An Optimal Privacy-Preserving Mechanism for Crowdsourced Traffic Monitoring [*]

Yunhua He
Institute of Information
Engineering, CAS
Beijing, China
and Xidian University
Xi'an, China
heyunhua610@163.com

Limin Sun[†]
Institute of Information
Engineering, CAS
Beijing, China
and Xidian University
Xi'an, China
sunlimin@iie.ac.cn

Zhi Li
Institute of Information
Engineering, CAS
Beijing, China
lizhi@iie.ac.cn

Hong Li
Institute of Information
Engineering, CAS
Beijing, China
lihong@iie.ac.cn

Xiuzhen Cheng
Department of Computer
Science
The George Washington
University
Washington DC, USA
cheng@gwu.edu

ABSTRACT

Crowdsourced traffic monitoring employs ubiquitous smartphone users to upload their GPS samples for traffic estimation and prediction. The accuracy of traffic estimation and prediction depends on the number of uploaded samples; but more samples from a user increases the probability of the user being tracked or identified, which raises a significant privacy concern. In this paper, we propose a privacy-preserving upload mechanism that can meet users' diverse privacy requirements while guaranteeing the traffic estimation quality. In this mechanism, the user upload decision process is formalized as a mutual objective optimization problem (user location privacy and traffic service quality) based on an incomplete information game model, in which each player can autonomously decide whether to upload or not to balance the live traffic service quality and its own location privacy for utility maximization. We theoretically prove the incentive compatibility of our proposed mechanism, which can motivate users to follow the game rules. The effectiveness of the proposed mechanism is verified by a simulation study based on real world traffic data.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; K.4.1 [**Computer and Society**]: Public Policy Issues—*Privacy*

General Terms

Design, Theory

Keywords

Crowdsourcing, location privacy, game theory

1. INTRODUCTION

Crowdsourced traffic monitoring is an important application of mobile crowdsourcing for estimating real-time traffic status and providing live traffic and navigation services. Ubiquitous GPS-enabled smartphone users when traveling along road segments in cities or rural areas are utilized to collect traffic data by uploading their GPS samples containing position, velocity, and time information to a central server [1] [2]. Based on the collected GPS samples, the server can estimate real time traffic status on the corresponding road segments and quickly provide live traffic services such as congestion warning, path planning, and navigation.

Typically, the quality of traffic services for a road segment provided by a crowdsourced traffic monitoring system depends on the number of uploaded GPS samples for that segment [3] [4] – more samples provide a more accurate traffic estimation and prediction. However, the uploaded fine-grained information about location and velocity from GPS-enabled smartphones may reveal users' sensitive information such as traffic law violations, political affiliations, and medical conditions [5]. Therefore, a smartphone user would not use the crowdsourced traffic monitoring system unless its location privacy can be fully protected.

Although smartphone users can upload GPS samples in an anonymous way, the spatio-temporal characteristics of

[*]This Work was supported by the Strategic Priority Research Program of the Chinese Academy of Science (XDA06040100).

[†]Limin Sun is the corresponding author.

FOMC'14, August 11, 2014, Philadelphia, PA, USA.
Copyright 2014 ACM 978-1-4503-2984-2/14/08 ...$15.00.
http://dx.doi.org/10.1145/2634274.2634275 .

the uploaded samples from a vehicle and the vehicular mobility constraints still allow vehicles to be tracked [6] [7]. In order to reduce the spatio-temporal correlations, a distributed approach is sought in this paper, in which smartphone users decide when and where to update GPS samples. The grant challenge is to meet the diverse privacy requirements of smartphone users as different users may have different privacy requirements and these requirements may vary with time and location.

We propose a privacy-preserving upload mechanism that can meet users' diverse privacy requirements while guaranteeing the overall traffic estimation quality. The challenge of our design lies in finding the upload strategy profile of the users, who may not know others' privacy levels, to meet the basic requirement of traffic estimation. We formalize a mutual objective optimization problem (user location privacy and traffic service quality) to characterize the user upload decision process based on an incomplete information game model. In this game, each user is assigned a type, whose probability density function (pdf) captures the distribution of the user's privacy level. A user, according to its belief about its opponent's type, balances the live traffic service quality and its own location privacy to maximize its utility. Base on the analytical results of the Nash equilibrium, we take advantage of the feedback from the server to design a mechanism that meets the basic requirements of traffic estimation. Our major contributions are listed as follows:

- We quantify live traffic service quality and location privacy by modeling the live traffic service quality as a function of the number of upload users on a road segment, and modeling location privacy as a function of tracking incorrectness and identity uncertainty.
- We propose a privacy-preserving traffic data collection mechanism based on a game theoretic model, and demonstrate that our mechanism achieves the dual goal of traffic estimation quality guarantee and user privacy protection. The incentive compatibility of our mechanism also motivates the users to follow the game rules.

This paper is organized as follows. Section 2 outlines the related work. We present our crowdsourced traffic monitoring system model in Section 3 and describe our problem formulation in Section 4. In Section 5, we present our incomplete information game model for analyzing users' upload behaviors. We then propose an optimal privacy-preserving upload mechanism in Section 6. In Section 7, we evaluate our mechanisms through a simulation study based on real world traffic data. We conclude this paper in Section 8.

2. RELATED WORK

In crowdsourced traffic monitoring systems, smartphone users may upload GPS samples in an anonymous way to protect their location privacy. However, anonymization techniques are not sufficient for such a purpose [6–8]. Montjoye *et al.* [6] studied a fifteen-month mobility trace data of one and half million individuals and found that four spatio-temporal points are enough to uniquely identify 95% of them. Though anonymization can hide obvious identifiers, vehicular mobility constraints and spatio-temporal characteristics of the samples from an anonymous vehicle allow itself to be traced.

Various methods to reduce the spatio-temporal correlation against the tracking attack were proposed [8]. These techniques can be classified as either centralized or distributed. In the centralized approaches [9–11], GPS samples are being processed, i.e., reducing the number of recorded samples, integrating the recorded samples, or introducing noise deliberately to the samples, at a trusted centralized privacy server before they are used to estimate traffic. An obvious drawback of centralized approaches is their dependence on the trusted privacy server. Once a server is compromised, the privacy of all associated users is disclosed [12].

Distributed approaches [4, 13, 14] do not depend on any centralized server, but allow smartphone users to determine when or where to update GPS samples at their own wills. As a distributed approach, mix-zone anonymizes user identity by enforcing that a set of users enter, change pseudonyms, and exit a mix-zone in a way such that the mappings between their old and new pseudonyms are not revealed. Palanisamy *et al.* [13] proposed a mix-zone framework to protect location privacy of mobile users traveling on road networks. Liu *et al.* [14] aimed to address the problem of optimal multiple mix-zone placement. We claim that mix-zones can hardly support traffic monitoring because users can not upload their locations before exiting a mix-zone. In [4], Hoh *et al.* proposed a system to specify geographic markers that indicate where vehicles should provide location updates. These markers can be placed to guarantee the maximum tracking uncertainty and to avoid particular privacy sensitive locations. Nevertheless, the markers can hardly meet the diverse privacy requirements of all users. Our approach not only allows users to control their own privacy, but also achieves a dual goal of traffic estimation quality and user privacy.

As game theory is suitable for investigating strategic decision making of multiple players with different objectives, there has been a growing interest in applying game-theoretic approaches to study the issues of mobile network security and privacy [15–18]. Freudiger *et al.* [16] analyzed the non-cooperative behaviors of mobile nodes in a popular location privacy protection mechanism (mix-zone) with a game-theoretic model. Yang *et al.* [17] provided a truthful auction-based incentive mechanism for mobile users to join an anonymous set so that k-anonymity can be achieved. Shokri *et al.* [18] studied the location-privacy of mobile users in location-based services (LBSs) by using the framework of Stackelberg Bayesian games. In our approach, we adopt an incomplete information game to analyze the behaviors of smartphone users with mutual objectives (location privacy *vs.* traffic service quality) in a crowdsourced traffic monitoring system, and propose a privacy-preserving mechanism that possesses an important property of incentive compatibility.

3. SYSTEM MODEL

In crowdsourced traffic monitoring, each user is required to periodically upload its GPS samples, which can be used to estimate the real-time traffic condition by a server [2]. In return, the user can get traffic services such as live traffic and navigation from the server. In practice, the accuracy of the traffic estimation, i.e., the QoS Q of the traffic services in a period of time, depends on the number k of the involved smartphone users who upload GPS samples periodically [4]. For clarity and simplify, we consider individual road segments, which can be easily extended to road networks. Assume a set of smartphone users $P = \{1, 2, \cdots, n\}$

on a road segment are willing to provide their GPS samples because they expect to get a better Q. For the problem of malicious users lying about their values, reputation mechanism is one possible solution, which is outside the scope of the paper. Since smartphones are owned by different individuals, it is reasonable to assume that users have different location privacy levels LP. The location privacy loss caused by uploading a GPS sample is denoted by c. We also assume that the server is curious but honest, it honestly provides traffic services but intends to disclose the users' private information.

3.1 Traffic Service Model

As described above, the accuracy of traffic estimation Q depends on the number k of involved smartphone users on a road segment [4]. A larger k leads to a larger value of Q. Our empirical study[1] indicates that the root-mean-square (RMS) error of speed estimates on road segment i exhibits a monotonic decrease with the number of upload users, denoted by k_i, on the segment (Fig. 1(a)). Let $Q_i = 1 - 1_{RMS}$ denote the accuracy of the traffic estimation on road segment i, where 1_{RMS} represents the normalized RMS error. Then we can fit an empirical pdf of Q_i for the given road segment i with a logarithmic function:

$$Q_i = \log_\alpha(1 + k_i\beta), \qquad (1)$$

where α and β are system parameters, and the $\log_\alpha(1+k_i\beta)$ term reflects the Q_i's diminishing return on k_i, the number of upload users. In this study, we obtain the values of α and β from an empirical pdf of Q by using unconstrained nonlinear minimization over real world data (Fig. 1(b)). In our experiment, we consider three kinds of road segments: straight sections, ramps, and intersections. To achieve a spatial accuracy of 83% ($RMS < 10$), the server needs a minimum of 3 uploads per km for straight sections, 5 uploads per km for ramps, and 8 uploads per km for intersections.

3.2 Privacy Threat Model

The threats to the location privacy of users by an adversary can be classified into two types: i) tracking individual vehicles from a mix of anonymous uploaded samples, and ii) finding the true identities of a given uploaded sample.

3.2.1 Tracking Inference Attack

In this attack, the adversary's goal is to extract a subset of samples generated by the same vehicle, given a series of GPS samples mixed from multiple users. The adversary associates a prior uploaded sample with the next one closest to its prediction, or with the most likely sample. The formulation of such a procedure is described by:

$$\arg\max_x p(x|x_{i-1}), \qquad (2)$$

where the conditional probability $p(x|x_{i-1})$ is defined as the probability of the next uploaded sample at location x given the prior sample at x_{i-1}. In a straight section, an adversary can predict the next uploaded location x according to $\hat{x} = v\Delta t + x_{i-1}$, where v is the reported speed at the prior time or the average speed in historical samples, and Δt is the difference between the timestamps of the two uploaded samples. In a ramp or intersection, an adversary, knowing the mobility profile P_{ij} and the speed profile V_{ij} (as shown in Fig. 2), which can be obtained by historical traces and general mobility constraints [5], can infer the location where the vehicle will reside at the next time.

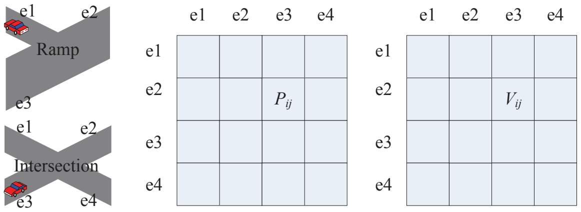


Figure 2: The mobility profile and speed profile in a ramp or an intersection.

3.2.2 Identity Inference Attack

In an identity inference attack, a GPS sample with unknown true identity is given, and the adversary's goal is to infer the most likely owner of the sample with the collected side information. The formulation of such a procedure is described below. Given a specific sample, compute

$$\arg\max_i \Pr(r_i(t)|x(t)), \qquad (3)$$

where $r_i(t)$ is the side information about the location of user i at time t collected by the adversary, and $x(t)$ is the recorded location at time t in the sample. In practice, the side information may be obtained through a number of practical means as follows: i) the locations of the users can be extracted from other public databases [9] such as attendance records and automatic payment records; and ii) users may disclose information on their whereabouts either voluntarily or inadvertently, i.e., by a casual conversation, or by published media [7] [19].

4. PROBLEM DESCRIPTION

4.1 Traffic Service Quality

According to Section 3.1, the accuracy of traffic estimation depends on the number of upload users. Let s_i be the upload strategy of user i, with two possible values: upload (Y) or not (N). Then (1) can be rewritten as:

$$Q = \log_\alpha(1 + \beta\sum_{i=1}^{n} I(s_i, Y)), \qquad (4)$$

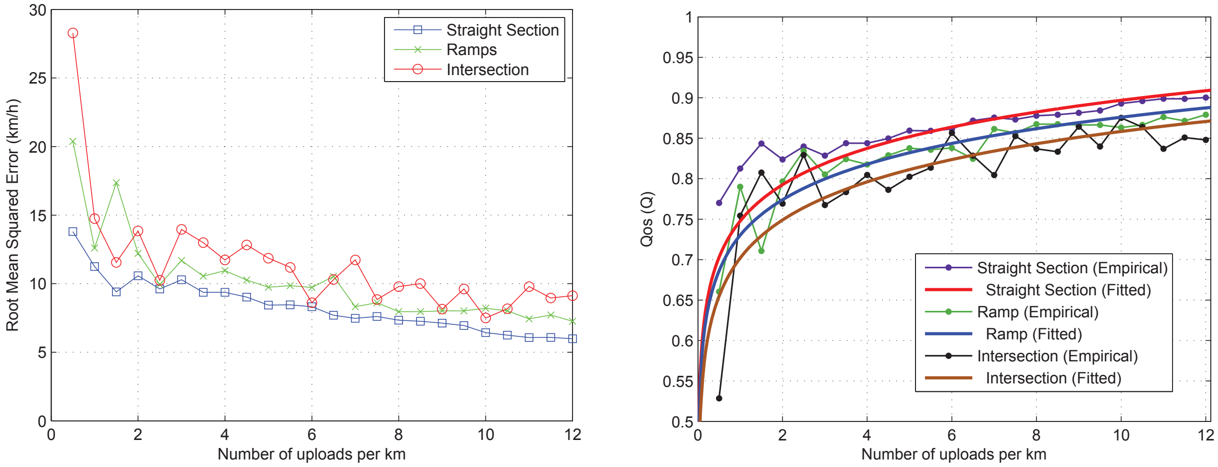
where $I(x, y) = 1$ if $x = y$ and 0 otherwise.

The objective is to guarantee the upload strategy profile $(s_1, s_2, \cdots, s_n)$ of the users such that $Q \geq Q_{min}$, where Q_{min} is the minimum service quality requirement. Since users are more concerned about the change of traffic status at rush hours, the update frequency of live traffic information should be increased, and the number of upload users should be limited in order to improve the server's process efficiency. At the light traffic case, there is still a need for one or two users to upload their GPS samples in order to handle unexpected situations such as accidents.

4.2 Location Privacy

The location privacy of a user is determined by tracking incorrectness [5] and identity uncertainty of the user (see

[1] 1. A month of GPS data from 28,000 taxis in Beijing were used to simulate the upload behavior of smartphone users on road segments. http://www.datatang.com/data/2987



(a) The empirical distribution of speed estimation RMS error *vs.* the number of uploaded samples for the three types of road segments.

(b) The empirical distribution of traffic estimation accuracy *vs.* the number of uploaded samples for the three types of road segments.

Figure 1: Fitting estimation accuracy using logarithmic functions, where $\alpha = 5.0239 \times 10^6$ and $\beta = 1.024 \times 10^5$ for straight sections, $\alpha = 7.1693 \times 10^6$ and $\beta = 1.011 \times 10^5$ for ramps, and $\alpha = 2.4625 \times 10^6$ and $\beta = 3.060 \times 10^4$ for intersections. To achieve a spatial accuracy of 83% ($RMS < 10$), the server needs a minimum of 3 uploads for straight sections, 5 for ramps, and 8 for intersections, per kilometer.

Section 3.2). The incorrectness of the tracking attack is defined to be the expected distance between the true location x_i and its estimate based on $\hat{p}(x|x_{i-1})$, which can be computed by the following sum:

$$\sum_x \hat{p}(x|x_{i-1}) I_\varepsilon(x, x_i), \qquad (5)$$

where $I_\varepsilon(x, x_i)$ equals 0 if and only if $\|x - x_i\| < \varepsilon$, with ε being a small positive real number, and 1 otherwise.

We quantify the uncertainty of the identity inference using the entropy of the distribution $\hat{p}(P = ID_i|x)$:

$$H = \sum_i \hat{p}(P = ID_i|x) \log_2 \frac{1}{\hat{p}(P = ID_i|x)} \qquad (6)$$

The entropy H shown above indicates how hard to pinpoint a single outcome ID_i out of P at location x. The higher the entropy, the higher the adversary's uncertainty about an identify.

By combining (5) and (6), we obtain the normalized location privacy of user i immediately before it makes a decision regarding whether to upload or not:

$$LP_i^- = \frac{1}{2}\left(\frac{H}{\log_2 n} + \sum\nolimits_{x \in R} p(x|x_{i-1}) I_\varepsilon(x, x_i)\right) \qquad (7)$$

Notice that uploading GPS samples suffers from location privacy loss because the adversary can get more information about users' location to obtain more accurate inference outcomes. Let c_i be the upload cost of user i, $0 < c_i < 1$, then the location privacy level according to user i's strategy can be computed by

$$LP_i(s_i) = \begin{cases} LP_i^- - c_i, & s_i = Y, \\ LP_i^-, & s_i = N. \end{cases} \qquad (8)$$

Typically, the higher the privacy level LP_i^-, the lower the probability of being traced and identified, the lower the cost c_i.

The objective of user privacy protection is to let each user determine its upload strategy s_i so as to maximize (8). Since user privacy is in conflict with service quality, users should consider a tradeoff between traffic service quality and their own privacy to make a decision.

4.3 Optimization Problem

Given the minimum service quality requirement Q_{min} and the privacy level LP_i^- of each user on a road segment, the optimization problem is to find the upload strategy profile $s = (s_1, s_2, \cdots, s_n)$ that maximizes the total privacy level $\sum_i LP_i$ such that $Q \geq Q_{min}$. The solution approach must consider the following two challenges: 1) user i may not know others' privacy level; and 2) how to estimate the minimum service quality requirement Q_{min}.

For the first challenge, we introduce an incomplete information game model [20] in which each user is assigned a type θ, whose probability density function $f(\theta)$ indicates the distribution of the user's privacy level. In other words, each user is aware of only the privacy level distribution, not the actual privacy level. For the second challenge, we exploit the server's global view (i.e., historical traffic status) to estimate the minimum service quality requirement.

5. GAME MODEL

We introduce the incomplete information game [20] in this section to model the upload decision process of smartphone users. In this game, players balance their location privacy requirements and traffic estimation accuracy to determine whether or not to upload. The set of players $P = \{1, 2, \cdots, n\}$ corresponds to the set of smartphone users on a specific road segment. Each player has two possible moves s_i: upload (Y) or not (N). The reward received by user i is determined by the traffic estimation quality and its privacy level, and the utility of user i is defined by:

$$u_i(s_i(\theta_i), s_{-i}(\theta_{-i})) = wQ_i(s_i(\theta_i), s_{-i}(\theta_{-i})) + LP_i(s_i(\theta_i)), \tag{9}$$

where $Q_i(s_i, s_{-i})$ is the traffic service quality determined by the moves of the user i and its opponents $-i$, $LP_i(s_i)$ is the location privacy of user i , w can be considered as the expectation degree of users to Q, and θ_i is the type of user i according to a common probability distribution $f(\theta_i)$ [20]. Note that θ_i can be considered as the location privacy level immediately before the game.

5.1 Nash Equilibrium

The concept of Bayesian Nash Equilibrium [16] for the incomplete information game is introduced as follows.

Definition 1. *A strategy profile $s^* = \{s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\}$ is a pure-strategy Bayesian Nash equilibrium (BNE) if, for each player i:*

$$s_i^*(\theta_i) \in \arg\max_{s_i \in \{Y,N\}} \sum_{\theta_{-i}} f(\theta_{-i}) u_i(s_i, s_{-i}^*(\theta_{-i})), \quad \forall \theta_i \tag{10}$$

The BNE in our user upload game can be obtained by comparing the average utility of Y with that of N, as follows:

$$\begin{aligned} E[u_i(Y, s_{-i})] &= wE[Q(Y, s_{-i}(\theta_{-i}))] + LP_i^- - c_i \\ E[u_i(N, s_{-i})] &= wE[Q(N, s_{-i}(\theta_{-i}))] + LP_i^- \end{aligned} \tag{11}$$

where Y is the NE strategy of user i for $c_i < w(E[Q(Y, s_{-i}(\theta_{-i}))] - E[Q(N, s_{-i}(\theta_{-i}))])$, and N is the NE strategy of user i for $c_i \geq w(E[Q(Y, s_{-i}(\theta_{-i}))] - E[Q(N, s_{-i}(\theta_{-i}))])$.

We denote the upload probability of user i by $p_i = \int_{\tilde{\theta}_i}^{1} f(\theta_i) d\theta_i$, where $\tilde{\theta}_i$ is the minimum privacy level at which user i is willing to upload. Let P_Y be a subset of k upload users in the given set P; thus the probability that the number of upload users is equal to k is $\Pr(K = k) = \prod_{i \in P_Y} p_i \prod_{j \in P - P_Y} (1 - p_j)$. Therefore, the average quality of traffic estimation is shown as follows:

$$E(Q) = \sum_{k=1}^{n} \Pr(K = k) \log_\alpha(1 + \beta k), \tag{12}$$

and there exists $\hat{k}$ such that $\log_\alpha(1 + \hat{k}\beta) \approx E(Q)$. Hence we have

$$E[Q(Y, s_{-i}(\theta_{-i}))] - E[Q(N, s_{-i}(\theta_{-i}))] \approx \log_\alpha \frac{1 + \beta(1 + \hat{k})}{1 + \beta\hat{k}}. \tag{13}$$

From (11) and (13), we can rewrite the upload threshold as $w\log_\alpha \frac{1+\beta(1+\hat{k})}{1+\beta\hat{k}}$.

6. THE UPLOAD MECHANISM

Our design goals are to provide users with an appropriate level of privacy preservation and to achieve an overall optimality of the traffic service quality and user privacy.

6.1 Upload Algorithm

We propose our privacy-preserving traffic data collection algorithm in this section, which is called UploadGame. Illustrated in Algorithm 1, UploadGame consists of two phases: the k determination phase and the upload user selection phase.

Algorithm 1 UploadGame

```
//Phase 1: Server determines $k$, the number of upload users.
    Calculate $Q(v)$ according to the estimated average speed:
        $Q(v) = \frac{\rho}{\sigma\sqrt{2\pi}} e^{-\frac{(v-\mu)^2}{2\sigma^2}}$
    Calculate $k$ according to $Q(v)$:
        $k = \frac{\alpha^{Q(v)} - 1}{\beta}$
//Phase 2: Users make a decision regarding whether to upload or not.
    Calculate $w$ according to $k$:
        $w = \frac{\lambda}{F^{-1}(1-k/n)\log_\alpha(1+\beta(k+1))/(1+\beta k)}$
if $c_i < w\log_\alpha \frac{1+\beta(1+k)}{1+\beta k}$ then
    Play $Y$
else
    Play $N$
end if
```

In the k determination phase, the server estimates the required number of upload users according to the historical traffic status. More specifically, we depict the function relationship between the required quality of traffic estimation on a road segment and the historical estimates of the average speed based on the users' expectation of the traffic estimation quality (see Section 3.1). Here, we assume the function obeys a normal distribution:

$$Q(v) = \frac{\rho}{\sigma\sqrt{2\pi}} e^{-\frac{(v-\mu)^2}{2\sigma^2}}, \tag{14}$$

where $\rho > 0$ is a system parameter, μ and σ are respectively the mean and standard deviation, and v is the historical estimate of the average speed. This equation reflects that the required quality is not high for a low speed because the low speed at rush hours does not change abruptly; and the required quality is also not high for a high speed because the high speed in free flowing traffics does not need an accurate estimation. Of course, we can use other function relationships, i.e., if users more concern about accident occurrence, we use the function relationship between Q and Δv to capture the changes in the traffic. According to the required traffic estimation quality in (14), we have $k = (\alpha^{Q(v)} - 1)/\beta$, where k is the required number of upload users we need to get.

In the upload user selection phase, each user calculates the value of w for which the NE can be achieved, and then decides whether to upload or not based on w. Recall that the desired NE is achieved if and only if we choose an appropriate value of w. If the players are aware of the upload costs of their opponents, i.e., $c_1 \leq c_2 \leq \cdots \leq c_n$, it is easy to get $w = \frac{c_k}{\log_\alpha(1+\beta(k+1))/(1+\beta k)}$. However, since each player does not know the privacy levels and privacy costs of others, we need to estimate the value of c_k. Typically, the higher the current privacy level, the lower the upload cost. Therefore we assume $LP_i = \lambda/c_i$. As the privacy level obeys the distribution $f(\theta_i)$, we have

$$\frac{k}{n} = \int_{\underline{\theta}}^{1} f(\theta_i)\, d\theta_i, \tag{15}$$

where n is the number of smartphone users in the road segment. According to (15), we have $\underline{\theta} = F^{-1}(1 - k/n)$. Then the estimate value of c_k is:

$$\hat{c}_k = \lambda/\underline{\theta} = \frac{\lambda}{F^{-1}(1-k/n)} \tag{16}$$

Furthermore, we obtain the value of w by:

$$w = \frac{\lambda}{F^{-1}(1-k/n)\log_\alpha(1+\beta(k+1))/(1+\beta k)} \tag{17}$$

Then users decide whether to upload or not according to the threshold defined in Section 5.1.

6.2 Game Analysis

6.2.1 Nash Equilibrium Result

The existence and uniqueness of a NE in the incomplete information game imply the convergence of Algorithm 1, given by Theorem 1.

THEOREM 1. *The strategy profile computed by Algorithm 1 is a unique NE of UploadGame.*

PROOF. Without loss of generality, assume that $c_1 \le c_2 \le \cdots \le c_n$, and $s^* = \{Y^1, Y^2, \cdots, Y^k, N^1, \cdots, N^{n-k}\}$ is an upload strategy profile computed by Algorithm 1. We first prove that the strategy profile s^* is a NE. Let $g(x) = \log_\alpha(1+\beta x) - \log_\alpha(1+\beta(x-1))$ be a real-valued continuous function on the interval $[1, n]$, where $\alpha, \beta > 0$ are defined according to Fig. 1. Then the derivative of $g(x)$ with respect to x is given by:

$$g'(x) = \frac{-\beta^2}{(1+\beta x)(1+\beta(x-1))\ln a} < 0 \tag{18}$$

Since the derivative of $g(x)$ is negative, $g(x)$ is a monotonically decreasing function. Let $c(x)$ be a real-valued continuous function on the interval $[1, n]$ such that $c(i) = c_i$ and $c'(x) \ge 0$. Then the derivative of $G(x) = wg(x) - c(x)$ with respect to x is $G'(x) = wg'(x) - c'(x) < 0$. Note that from Algorithm 1, $G(1) < 0 < G(n)$. According to the intermediate value theorem, there exists $\tilde{k}$ such that $G(\tilde{k}) = 0$. When $i \le k = \lfloor \tilde{k} \rfloor$, we have

$$\begin{aligned} & u_i(Y, s_{-i}) - u_i(N, s_{-i}) \\ & = w\log_\alpha \tfrac{1+\beta(\tilde{k})}{1+\beta(\tilde{k}-1)} - c(i) \\ & = w\log_\alpha \tfrac{1+\beta(\tilde{k})}{1+\beta(\tilde{k}-1)} - c(\tilde{k}) + c(\tilde{k}) - c(i) \\ & = G(\tilde{k}) + (c(\tilde{k}) - c(i)) > 0 \end{aligned}$$

Thus Y is the dominant strategy for $i \le k$. When $i > k = \lfloor \tilde{k} \rfloor$, we have

$$\begin{aligned} & u_i(Y, s_{-i}) - u_i(N, s_{-i}) \\ & = w\log_\alpha \tfrac{1+\beta(\tilde{k}+1)}{1+\beta\tilde{k}} - c(i) \\ & = G(\tilde{k}+1) + (c(\tilde{k}+1) - c(i)) > 0 \end{aligned}$$

Thus N is the dominant strategy for $i > k$.

Therefore, $s^* = \{Y^1, Y^2, \cdots, Y^k, N^1, \cdots, N^{n-k}\}$ is a NE of the upload game.

We next prove the uniqueness. Assume that $s^* = \{Y^1, \cdots, Y^h, N^1, \cdots, N^{n-h}\}$ is also a NE of the upload game. From Algotithm 1, we conclude that $h \le k$. For $q < h$, there exists c_{h+1} such that $c_{h+1} < w\log_\alpha \frac{1+\beta(1+\tilde{k})}{1+\beta\tilde{k}}$. In other words, user $h+1$ can increase its utility by unilaterally changing its strategy from N to Y, contradicting the NE assumption. Therefore $k = h$, which proves the uniqueness of s^*. □

6.2.2 Incentive Compatibility of UploadGame

We analyze the incentive compatibility of UploadGame in this section. Incentive compatibility is a characteristic of a mechanism whereby each participant knows that its best strategy is to follow the rules, no matter what other participants would do [17]. Being incentively compatible, the mechanism can eliminate users' fears about market manipulation.

DEFINITION 2. *A mechanism is incentively compatible if the direct-revelation mechanism can induce a Bayesian-Nash equilibrium $s^*() = (s_1^*(), \cdots, s_n^*())$ such that $s_i^*(\theta_i) = \theta_i$, $\forall \theta_i \in \Theta_i, \forall i \in N$.*

In other words, a user achieves its optimal utility if it reports a true valuation (the cost in this paper).

THEOREM 2. *UpdateGame is incentively compatible.*

PROOF. Let the upload cost of user i be c_i, while user i chooses c_i'. Consider the following two cases.

Case 1: $c_i < w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$. If $c_i' < c_i$ or $c_i < c_i' < w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$, it does not affect the upload outcomes of the users and their utilities do not change. If $c_i' \ge w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$, user i does not upload its GPS samples, and its utility becomes $u_i(N, s_{-1}^*) = w\log_\alpha(1+\beta k) + LP_i^- \le w\log_\alpha(1+\beta(k+1)) + LP_i^- - c_i = u_i(Y, s_{-1}^*)$. User i can increase its utility by unilaterally changing its strategy from N to Y, so user i 's utility is reduced by cheating with c_i'.

Case 2: $c_i \ge w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$. If $c_i' > c_i$ or $c_i > c_i' \ge w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$, user i does not upload and its utility does not change. If $c_i' < w\log_\alpha \frac{1+\beta(k+1)}{1+\beta k}$, user i chooses to upload and its utility becomes $u_i(Y, s_{-1}^*) = w\log_\alpha(1+\beta(k+1)) + LP_i^- - c_i < w\log_\alpha(1+\beta k) + LP_i^- = u_i(N, s_{-1}^*)$. Cheating would reduce user i's utility.

Therefore, UpdateGame is incentively compatible. □

7. EXPERIMENTAL EVALUATION

We verify the effectiveness and incentive compatibility of UploadGame in this section. The traffic data used in the experiments contains one month of GPS coordinates for 28,000 taxis in Beijing, and the GPS coordinate update period of each taxi is less than 30 sec [21]. In the simulation, we consider two specific scenarios: free-flowing traffic and stop-and-go traffic (rush hours). As shown in Fig. 3, we recursively divide the geographic region of interest into four smaller rectangles (or quadrants). Each rectangle is defined by a triplet $< level, x, y >$, where $level$ is the depth of recursion, x and y are the offsets from the top left corner of the region. Here we set $level = 3$, which corresponds to $8 \times 8 = 64$ rectangles.

7.1 Tradeoff between Location Privacy and Traffic Estimation Quality

We experimentally compare our mechanism UploadGame with a naive upload mechanism, in which the users decide whether or not to upload according to the strategy threshold with a fixed value of w. In this study, user 1 and user 3 (correspondingly user 5 and user 7) are driving on route 1 from $< 3, 8, 1 >$ to $< 3, 2, 7 >$ with a free flowing traffic (correspondingly stop-and-go traffic); user 2 and user 4 (correspondingly user 6 and user 8) are driving on route 2 from

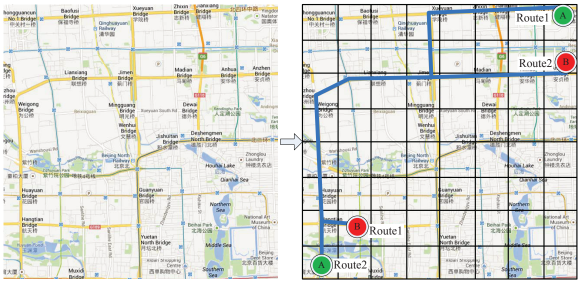


Figure 3: Region partition.

$< 3, 1, 8 >$ to $< 3, 8, 3 >$ with a free flowing traffic (correspondingly stop-and-go traffic). User location privacy and the estimated traffic quality are computed once every two minutes, and the results are reported in Fig. 4 and Fig. 5, respectively.

We observe that the user location privacy level in the stop-and-go traffic is higher than that in the free flowing traffic (Fig. 4). This is because more number of users increases the tracking incorrectness and the identity uncertainty in stop-and-go traffic. As expected, UploadGame improves about 25% of user privacy in stop-and-go traffic compared to the naive scheme, while maintaining an appropriate privacy preservation in the free flowing traffic. Notice that there exist sudden fallings and risings in the curves. These are attributed to the facts that a user's upload at the corresponding time increases the possibility of being tracked and identified (causing fallings), and that it is hard to trace a user when it passes a ramp or an intersection (causing risings).

Fig. 5 demonstrates the QoS of traffic estimation received by the users. The QoS in stop-and-go ($> 83\%$) is higher than that in free flowing traffic ($< 83\%$) because of the larger number of upload users in the stop-and-go traffic case. Fortunately, there is no need to have a high QoS in the free flowing traffic case as users have less concerns. However, in order to handle unexpected events, UploadGame encourages users to upload their GPS samples, resulting in a higher QoS (Fig. 5(a) and Fig. 5(b)) in free flowing traffics.

7.2 Incentive Compatibility of UploadGame

We also verify the incentive compatibility of UploadGame by randomly picking 4 users in free flowing traffic and 4 users in stop-and-go traffic at the region $< 3, 4, 3 >$. These users are allowed to choose upload costs that are different from their true costs. We illustrate the results in Fig. 6. As one can see, the users achieve their optimal utility if they play truthfully. The markers in the graph are the users' utilities according to their true upload costs. We notice that users choosing other privacy costs can not increase their utilities.

8. CONCLUSION

In this paper, we propose an upload mechanism to protect user location privacy in crowdsourced traffic monitoring. Our mechanism is user-centric, and can achieve the dual goal of traffic estimation quality guarantee and user privacy protection. We first quantify the traffic service quality and user location privacy. Then we design the user upload algorithm UploadGame that meets the basic traffic service quality requirement, and meanwhile provides with a strong privacy guarantee. Through the incomplete information game analysis, we theoretically prove the convergence and incentive compatibility of the UploadGame mechanism. Finally, we verify the effectiveness of UploadGame with a real world traffic data.

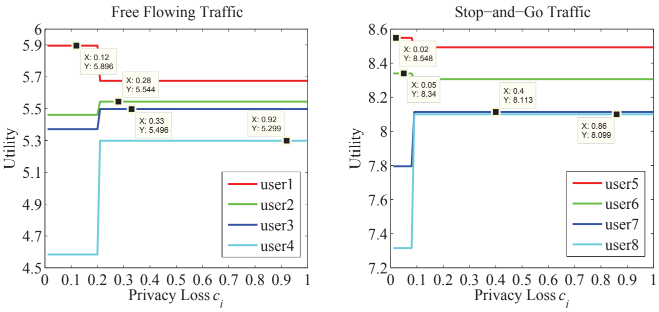


Figure 6: Incentive compatibility of UploadGame.

9. ACKNOWLEDGMENTS

The authors would like to thank Weidong Yang and Xiang Lu from Institute of Information Engineering, Chinese Academy of Science for discussions and comments on the manuscript. This study is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under grant No. XDA06040100.

10. REFERENCES

[1] Google Map. Location source and accuracy. `https://support.google.com/gmm/answer/3144282?hl=en&reftopic=3137371`.

[2] Waze. How to use waze. `http://www.waze.com/guidedtour`.

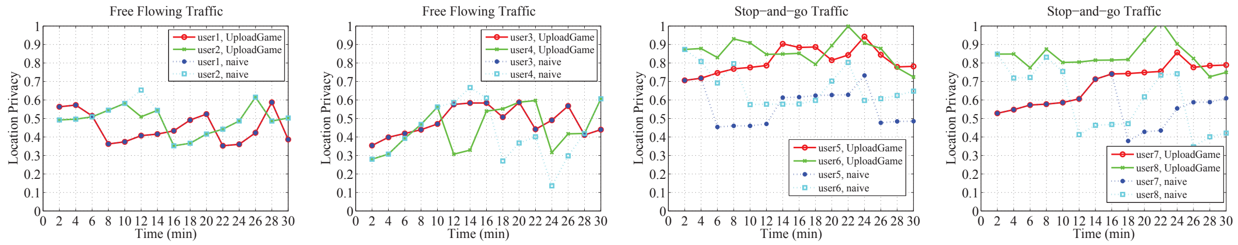
[3] Prashanth Mohan, Venkata N Padmanabhan, and Ramachandran Ramjee. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 323–336. ACM, 2008.

[4] Baik Hoh, Toch Iwuchukwu, Quinn Jacobson, Daniel Work, Alexandre M Bayen, Ryan Herring, J-C Herrera, Marco Gruteser, Murali Annavaram, and Jeff Ban. Enhancing privacy and accuracy in probe vehicle-based traffic monitoring via virtual trip lines. *Mobile Computing, IEEE Transactions on*, 11(5):849–864, 2012.

[5] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *IEEE Symposium on Security and Privacy (S&P), Oakland*, pages 247–262. IEEE, 2011.

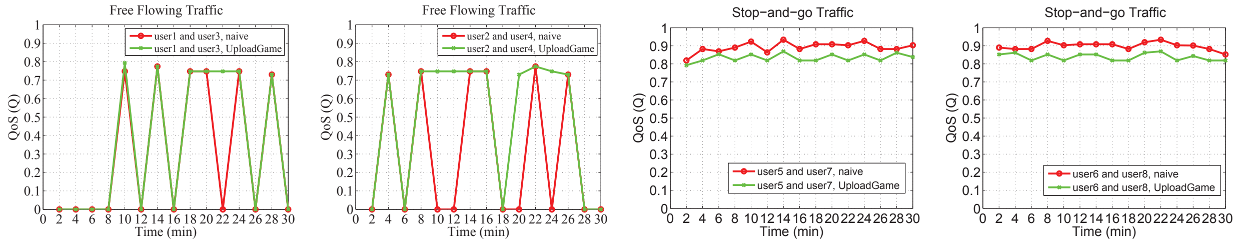
[6] Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Nature srep.*, 3, 2013.

[7] Chris Y.T. Ma, David K.Y. Yau, Nung Kwan Yip, and Nageswara S.V. Rao. Privacy vulnerability of published anonymous mobility traces. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, MobiCom '10. ACM, 2010.



(a) Location privacy of user 1 and user 2 (b) Location privacy of user 3 and user 4 (c) Location privacy of user 5 and user 6 (d) Location privacy of user 7 and user 8

Figure 4: Location Privacy of users for both the free flowing traffic and the stop-and-go traffic.



(a) QoS received by user 1 and user 3 (b) QoS received by user 2 and user 4 (c) QoS received by user 5 and user 7 (d) QoS received by user 6 and user 8

Figure 5: Traffic service quality for both the free flowing traffic and the stop-and-go traffic.

[8] Laurent Bindschaedler, Murtuza Jadliwala, Igor Bilogrevic, Imad Aad, Philip Ginzboorg, Valtteri Niemi, and Jean-Pierre Hubaux. Track me if you can: On the effectiveness of context-based identifier changes in deployed mobile networks. In *NDSS*. The Internet Society, 2012.

[9] Baik Hoh, Marco Gruteser, Hui Xiong, and Ansaf Alrabady. Achieving guaranteed anonymity in gps traces via uncertainty-aware path cloaking. *Mobile Computing, IEEE Transactions on*, 9(8):1089–1107, 2010.

[10] Mehmet Ercan Nergiz, Maurizio Atzori, Yucel Saygin, and Bar Guc. Towards trajectory anonymization: A generalization-based approach. *Trans. Data Privacy*, 2009.

[11] Karen P Tang, Pedram Keyani, James Fogarty, and Jason I Hong. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 93–102. ACM, 2006.

[12] Elaine Shi, Richard Chow, T h. Hubert Chan, Dawn Song, and Eleanor Rieffel. Privacy-preserving aggregation of time-series data. In *In NDSS*, 2011.

[13] Balaji Palanisamy and Ling Liu. Mobimix: Protecting location privacy with mix-zones over road networks. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*, pages 494–505. IEEE, 2011.

[14] Xinxin Liu, Han Zhao, Miao Pan, Hao Yue, Xiaolin Li, and Yuguang Fang. Traffic-aware multiple mix zone placement for protecting location privacy. In *The 31st IEEE International Conference on Computer Communications (INFOCOM 2012)*, 2012.

[15] Tansu Alpcan and Sonja Buchegger. Security games for vehicular networks. *Mobile Computing, IEEE Transactions on*, 10(2):280–290, 2011.

[16] Julien Freudiger, Mohammad Hossein Manshaei, Jean-Pierre Hubaux, and David C Parkes. On non-cooperative location privacy: a game-theoretic analysis. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 324–337. ACM, 2009.

[17] Dejun Yang, Xi Fang, and Guoliang Xue. Truthful incentive mechanisms for k-anonymity location privacy. In *INFOCOM, IEEE International Conference on Computer Communications*, pages 3094–3102, 2013.

[18] Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of the ACM conference on Computer and Communications Security (CCS)*, pages 617–627. ACM, 2012.

[19] Mudhakar Srivatsa and Mike Hicks. Deanonymizing mobility traces: Using social network as a side-channel. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 628–637. ACM, 2012.

[20] John C. Harsanyi. Games with incomplete information played by "bayesian" players, i-iii. *Manage. Sci.*, 50(12 Supplement):1804–1817, 2004.

[21] Datatang Company. Taxi gps data of one city in north of china (200903). `http://dx.doi.org/10.1287/mnsc.1040.0270`.