Fundamental Limits of Location Privacy using Anonymization

Nazanin Takbiri
Electrical and Computer Engineering
UMass-Amherst
ntakbiri@umass.edu

Amir Houmansadr
Information and Computer Science
UMass-Amherst
amir@cs.umass.edu

Dennis L. Goeckel
Electrical and Computer Engineering
UMass-Amherst
goeckel@ecs.umass.edu

Hossein Pishro-Nik
Electrical and Computer Engineering
UMass-Amherst
pishro@ecs.umass.edu

***Abstract*—In [1]–[3], the concept of perfect location privacy is defined and sufficient conditions for achieving it were obtained when anonymization is used. In this paper, necessary conditions for perfect privacy are obtained. Specifically, we prove that the previous sufficient bounds are tight, and thus we obtain the threshold for achieving perfect location privacy using anonymization. First, we assume that a user's current location is independent from her past locations. Using this i.i.d model, we show that if the adversary collects more than $\Omega(n^{\frac{2}{r-1}})$ anonymous observations, then the adversary can successfully recover the users' locations with high probability. Here, n is the number of users in the network and r is the number of all possible locations that users can go to. Next, we model users' movements using Markov chains to better model real-world movement patterns. We show similar results if the adversary collects more than $\Omega(n^{\frac{2}{|E|-r}})$ observations, where $|E|$ is the number of edges in the user's Markov chain model.**

***Index Terms*—Location Based Service (LBS), Location Privacy Protecting Mechanism (LPPM), Mobile Networks, Information Theoretic Privacy, Anonymization, Markov Chains.**

I. Introduction

MOBILE devices, ranging from smart phones to connected automobiles, offer a wide range of *location-based services (LBS)* such as navigation, ride-sharing, dining recommendations, and auto collision warnings. LBS applications are exploding in popularity, e.g., Uber, Google Maps, Yelp, and connected vehicles serve tens to hundreds of millions of users per day. However, these popular, important services impose significant privacy threats to their users because they require access to the location information of mobile devices. Aggregated with other collected personal data, this information allows adversaries to infer sensitive information that goes far beyond user location: their habits, relationships, employments, hobbies, etc. Such privacy compromises can be launched by various types of adversaries: the LBS system may compromise users' privacy by selling private location information to advertisers; malevolent staff of LBS systems can access users' information for fun or profit (as exemplified in a recent Uber scandal [4], [5]); and cybercriminals may break into the location database of an LBS system [6] or launch Sybil attacks [7], [8].

This work was supported by National Science Foundation under grants CCF 0844725 and CCF 1421957

Because of the importance of privacy in LBS systems, researchers have devised *location privacy protection mechanisms (LPPMs)* [9]–[15]. Existing LPPMs are often tailored to specific LBS systems and can be classified into two main classes: *identity perturbation* LPPMs [12], [14], [15] (e.g. through anonymization techniques), and *location perturbation* LPPMs [9]–[13] (e.g. purposeful obfuscation by adding noise to mobile users' coordinates). Despite extensive previous studies on location privacy and LPPM mechanisms, the theoretical foundations of location privacy have not been established.

In [1]–[3], users are characterized by the statistics of their locations, and the adversary then tries to match traces to those statistics to attempt identification. So anonymization technique is used to apply identity perturbation, and then concept of perfect location privacy is defined and sufficient conditions for achieving it is discussed. More specifically, it was shown that if the number of observations by the adversary is smaller than a critical number, then all users have perfect location privacy. In this paper, the converse result is proved for the same critical value. That is, we prove that if the number of observations by the adversary is larger than the critical value, then the adversary can find an algorithm to successfully estimate the location of users with arbitrary small error probability. Thus essentially a fundamental threshold for location privacy is established.

In the first step, we assume users' movements are modeled as independent and identically distributed ($i.i.d$) random variables. That is, we assume their locations are independent from their previous locations. In the next step, we assume users' movements are modeled by Markov chains to be more realistic. For both models, we obtain the critical threshold for location privacy.

II. Related Work

A common approach used by *identity perturbation* LPPMs is to obfuscate user identities within groups of users, an approach known as k-anonymity [16], [17]. A second common approach to identity perturbation LPPMs is to exchange users' pseudonyms within areas called mix-zones [18], [19]. Freudiger et al. show that combining techniques from cryptography with mix-zones can result in higher levels of location privacy [14]. Game theoretic approaches [20], [21] and location

cryptography [22], [23] approaches have also been taken. Several *location perturbation* LPPMs work by replacing each user's location information with a larger region, a technique known as cloaking [24], [25].

Another direction to location perturbation is including dummy locations in the set of possible locations of users [26], [27].

Several anonymization works [13], [28] employ differential privacy approaches. For instance, Ho et al. [29] propose a differentially private location pattern mining algorithm using quadtree spatial decomposition. Dewri [30] combines k-anonymity and differential privacy to improve location privacy. In addition, several location perturbation LPPMs are based on ideas from differential privacy [13], [31]–[34]. For instance, Andres et al. hides the exact location of the user in a region by adding Laplacian distributed noise to achieve a desired level of geo-indistinguishability [34]. The focus of some works is on trajectory privacy [35]. The idea comes from the fact that even if the privacy of individual locations are revealed, it is important that no meaningful behavior can be inferred from the corresponding trajectory.

Several works aim at *quantifying* location privacy protection. Shokri et al. [12], [36] define the expected estimation error of the adversary as a metric to evaluate LPPM mechanisms. Ma et al. [15] use uncertainty about users' location information to quantify user location privacy in vehicular networks. Li et al. [37] define metrics to quantify the trade off between privacy and utility of LPPM systems. In [38], the authors propose a user-centric location-based service architecture where a user can observe the impact of location inaccuracy on the service accuracy. Shokri et al. [9] design LPPM mechanisms that will defeat localization attacks. In [39] and [40], privacy leakage of location sharing and interdependent location privacy risks are quantified, respectively. A similar idea is proposed in [41] where the quantification model is based on the Bayes conditional risk.

These above-mentioned studies confirm the growing importance of location privacy. What is missing from the current literature is a solid theoretical framework for location privacy that is general enough to encompass the various location privacy preserving methods in the existing literature. As we will see, the proposed framework allows us to establish the fundamental limits and trade-offs of such LPPMs as well as to achieve *provable location privacy*.

Previously, the mutual information has been used as a privacy metric in a number of settings, [42]–[45]. However, the framework and problem formulation for location privacy is quite different from those encountered in previous works. More specifically, the location privacy problem is based on a large set of time-series data that belong to different users with different movement dynamics that has gone through an LPPM, and the adversary is aiming at de-anonymizing and denoising the data. To the best of our knowledge, no prior work has studied the *fundamental limits and trade-offs* in such a setting.

Finally, [46] studies asymptotically optimal matching of time series to source distributions. However, there are significant differences between the the settings of [46] and this paper: First, [46] does not consider non-i.i.d cases (i.e., the Markov chain case). Second, fitting to the location privacy problem, we assume the existence of a general (but possibly unknown) prior distribution for the sources (i.e. a Bayesian setting). This implies that the crucial factor in our analysis will be obtaining the privacy thresholds as a function of the number of users.

III. Framework

Here we use a similar framework to [1]–[3]. Specifically, the locations of n users which are in a specific region are recorded, and we define $X_u(k)$ as location of user u at time k. We also consider the strongest adversary $\mathcal{A}$ that has complete statistical knowledge of the users' movements based on the previous observations or other resources, and in order to secure location privacy of users, anonymization technique is used. In other words, the adversary can observe the anonymized version of users' locations. The adversary obtains m observations per user, where m is a function of n, $m = m(n)$, and then tries to estimate $X_u(k)$ by using those observations. $\mathbf{Y}^{(m)}$ is the anonymized version of users' locations which the adversary can observe.

Anonymization can be modeled by a random permutation $\Pi^{(n)}$ on the set of n users. The user u is assigned the pseudonym $\Pi^{(n)}(u)$. In this paper, $\Pi(u)$ is used instead of $\Pi^{(n)}(u)$ for simplicity. Let $\mathbf{X}_u^{(m)}$ be the vector which contains m number of locations of user u, and $\mathbf{X}^{(m)}$ is a collection which contains $\mathbf{X}_u^{(m)}$ for all users,

$$\mathbf{X}_u^{(m)} = \begin{bmatrix} X_u(1) \\ X_u(2) \\ \vdots \\ X_u(m) \end{bmatrix}, \quad \mathbf{X}^{(m)} = \left[\mathbf{X}_1^{(m)}, \mathbf{X}_2^{(m)}, \cdots, \mathbf{X}_n^{(m)}\right].$$

Now, we apply the anonymization function Perm(.) to support location privacy. In other words, we anonymize the users and so the adversary observes

$$\begin{aligned} \mathbf{Y}^{(m)} &= \text{Perm}\left(\mathbf{X}_1^{(m)}, \mathbf{X}_2^{(m)}, \cdots, \mathbf{X}_n^{(m)}; \Pi^{(n)}\right) \\ &= \left(\mathbf{X}_{\Pi^{-1}(1)}^{(m)}, \mathbf{X}_{\Pi^{-1}(2)}^{(m)}, \cdots, \mathbf{X}_{\Pi^{-1}(n)}^{(m)}\right) \\ &= \left(\mathbf{Y}_1^{(m)}, \mathbf{Y}_2^{(m)}, \cdots, \mathbf{Y}_n^{(m)}\right). \end{aligned}$$

So

$$\mathbf{Y}_u^{(m)} = \mathbf{X}_{\Pi^{-1}(u)}^{(m)}, \ \mathbf{Y}_{\Pi(u)}^{(m)} = \mathbf{X}_u^{(m)}.$$

Note that the permutation $\Pi^{(n)}$ is the only piece of the information that is required for the adversary, so that he can successfully de-anonymize the location data. In this paper, we prove that if $m(n)$ is bigger than the threshold we obtained, the adversary can successfully de-anonymize the location data. That is, the adversary can invert the permutation $\Pi^{(n)}$, and thus recovers the true locations of the users.

IV. I.I.D Model

A. Two-State Model

We first consider the i.i.d two-state model. We assume users move independently from their previous locations and other users' locations, and can only go to states 0 and 1. Therefore, we can consider location of users at any time as a Bernoulli random variable with parameter p_u, which is the probability of user u being at location 1.

We also assume that p_u's are drawn independently from some continuous density function, $f_P(p_u)$, on the $(0,1)$ interval. Specifically, there is $\delta > 0$ such that[1]:

$$\begin{cases} f_P(p_u) < \delta & p_u \in (0,1) \\ f_P(p_u) = 0 & p_u \notin (0,1) \end{cases}$$

We consider the strongest adversary who knows the values of p_u's , and use this knowledge to identify users.

Theorem 1. For a simple two-state model, if $\mathbf{Y}^{(m)}$ is anonymized version of $\mathbf{X}^{(m)}$, and $m = cn^{2+\alpha}$ for $c > 0$ and $0 < \alpha < 1$, then user 1 has no location privacy as n goes to infinity. In other words,

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0.$$

where $X_u(k)$ is the actual location of user u, $\widetilde{X_u(k)}$ is location of user u which the adversary obtains from her algorithm, and $P_e(u)$ is error probability.

Note that due to the symmetry of the problem, we can restate the theorem for all users. Since this is a converse result, we give an explicit detector at the adversary and show that it can be used by the adversary to recover the true location of user 1.

Proof. We provide an explicit method for the adversary to recover $X_1(k)$. The basic idea is that the adversary calculates the empirical averages for the presence of users at location 1 and then assigns the string with the empirical average closest to p_1 to user 1.

Formally, for $u = 1, 2, \cdots, n$, the adversary computes $\overline{Y_u^{(m)}}$ as follows

$$\overline{Y_u^{(m)}} = \frac{Y_u^{(m)}(1) + Y_u^{(m)}(2) + \cdots + Y_u^{(m)}(m)}{m}.$$

Therefore, we can conclude

$$\overline{Y_{\Pi(u)}^{(m)}} = \frac{X_u^{(m)}(1) + X_u^{(m)}(2) + \cdots + X_u^{(m)}(m)}{m}.$$

Let's define

$$B^{(n)} \triangleq \{x \in (0,1); p_1 - \Delta_n < x < p_1 + \Delta_n\},$$

where $\Delta_n = \frac{1}{n^{1+\frac{\alpha}{4}}}$, we claim that for $m = cn^{2+\alpha}$ and large n:

1) $P\left(\overline{Y_{\Pi(1)}^{(m)}} \in B^{(n)}\right) \to 1$

2) $P\left(\bigcup_{j=2}^{n}\left(\overline{Y_{\Pi(j)}^{(m)}} \in B^{(n)}\right)\right) \to 0$

Thus, the adversary can identify $\Pi(1)$ by examining $\overline{Y_u^{(m)}}$'s and choosing the only one that belongs to $B^{(n)}$.

First we want to show that as n goes to infinity,

$$P\left(\overline{Y_{\Pi(1)}^{(m)}} \in B^{(n)}\right) \to 1.$$

Note

$$\begin{aligned} P\left(Y_{\Pi(1)}^{(m)} \in B^{(n)}\right) &= P\left(\frac{\sum_{i=1}^{m} X_1^{(m)}(i)}{m} \in B^{(n)}\right) \\ &= P\left(p_1 - \Delta_n < \frac{\sum_{i=1}^{m} X_1^{(m)}(i)}{m} < p_1 + \Delta_n\right) \\ &= P\left(\left|\sum_{i=1}^{m} X_1^{(m)}(i) - mp_1\right| < m\Delta_n\right). \end{aligned}$$

According to Chernoff bound,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{m} X_1^{(m)}(i) - mp_1\right| < m\Delta_n\right) &\geq 1 - 2e^{-\frac{m\Delta_n^2}{3}} \\ &\geq 1 - 2e^{-\frac{c}{3}n^{2+\alpha}\cdot\left(\frac{1}{n^{1+\frac{\alpha}{4}}}\right)^2} \\ &\geq 1 - 2e^{-\frac{c}{3}n^{\frac{\alpha}{2}}}. \end{aligned}$$

We also know that as n goes to infinity, $2e^{-\frac{c}{3}n^{\frac{\alpha}{2}}}$ goes to zero, as a result

$$P\left(Y_{\Pi(1)}^{(m)} \in B^{(n)}\right) \to 1.$$

Now, we need to show that as n goes to infinity,

$$P\left(\bigcup_{j=2}^{n}\left(\overline{Y_{\Pi(j)}^{(m)}} \in B^{(n)}\right)\right) \to 0.$$

First, we define

$$C^{(n)} = \{X \in (0,1); p_1 - 2\Delta_n < X < p_1 + 2\Delta_n\},$$

and claim as n goes to infinity,

$$P\left(\bigcup_{j=2}^{n}\left(p_j \in C^{(n)}\right)\right) \to 0.$$

Note

$$P\left(p_j \in C^{(n)}\right) < 4\Delta_n\delta,$$

and according to the union bound,

$$\begin{aligned} P\left(\bigcup_{j=2}^{n}\left(p_j \in C^{(n)}\right)\right) &\leq \sum_{j=2}^{n} P\left(p_j \in C^{(n)}\right) \\ &\leq 4n\Delta_n\delta \\ &\leq 4n\frac{1}{n^{1+\frac{\alpha}{4}}}\delta \\ &\leq 4n^{\frac{-\alpha}{4}}\delta \to 0. \end{aligned} \tag{1}$$

[1]The condition $f_P(p_u) < \delta$ is not actually necessary for the results and can be relaxed; however, we keep it here to avoid unnecessary technicalities.

Thus all p_j's are outside of $C^{(n)}$ for $j \in \{2, 3, \cdots, n\}$ with high probability.

Now, we claim that given all p_j's are outside of $C^{(n)}$, $P\left(\overline{Y_{\Pi(j)}} \in B^{(n)}\right)$ is small. Note,

$$P\left(Y_{\Pi(j)}^{(m)} \in B^{(n)}\right) < P\left(|\overline{Y_{\Pi(j)}^{(m)}} - p_j| > \Delta_n\right)$$
$$= P\left(\left|\sum_{j=2}^{n} X_{\Pi(j)}^{(m)}(i) - mp_j\right| > m\Delta_n\right)$$

According to the Chernoff bound,

$$P\left(\left|\sum_{j=2}^{n} X_{\Pi(j)}^{(m)}(i) - mp_j\right| > m\Delta_n\right) < 2e^{-\frac{m\Delta_n^2}{3}}$$
$$< 2e^{-\frac{c}{3}n^{2+\alpha}.\left(\frac{1}{n^{1+\frac{\alpha}{4}}}\right)^2}$$
$$< 2e^{-\frac{c}{3}n^{\frac{\alpha}{2}}}.$$

As a result, by using union bound, we have

$$P\left(\bigcup_{j=2}^{n}\left(\overline{Y_{\Pi(j)}^{(m)}} \in B^{(n)}\right)\right) < n\left(2e^{-\frac{c}{3}n^{\frac{\alpha}{2}}}\right),$$

and thus as n goes to infinity,

$$P\left(\bigcup_{j=2}^{n}\left(\overline{Y_{\Pi(j)}^{(m)}} \in B^{(n)}\right)\right) \to 0.$$

Thus, we have proved that if $m = n^{2+\alpha}$, there exists an algorithm for the adversary to successfully recover $X_u^{(m)}(k)$. □

B. Extension to r- States

Now, assume users can go to r states (locations), $0, 1, \cdots, r-1$, and $p_u(i)$ shows the probability of user u being at location i. We define the vector $\mathbf{p}_u$ and the matrix $\mathbf{P}$ as

$$\mathbf{p}_u = \begin{bmatrix} p_u(1) \\ p_u(2) \\ \vdots \\ p_u(r-1) \end{bmatrix}, \quad \mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n).$$

Note that the dimension is $r-1$ as we need to have

$$\sum_{i=0}^{r-1} p_u(i) = 1.$$

We assume $\mathbf{p}_u$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, on the $(0,1)^{r-1}$ hypercube. In particular, define the range of distribution as

$$R_{\mathbf{P}} = \{(x_1, x_2, \cdots, x_{r-1}) \in (0,1)^{r-1} : x_i > 0, x_1 + x_2 + \cdots + x_{r-1} < 1\}.$$

Then, we assume there is $\delta > 0$ such that:

$$\begin{cases} f_{\mathbf{P}}(\mathbf{p}_u) < \delta & \mathbf{p}_u \in R_{\mathbf{P}} \\ f_{\mathbf{P}}(\mathbf{p}_u) = 0 & \mathbf{p}_u \notin R_{\mathbf{P}} \end{cases}$$

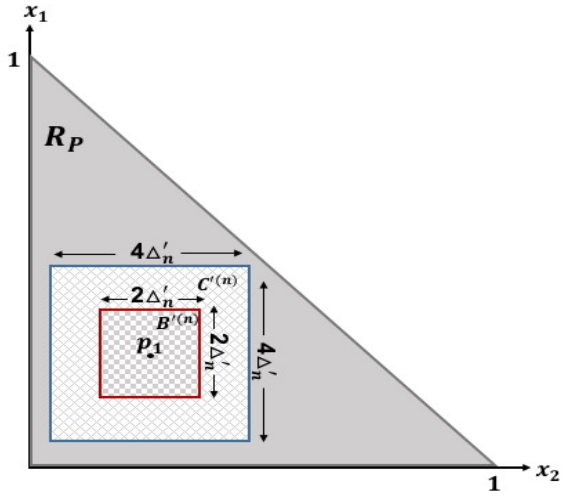

Fig. 1: $\mathbf{p}_1$, sets $B'(n)$ and $C'(n)$ in R_P for case $r = 3$.

Theorem 2. For an r-state model, if $\mathbf{Y}^{(m)}$ is anonymized version of $\mathbf{X}^{(m)}$, and $m = cn^{\frac{2}{r-1}+\alpha}$ for $c > 0$ and $0 < \alpha < 1$, then the adversary can successfully identify the location of user 1 as n goes to infinity. It means

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0.$$

Proof of Theorem 2 is similar to the proof of Theorem 1, so we just provide the general idea.

Let's define sets $B'^{(n)}$ and $C'^{(n)}$ as

$$B'^{(n)} \triangleq \{(x_1, ..., x_{r-1}) \in R_{\mathbf{P}} : p_1(i) - \Delta_n' < x_i < p_1(i) + \Delta_n', i = 1, ..., r-1\},$$

$$C'^{(n)} \triangleq \{(x_1, ..., x_{r-1}) \in R_{\mathbf{P}} : p_1(i) - 2\Delta_n' < x_i < p_1(i) + 2\Delta_n', i = 1, ..., r-1\},$$

where $\Delta_n' = \frac{1}{n^{\frac{1}{r-1}+\frac{\alpha}{4}}}$. Figure 1 shows $\mathbf{p}_1$, sets $B'(n)$ and $C'(n)$ in range of $R_{\mathbf{P}}$ for case $r = 3$.

We claim for $m = cn^{\frac{2}{r-1}+\alpha}$ and large n,

1) $P\left(\overline{\mathbf{Y}_{\Pi(1)}^{(m)}} \in B'^{(n)}\right) \to 1$
2) $P\left(\bigcup_{j=2}^{n}\left(\overline{\mathbf{Y}_{\Pi(j)}^{(m)}} \in B'^{(n)}\right)\right) \to 0$

This can be shown similar to the above proof that we provided for the two-state case. Thus, the adversary can de-anonymize the locations with vanishing error probability.

V. Markov Chain Model

In Section 4.2, we assumed there are r locations which users can go and users' movements are i.i.d. In this section, we model users' movements by using Markov chains, in which a user's movements are dependent over time. In this model, we again assume there are r possible locations. Let E be the set of edges. More specifically, $(i, j) \in E$ if there exists an edge from i to j with probability $p(i, j) > 0$. Different users can have different transition probability matrices.

Here, we again show that the adversary will be able to de-anonymize the locations if the number of observations is larger than a threshold. The key idea is that the adversary can focus on a subset of transition probabilities that are sufficient for recovering the entire transition probability matrix. In particular, note that for each state i, we must have

$$\sum_{j=1}^{r} p(i,j) = 1,$$

so Markov chain of user u is completely determined by a subset of size $d = |E| - r$ of transition probabilities. Let's write this subset in a vector as follows:

$$\mathbf{p}_u = \begin{bmatrix} p_u(1) \\ p_u(2) \\ \vdots \\ p_u(|E|-r) \end{bmatrix}, \quad \mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n).$$

We also consider $p_u(i)$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, on the $(0,1)^{|E|-r}$ hypercube. Let $R_{\mathbf{p}} \subset \mathbb{R}^d$ be the range of acceptable values for $\mathbf{P}_u$. As before, we assume there are $\delta > 0$, such that

$$\begin{cases} f_{\mathbf{P}}(\mathbf{p}) < \delta & \mathbf{p} \in R_{\mathbf{p}} \\ f_{\mathbf{P}}(\mathbf{p}) = 0 & \mathbf{p} \notin R_{\mathbf{p}} \end{cases}$$

Using the above observations, we can now repeat the same reasoning as the last sections to show the following theorem.

Theorem 3. For the an irreducible, aperiodic Markov chain model, if $\mathbf{Y}^{(m)}$ is anonymized version of $\mathbf{X}^{(m)}$, and $m = cn^{\frac{2}{|E|-r}+\alpha}$ for $c > 0$ and $0 < \alpha < 1$, then the adversary can successfully identify the location of user 1 as n goes to infinity. In other words,

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0.$$

Considering the fact that the vector $\mathbf{p}_u$ uniquely determines the user u, the proof is now analogous to the i.i.d case. The basic idea is that the adversary computes the empirical averages for $p_u(i)$ for each anonymized user based on his observations. The adversary can the invert the anonymization permutation function in a similar fashion to the i.i.d case:

In particular, Let's define sets $B^{''(n)}$ and $C^{''(n)}$ as

$$B^{''(n)} \triangleq \{(x_1, ..., x_d) \in R_{\mathbf{P}} : p_1(i) - \Delta_n'' < x_i < p_1(i) + \Delta_n'', i = 0, 1, ..., d\},$$

$$C^{''(n)} \triangleq \{(x_1, ..., x_d) \in R_{\mathbf{P}} : p_1(i) - 2\Delta_n'' < x_i < p_1(i) + 2\Delta_n'', i = 0, 1, ..., d\},$$

where $\Delta_n'' = \frac{1}{n^{\frac{1}{|E|-r}+\frac{\alpha}{4}}}$, and $d = |E| - r$.

We claim for $m = cn^{\frac{2}{|E|-r}+\alpha}$ and afor large n,

1) $P\left(\overline{\mathbf{Y}^{(m)}_{\Pi(1)}} \in B^{''(n)}\right) \to 1$
2) $P\left(\bigcup_{j=2}^{n}\left(\overline{\mathbf{Y}^{(m)}_{\Pi(j)}} \in B^{''(n)}\right)\right) \to 0$

VI. Conclusion

In this paper, we proved converse results for the concept of perfect location privacy which was defined in [1]–[3], and expanded their results. We proved that there is a threshold for m, number of adversary's observation, and if adversary's observation is bigger than that threshold, she can successfully recover the locations of the users with vanishing error probability.

In the first step, we proved this claim for users whose locations are independent from their previous locations and other users' movements. For this model, we obtained the privacy threshold as $m = n^{\frac{2}{r-1}}$ (n is the number of users and r is the numbers of possible states that users can go). In the next step, we modeled users' movements by using Markov chain models and proved the threshold is equal to $m = n^{\frac{2}{|E|-r}}$ (where $|E|$ is the number of edges in the users' Markov chain model).

References

[1] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Defining perfect location privacy using anonymization," in *2016 Annual Conference on Information Science and Systems (CISS)*. IEEE, 2016, pp. 204–209.

[2] Z. Montazeri, A. Houmansadr, and H.Pishro-Nik, "Achieving perfect location privacy in markov models using anonymization," in *2016 International Symposium on Information Theory and its Applications (ISITA2016)*, Monterey, USA, oct 2016.

[3] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Achieving Perfect Location Privacy in Wireless Devices Using Anonymization," *under revision in IEEE Transaction on Information Forensics and Security*, 2016.

[4] ""God View": Uber Investigates Its Top New York Executive For Privacy Violations," November 2014, https://www.washingtonpost.com/news/the-switch/wp/2014/12/01/is-ubers-rider-database-a-sitting-duck-for-hackers/.

[5] C. Timberg, "Is Uber's rider database a sitting duck for hackers?" December 2014, https://www.washingtonpost.com/news/the-switch/wp/2014/12/01/is-ubers-rider-database-a-sitting-duck-for-hackers/.

[6] "Uber Statement," February 2015, http://newsroom.uber.com/2015/02/uber-statement/.

[7] T. Zhou, R. R. Choudhury, P. Ning, and K. Chakrabarty, "Privacy-preserving detection of sybil attacks in vehicular ad hoc networks," in *Mobile and Ubiquitous Systems: Networking & Services, 2007. MobiQuitous 2007. Fourth Annual International Conference on*. IEEE, 2007, pp. 1–8.

[8] S. Chang, Y. Qi, H. Zhu, J. Zhao, and X. S. Shen, "Footprint: Detecting sybil attacks in urban vehicular networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 6, pp. 1103–1114, 2012.

[9] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 617–627.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st international conference on Mobile systems, applications and services*. ACM, 2003, pp. 31–42.

[11] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in gps traces via uncertainty-aware path cloaking," in *Proceedings of the 14th ACM conference on Computer and communications security*. ACM, 2007, pp. 161–171.

[12] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 247–262.

[13] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 251–262.

[14] J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," 2007.

[15] Z. Ma, F. Kargl, and M. Weber, "A location privacy metric for v2x communication systems," in *Sarnoff Symposium, 2009. SARNOFF'09. IEEE*. IEEE, 2009, pp. 1–6.

[16] G. P. Corser, H. Fu, and A. Banihani, "Evaluating location privacy in vehicular communications and applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 9, pp. 2658–2667, 2016.

[17] Y. Zhang, W. Tong, and S. Zhong, "On designing satisfaction-ratio-aware truthful incentive mechanisms for< inline-formula>< tex-math notation=," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2528–2541, 2016.

[18] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*. IEEE, 2005, pp. 194–205.

[19] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive computing*, no. 1, pp. 46–55, 2003.

[20] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, p. 25, 2013.

[21] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes, "Non-cooperative location privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 10, no. 2, pp. 84–98, 2013.

[22] R. Paulet, M. G. Kaosar, X. Yi, and E. Bertino, "Privacy-preserving and content-protecting location based queries," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 5, pp. 1200–1210, 2014.

[23] A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," in *Advances in Spatial and Temporal Databases*. Springer, 2007, pp. 239–257.

[24] R. Shokri, G. Theodorakopoulos, P. Papadimitratos, E. Kazemi, and J.-P. Hubaux, "Hiding in the mobile crowd: Locationprivacy through collaboration," *IEEE transactions on dependable and secure computing*, vol. 11, no. 3, pp. 266–279, 2014.

[25] M. A. Zurbaran, K. Avila, P. Wightman, and M. Fernandez, "Near-rand: Noise-based location obfuscation based on random neighboring points," *IEEE Latin America Transactions*, vol. 13, no. 11, pp. 3661–3667, 2015.

[26] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Data Engineering Workshops, 2005. 21st International Conference on*. IEEE, 2005, pp. 1248–1248.

[27] J. Krumm, "A survey of computational location privacy," *Personal and Ubiquitous Computing*, vol. 13, no. 6, pp. 391–399, 2009.

[28] J. Lee and C. Clifton, "Differential identifiability," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1041–1049.

[29] S.-S. Ho and S. Ruan, "Differential privacy for location pattern mining," in *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*. ACM, 2011, pp. 17–24.

[30] R. Dewri, "Local differential perturbations: Location privacy under approximate knowledge attackers," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 12, pp. 2360–2372, 2013.

[31] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics." in *Privacy Enhancing Technologies*. Springer, 2013, pp. 82–102.

[32] R. Shokri, "Optimal user-centric data obfuscation," *arXiv preprint arXiv:1402.3426*, 2014.

[33] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "Location privacy via geo-indistinguishability," *ACM SIGLOG News*, vol. 2, no. 3, pp. 46–69, 2015.

[34] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 901–914.

[35] R.-H. Hwang, Y.-L. Hsueh, and H.-W. Chung, "A novel time-obfuscated algorithm for trajectory privacy protection," *IEEE Transactions on Services Computing*, vol. 7, no. 2, pp. 126–139, 2014.

[36] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, "Quantifying location privacy: the case of sporadic location exposure," in *Privacy Enhancing Technologies*. Springer, 2011, pp. 57–76.

[37] T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 517–526.

[38] R. Dewri and R. Thurimella, "Exploiting service similarity for privacy in location-based search queries," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 374–383, 2014.

[39] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. PP, no. 99, 2016.

[40] A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux, "Quantifying interdependent privacy risks with location data," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[41] X. Zhang, X. Gui, F. Tian, S. Yu, and J. An, "Privacy quantification model based on the bayes conditional risk in location-based services," *Tsinghua Science and Technology*, vol. 19, no. 5, pp. 452–462, 2014.

[42] S. Salamatian, A. Zhang, F. du Pin Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft, "How to hide the elephant-or the donkey-in the room: Practical privacy against statistical inference for large data." in *GlobalSIP*, 2013, pp. 269–272.

[43] I. Csiszár, "Almost independence and secrecy capacity," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 48–57, 1996.

[44] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 1796–1800.

[45] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 6, pp. 838–852, 2013.

[46] J. Unnikrishnan, "Asymptotically optimal matching of multiple sequences to source distributions and training sequences," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 452–468, 2015.