGRS. MGRS (Military Grid Reference System) (Lampinen, 2001) is a coordinate systems that divides the world in square areas, each of them having an x and a y coordinate. The division is hierarchical: each square contains 100 smaller squares which are identified by adding a digit to both its x and y coordinates. In this way it is easy to see whether a square A lies within a square B, in such case A is obtained by removing a number of digits from both coordinates of B. More in detail, the world map is divided into 100 km × 100 km squares that are uniquely identified by a label, forming the first part of an MGRS coordinate (e.g. 32TPQ). The second part is then given by the x and y coordinates of a square that, depending on the number of digits of both coordinates, identifies the precision, spanning from 1 to 5, where precision 1 stands for a 10 km-sided square, while precision 5 stands for a 1 m-sided square. For instance, 32TPQ 56 78 has precision 2 and identifies a 1 km-sided square, and is a subsquare of 32TPQ 5 7, which is obtained by removing two digits from the x and y coordinates.

Given that a single MGRS coordinate can be expressed with different precisions, in this paper we will then use the notation $l_{j,p}^{\{i\}}$ to identify the MGRS location of worker i at time slot t_j expressed with precision p, where $1 \leq p \leq 5$. In this case, 1 identifies the coarser granularity (i.e. 10 km), while 5 identifies the finest granularity (i.e. 1 m). Note that it is always possible to translate an MGRS coordinate to any coarser granularity, while it is not possible the opposite. At any time slot t_j, a worker W_i is aware of his/her MGRS position $l_{j,p}^{\{i\}}$, for all values of p. Over time, a worker can send sensor data relevant to the MCS campaign to the central server in exchange for a reward. More formally, a data point is defined as a tuple $D = \langle t_j^{\{i\}}, l_{j,p}^{\{i\}}, d\rangle$, where d is the data payload, $t_j^{\{i\}}$ is the data timestamp and $l_{j,p}^{\{i\}}$ is the data location. The sender W_i may freely decide which precision p to use for sending data: the higher the *precision*, the higher the *reward* received in exchange, the lower is the *privacy* guaranteed to the worker. Ideally, W_i aims to maximize p (therefore maximizing the reward) as long as the MGRS area revealed $l_{j,p}^{\{i\}}$ guarantees a sufficient anonymity, ensured by the presence of other workers within $l_{j,p}^{\{i\}}$. The problem, however, cannot be modeled as an optimization strategy, as workers do not know the location of other workers. For such reason, we introduce an automatic negotiation phase between a worker and the central server before sending the data. Such phase is outlined in detail in Section 4.

The architecture of the proposed model is represented in Fig. 1. The figure shows the set of workers scattered onto a square area. For simplicity, we graphically represented a simple hierarchical world map that observes the same concept as MGRS, yet more suitable for this graphic example. In this case, the world map is divided into four square areas of the same size: a, b, c and d. Every square area is divided into other four square areas that belong to the immediately higher precision level, until reaching three levels of precision. Adding a precision level to a square implies adding a letter on top of the coordinate of such square for every square of higher precision that are contained into it. For instance, as shown in the figure, square bcb has the highest possible precision in our example ($p = 3$) and it is contained into square bc with $p = 2$ and square b with $p = 1$. Two of the workers in the figure are about to send data to the central server. Let us assume that their requirement is to have at least two other workers located within their same area, therefore aiming for a k-anonymity of at least 3. The sender in bcc would be then better off sending its precise location bcc as it contains exactly three workers, fulfilling the privacy requirement as also the other two workers would share the same measurement location. Such worker could also have been sending bc or b, however the choice of bcc yields the highest rewards as it is more precise, hence more valuable for the campaign owner. Conversely, the worker in cda is alone within the cda square, therefore he/she should send the location cd reducing p by 1, thus fulfilling the privacy requirement as cd contains five workers. Nevertheless, this scenario is obviously simplified, as neither of the two workers know how many other workers are located nearby, hence it has to be heuristically determined. The negotiation phase described in the next section has the goal of producing an estimate, helping the workers themselves to choose which MGRS precision they shall use.

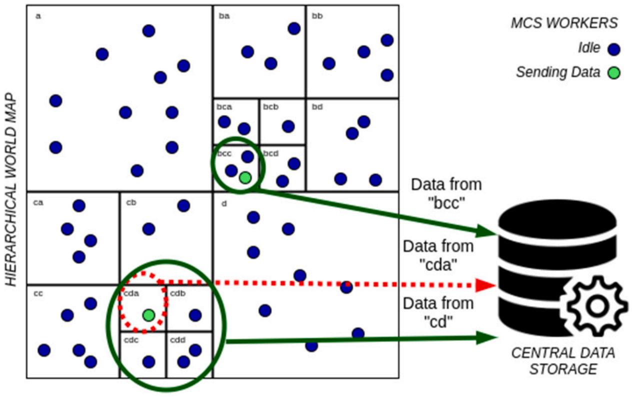


Fig. 1. Architecture of the proposed privacy-preserving system. Workers send measurements to the central data storage. A higher precision yields a higher reward, at the cost of a reduced privacy.

3.2. Privacy patterns

To guarantee PbD we leverage on the use of Privacy Patterns (Notario et al., 2015), which define a set of design patterns aimed at guaranteeing resilience of the systems against various attacks. More specifically, we have applied the following patterns to our model:

- **Location Granularity**: this privacy pattern aims at minimizing the location granularity of collected information by users. It states that users should be allowed with diverse granularity of locations when sharing information, as providing coarser data better protects their routine habits. We implement this in our system by allowing users to send their location with an increased precision, depending on their wanted privacy levels and depending on the other information available to the system.
- **Minimal Information Asymmetry**: this privacy pattern relates to the fact that when users interact with an agent, they may know less information about themselves than the agent itself. In our system, this pattern is implemented through the use of transparent messages about the interactions. Moreover, no identifiers about the users are maintained on the server, which consequently cannot correlate data reported at different times together.
- **Incentivized Participation or Reciprocity**: this pattern states to pay back users that take part into a system, and is known under different names. This is implemented in our system intrinsically since our aim is exactly to reward users for their data. Moreover the pattern also describes that users may have different participation in a computing system, hence they may be rewarded differently. In our system this is implemented by the variable reward each user can obtain, depending on the information they share. A more precise location data allows for a more detailed representation of the environment by the central server, with a higher associated reward, while coarser reports are associated with a lower reward.

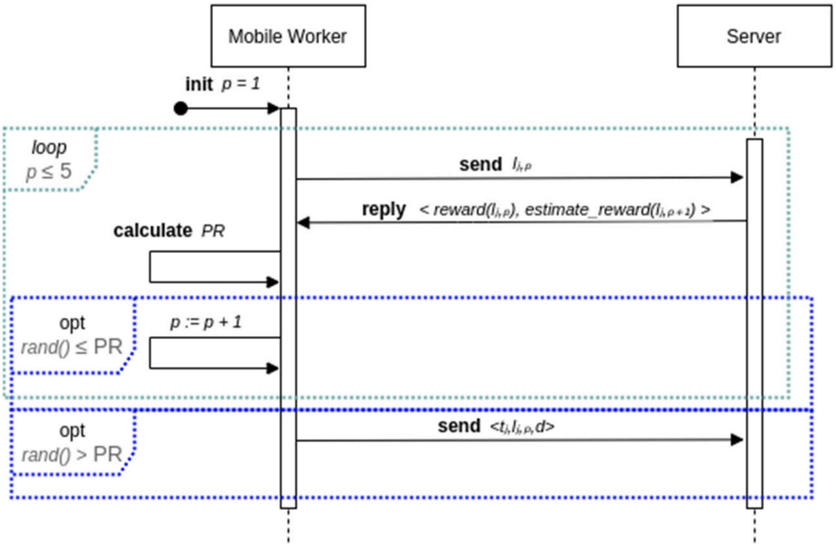


Fig. 2. Sequence diagram of the negotiation. Workers and the Server communicate to achieve each one their objective.

- **Identity Federation — Do not track pattern**: this privacy pattern states that the Identity Provider cannot learn the correlation between the user and its data. In our platform this is implemented since no personal information nor pseudo-identifiers are maintained together with the data provided.
- **Personal Data Store**: this privacy pattern states that personal data can be kept on the device rather than on a central server, so that data cannot be used to identify users. In our system this is implemented by the fact that the server does not maintain any personal data about the user, which is never shared by her.
- **User data confinement pattern**: this privacy pattern is similar to the previous one, and aims at moving some computation to the device of the user, so that less data is shared to the central server. Specifically this pattern is tackled by the device which takes autonomous decisions about sending more precise information. The server only advertises the rewards the user can get, and the user decides on its own whether to increase her measurement privacy or leave it as it is.
- **Anonymity Set**: this pattern aims at removing identifying information from data, so that multiple users can be part of a set without specific data which can link to any of them. In our platform this is implemented on the server, since the data maintained by it do not contain any specific information about the user, so all users belong to the same anonymity set.

4. Negotiation phase

Whenever a mobile worker is about to send out sensor data, it initiates a negotiation phase with the server in order to establish the best precision at which to disclose his/her location. The process is described at a high level in Fig. 2 in the form of a sequence diagram, while the steps are described in detail in this section. This process is executed each time a worker needs to send data to the remote server and it does not involve any human interaction, in fact it runs automatically in the worker's device and takes decisions upon specific parameters. In our scenario, we run the negotiation phase every Δt seconds by every $W \in \mathbf{W}$.

4.1. Initialization

Suppose that the mobile worker W_i wishes to send sensor data d at timestamp t_j. He/she first detects its own position and translates it into an MGRS coordinate with the maximum precision ($l^{\{i\}}_{j,5}$). This means that W_i knows at all times its own position $l^{\{i\}}_{j,p}$ for each p.

4.2. Advertise position

Subsequently, W_i sends its MGRS location $l^{\{i\}}_{j,p}$ to the server, where p is given in input. If this is the first time within t_j, then $p = 1$, therefore the *least* precision. This lets the server know the *rough* position of W_i, but does not disclose the precise location of it.

4.3. Return estimated reward

Once the server receives the location of W_i it computes two values: (i) the reward that W_i would earn by only disclosing its location with precision p, defined as $R(l^{\{i\}}_{j,p})$, and (ii) the estimated reward that W_i would earn by increasing its precision by 1, defined as $\tilde{R}_{p+1}(l^{\{i\}}_{j,p})$. Both values are then sent back to W_i.

More in detail, we assume that R is a function of both the freshness of existing data for the considered MGRS location and the existing number of measurements. In this paper we define the reward as:

$$R(l^{\{i\}}_{j,p}) = \frac{t_j - t_{last}(l^{\{i\}}_{j,p})}{N(l^{\{i\}}_{j,p}) + 1}, \tag{1}$$

where $t_{last}(l)$ is a function that returns the number of seconds elapsed since the last data point was uploaded within location l, and $N(l)$ is the number of data points uploaded from l. Note that R is computed by the server, which has a complete view over the data already uploaded by all $W \in \mathbf{W}$. The idea behind the definition of Eq. (1) is that R tends to be higher if the data within $l^{\{i\}}_{j,p}$ is not fresh, meaning that fresher data has a higher importance. Conversely, this is mitigated if $N(l^{\{i\}}_{j,p})$ is high, meaning that the systems rewards more when the location has been less monitored previously. Clearly different rewards can be adopted in this step, depending on the scenario and on the campaign owner objectives.

The server also computes $\tilde{R}$, which estimates the reward that would be generated from a higher precision. Let us first define the notion of inclusiveness between areas: let l_{1,p_1} and l_{2,p_2} be two MGRS coordinates. We say the $l_{2,p_2} \subset l_{1,p_1}$ if and only if l_{2,p_2} is fully contained within l_{1,p_1}. This also implies that if $l_{2,p_2} \subset l_{1,p_1}$, then $l_{1,p_1} = l_{2,p_1}$ and $p_1 < p_2$. Then we define the estimated reward as:

$$\tilde{R}_{p+1}(l^{\{i\}}_{j,p}) = \frac{\sum_{l'_{j,p+1} \subset l^{\{i\}}_{j,p}} R(l'_{j,p+1})}{|l'_{j,p+1} \subset l^{\{i\}}_{j,p}|}. \tag{2}$$

Eq. (2) therefore calculates the mean over the rewards of all the squares with precision $p + 1$ that are contained within $l^{\{i\}}_{j,p}$.

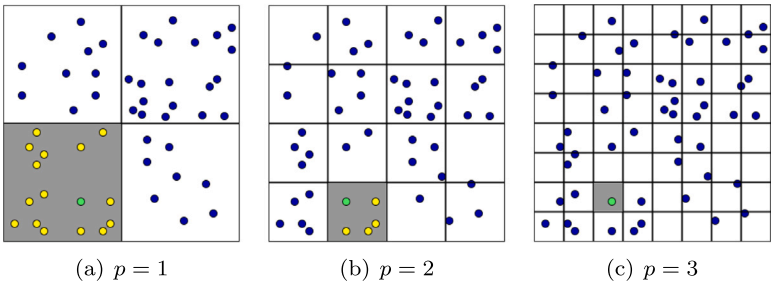


Fig. 3. Negotiation example.

4.4. Calculate probability

At this point in time, W_i receives the information about the reward that he/she would earn if submitting the data with precision p ($R(l^{\{i\}}_{j,p})$) as well as the mean reward that he/she would earn if submitting the data with precision $p+1$ ($\tilde{R}_{p+1}(l^{\{i\}}_{j,p})$). W_i then by using a higher precision would expect a mean gain $\tilde{R}_{p+1}(l^{\{i\}}_{j,p}) - R(l^{\{i\}}_{j,p})$. Such a gain is always positive, as proven by Theorem 1 (Theorems and proofs are presented at the end of this paper). Based on the gain calculated, W_i can decide whether to disclose his/her position with a higher precision, promoting then the gain, or keeping the precision as it is, promoting the privacy preservation. As every mobile worker may behave differently depending on their preferences, we model such a behavior by defining two probabilities: $PR^{\{gain\}}$ (probability promoting gain) and $PR^{\{privacy\}}$ (probability promoting privacy). They are defined as follows in Eq. (3) and in Eq. (4):

$$PR^{\{gain\}} = \frac{\tilde{R}_{p+1}(l^{\{i\}}_{j,p}) - R(l^{\{i\}}_{j,p})}{\tilde{R}_{p+1}(l^{\{i\}}_{j,p})} \tag{3}$$

$$PR^{\{privacy\}} = \frac{R(l^{\{i\}}_{j,p})}{\tilde{R}_{p+1}(l^{\{i\}}_{j,p})}. \tag{4}$$

In order to model different worker preferences, we then set the probability of choosing a finer precision as defined in Eq. (5):

$$PR = \alpha \cdot PR^{\{gain\}} + (1-\alpha) \cdot PR^{\{privacy\}}. \tag{5}$$

The α parameter, with $0 < \alpha < 1$, is tuning the tendency of the mobile worker to be more conservative or seeking a higher reward. A higher value of α indicates a major willingness to use a higher precision, a lower value of α indicates the opposite.

4.5. Send data or reiterate

Once W_i calculates its value of PR, he/she now performs a conditional choice (shown in the `opt` block in Fig. 2). With probability PR, W_i increases its MGRS precision by 1 and goes back to the "Advertise Position" step using the new precision value. This process eventually stops W_i to perform the second branch in the `opt` block with probability $1 - PR$ or p reaches 5 (in such case PR would be 0). Under this choice, W_i sends the tuple $D = \langle t^{\{i\}}_j, l^{\{i\}}_{j,p}, d\rangle$ with precision p.

In Fig. 3 we show a sample negotiation of a client. Note that the server has already stored the blue and yellow measurements, and is now negotiating for the green one. As we explained, the worker starts from a coarse precision, which ends up in a scenario like Fig. 3(a), in which the location of the client is sent as the gray MGRS square, which is shared by all the other 15 yellow measurements, which means that the data point would be fairly anonymized. At this point the server replies with the reward for the measurement, and advertises the possible increased rewards in case the clients decides to increase the precision. The client decides to continue and ends up in the scenario depicted in Fig. 3(b), where the gray MGRS square is smaller hence shared only by 4 other clients' measurements. The same negotiation is repeated again, and in case the client indefinitely continues to increase even more the precision of the measurement, it will end up in the scenario depicted in Fig. 3(c), where the green measurement is the only one in a given MGRS square, thus uniquely identifying it in the dataset and compromising anonymization.

5. Privacy metrics

The whole negotiation process described in Section 4 is executed between every $W \in \mathbf{W}$ and the Server. Since mobile workers can only roughly estimate the number of other mobile workers nearby from the values of R and $\tilde{R}$ received by the Server, they can only estimate their own privacy preservation. In this paper we quantify such a metric by adapting the well-known concept of k-anonymity (Sweeney, 2002) to the present problem. We recall that a scenario is k-anonymous, with k known, if and only if each and every user of the system is indistinguishable from at least other $k-1$ other users. In our case, this translates into verifying that every MGRS coordinate recorded onto the Server is repeated at least k times over a certain time interval T. We note that in this work we are not focusing on the k-anonymity of the user trace, but rather on the anonymity of each reported measurement. Extending our framework to the user trace is left as a future work. More formally, it should stick to the following constraint defined in Eq. (6):

$$\begin{aligned}
&\forall l^{\{i\}}_{j,p}.\exists l^{\{u_1\}}_{j_1,p_1}, l^{\{u_2\}}_{j_2,p_2}, \dots, l^{\{u_k\}}_{j_k,p_k} \quad \text{s.t.}\\
&u_1 \neq u_2 \neq \dots \neq u_k\\
&|j-j_1| \leq T, |j-j_2| \leq T, \dots, |j-j_k| \leq T\\
&l^{\{u_1\}}_{j_1,p_1} \subseteq l^{\{i\}}_{j,p}, l^{\{u_2\}}_{j_2,p_2} \subseteq l^{\{i\}}_{j,p}, \dots, l^{\{u_k\}}_{j_k,p_k} \subseteq l^{\{i\}}_{j,p}
\end{aligned} \tag{6}$$

Given the probabilistic nature of the negotiation phase, it is highly unrealistic that a fully-fledged k-anonymity would be fulfilled. For this reason, assuming that a tolerance for k-anonymity exists in our scenario, then we adopt an alternative metric named k-quasi-anonymity, which aims to fulfill its correspondent k-anonymity as much as possible.

Suppose a scenario in which k-quasi-anonymity is required for a certain value of k. For a certain T, we will then have a set of data points uploaded in different locations by different mobile workers. More formally, let us define such a set as $\mathbf{L_T} = \{l^{\{u_1\}}_{j_1,p_1}, l^{\{u_2\}}_{j_2,p_2}, \dots, l^{\{u_k\}}_{j_k,p_k}\}$. If, for our fixed k, L_T does not satisfy Eq. (6), then it does not satisfy k-anonymity. In order to estimate to what extent it satisfies k-quasi-anonymity, we apply iteratively a transformation to L_T which reduces the precision of a number of its locations. Once the transformed set L'_T satisfies the k-anonymity, then the transformation stops and the number of locations that have been updated with a new precision is the k-quasi-anonymity score (QS_k) of L_T. A lower value of $QS_k(L_T)$ means a higher privacy preservation, with $QS_k(L_T) = 0$ meaning that L_T satisfies k-anonymity. The transformation is outlined in Algorithm 1.

The algorithm takes as input the level of anonymity k and the set of locations $\mathbf{L_T}$. It is important to point out that Algorithm 1 works if L_T has no duplicate MGRS location belonging to the same worker, as k-anonymity is only significant among k different subjects to anonymize. If this happens, the algorithm drops the duplicates before running. The

Algorithm 1: Calculate k-quasi-anonymity.

```
   Input  : $k$, $L_T$
   Output: $k$-quasi-anonymity score: $QS_k(L_T)$
 1 $anonymous = false$
 2 $pcheck = 5$
 3 $QS_k(L_T) = 0$
 4 while not anonymous do
 5     $C_T = l \in L_T$ s.t. $l$ does not fulfill $k$-anonymity
 6     $k_{current} = max(k')$ s.t. $L_T$ fulfills $k'$-anonymity
 7     if $k_{current} \geq k$ then
 8         $anonymous = true$
 9     else
10         Reduce precision of $l \in C_T$ that have $p = pcheck$ by 1
11         $QS_k(L_T) = QS_k(L_T)$+ number of reductions
12         $pcheck = pcheck - 1$
13     end
14 end
15 return $QS_k(L_T)$
```

algorithm starts by initializing $QS_k(L_T)$ to 0 and the precision to be checked $pcheck$ as the maximum value of precision (lines 2–3). Next it cycles until L_T fulfills the level of anonymity k (line 4). Within the cycle, first a set of *candidates* is generated. Such a set is indicated with C_T and contains all the locations in L_T that do not fulfill Eq. (6) (line 5). They are called candidates because the algorithm will select among them a number of locations for which reducing the precision. Next, we calculate the current anonymity level of L_T, i.e. the maximum k' for which L_T fulfills k'-anonymity (line 6). If that is greater or equal than the desired anonymity level k, then it exits the cycle (line 7–8). Otherwise, the precision of all the locations in the set of candidates C_T that have their precision equal to $pcheck$ (which starts off at 5) is reduced by 1. $QS_k(L_T)$ is then increased by the number of such reductions (11). However, if a candidate has been reduced more than once during the whole process, it only counts once, therefore it holds that $QS_k(L_T) \leq |L_T|$. Finally, $pcheck$ is decreased by 1 and the whole cycle repeats (line 12). The idea behind this procedure is to (i) select the locations that are not anonymized according to k (C_T), (ii) gradually reduce the MGRS precision of the locations in C_T starting with the most precise one, and (iii) stop this process whenever the k-anonymity is achieved. In this way, the metric QS gives us a "distance" between the desired anonymity level and the one existing in the scenario. A more informative and less quantitative version of this metric is defined as the Relative k-quasi-anonymity, which is defined as follows in Eq. (7):

$$RQS_k(L_T) = \frac{QS_k(L_T)}{|L_T|} \tag{7}$$

The latter definition ensures that $0 \leq RQS_k(L_T) \leq 1$ and represents our main metric used within the next section. On a side note, we acknowledge that RQS cannot capture efficiently the level of privacy in cases where workers are extremely sparse. In such extreme situations, however, workers will always submit measurements with the coarsest possible granularity by stopping at the very first step of the negotiation phase. This implies that privacy is sort of natively guaranteed by the extremely vague localization, in fact these cases are less relevant for the purpose of this study.

6. Performance evaluation

In this section we present how we evaluated the proposed system, both against the negotiation phase and the k-quasi-anonymity metric. We first outline our experimental setup, which frames all the experiments into a common storyline, then we present singularly each result.

6.1. Experimental setup

The evaluation process has taken place through the usage of the well-known TLC Trip Record Dataset[1] – we will hereafter refer to it as TLC – which contains the traces of all the green and yellow taxis in New York City. The traces can be downloaded daily, in our case, we used the full set of traces from January 6th, 2015. We first separated the traces depending on the NYC boroughs, in order to run localized simulations. In order to do so we isolated the traces that lie entirely within a borough and split those that are crossing the borders. From this step we obtained 20 different datasets, one for each borough. We then further split each of such dataset over time, dividing each of them onto different 30 [min] chunks. From these, we erased the ones with less than 100 vehicles, as they would be too sparse for our evaluation, obtaining then 401 different datasets, on top of which we ran a different simulation each. In each simulation vehicles are moving within the area of interest and are triggering sensing events at a constant pace, specifically $\Delta t = 10$ s. Each sensing event will then trigger a negotiation phase with the server, as outlined in Section 4, which will eventually result in uploading a single data point with a defined spatial precision. We also set the value for T to be 30 min. Given the probabilistic nature of our framework, we ran each simulation 50 times. The simulator has been developed ad-hoc and written in Python 3, its code is released open source.[2]

The above described simulation setup enables us to thoroughly validate and evaluate both the negotiation mechanism and the k-quasi-anonymity, as results from different boroughs and times of the day can run independently, while we aggregate them in order to show the final results. Such results, shown in the next section, have a twofold rationale: first of all we examine all the aspects of the proposed framework under different conditions to grasp the behavior of each of the parameters involved and how they influence the metrics, second, we aim to give a model for individual workers on how to tune the parameters themselves in order to achieve a certain level of privacy or a certain reward or, better, how to minimize the disclosure of sensitive location data while aiming to maximize the personal reward. Many of the results turned out to be dependent on a number of environmental parameters, namely the number of vehicles and morphological aspects of the borough itself (the density of the streets, the number of junctions etc.). This is positive, because it might give workers a guideline on how to orient their parameters on top of such conditions. In this paper we choose to borrow the LWVD metric (Layout-Weighted Vehicle Density) from literature (Bedogni et al., 2018). This metric is, given a map M, calculated as follows in Eq. (8):

$$LWVD_M = \frac{|V_M| \cdot |E_M| \cdot N(M)}{Len^2_M}, \tag{8}$$

where V_M is the set of nodes in M, E_M is the set of edges in M, $N(M)$ is the average number of vehicles in M and Len_M is the *sum* of the length of the edges in M.

More in detail, next section will show the following: (i) the overall behavior of $RQS_k(L_T)$ for different values of α and k as well as at different times of the day, (ii) how different values of the privacy requirement k would affect the precision of the submitted measurements (i.e. how many measurements would not fit the requirement), (iii) how the LWVD parameter can be used to drive workers into choosing a near-optimal α parameter, and (iv) how other aspects such as the road type could also influence the choice.

[1] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

[2] https://github.com/lbedogni/variable-reward

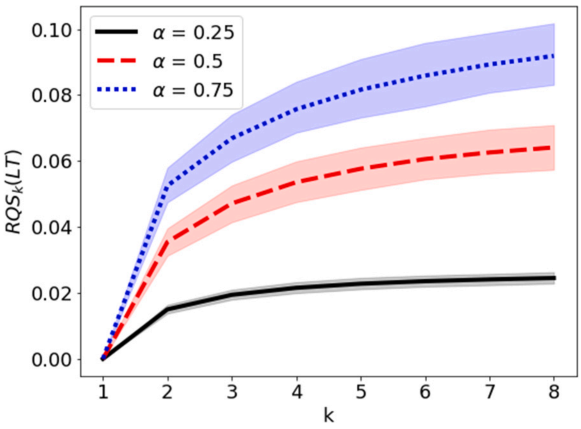


Fig. 4. $RQS_k(L_T)$ for different values of k and different α.

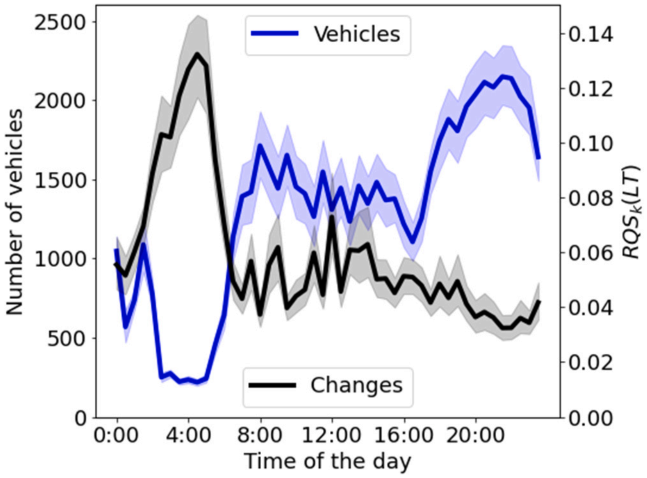


Fig. 5. Number of vehicles and $RQS_k(L_T)$ over time for a simulation covering 24 h.

6.2. Comparison

To better evaluate our proposal, we compare it against a baseline algorithm that is equivalent to the ones found in many different contributions (Andrés et al., 2013; Micinski et al., 2013). To better frame this comparison we also introduce the difference between spatial cloaking and location obfuscation, which share some common aspects (Kim et al., 2022). Mainly, location cloaking aims at reducing the precision of the location measurement, to maximize the number of users that share the same location resulting in a better protection against re-identification attacks. However, this requires a trusted server, which feeds back to the user the precision that they have to select in order to meet the desired privacy levels. Location obfuscation instead alters the precise location of the user with noise without the need for a trusted entity (Andrés et al., 2013) (Micinski et al., 2013). Comparing our proposal with spatial cloaking alternatives would clearly lead to reduced performance, as it is clearly impossible to achieve better results than those which can be obtained with a centralized optimization process and precise information from the server. Instead, we compare our work with obfuscation techniques, which do not require a trusted server and can then provide the location precision they envision to be the most suitable one. Since current works in literature reduce the location precision in an empirical way, a baseline obfuscation that uses a constant precision reduction is equivalent. For this reason, by using MGRS as reference system, we then compare our proposal against obfuscation algorithms that keep the precision reduction constant at different granularities, using the MGRS squares of different size in which the real location lies within.

6.3. Results

The results from the experiments introduced in the previous section are presented here. If not otherwise specified, the results are averaged out over all the boroughs and all the time chunks.

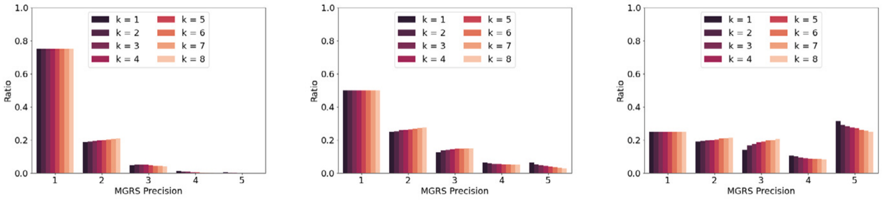
Fig. 4 shows the value of $RQS_k(L_T)$ over all the simulations by varying the number of k. The figure clearly validates the expected monotonic increase of $RQS_k(L_T)$: in fact, a higher k demands more precision reductions within L_T in order to meet the privacy requirement. Secondly, we also observe how a higher α causes more measurements to undergo a precision reduction, as it means a higher tendency to promote rewards over privacy.

This is validated further by Fig. 5, in which the $RQS_k(L_T)$ is shown over a time span of 24 simulated hours. In the same plot we also shown how the number of vehicles changes during the day and how these two parameters are correlated: with a considerably high number of vehicles (e.g. at 8:00 PM), under the same working conditions, the value of $RQS_k(L_T)$ decreases, while it increases in the opposite scenario (e.g. at 4:00 AM). This is again expected, as a higher number of vehicles gives a higher chance for a single vehicle to be obscured by the presence of others in its proximity.

Fig. 6 shows three histograms, each for a different value of α, that display the ratio of measurements having a certain MGRS precision. This is calculated after k-quasi-anonymity has been applied, with k ranging from 1 to 8. Note that $k = 1$ means that k-quasi-anonymity returns exactly the same L_T, therefore it resembles k-quasi-anonymity not being applied. For each MGRS precision, we notice that from left to right the bars adjust slightly to meet the desired k-anonymity. In particular, for $\alpha = 0.25$ we observe that measurement with $p \geq 3$ tend to decrease while increasing the measurements with $p = 2$. For $\alpha = 0.5$ this happens for $p \geq 4$. This is expected as $\alpha = 0.5$ means that workers are generally less conservative, therefore more willingly submitting precise measurements. For the same reason, we observe a similar behavior for $\alpha = 0.75$, although the measurements having a high p are more numerous.

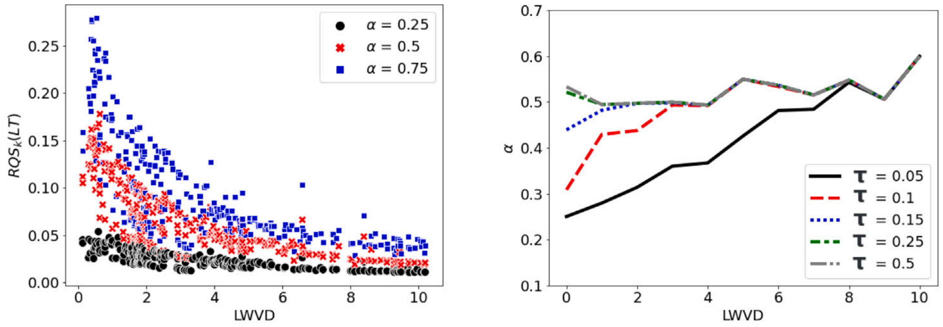
An important outcome is given by the evaluation in Fig. 7, which studies the relationship of several parameters with the LWVD metric. We observe that such a metric causes parameters to change almost linearly, thus giving us a clear characterization of a location against the framework presented in this paper. In detail, we first validate the $RQS_k(L_T)$ against the LWVD metric in Fig. 7(a). The figure shows the high values of LWVD are more likely to cause less precision drops, because a high LWVD implies a higher density of the population, which also leads to workers being unlikely alone and easily anonymizable. On a side note, we observe that $RQS_k(L_T)$ gets more boosted in absolute for higher values of α, which confirms the previous considerations. This led us to use LWVD as a driving parameter: in Fig. 7 we are fixing a threshold value τ for $RQS_k(L_T)$. This threshold τ accounts for the maximum value of $RQS_k(L_T)$ that the system is going to tolerate. This is crucial, because this way we can give a hint on how potentially adjust the α parameter in order for $RQS_k(L_T)$ to stay below such a threshold τ on top of other environmental parameters. In the case of Fig. 7(b), the environmental parameter is the LWVD, and we can observe that, for most of the values of τ(intentionally low values, to ensure a decent privacy level), the suggested α value tends to increase as LWVD increases. This is expected: the denser the environment is, the higher α can be used safely. We also notice that, for values of τ over 0.25 the curve does not change, as a higher α would cause a privacy loss nonetheless.

A consequence of this finding is also reflected in Fig. 8, where a number of parameters is studied on top of τ . All these analyses are compared against a baseline, which keeps the MGRS precision of the data sent constant. In particular, for each plot in Fig. 8, the



(a) $\alpha = 0.25$ (b) $\alpha = 0.5$ (c) $\alpha = 0.75$

Fig. 6. Histograms showing the ratio of measurements with a certain MGRS precision after applying k-quasi-anonymity for different values of k.



(a) Scatterplot displaying the value of $RQS_k(L_T)$ for different values of LWVD.

(b) Lineplot displaying the maximum α for which a certain $RQS_k(L_T)$ is tolerated.

Fig. 7. Plots showing the influence of LWVD and how such parameter can be beneficial for the individual workers to tune internally their parameters.

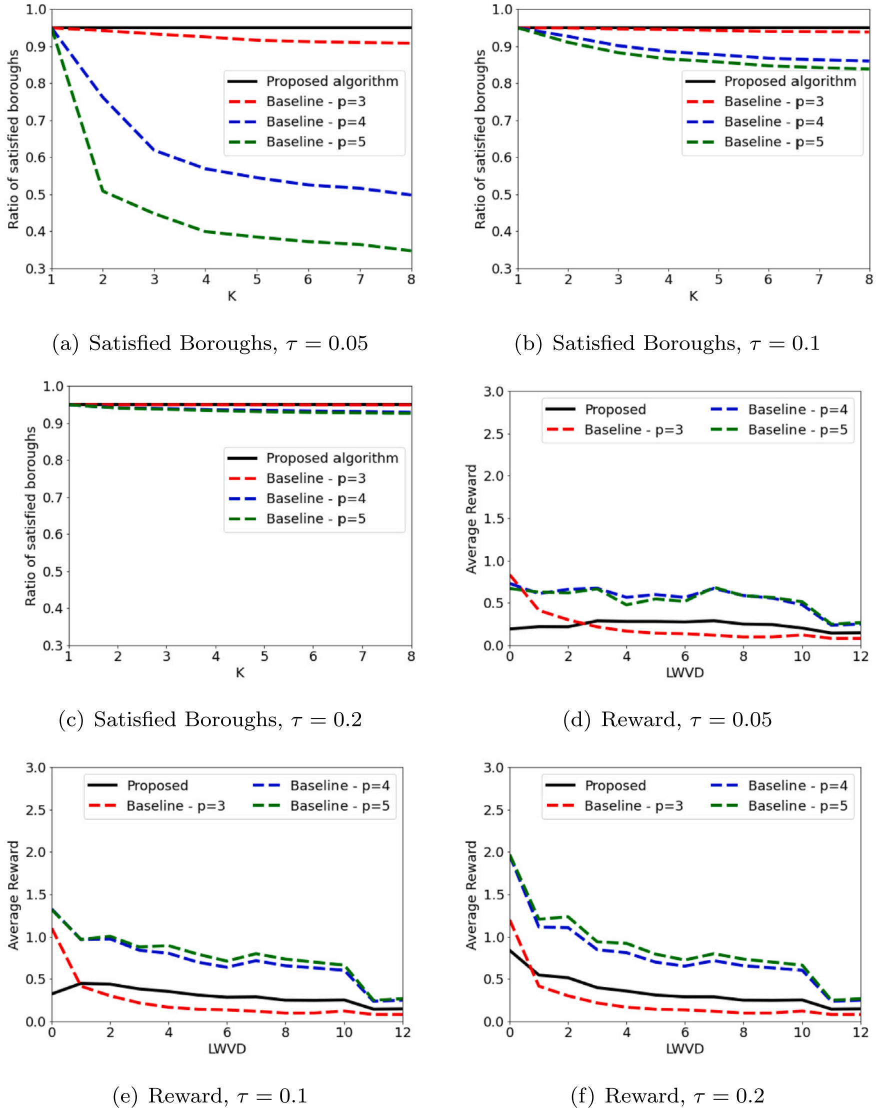
solid line represents our proposal, whereas the dashed lines represent the baseline with $p = 5$, $p = 4$ and $p = 3$ – lower values were not plotted for displaying purposes, as their behavior is not relevant. Figs. 8(a)–8(c) show, for different values of k, the "ratio of satisfied boroughs", that is the ratio of boroughs in which measurements sent by the workers fulfill the established threshold – i.e. $RQS_k(L_T) \leq \tau$ – during the whole simulation, averaged out considering all three values of α used for Fig. 7(a). We notice how our solution keeps the value constantly to almost 1, whereas constant values of MGRS precision cause many measurements to be exposed. By comparing these results with Figs. 8(d)–8(f), we observe that the baseline with $p = 5$ and $p = 4$ tend to reward the users more that our proposal, whilst $p = 3$ keeps a generally lower reward, as the behavior is far more conservative. Nonetheless, all the three baselines still cause some of the simulations to not meet the threshold τ . Our proposal instead almost always meets the the condition $RQS_k(L_T) \leq \tau$, while guaranteeing a relatively high reward at the price of a higher $RQS_k(L_T)$ that, however, is always below the tolerated value. We notice how, for our solution, no matter the value of α, the only affected boroughs are the very few ones with a fairly low LWVD. We assume that a widespread deployment of our system would also affect a tiny portion of less crowded areas, for which the risk for exposure is higher. Looking at the plots in Fig. 8, we can see that the value of τ only affects such few boroughs, while for the vast majority of the others it does not. For this reason, a general guideline could be setting globally a low τ (i.e. 0.05), because for crowded areas the outcome does not change, while it guarantees a higher privacy preservation for less crowded ones.

A further interesting outcome is shown in Fig. 9(a). The boxplot shows the MGRS precision of each measurement submitted, after k-quasi-anonymity has been applied. In this case we used $k = 8$. The figure displays the effect of the framework over three types of road, as reported on OpenStreetMap: primary, secondary and tertiary. The pattern displays clearly that primary roads are likely to allow a higher precision to workers, because of their higher people density. Such a precision is slightly but steadily decreasing as roads get smaller and, thus, less crowded. This is an interesting outcome, as it shows that even the road type could be significant for choosing the MGRS granularity, therefore, in an enhanced version of the framework it could offer an additional feature to help workers in discriminating the precision themselves.

Finally Fig. 9(b) shows an example on a constrained area of NYC which plots an heatmap regarding the average reward which can be obtained in different locations. It is immediately evident that bigger roads and, more importantly, cross roads are areas in which the density of vehicles is higher, hence workers may provide higher precision measurement locations since they can better hide in the crowd.

The analysis provided in this section, although performed onto the specific case of the city of New York, highlights a number of important takeaway lessons that can be applied to the general case.

First, our method is superior to the considered baseline, which, to the best of our knowledge, represents the equivalent to existing solutions that work empirically upon location obfuscation without trusted entity (Micinski et al., 2013; Andrés et al., 2013). The evaluation shows that even setting a very low tolerance threshold for k-quasi-anonymityhas a negligible negative impact on the overall reward per user, while ensuring a higher privacy preservation even in low populated areas. Second, while on the one hand our method has shown to be equally effective for crowded areas (i.e. with a high LWVD) no matter the configuration, on the other hand its effectiveness is more subject to the user/crowd-specific parameters α and τ for less crowded ones. In such areas, correctly setting such parameters might be crucial. While we showed that τ can be set to be globally low, α is instead dynamically (and automatically) set by the user's device. In Fig. 7(b) we showed how, given τ, we can automatically set α based on environmental parameters, that are supposed to be constant and known (the LWVD of an area and the road type are only examples). Third, this study suggests that future works could investigate more environmental parameters and how each of them influences the $RQS_k(L_T)$. Furthermore, we could envision automated adjustments techniques based on, for instance, reinforcement learning for fine-tuning α as to achieve the best privacy-reward tradeoff.



(a) Satisfied Boroughs, $\tau = 0.05$

(b) Satisfied Boroughs, $\tau = 0.1$

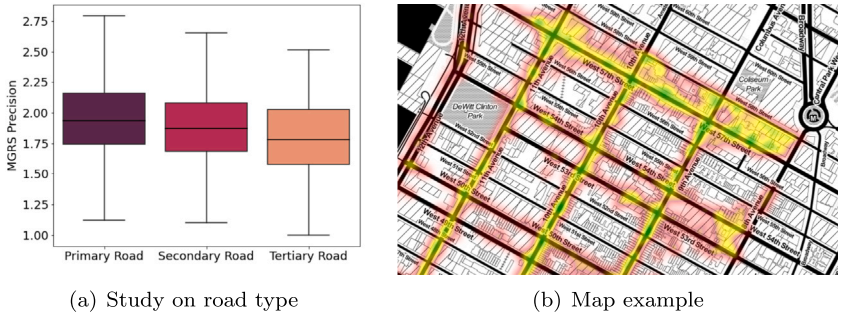
(c) Satisfied Boroughs, $\tau = 0.2$

(d) Reward, $\tau = 0.05$

(e) Reward, $\tau = 0.1$

(f) Reward, $\tau = 0.2$

Fig. 8. Plots showing the performance of our proposal against a baseline with fixed MGRS precision.



(a) Study on road type

(b) Map example

Fig. 9. Fig. 9(a) shows the average MGRS precision of each measurement after k-quasi-anonymity with $k = 8$ has been applied for different types of road, while Fig. 9(b) presents an example which highlights where user are able to achieve better rewards thanks to a higher density of vehicles.

7. Conclusion

In this paper we have proposed a novel framework to reward MCS workers which jointly considers the data quality and the privacy of users. We have shown how our framework can effectively reward users which are willing to report more precise measurements, also taking into account the user preference regarding the privacy protection. Our results obtained considering real data highlight that there are many factors to take into account, although certainly a larger user base enables a better protection for users, which can send more precise locations while still being protected by similar measurements. To achieve this, we have also presented the k-quasi-anonymity metric, focused on understanding how frequent is a location in the remote server, so that the user reporting such measurement is more exposed than others.

We have also shown how urban dynamics impact on the privacy of users, such as the kind of road in which the measurement took place, which directly relates to the density of vehicles which may provide measurements with similar locations. Clearly, this also depends on the mobility of users, and on the set of possible locations which may be sent.

The main future work on this topic is to enable workers to automatically determine the optimal precision to use when sending data, which can be done by examining the city dynamics and statistical historical data.

Theorems and proofs

Theorem 1. *Let l_p be an MGRS location with precision p, then $\forall l'_{p'}$ such that $l'_{p'} \subset l_p$, then $R(l_p) \leq R(l'_{p'})$.*

Recall the definition of R given in Eq. (1), we will have $R(l_p) = \frac{t_j - t_{last}(l_p)}{N(l_p)+1}$ and $R(l'_{p'}) = \frac{t_j - t_{last'}(l'_{p'})}{N(l'_{p'})+1}$ for a certain time slot t_j. If we define $t_{last'}$ as the timestamp of the last measurement in $l'_{p'}$, we will certainly have that $t_{last} \geq t_{last'}$, as the last measurement occurred in $l'_{p'}$ has also occurred in l_p, while the opposite is not necessarily true. Furthermore, we have that $N(l'_{p'}) \leq N(l_p)$, as all measurements in $l'_{p'}$ also belong to l_p. Upon these premises we can easily verify that if $N(l'_{p'}) \leq N(l_p)$ and $t_{last} \geq t_{last'}$, then $R(l_p) \leq R(l'_{p'})$.

Corollary 1. *Let l_p be an MGRS location with precision p, then $\forall p' > p$ we have that $R(l_p) < \tilde{R}_{p'}(l_p)$.*

Recall the definition of $\tilde{R}$ given in Eq. (2), we have that $\tilde{R}_{p'}(l_p)$ is the average over a set of rewards $R(l'_{p'})$ such that $l'_{p'} \subset l_p$. Then, by Theorem 1, for all the $l'_{p'}$ it holds $R(l_p) \leq R(l'_{p'})$, hence their average is also greater or equal to $R(l_p)$.

CRediT authorship contribution statement

Luca Bedogni: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Federico Montori:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Dataset used is public available.

References

Alagha, A., Mizouni, R., Singh, S., Otrok, H., Ouali, A., 2021. SDRS: A stable data-based recruitment system in IoT crowdsensing for localization tasks. J. Netw. Comput. Appl. 177, 102968 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804520304197.

An, J., Cheng, J., Gui, X., Zhang, W., Liang, D., Gui, R., Jiang, L., Liao, D., 2020. A lightweight blockchain-based model for data quality assessment in crowdsensing. IEEE Trans. Comput. Soc. Syst. 7 (1), 84–97.

Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C., 2013. Geo-indistinguishability: Differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. pp. 901–914.

Arkian, H.R., Diyanat, A., Pourkhalili, A., 2017. MIST: Fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowdsensing applications. J. Netw. Comput. Appl. 82, 152–165 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804517300188.

Azzam, R., Mizouni, R., Otrok, H., Ouali, A., Singh, S., 2016. GRS: A group-based recruitment system for mobile crowd sensing. J. Netw. Comput. Appl. 72, 38–50 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804516300170.

Bedogni, L., Fiore, M., Glacet, C., 2018. Temporal reachability in vehicular networks. In: IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, pp. 81–89.

Bou Abdo, J., Bourgeau, T., Demerjian, J., Chaouchi, H., 2016. Extended Privacy in Crowdsourced Location-Based Services Using Mobile Cloud Computing. Mob. Inf. Syst. 2016, 1–13 [Online]. Available: http://www.hindawi.com/journals/misy/2016/7867206/.

Capponi, A., Fiandrino, C., Kantarci, B., Foschini, L., Kliazovich, D., Bouvry, P., 2019. A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. IEEE Commun. Surv. Tutor. 21 (3), 2419–2465.

Cheng, Y., Ma, J., Liu, Z., 2022a. A lightweight privacy-preserving participant selection scheme for mobile crowdsensing. In: 2022 IEEE Wireless Communications and Networking Conference. WCNC, pp. 1509–1514.

Cheng, Y., Ma, J., Liu, Z., Wu, Y., Wei, K., Dong, C., 2022b. A lightweight privacy preservation scheme with efficient reputation management for mobile crowdsensing in vehicular networks. IEEE Trans. Dependable Secure Comput.

Cheng, L., Niu, J., Kong, L., Luo, C., Gu, Y., He, W., Das, S.K., 2017. Compressive sensing based data quality improvement for crowd-sensing applications. J. Netw. Comput. Appl. 77, 123–134 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804516302338.

Hu, J., Yang, K., Wang, K., Zhang, K., 2020. A blockchain-based reward mechanism for mobile crowdsensing. IEEE Trans. Comput. Soc. Syst. 7 (1), 178–191.

Jin, W., Xiao, M., Li, M., Guo, L., 2019. If you do not care about it, sell it: Trading location privacy in mobile crowd sensing. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. IEEE, pp. 1045–1053 [Online]. Available: https://ieeexplore.ieee.org/document/8737457/.

Kim, J.W., Edemacu, K., Jang, B., 2022. Privacy-preserving mechanisms for location privacy in mobile crowdsensing: A survey. J. Netw. Comput. Appl. 200, 103315 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804521003039.

Klopfenstein, L.C., Delpriori, S., Aldini, A., Bogliolo, A., 2019. "Worth one minute": An anonymous rewarding platform for crowd-sensing systems. J. Commun. Netw. 21 (5), 509–520.

Lampinen, R., 2001. Universal transverse mercator (UTM) and military grid reference system (MGRS).

Liu, S., Zheng, Z., Wu, F., Tang, S., Chen, G., 2017. Context-aware data quality estimation in mobile crowdsensing. In: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. pp. 1–9.

Luo, T., Huang, J., Kanhere, S.S., Zhang, J., Das, S.K., 2019. Improving IoT data quality in mobile crowd sensing: A cross validation approach. IEEE Internet Things J. 6 (3), 5651–5664.

Marjanović, M., Skorin-Kapov, L., Pripužić, K., Antonić, A., Podnar Žarko, I., 2016. Energy-aware and quality-driven sensor management for green mobile crowd sensing. J. Netw. Comput. Appl. 59, 95–108 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804515001678.

Micinski, K., Phelps, P., Foster, J.S., 2013. An empirical study of location truncation on android. Weather 2, 21.

Montori, F., Bedogni, L., 2020. A Privacy Preserving Framework for Rewarding Users in Opportunistic Mobile Crowdsensing. In: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops. PerCom Workshops 2020.

Montori, F., Bedogni, L., 2023. Privacy preservation for spatio-temporal data in Mobile Crowdsensing scenarios. Pervasive Mob. Comput. 90, 101755 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574119223000135.

Montori, F., Jayaraman, P.P., Yavari, A., Hassani, A., Georgakopoulos, D., 2018. The curse of sensing: Survey of techniques and challenges to cope with sparse and dense data in mobile crowd sensing for internet of things. Pervasive Mob. Comput. 49, 111–125.

Mota, V.F., Silva, T.H., Macedo, D.F., Ghamri-Doudane, Y., Nogueira, J.M., 2018. Towards scalable mobile crowdsensing through device-to-device communication. J. Netw. Comput. Appl. 122, 99–106 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804518302674.

Ni, J., Zhang, K., Lin, X., Xia, Q., Shen, X.S., 2017. Privacy-preserving mobile crowdsensing for located-based applications. In: IEEE International Conference on Communications.

Notario, N., Crespo, A., Martin, Y.-S., Del Alamo, J.M., Metayer, D.L., Antignac, T., Kung, A., Kroener, I., Wright, D., 2015. PRIPARE: Integrating privacy best practices into a privacy engineering methodology. In: 2015 IEEE Security and Privacy Workshops. pp. 151–158.

Pournajaf, L., Xiong, L., Garcia-Ulloa, D.A., Sunderam, V., 2014a. A Survey on Privacy in Mobile Crowd Sensing Task Management. Tech. Rep. TR-2014-002, Dept. Math. Comput. Sci., Emory Univ., Atlanta, GA, USA.

Pournajaf, L., Xiong, L., Sunderam, V., Goryczka, S., 2014b. Spatial task assignment for crowd sensing with cloaked locations. In: 2014 IEEE 15th International Conference on Mobile Data Management, Vol. 1. IEEE, pp. 73–82.

Sun, G., Sun, S., Sun, J., Yu, H., Du, X., Guizani, M., 2019. Security and privacy preservation in fog-based crowd sensing on the internet of vehicles. J. Netw. Comput. Appl. 134, 89–99 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804519300694.

Sweeney, L., 2002. K-anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 10 (05), 557–570.

Wang, Z., Hu, J., Zhao, J., Yang, D., Chen, H., Wang, Q., 2018. Pay on-demand: Dynamic incentive and task selection for location-dependent mobile crowdsensing systems. In: 2018 IEEE 38th International Conference on Distributed Computing Systems. ICDCS, pp. 611–621.

Wang, L., Yang, D., Han, X., Wang, T., Zhang, D., Ma, X., 2017. Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation. In: Proceedings of the 26th International Conference on World Wide Web. pp. 627–636.

Wang, L., Zhang, D., Yang, D., Lim, B.Y., Ma, X., 2016. Differential location privacy for sparse mobile crowdsensing. In: 2016 IEEE 16th International Conference on Data Mining. ICDM, IEEE, pp. 1257–1262.

Wu, F.-J., Luo, T., 2020. CrowdPrivacy: Publish More Useful Data with Less Privacy Exposure in Crowdsourced Location-Based Services. ACM Trans. Priv. Secur. 23 (1), 1–25 [Online]. Available: https://dl.acm.org/doi/10.1145/3375752.

Yan, K., Luo, G., Zheng, X., Tian, L., Sai, A.M.V.V., 2019. A comprehensive location-privacy-awareness task selection mechanism in mobile crowd-sensing. IEEE Access 7, 77541–77554.

Zhao, B., Liu, X., Chen, W.-N., Deng, R., 2022a. CrowdFL: Privacy-preserving mobile crowdsensing system via federated learning. IEEE Trans. Mob. Comput.

Zhao, S., Qi, G., He, T., Chen, J., Liu, Z., Wei, K., 2022b. A survey of sparse mobile crowdsensing: Developments and opportunities. IEEE Open J. Comput. Soc. 3, 73–85.

Zhao, B., Tang, S., Liu, X., Zhang, X., 2021. PACE: Privacy-preserving and quality-aware incentive mechanism for mobile crowdsensing. IEEE Trans. Mob. Comput. 20 (5), 1924–1939.

Luca Bedogni is an Associate Professor at the University of Modena and Reggio Emilia, Italy. He received his Bachelor Degree and Master Degree (Summa Cum Laude) in Computer Science from the University of Bologna, Italy. His Master thesis developed a MAC Protocol for Vehicular Ad Hoc Networks (VANETs). In 2015, he received the Ph.D. degree from the University of Bologna. His current research interest span from the Internet of Things for heterogeneous devices, to context aware computing.

Federico Montori is a tenured Assistant Professor at the University of Bologna since 2019, focuses his research activities in the field of the Internet of Things (IoT). His research topics deal primarily with service-oriented IoT architectures, distributed systems for the collaborative IoT, integration of different wireless technologies and Cloud and Fog Computing. Currently his research interest include the crowdsensing for environmental and urban monitoring and heterogeneous IoT data analytics.