University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Masters Theses | Dissertations and Theses

March 2017

Achieving Perfect Location Privacy in Wireless Devices Using Anonymization

Zarrin Montazeri
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2

 Part of the Computer Engineering Commons, and the Information Security Commons

Recommended Citation

Montazeri, Zarrin, "Achieving Perfect Location Privacy in Wireless Devices Using Anonymization" (2017). *Masters Theses*. 478.
https://doi.org/10.7275/9501507 https://scholarworks.umass.edu/masters_theses_2/478

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

ACHIEVING PERFECT LOCATION PRIVACY IN LOCATION BASED SERVICES USING ANONYMIZATION

A Thesis Presented

by

ZARRIN MONTAZERI

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

February 2017

Electrical and Computer Engineering

ACHIEVING PERFECT LOCATION PRIVACY IN LOCATION BASED SERVICES USING ANONYMIZATION

A Thesis Presented

by

ZARRIN MONTAZERI

Approved as to style and content by:

Hossein Pishro-Nik, Chair

Dennis Goeckel, Member

Amir Houmansadr, Member

Christopher Hollot, Department Chair
Electrical and Computer Engineering

DEDICATION

To my beloved family.

ACKNOWLEDGMENTS

I wish to thank my advisor, Professor Hossein Pishro-Nik, for his great support and guidance through every step of this work. I am very fortunate to have had the chance to work with such a knowledgeable person from whom I have learned many things through the past years.

I would also like to express my gratitude towards my committee members Professor Dennis Goeckel and Professor Amir Houmansadr, who provided very helpful feedbacks for this thesis.

I would like to thank Shirin Montazeri and Mohammad Ghadiri-Sadrabadi for their great help and unconditional support over the past years.

This work was supported by National Science Foundation.

ABSTRACT

ACHIEVING PERFECT LOCATION PRIVACY IN LOCATION BASED SERVICES USING ANONYMIZATION

FEBRUARY 2017

ZARRIN MONTAZERI

B.Sc., SHARIF UNIVERSITY OF TECHNOLOGY

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hossein Pishro-Nik

The popularity of mobile devices and location-based services (LBS) have created great concerns regarding the location privacy of the users of such devices and services. Anonymization is a common technique that is often being used to protect the location privacy of LBS users. This technique assigns a random pseudonym to each user and these pseudonyms can change over time. Here, we provide a general information theoretic definition for perfect location privacy and prove that perfect location privacy is achievable for mobile devices when using the anonymization technique appropriately. First, we assume that the user's current location is independent from her past locations. Using this i.i.d model, we show that if the pseudonym of the user is changed before $O(n^{\frac{2}{r-1}})$ number of anonymized observations is made by the adversary for that user, then she has perfect location privacy, where n is the number of users in the network and r is the number of all possible locations that the user might occupy. Then, we model each user's movement by a Markov chain so that a user's current location depends on his previous locations, which is a more realistic

model when approximating real world data. We show that perfect location privacy is achievable in this model if the pseudonym of the user is changed before $O(n^{\frac{2}{|E|-r}})$ anonymized observations is collected by the adversary for that user where $|E|$ is the number of edges in the user's Markov model.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

Over the past decades, the number of cellphones, laptops and other electronic devices capable of network communications have increased significantly. In the past few years, mobile devices have started to be equipped with high-precision localization capabilities such as Global Positioning System (GPS) or Global System for Mobile (GSM) technologies, such as WiFi or Bluetooth. Communication between these devices range from communicating over the Web to automobiles connected to various types of traffic safety networks. Such communicating mobile devices offer a wide spectrum of services based on their geographic location such as navigation, ride-sharing, dining recommendation, auto collision warning and advertisement. These applications that utilize the geographic location of their users to provide them with services are called location-based services (LBS). LBS applications, as a part of Ubiquitous Computing (ubicomp), have attracted a lot of attention in recent years, e.g., Uber [66], Google Maps [1, 21], and Yelp [67] serve tens to hundreds of millions of mobile users per day.

While LBSs provide so many services to their users, considering their unrestricted access to the user's location information, they impose significant privacy threats to their users. These services are mostly offered at no money expense to the users but users have to pay with their private information to enjoy these services. Such privacy compromises can also be launched by various types of adversaries including third-party applications, nearby mobile users and cellular service providers. Based on the adversary's purposes, the leaked private information goes far beyond just the geographic location of the users. By sending location information to such services, potential adversaries could establish profiles for the users about their daily routines and habits. The inte-

gration of LBS in online social networks leads to more privacy risks. By aggregating the leaked information over time and combining them with the information that users publish on social networks, the adversary can infer a wide range of other sensitive information about the users such as their habits, relationships, employments, hobbies and even sensitive private or corporate secrets.

Advanced data storage gives governments and other corporations the power to profile the growing number of users and keep their traces for a long period of time. There may be various incentives behind such tracking such as financial, strategic or security reasons or even in order to provide other useful services to the users. On the other hand, with the continuous cost reduction for such storage systems and the probable benefit of keeping such data in the future, their data never gets erased. The tools required to analyze trace data have also made progress. Sophisticated data mining algorithms can leverage on fast growing storage and processing power, thus facilitating the joint analysis of multiple data-sets in parallel. The privacy risks are getting amplified using such low-cost data storages and empowering computation algorithms as well as the public access to powerful data analytics software such as Google's recently published [34] TensorFlow machine learning software [2].

Such privacy compromises can be launched by various types of adversaries. The LBSs can learn users' personal or corporate secrets by using various inference algorithms. They may also compromise users' privacy by selling private location information to advertisers; malevolent staff of LBS systems can access users' private information for fun or profit (as exemplified in a recent Uber scandal [3, 70]); and cybercriminals may break into the location database of an LBS system [4, 5] or launch a Sybil attack [12, 80] on an LBS system to steal large amounts of private location information. These can expose people to unwanted advertisements and location-based scams, it can affect their social reputation or even make them victims of blackmail or physical violence [60]. More importantly, leaked information gives power to the informed corporation or government which they may use against those individuals. In order to protect the privacy of users, we need to protect their private information from being reached by any irrelevant entity.

Some mechanisms have been proposed in order to protect location privacy of LBS users, [9, 28, 33, 37, 49, 63, 65] , generally referred to as location privacy protection mechanisms (LPPM). An LPPM is a system which perturbs users' location information before it reaches the operator of the LBS system or other users. Todays LPPMs can be classified in to two different categories. First, identity perturbation LPPMs [28, 49, 63], which modify the identity of the mobile users in order to protect their location privacy (e.g., through anonymization techniques like mix-zone). In other words, they aim at improving location privacy by concealing the mapping between users and location observations. Second, location perturbation LPPMs [9, 33, 37, 63, 65], which add noise to mobile users' location coordinates to protect their location privacy (e.g., adding dummy locations or hiding locations for periods of time). This can potentially improve the location privacy of the user by returning an inaccurate location information to the LBS application. Some LPPMs combine the two approaches. These mechanisms tend to deliberately decrease the quality of information in some way to protect the privacy of the individual to whom that information refers. The improvement in location privacy by these LPPMs usually comes at the price of performance degradation for the underlying LBS systems, e.g., the service it is offering to the users based on their location information. For instance, an LPPM that perturbs automobile traffic safety messages to increase the privacy of the user, will degrade the effectiveness of the underlying collision prevention LBS system. Finding an optimal choice of LPPM which provide both satisfying performance and adequate privacy is still problem.

In this master thesis, we provide a mathematical framework for the location privacy of mobile devices using information theory by defining an information theoretic notions of Perfect Location Privacy. In the proposed framework, we employ the anonymization technique to hide the identity of users over time. First, we assume that each user's current location is independent from her past locations to simplify the derivations. Then, we model the user's movements by Markov chain which is a more realistic setting by considering the dependencies between locations over time. Also, we assume the strongest model for the adversary, i.e., we assume that the adversary has complete statistical knowledge of the users' movements. We formulate a user's location privacy

based on the mutual information between the adversary's anonymized observations and the user's actual location information. We define the notion of *perfect location privacy* and show that with a properly designed anonymization method, users can achieve perfect location privacy.

1.2 Related Work

Over the past few years, researchers have tried to improve LPPMs to protect location privacy of the users. LPPMs have been classified in to two different categories, location perturbation mechanisms and identity perturbation mechanism. The former methods try to hide the identity information of the users sending the data and the latter methods try to confuse the adversary by either adding noise to the location coordinates of the user or other techniques such as hiding their location for a period of time or adding dummy locations.

In identity perturbation methods, the common approach is to hide the identity of the user within a group of users in the area. In this approach, the adversary gets confused between all the users in the region and cannot distinguish between them. Bugra Gedik et al., [31], defined a framework in which users are able to set the minimum level of anonymity and also add levels of spatial and temporal noise that is acceptable by the LBSs. Another common approach in identity perturbation LPPMs is called mix-zone, [8, 29, 36, 55]. In this approach users have pseudonyms and they exchange their assigned pseudonyms in specific areas called *mix-zones*. Mix-zones are pre-determined regions in which each user in the mix-zone can exchange her pseudonym with another user in the same mix-zone. In order to be well protected, some cryptography mechanisms has been used in the exchanging areas to encrypt messages passing through the mix-zones so that the adversary would not be able to access those messages [76]. The mix-zone strategy may be costly in managing pseudonyms and may not be efficient in an area with low user density. In order to measure the effectiveness of mix-zone approach the measure *anonymity* has been proposed which shows how unidentifiable a user is within a set of users, called an anonymity set, [24]. K-anonymity is the most well known approach that hides a user's identity within $k-1$ other users [9, 19, 23, 30, 33, 39, 47, 51, 53, 68, 69, 77, 79]. Game theoretic approaches [27, 52] and

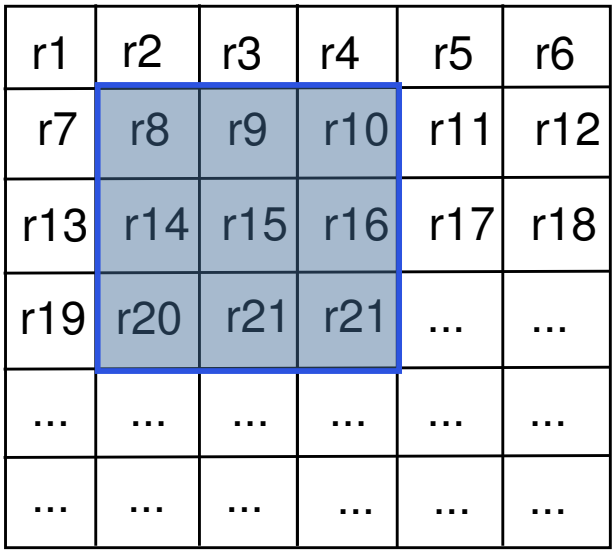

Figure 1.1: Reducing location precision by reporting larger area.

user's actual location is r_{15}, reported location is $r = \{r_8, r_9, r_{10}, r_{14}, r_{15}, r_{16}, r_{20}, r_{21}, r_{22}\}$.

location cryptography [32, 40, 56] approaches have also been taken. Using game theoretic algorithms and combining them with the mix-zone, Manshaei et al. [52] enhanced the privacy of the users in vehicular networks. Also, Shokri et al. [65] utilize Stackelberg Bayesian game to formalize the users' location privacy and adversary's correctness of localization. They show that this optimal LPPM works better in the face of a localization attack.

In location perturbation methods, different approaches has been proposed. In order to protect users' privacy, techniques beyond simply omitting the identifier of the user is required. Since the spatial and temporal characteristics in location data can give useful information about the user, some techniques protect the re-identification of users by modifying these characteristics of the data in traces. In peer-to-peer mobile communication, spatial path cloaking has been used to protect the mobile users' privacy, [7, 9, 10, 16, 17, 25, 26, 30, 31, 33, 37, 38, 41, 53, 64, 71, 73, 74, 78, 81]. Several location perturbation LPPMs work by replacing each users location information with a larger region (e.g., Figure 1.1), [10, 33, 37, 73], and some by dummy locations, [18, 42–44, 48, 59].

Differential privacy is an approach which protects the location privacy of users in the location information datasets [9, 14, 45, 50, 54, 72]. This technique insures that the presence of no single user could significantly change the outcome of the aggregated location information. For instance, Ho et al. [35] proposed a differentially private location pattern mining algorithm using quadtree spatial decomposition. Dewri [22] combined k-anonymity and differential privacy to improve location privacy. Some location perturbation LPPMs are based on ideas from differential privacy [6, 9, 13, 15, 61]. For instance, Andres et al. hide the exact location of each user in a region by adding Laplacian distributed noise to achieve a desired level of geo-indistinguishability [6].

Shokri et al. [62, 63] define the expected estimation error of the adversary as a metric to evaluate LPPMs. On the other hand, Ma et al. [49] use uncertainty about users' location information to quantify users' location privacy in vehicular networks. Li et al. [46] define metrics to quantify the tradeoff between the privacy and utility of LPPM systems. Shokri et al. [65] design LPPMs that will defeat localization attacks.

Previously, the mutual information has been used as a privacy metric in different topics, [11, 20, 57, 58, 75]. However, in this thesis we specifically use the mutual information for location privacy. We provide an information theoretic definition for location privacy using the mutual information. We show that mobile devices can achieve provable perfect location privacy by using the anonymization method in the suggested way.

1.3 Contribution

Different frameworks aim to improve the location privacy of users. These different privacy preserving methods have different impacts on the LBSs' performance. Based on the services that LBSs are providing to their users, they utilize different methods to protect privacy of their users.

In this master thesis, a new theoretical framework is proposed to protect the location privacy of mobile users when using LBSs. Using information theory, the notion of perfect location privacy is defined for mobile users. An identity perturbation LPPM, known as anonymization technique, allows users to change their pseudonyms over time. However, changing pseudonyms is costly and

overusing it may degrade the performance of the service. In this framework, using the anonymization technique in the proposed way allows users to reach perfect location privacy. The upper bound on the frequency of changing pseudonyms is derived and proven here so that the strongest adversary, who has all the statistical information about users' movement, would not be able to distinguish between users by observing their anonymized locations over time. It is proven here that perfect location privacy is indeed achievable if the LPPMs are designed appropriately.

CHAPTER 2

PRELIMINARIES

2.1 Defining Perfect Location Privacy

Let us consider a network with a large number of users. In the proposed framework, an identity perturbation LPPM known as anonymization technique is used to protect the privacy of the users which assigns random pseudonyms to users over time. An adversary is observing this network over time and her intention is to identify anonymized users by tracking their traces over time. In this framework, the strongest adversary is assumed to be observing the network. This adversary has the complete statistical knowledge of the users' movement from her past observations or other sources and she can describe the users' movement as a random process on the corresponding geographic area.

Users start moving at time zero while the adversary starts observing the network. Over time, users move from one place to another. Let $X_u(k)$ be the actual location of user u at time k.

$$\begin{array}{ccccccc}
X_1(1) & X_1(2) & X_1(3) & \cdots & X_1(m) & X_1(m+1) & \cdots \\
X_2(1) & X_2(2) & X_2(3) & \cdots & X_2(m) & X_2(m+1) & \cdots \\
X_3(1) & X_3(2) & X_3(3) & \cdots & X_3(m) & X_3(m+1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_n(1) & X_n(2) & X_n(3) & \cdots & X_n(m) & X_n(m+1) & \cdots \\
X_{n+1}(1) & X_{n+1}(2) & X_{n+1}(3) & \cdots & X_{n+1}(m) & X_{n+1}(m+1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

The adversary is observing this network and what she collects is the anonymized version of $X_u(k)$ over time produced by the anonymization method.

Let us assume we have n number of users in our network, $u = 1, 2, \cdots, n$ and we just have access to these n users' locations. Also, The adversary observes m anonymized locations for all the n users over time before they change their pseudonyms.

$$\begin{array}{|ccccc|cc}
\hline
X_1(1) & X_1(2) & X_1(3) & \cdots & X_1(m) & X_1(m+1) & \cdots \\
X_2(1) & X_2(2) & X_2(3) & \cdots & X_2(m) & X_2(m+1) & \cdots \\
X_3(1) & X_3(2) & X_3(3) & \cdots & X_3(m) & X_3(m+1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_n(1) & X_n(2) & X_n(3) & \cdots & X_n(m) & X_n(m+1) & \cdots \\
\hline
X_{n+1}(1) & X_{n+1}(2) & X_{n+1}(3) & \cdots & X_{n+1}(m) & X_{n+1}(m+1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

The adversary's intention is to assign the path that she observed to the user that created it. Let $\mathbf{Y}^{(m)}$ be the collection of anonymized observations available to the adversary. We define *perfect location privacy* as follows

Definition 1. User u has perfect location privacy at time k with respect to the adversary, if and only if

$$\lim_{n\to\infty} I\left(X_u(k); \mathbf{Y}^{(m)}\right) = 0,$$

where $I(.)$ shows the mutual information and m is the number of previous observations of the adversary.

The above definition shows that over time, the adversary's observations does not give any information about the user's location. The assumption of $n \to \infty$ is valid for all the applications that we consider since the number of users in those applications are significantly high.

2.2 Defining the Anonymization Technique

In this framework, to achieve location privacy, the LPPM performs an anonymization method and changes the identifier of each user with a random pseudonym. That is, it performs a random permutation $\Pi^{(n)}$ on the set of n users and then assigns the pseudonym $\Pi^{(n)}(u)$ to user u.

$$\Pi^{(n)} : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, n\}$$

Here, the permutation $\Pi^{(n)}$ is chosen uniformly at random among all $n!$ possible permutations on the set of all n users, $\{1, 2, \cdots, n\}$.

For $u = 1, 2, \cdots, n$ and time $k = 1, 2, \cdots, m$, let $\mathbf{X}_u^{(m)} = (X_u(1), X_u(2), \cdots, X_u(m))^T$ be a vector which shows the u^{th} user's locations up to time k. Using the permutation function $\Pi^{(n)}$, the adversary observes a permutation of users' location vectors, $\mathbf{X}_u^{(m)}$'s. In other words, the adversary observes

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

where Perm(.) shows the applied permutation function. Then,

$$\mathbf{Y}_{\Pi(u)}^{(m)} = \mathbf{X}_u^{(m)} = (X_u(1), X_u(2), \cdots, X_u(m))^T .$$

2.3 Example

Here we provide a simple example to further elaborate the problem setting. Assume that we have only three users, $n = 3$, and five locations, $r = 5$, that users can occupy (Figure 2.1). Also,

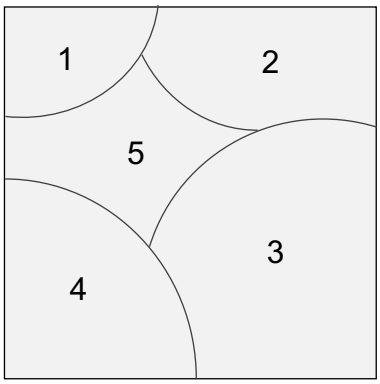

Figure 2.1: An area is divided into five regions that users can occupy.

let us assume that the adversary can collect $m = 4$ observations per user. Each user creates a path as below:

user	path
user 1	$1 \to 2 \to 3 \to 4$
user 2	$2 \to 1 \to 3 \to 5$
user 3	$4 \to 5 \to 1 \to 3$

$$\mathbf{X}_1^{(4)} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{X}_2^{(4)} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 5 \end{bmatrix}, \quad \mathbf{X}_3^{(4)} = \begin{bmatrix} 4 \\ 5 \\ 1 \\ 3 \end{bmatrix}, \; \mathbf{X}^{(4)} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 3 & 1 \\ 4 & 5 & 3 \end{bmatrix}$$

To anonymize the users, we will assign a pseudonym to each. The pseudonyms are determined by the function defined by a random permutation on the user set:

$$\Pi^{(3)} : \{1, 2, 3\} \mapsto \{1, 2, 3\}$$

For this example, suppose that the permutation function is given by $\Pi(1) = 3$, $\Pi(2) = 1$, and $\Pi(3) = 2$. The choice of the permutation is the only piece of information that is not available to the adversary. So here, the adversary observes anonymized users and their paths:

pseudonym	observation
user 1	$2 \to 1 \to 3 \to 5$
user 2	$4 \to 5 \to 1 \to 3$
user 3	$1 \to 2 \to 3 \to 4$

$$\mathbf{Y}^{(4)} = \begin{bmatrix} 2 & 4 & 1 \\ 1 & 5 & 2 \\ 3 & 1 & 3 \\ 5 & 3 & 4 \end{bmatrix}$$

and she wants to find which user (with the pseudomym $user3$) actually made $1 \to 2 \to 3 \to 4$, and so on for the other users. Based on the number of observations that the adversary collects for each user, $m = 4$, and also the user profiles which are the statistical knowledge of the users' movements, she aims at breaking the anonymization function and de-anonymizing the users. The accuracy of this method depends on the number of observations that the adversary collects, and thus our main goal in this paper is to find the function $m(n)$ in a way that the adversary is unsuccessful and the users have perfect location privacy.

CHAPTER 3

ACHIEVING PERFECT LOCATION PRIVACY

3.1 Perfect Location Privacy for Two-State i.i.d Model

To get a better insight about the location privacy problem, here we consider a simple scenario where there are two locations, location 0 and 1. At any time $k \in \{0, 1, 2, \cdots\}$, user u has probability $p_u \in (0, 1)$ to be at location 1, independent from previous locations and other users' locations. Therefore, $X_u(k) \sim Bernoulli(p_u)$.

To keep things generic, we assume that p_u's are drawn independently from some continuous density $f_P(p)$ on the $(0, 1)$ interval. Specifically, there are $\delta_2 > \delta_1 > 0$ such that

$$\begin{cases} \delta_1 < f_P(p) < \delta_2 & p \in (0, 1) \\ f_P(p) = 0 & p \notin (0, 1) \end{cases}$$

The values of p_u's are each user's profile that are known to the adversary. Note that our results do not depend on the choice of $f_P(p)$.

Theorem 1. For two locations with the above definition and anonymized observation vector of the adversary, $\mathbf{Y}^{(m)}$, if all the following holds

1. $m = cn^{2-\alpha}$, which $c, \alpha > 0$ and are constant
2. $p_1 \in (0, 1)$
3. $(p_2, p_3, \cdots, p_n) \sim f_P$, $0 < \delta_1 < f_P < \delta_2$
4. $P = (p_1, p_2, \cdots, p_n)$ be known to the adversary

then, we have

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} I\left(X_1(k); \mathbf{Y}^{(m)}\right) = 0$$

i.e., user 1 has perfect location privacy.

3.1.1 The Intuition Behind the Proof

Here we provide the intuition behind the proof. The formal proof for Theorem 1 is given in 3.2. Let us look from the adversary's perspective. The adversary is observing anonymized locations of the first user and she wants to figure out the index of the user that she is observing, in other words she wants to obtain the mapping between users and observations. Note that the adversary knows the values of $p_1, p_2, \cdots, p_n$. To obtain $X_1(k)$, it suffices that the adversary obtains $\Pi(1)$. Since $X_u(k)$ is a Bernoulli random variable with parameter p_u, to do so, the adversary can look at the averages

$$\overline{Y}_{\Pi(u)} = \frac{Y_{\Pi(u)}(1) + Y_{\Pi(u)}(2) + ... + Y_{\Pi(u)}(m)}{m}.$$

In fact, $\overline{Y}_{\Pi(u)}$'s provide sufficient statistics for this problem. Now, intuitively, the adversary is successful in recovering $\Pi(1)$ if two conditions hold:

1. $\overline{Y}_{\Pi(1)} \approx p_1$.
2. For all $u \neq 1$, $\overline{Y}_{\Pi(u)}$ is not too close to p_1.

Now, note that by the Central Limit Theorem (CLT),

$$\frac{\overline{Y}_{\Pi(u)} - p_u}{\sqrt{\frac{p_u(1-p_u)}{m}}} \to N(0,1).$$

That is, loosely speaking, we can write

$$\overline{Y}_{\Pi(u)} \to N\left(p_u, \frac{p_u(1-p_u)}{m}\right).$$

Consider an interval $I \subset (0,1)$ such that $p_1 \in I$ and the length of I is equal to $\ell^n = \frac{1}{n^{1-\eta}}$ where $0 < \eta < \frac{\alpha}{2}$. Note that for any $u \in 1, 2, \cdots, n$ the probability that p_u is in I is larger than $\delta_1 \ell^n = \frac{\delta_1}{n^{1-\eta}}$. In other words, since there are n users, we can guarantee that a large number of p_u's be in I. On the other hand, we have

$$\frac{\sqrt{\text{Var}(\overline{Y}_{\Pi(u)})}}{\ell^n} = \frac{\sqrt{\frac{p_u(1-p_u)}{m}}}{\frac{1}{n^{1-\eta}}} = n^{\frac{\alpha}{2}-\eta} \to \infty.$$

Note that here, we will have a large number of normal random variables $\overline{Y}_{\Pi(u)}$ whose expected values are in interval I (that has a vanishing length) with high probability and their standard deviation is much larger than the interval length. Thus, distinguishing between them will become impossible for the adversary. In other words, the probability that the adversary will correctly identify $\Pi(u)$ goes to zero as n goes to infinity. That is, the adversary will most likely choose an incorrect value j for $\Pi(u)$. In this case, since the locations of different users are independent, the adversary will not obtain any useful information by looking at $X_j(k)$.

3.2 Proof of Theorem 1 (Perfect Location Privacy for Two-State Model)

Here, we provide a formal proof for Theorem 1. In the proposed setting, we assume we have an infinite number of potential users indexed by integers, and at any step we consider a network consisting of n users, i.e., users $1, 2, \cdots, n$. We would like to show perfect location privacy when n goes to infinity. Remember that $X_u(t)$ shows the location of user u at time t.

In a two-state model, let us assume we have state 0 and state 1. There is a sequence $p_1, p_2, p_3, \cdots$ for the users. In particular, for user u we have $p_u = P(X_u(k) = 1)$ for times $k = 1, 2, \cdots$. Thus, the locations of each user u are determined by a $\text{Bernoulli}(p_u)$ process.

When we set $n \in \mathbb{N}$ as the number of users, we assume m to be the number of adversary's observations per user,

$$m = m(n) = cn^{2-\alpha} \quad \text{where} \quad 0 < \alpha < 1.$$

So, we have $n \to \infty$ if and only if $m \to \infty$.

As defined previously, $\mathbf{X}_u^{(m)}$ contains m number of user u's locations and $\mathbf{X}^{(m)}$ is the collection of $\mathbf{X}_u^{(m)}$'s for all users,

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

The permutation function applied to anonymize users is $\Pi^{(n)}$ (or simply Π). For any set $A \subset \{1, 2, \cdots, n\}$, we define

$$\Pi(A) = \{\Pi(u) : u \in A\}.$$

The adversary who knows all the p_u's, observes n anonymized users for m number of times each and collects their locations in $\mathbf{Y}^{(m)}$

$$\mathbf{Y}^{(m)} = \text{Perm}\left(\mathbf{X}_1^{(m)}, \mathbf{X}_2^{(m)}, \cdots, \mathbf{X}_n^{(m)}; \Pi\right) = \left(\mathbf{Y}_1^{(m)}, \mathbf{Y}_2^{(m)}, \cdots, \mathbf{Y}_n^{(m)}\right)$$

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

Based on the assumptions of Theorem 1, if the following holds

1. $m = cn^{2-\alpha}$, which $c > 0, 0 < \alpha < 1$ and are constant
2. $p_1 \in (0, 1)$
3. $(p_2, p_3, \cdots, p_n) \sim f_P$, $0 < \delta_1 < f_P < \delta_2$
4. $P = (p_1, p_2, \cdots, p_n)$ be known to the adversary,

then we want to show

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} I\left(X_1(k); \mathbf{Y}^{(m)}\right) = 0$$

i.e., user 1 has perfect location privacy and the same applies for all other users.

3.2.1 Proof procedure

Steps of the proof are as follows:

1. We show that there exists a sequence of sets $J^{(n)} \subseteq \{1, 2, \cdots, n\}$ with the following properties:

   - $1 \in J^{(n)}$
   - if $N^{(n)} = |J^{(n)}|$ then, $N^{(n)} \to \infty$ as $n \to \infty$
   - let $\{j_n\}_{n=1}^{\infty}$ be any sequence such that $j_n \in \Pi(J^{(n)})$ then

   $$P\left(\Pi(1) = j_n | \mathbf{Y}^{(m)}, \Pi(J^{(n)})\right) \to 0$$

2. We show that

   $$X_1(k) | \mathbf{Y}^{(m)}, \Pi(J^{(n)}) \xrightarrow{d} Bernoulli(p_1).$$

3. Using 2, we conclude

   $$H\left(X_1(k) | \mathbf{Y}^{(m)}, \Pi(J^{(n)})\right) \to H\left(X_1(k)\right)$$

   and in conclusion,

   $$I\left(X_1(k); \mathbf{Y}^{(m)}\right) \to 0.$$

3.2.2 Detail of the proof

We define $S_u^{(m)}$ for $u = 1, 2, \cdots, n$ to be the number of times that user u was at state 1,

$$S_u^{(m)} = X_u(1) + X_u(2) + \cdots + X_u(m).$$

Based on the assumptions, we have $S_u^{(m)} \sim \text{Binomial}(m, p_u)$. One benefit of $S_u^{(m)}$'s is that they provide a sufficient statistic for the adversary when the adversary's goal is to obtain the permutation $\Pi^{(n)}$. To make this statement precise, let's define $\mathbf{S}^{(m)}$ as the vector containing $S_u^{(m)}$, for $u = 1, 2, \cdots, n$:

$$\mathbf{S}^{(m)} = \left(S_1^{(m)}, S_2^{(m)}, \cdots, S_n^{(m)}\right)$$

Note that

$$\begin{aligned} S_u^{(m)} &= X_u(1) + X_u(2) + \cdots + X_u(m) \\ &= Y_{\Pi(u)}(1) + Y_{\Pi(u)}(2) + \cdots + Y_{\Pi(u)}(m) \quad \text{for } u = 1, 2, \cdots, n. \end{aligned}$$

Thus, the adversary can obtain $\text{Perm}\big(\mathbf{S}^{(m)}, \Pi^{(n)}\big)$, a permuted version of $\mathbf{S}^{(m)}$, by adding the elements in each column of $\mathbf{Y}^{(m)}$. In particular, we can write

$$\begin{aligned} \text{Perm}\big(\mathbf{S}^{(m)}, \Pi^{(n)}\big) &= \text{Perm}\left(S_1^{(m)}, S_2^{(m)}, \cdots, S_n^{(m)}; \Pi^{(n)}\right) \\ &= \left(S_{\Pi^{-1}(1)}^{(m)}, S_{\Pi^{-1}(2)}^{(m)}, \cdots, S_{\Pi^{-1}(n)}^{(m)}\right). \end{aligned}$$

We now state and prove a lemma that confirms $\text{Perm}\big(\mathbf{S}^{(m)}, \Pi^{(n)}\big)$ is a sufficient statistic for the adversary when the adversary's goal is to recover $\Pi^{(n)}$. The usefulness of this lemma will be clear since we can use the law of total probability to break the adversary's decision problem into two steps of (1) obtaining the posterior probability distribution for $\Pi^{(n)}$ and (2) estimating the locations $X_u(k)$ given the choice of $\Pi^{(n)}$.

Lemma 1. Given Perm$\left(\mathbf{S}^{(m)}, \Pi^{(n)}\right)$, the random matrix $\mathbf{Y}^{(m)}$ and the random permutation $\Pi^{(n)}$ are conditionally independent. That is

$$P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}, \text{Perm}\left(\mathbf{S}^{(m)}, \Pi^{(n)}\right) = \mathbf{s}\right) = P\left(\Pi^{(n)} = \pi \;\middle|\; \text{Perm}\left(\mathbf{S}^{(m)}, \Pi^{(n)}\right) = \mathbf{s}\right) \tag{3.1}$$

Proof. Remember

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

Note that $\mathbf{Y}^{(m)}$ (and therefore $\mathbf{y}$) is an m by n matrix, so we can write

$$\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n),$$

where for $u = 1, 2, \cdots, n$, we have

$$\mathbf{y}_u = \begin{bmatrix} y_u(1) \\ y_u(2) \\ \vdots \\ y_u(m) \end{bmatrix}.$$

Also, $\mathbf{s}$ is a 1 by n vector, so we can write

$$\mathbf{s} = (s_1, s_2, \cdots, s_n).$$

We now show that the two sides of Equation 3.7 are equal. The right hand side probability can be written as

$$P\left(\Pi^{(n)}=\pi \;\middle|\; \text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s}\right)=\frac{P\left(\text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s} \;\middle|\; \Pi^{(n)}=\pi\right)P\left(\Pi^{(n)}=\pi\right)}{P\left(\text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s}\right)}$$
$$=\frac{P\left(\text{Perm}\big(\mathbf{S}^{(m)},\pi\big)=\mathbf{s} \;\middle|\; \Pi^{(n)}=\pi\right)}{n!P\left(\text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s}\right)}$$
$$=\frac{P\left(\text{Perm}\big(\mathbf{S}^{(m)},\pi\big)=\mathbf{s}\right)}{n!P\left(\text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s}\right)}.$$

Now note that

$$P\left(\text{Perm}\big(\mathbf{S}^{(m)},\pi\big)=\mathbf{s}\right)=P\left(\bigcap_{j=1}^{n}\left(S^{(m)}_{\pi^{-1}(j)}=s_j\right)\right)$$
$$=P\left(\bigcap_{u=1}^{n}\left(S^{(m)}_{u}=s_{\pi(u)}\right)\right)$$
$$=\prod_{u=1}^{n}P\left(S^{(m)}_{u}=s_{\pi(u)}\right)$$
$$=\prod_{u=1}^{n}\binom{m}{s_{\pi(u)}}p_u^{s_{\pi(u)}}(1-p_u)^{m-s_{\pi(u)}}$$
$$=\prod_{k=1}^{n}\binom{m}{s_k}\prod_{u=1}^{n}p_u^{s_{\pi(u)}}(1-p_u)^{m-s_{\pi(u)}}$$

Similarly, we obtain

$$P\left(\text{Perm}\big(\mathbf{S}^{(m)},\Pi^{(n)}\big)=\mathbf{s}\right)=\sum_{\text{all permutations }\pi'}P\left(\text{Perm}\big(\mathbf{S}^{(m)},\pi'\big)=\mathbf{s} \;\middle|\; \Pi^{(n)}=\pi'\right)P\left(\Pi^{(n)}=\pi'\right)$$
$$=\frac{1}{n!}\sum_{\text{all permutations }\pi'}\prod_{k=1}^{n}\binom{m}{s_k}\prod_{u=1}^{n}p_u^{s_{\pi'(u)}}(1-p_u)^{m-s_{\pi'(u)}}$$
$$=\frac{1}{n!}\prod_{k=1}^{n}\binom{m}{s_k}\sum_{\text{all permutations }\pi'}\prod_{u=1}^{n}p_u^{s_{\pi'(u)}}(1-p_u)^{m-s_{\pi'(u)}}.$$

Thus, we conclude that the right hand side of Equation 3.7 is equal to

$$\frac{\prod_{u=1}^{n} p_u^{s_{\pi(u)}}(1-p_u)^{m-s_{\pi(u)}}}{\sum_{\text{all permutations } \pi'} \prod_{u=1}^{n} p_u^{s_{\pi'(u)}}(1-p_u)^{m-s_{\pi'(u)}}}.$$

Now let's look at the left hand side of Equation 3.7. First, note that in the left hand side probability in Equation 3.7 we must have

$$s_u = \sum_{k=1}^{m} y_u(k) \quad \text{for } u = 1, 2, \cdots, n. \tag{3.2}$$

Next, we can write

$$P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}, \text{Perm}\big(\mathbf{S}^{(m)}, \Pi^{(n)}\big) = \mathbf{s}\right) = P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}\right).$$

This is because $\text{Perm}\big(\mathbf{S}^{(m)}, \Pi^{(n)}\big)$ is a function of $\mathbf{Y}^{(m)}$. We have

$$P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}\right) = \frac{P\left(\mathbf{Y}^{(m)} = \mathbf{y} \;\middle|\; \Pi^{(n)} = \pi\right) P\left(\Pi^{(n)} = \pi\right)}{P\left(\mathbf{Y}^{(m)} = \mathbf{y}\right)}$$

We have

$$\begin{aligned} P\left(\mathbf{Y}^{(m)} = \mathbf{y} \;\middle|\; \Pi^{(n)} = \pi\right) &= \prod_{u=1}^{n} p_u^{\sum_{k=1}^{m} y_{\pi(u)}(k)}(1-p_u)^{m-\sum_{k=1}^{m} y_{\pi(u)}(k)} \\ &= \prod_{u=1}^{n} p_u^{s_{\pi(u)}}(1-p_u)^{m-s_{\pi(u)}} \quad \text{Using Euqation (3.2)} \end{aligned}$$

Similarly, we obtain

$$P\left(\mathbf{Y}^{(m)} = \mathbf{y}\right) = \frac{1}{n!} \sum_{\text{all permutations } \pi'} \prod_{u=1}^{n} p_u^{s_{\pi'(u)}}(1-p_u)^{m-s_{\pi'(u)}}$$

Thus, we conclude that the left hand side of Equation 3.7 is equal to

$$\frac{\prod_{u=1}^{n} p_u^{s_{\pi(u)}}(1-p_u)^{m-s_{\pi(u)}}}{\sum_{\text{all permutations } \pi'} \prod_{u=1}^{n} p_u^{s_{\pi'(u)}}(1-p_u)^{m-s_{\pi'(u)}}},$$

which completes the proof.

□

Next, we need to turn our attention to defining the critical set $J^{(n)}$. First, remember that

$$m = cn^{2-\alpha} \quad \text{where} \quad 0 < \alpha < 1.$$

We choose real numbers θ and ϕ such that $0 < \theta < \phi < \frac{\alpha}{2(2-\alpha)}$, and define

$$\epsilon_m \triangleq \frac{1}{m^{\frac{1}{2}+\phi}} \quad \beta_m \triangleq \frac{1}{m^{\frac{1}{2}-\theta}}.$$

We now define the set $J^{(n)}$ for any positive integer n as follows: Set $J^{(n)}$ consists of the indices of users such that the probability of them being at state 1 is within a range with ϵ_m difference around p_1,

$$J^{(n)} = \{i \in \{1, 2, \ldots, n\} : p_1 - \epsilon_m < p_i < p_1 + \epsilon_m\}.$$

Clearly for all $n, 1 \in J^{(n)}$. The following lemma confirms that the number of elements in $J^{(n)}$ goes to infinity as $n \to \infty$.

Lemma 2. If $N^{(n)} \triangleq |J^{(n)}|$, then $N^{(n)} \to \infty$ as $n \to \infty$. More specifically, as $n \to \infty$,

$$\exists \lambda, c'' > 0 : \quad P(N^{(n)} > c'' n^{\lambda}) \to 1 \qquad \text{as} \quad n \to \infty.$$

Proof. Remember that we assume p_u's are drawn independently from some continuous density function, $f_P(p)$, on the $(0,1)$ interval which satisfies

$$\begin{cases} \delta_1 < f_P(p) < \delta_2 & p \in (0,1) \\ f_P(p) = 0 & p \notin (0,1) \end{cases}$$

So given $p_1 \in (0,1)$, for n large enough (so that ϵ_m is small enough), we have

$$P(p_1 - \epsilon_m < p_i < p_1 + \epsilon_m) = \int_{p_1-\epsilon_m}^{p_1+\epsilon_m} f_P(p)dp,$$

so we can conclude that

$$2\epsilon_m\delta_1 < P(p_1 - \epsilon_m < p_i < p_1 + \epsilon_m) < 2\epsilon_m\delta_2.$$

We can find a δ such that $\delta_1 < \delta < \delta_2$ and

$$P(p_1 - \epsilon_m < p_i < p_1 + \epsilon_m) = 2\epsilon_m\delta.$$

Then, we can say that $N^{(n)} \sim \text{Binomial}(n, 2\epsilon_m\delta)$, where

$$\epsilon_m = \frac{1}{m^{\frac{1}{2}+\phi}} = \frac{1}{(cn^{2-\alpha})^{(\frac{1}{2}+\phi)}}.$$

The expected value of $N^{(n)}$ is $n2\epsilon_m\delta$, and by substituting ϵ_m we get

$$E[N^{(n)}] = n2\epsilon_m\delta = \frac{n2\delta}{(c'n^{2-\alpha})^{(\frac{1}{2}+\phi)}} = c''n^{(\frac{\alpha}{2}+\alpha\phi-2\phi)}.$$

Let us set $\lambda = \frac{\alpha}{2} + \alpha\phi - 2\phi$. Since $\phi < \frac{\alpha}{2(2-\alpha)}$, we have $\lambda > 0$. Therefore, we can write

$$E[N^{(n)}] = c''n^{\lambda},$$

$$Var(N^{(n)}) = n(2\epsilon_m\delta)(1-2\epsilon_m\delta) \to n^{\lambda}(1+o(1)).$$

Using Chebyshev's inequality

$$P(|N^{(n)} - E[N^{(n)}]| > \frac{c''}{2}n^{\lambda}) < \frac{n^{\lambda}(1+o(1))}{\frac{c''^2}{4}n^{2\lambda}} \to 0$$

$$P(N^{(n)} > \frac{c''}{2}n^{\lambda}) \to 1 \quad \text{as } n \to \infty.$$

□

The next step in the proof is to show that users that are identified by the set $J^{(n)}$ produce a very similar moving process as user 1. To make this statement precise, we provide the following definition. Define the set $A^{(m)}$ as the interval in $\mathbb{R}$ consisting of real numbers which are within the $m\beta_m$ distance from mp_1 (the expected number of times that user 1 is at state 1 during the m number of observations),

$$A^{(m)} = \{x \in R, m(p_1 - \beta_m) \leq x \leq m(p_1 + \beta_m)\}.$$

Lemma 3. We have

$$P\left(\bigcap_{j \in J^{(n)}} \left(S_j^{(m)} \in A^{(m)}\right)\right) \to 1 \quad \text{as } n \to \infty$$

Proof. Let $j \in J^{(n)}$ and $p_1 - \epsilon_m < p_j < p_1 + \epsilon_m$. Since $S_j^{(m)} \sim \text{Binomial}(m, p_j)$, by the Large Deviation Theory (Sanvo's Theorem), we can write

$$P\left(S_j^{(m)} > m(p_1 + \beta_m)\right) < (m+1)2^{-mD(Bernoulli(p_1+\beta_m)\|Bernoulli(p_j))}$$

By using the fact that for all $p \in (0,1)$

$$D\left(Bernoulli(p+\epsilon)\|Bernoulli(p)\right) = \frac{\epsilon^2}{2p(1-p)\ln 2} + O(\epsilon^3),$$

we can write

$$D\left(Bernoulli(p_1+\beta_m)\|Bernoulli(p_j)\right)=\frac{(p_1+\beta_m-p_j)^2}{2p_j(1-p_j)\ln 2}+O\left((p_1+\beta_m-p_j)^3\right).$$

Note that $|p_1-p_j|<\epsilon_m$, so for large m we can write

$$|p_1+\beta_m-p_j|\geq\beta_m-\epsilon_m=\frac{1}{m^{\frac{1}{2}-\theta}}-\frac{1}{m^{\frac{1}{2}+\phi}}>\frac{\frac{1}{2}}{m^{\frac{1}{2}-\theta}}.$$

so we can write

$$D\left(Bernoulli(p_1+\beta_m)\|Bernoulli(p_j)\right)=\frac{1}{8p_j(1-p_j)m^{1-2\theta}\ln 2}+O((p_1+\beta_m-p_j)^3)$$

and for some constant $c'>0$

$$D\left(Bernoulli(p_1+\beta_m)\|Bernoulli(p_j)\right)>\frac{c'}{m^{1-2\theta}}\Rightarrow$$

$$mD\left(Bernoulli(p_1+\beta_m)\|Bernoulli(p_j)\right)>\frac{mc'}{m^{1-2\theta}}>c'm^{2\theta}\Rightarrow$$
$$P(S_j^{(m)}>m(p_1+\beta_m))<m2^{-c'm^{2\theta}}.$$

So in conclusion

$$P\left(\bigcup_{j\in J^{(n)}}S_j^{(m)}>m(p_1+\beta_m)\right)<|J^{(n)}|m2^{-c'm^{2\theta}}$$
$$|J^{(n)}|m2^{-c'm^{2\theta}}<nm2^{-c'm^{2\theta}}<m^2 2^{-c'm^{2\theta}}\to 0\quad\text{as}\quad m\to\infty.$$

Similarly we obtain

$$P\left(\bigcup_{j\in J^{(n)}}S_j^{(m)}<m(p_1-\beta_m)\right)\to 0\quad\text{as}\quad m\to\infty,$$

which completes the proof. This shows that for all users j for which p_j is within ϵ range around p_1, i.e. it is in set $J^{(n)}$, the average number of times that this user was at state 1 is within $m\beta_m$ from mp_1 with high probability. □

We are now in a position to show that distinguishing between the users in $J^{(n)}$ is not possible for an outside observer (i.e., the adversary) and this will pave the way in showing perfect location privacy.

Lemma 4. Let $\{a_m\}_{m=1}^{\infty}$, $\{b_m\}_{m=1}^{\infty}$ be such that a_m, b_m are in set $A^{(m)}$ and also $\{i_m\}_{m=1}^{\infty}$, $\{j_m\}_{m=1}^{\infty}$ be such that i_m, j_m are in set $J^{(n)}$. Then, we have

$$\frac{P\left(S_{i_m}^{(m)} = a_m, S_{j_m}^{(m)} = b_m\right)}{P\left(S_{i_m}^{(m)} = b_m, S_{j_m}^{(m)} = a_m\right)} \to 1 \quad \text{as} \ \ m \to \infty.$$

Proof. Remember that

$$A^{(m)} = \{x \in R, m(p_1 - \beta_m) \leq x \leq m(p_1 + \beta_m)\}$$

where $\beta_m = \frac{1}{m^{\frac{1}{2}-\theta}}$ and $S_j^{(m)} \sim \text{Binomial}(m, p_j)$. Thus, $S_{i_m}^{(m)} \sim \text{Binomial}(m, p_{i_m})$ and $S_{j_m}^{(m)} \sim \text{Binomial}(m, p_{j_m})$,

$$P(S_{i_m}^{(m)} = a_m) = \binom{m}{a_m} p_{i_m}^{a_m} (1 - p_{i_m})^{m - a_m},$$

$$P(S_{j_m}^{(m)} = b_m) = \binom{m}{b_m} p_{j_m}^{b_m} (1 - p_{j_m})^{m - b_m}.$$

In conclusion,

$$\Delta_m = \frac{P\left(S_{i_m}^{(m)} = a_m, S_{j_m}^{(m)} = b_m\right)}{P\left(S_{i_m}^{(m)} = b_m, S_{j_m}^{(m)} = a_m\right)} = (\frac{p_{i_m}}{p_{j_m}})^{a_m - b_m} (\frac{1 - p_{j_m}}{1 - p_{i_m}})^{a_m - b_m}$$

$$\ln \Delta_m = (a_m - b_m) \ln(\frac{p_{i_m}}{p_{j_m}}) + (a_m - b_m) \ln(\frac{1 - p_{j_m}}{1 - p_{i_m}})$$

and since $\{i_m, j_m\} \in J^{(n)}$ we have

$$|p_{i_m} - p_{j_m}| \leq 2\epsilon_m = \frac{2}{m^{\frac{1}{2}+\phi}}.$$

Also, since $\{a_m, b_m\} \in A^{(m)}$ we can say that

$$|a_m - b_m| \leq 2m\beta_m.$$

Since $p_{i_m} \leq p_{j_m} + 2\epsilon_m$ and $1 - p_{j_m} \leq (1 - p_{i_m}) + 2\epsilon_m$ and

$$\ln(1 + \epsilon_m) = \epsilon_m + O(\epsilon_m^2)$$

we can write

$$\ln \Delta_m \leq 2m\beta_m\epsilon_m + 2m\beta_m\epsilon_m + 2m\beta_m O(\epsilon_m^2)$$

and since $\phi > \theta$,

$$m\beta_m\epsilon_m = m\frac{1}{m^{\frac{1}{2}+\phi}}\frac{1}{m^{\frac{1}{2}-\theta}} = \frac{1}{m^{\phi-\theta}} \to 0,$$
$$\Rightarrow \ln \Delta_m \to 0$$
$$\Rightarrow \Delta_m \to 1.$$

Note that the convergence is uniform.

This shows that for two users i and j, if the probability of them being at state 1 is in set $J^{(n)}$, $p_i, p_j \in J^{(n)}$, and also the observed number of times for these users to be at state 1 is in set $A^{(m)}$, then distinguishing between these two users is impossible. □

Lemma 5. For any $j \in \Pi(J^{(n)})$, we define $W_j^{(n)}$ as follows

$$W_j^{(n)} = P(\Pi(1) = j | \mathbf{Y}^{(m)}, \Pi(J^{(n)})).$$

Then, for all $j^{(n)} \in \Pi(J^{(n)})$,

$$N^{(n)}W_j^{(n)} \xrightarrow{p} 1.$$

More specifically, for all $\gamma_1, \gamma_2 > 0$, there exists n_o such that if $n > n_o$:

$$\forall j \in \Pi(J^{(n)}): \quad P\left(\left|N^{(n)}W_j^{(n)} - 1\right| > \gamma_1\right) < \gamma_2.$$

Proof. This is the result of Lemma 4. First, remember that

$$\sum_{j \in \Pi(J^{(n)})} W_j^{(n)} = 1,$$

and also note that

$$|\Pi(J^{(n)})| = |J^{(n)}| = N^{(n)} \to \infty \quad \text{as} \quad n \to \infty.$$

Here, we show that for any $\{j_n\}_{n=1}^{\infty} \in \Pi(J^{(n)})$,

$$\frac{W_{j_n}^{(n)}}{W_1^{(n)}} = \frac{P\left(\Pi(1) = j | D\right)}{P\left(\Pi(1) = 1 | D\right)} \xrightarrow{p} 1$$

where $D = \left(\mathbf{Y}^{(m)}, \Pi(J^{(n)})\right)$.

Let a_i, for $i \in \Pi(J^{(n)})$, be the permuted observed values of $S_i^{(m)}$'s. Then note that

$$P(\Pi(1) = j | D) = \sum_{\substack{\text{permutation} \\ \text{such that } \Pi(1)=j}} \sum_{i \in \Pi(J)} P(S_i^{(m)} = a_i).$$

Then, in

$$\frac{W_{j_n}^{(n)}}{W_1^{(n)}} = \frac{P(\Pi(1) = j | D)}{P(\Pi(1) = 1 | D)}$$

the numerator and denominator have the same terms. In particular, for each term

$$P(S_j^{(m)} = a_{j_n}) \times P(S_1^{(m)} = b_{j_n})$$

in $W_j^{(n)}$, there is a corresponding term

$$P(S_j^{(m)} = b_{j_n}) \times P(S_1^{(m)} = a_{j_n})$$

in $W_1^{(n)}$. Since by Lemma 4

$$\frac{P(S_j^{(m)} = a_{j_n}) \times P(S_1^{(m)} = b_{j_n})}{P(S_j^{(m)} = b_{j_n}) \times P(S_1^{(m)} = a_{j_n})}$$

converges uniformly to 1, we conclude

$$\frac{W_{j_n}^{(n)}}{W_1} \to 1.$$

We conclude that for any $\zeta > 0$, we can write (for large enough n)

$$(1-\zeta) < \frac{W_{j_n}^{(n)}}{W_1} < (1+\zeta),$$

$$\sum_{j\in\Pi(J^{(n)})} (1-\zeta)W_1^{(n)} < \sum_{j\in\Pi(J^{(n)})} W_{j_n}^{(n)} < \sum_{j\in\Pi(J^{(n)})} (1+\zeta)W_1^{(n)}$$

and since $\sum_{j\in\Pi(J^{(n)})} W_{j_n}^{(n)} = 1$, $|\Pi(J^{(n)})| = N^{(n)}$, we have

$$(1-\zeta)N^{(n)}W_1^{(n)} < 1 < (1+\zeta)N^{(n)}W_1^{(n)}$$

so, we conclude that $N^{(n)}W_1^{(n)} \to 1$ as $n \to \infty$. We can repeat the same argument for all users in set $j \in J^{(n)}$ and we get $N^{(n)}W_j^{(n)} \to 1$ as $n \to \infty$. □

Now to finish the proof of Theorem 1,

$$P\left(X_1(k) = 1|\mathbf{Y}^{(m)}, \Pi(J^{(n)})\right) =$$

$$\sum_{j\in\Pi(J^{(n)})} P\left(X_1(k) = 1|\mathbf{Y}^{(m)}, \Pi(1) = j, \Pi(J^{(n)})\right) \times P\left(\Pi(1) = j|\mathbf{Y}^{(m)}, \Pi(J^{(n)})\right)$$

$$= \sum_{j\in\Pi(J^{(n)})} 1_{[Y_j^{(m)}(k)=1]} W_j^{(n)} \triangleq Z_n.$$

But, since $Y_j^{(m)}(k) \sim \text{Bernoulli}(p_j^{(n)})$ and $p_j^{(n)} \to p_1$ for all $j \in \Pi(J^{(n)})$, by the law of large numbers we have:

$$\frac{1}{N^{(n)}} \sum_{j\in\Pi(J^{(n)})} 1_{[Y_j^{(m)}(k)=1]} \to p_1$$

$$Z_n = \frac{1}{N^{(n)}} \sum_{j\in\Pi(J^{(n)})} (1_{[Y_j^{(m)}(k)=1]})(N^{(n)}W_j^{(n)}).$$

Using $N^{(n)}W_j^{(n)} \to 1$ in the previous equation, we obtain $Z_n \to p_1$.

In conclusion $X_1(k)|\mathbf{Y}^{(m)}, \Pi(J^{(n)}) \xrightarrow{d} \text{Bernoulli}(p_1)$ which means that

$$H\left(X_1(k)|\mathbf{Y}^{(m)}, \Pi(J^{(n)})\right) \to H(X_1(k))$$

$$\Rightarrow H\left(X_1(k)|\mathbf{Y}^{(m)}\right) \geq H\left(X_1(k)|\mathbf{Y}^{(m)}, \Pi(J^{(n)})\right) \to H(X_1(k))$$

$$\Rightarrow I\left(X_1(k); \mathbf{Y}^{(m)})\right) \to 0$$

3.3 Perfect Location Privacy for r-States i.i.d. Model

Here we extend the results to a scenario in which we have $r \geq 2$ locations or regions, locations $0, 1, \cdots, r-1$. At any time $k \in \{0, 1, 2, \cdots\}$, user u has probability $p_u(i) \in (0, 1)$ to be at location i, independent from previous locations and other users' locations. At any given time k, we show the probability of user u being at location i as follows:

$$p_u(i) = P(X_u(k) = i),$$

$$\mathbf{p}_u = (p_u(0), p_u(1), \cdots, p_u(r-1)).$$

We assume that $\mathbf{p}_u(i)$'s are drawn independently from some $r-1$ dimensional continuous density function $f_P(p)$ on $(0, 1)^{r-1}$. Let

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

Then, $\exists \delta_1, \delta_2 > 0$ such that

$$\begin{cases} \delta_1 < f_P(p) < \delta_2 & \mathbf{p}_u \in R_P \\ f_P(p) = 0 & \mathbf{p}_u \notin R_P \end{cases} \tag{3.3}$$

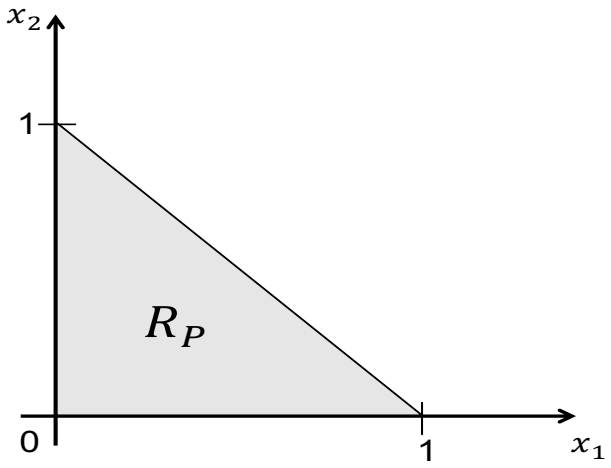

Figure 3.1: R_P for case $r = 3, (d = 2)$.

Theorem 2. For r locations with the above definition and the adversary's observation vector $\mathbf{Y}^{(m)}$ if all the following holds,

1. $m = cn^{\frac{2}{r-1}-\alpha}$, which $c, \alpha > 0$ and are constant
2. $\mathbf{p}_1 \in (0,1)^{(r-1)}$
3. $(\mathbf{p}_2, \mathbf{p}_3, \cdots, \mathbf{p}_n) \sim f_P$, $0 < \delta_1 < f_P < \delta_2$
4. $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n)$ be known to the adversary

then, we have

$$\forall k \in \mathbb{N}, \quad \lim_{N\to\infty} I\left(X_1(k); \mathbf{Y}^{(m)}\right) = 0$$

Proof of the Theorem 2 is analogous to the proof of Theorem 1. Here, we provide the general intuition.

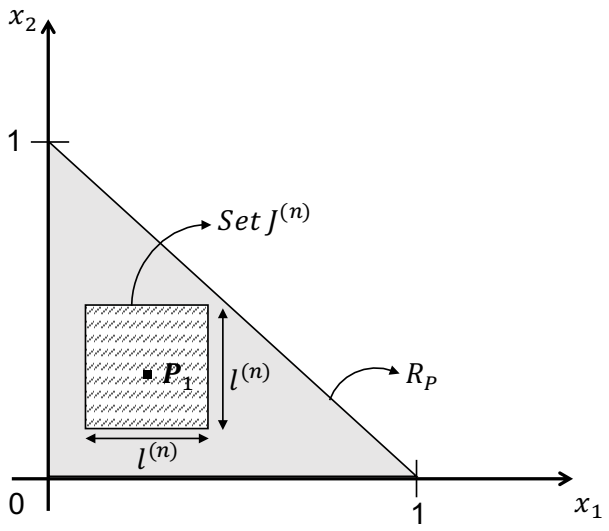

Figure 3.2: $\mathbf{p}_1 = (p_1(0), p_1(1), \cdots, p_1(r-1))$ is in set $J^{(n)}$ in R_P.

Let $\mathbf{p}_1 = (p_1(0), p_1(1), \cdots, p_1(r-1))$ and $d = r - 1$. As you can see in figure 3.2, there exists a set $J^{(n)}$ such that $\mathbf{p}_1$ is in this set and also we have:

$$Vol(J^{(n)}) = (l^{(n)})^d.$$

We choose $l^{(n)} = \frac{1}{n^{\frac{1}{d}-\eta}}$, where $\eta < \frac{\alpha}{2}$. Thus, the average number of users with their $\mathbf{p}$ vector in $J(n)$ is

$$n\frac{1}{\left(n^{\frac{1}{d}-\eta}\right)^d} = n^{d\eta} \to \infty. \tag{3.4}$$

So, we can guarantee a large number of users in $J^{(n)}$. Here, the number of times each user is at any location follows a multinomial distribution and in the long-term observations these numbers have a jointly gaussian distribution.

The standard deviation of these variables are in the form of $\frac{const}{\sqrt{m}}$. In particular, the standard deviation over the length of this interval is large.

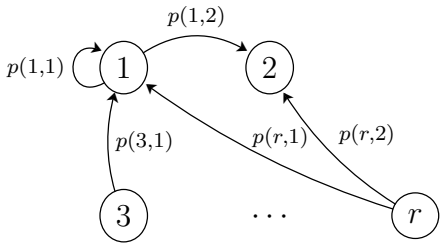

Figure 3.3: Markov chain model with r states and $|E|$ edges.

$$\frac{\frac{const.}{\sqrt{m}}}{l^{(n)}} = \frac{const.\ n^{\frac{1}{d}-\eta}}{\sqrt{m}} \sim \frac{n^{\frac{1}{d}-\eta}}{\left(n^{\frac{2}{d}-\alpha}\right)^{\frac{1}{2}}} = n^{\frac{\alpha}{2}-\eta} \to \infty \tag{3.5}$$

Again, we have a large number of asymptotically jointly normal random variables that have a much larger standard deviation compared to the differences of their means. Thus, distinguishing between them becomes impossible.

This proves that it is impossible for the adversary to find a specific user to map to the observations even by having $\mathbf{P}$ and $\mathbf{Y}^{(m)}$. So, all the users have perfect location privacy.

3.4 Perfect Location Privacy in Markov Chain Model

Assume there are r possible locations which users can occupy. We use a Markov chain with r states to model movements of each user. We define E, the set of edges in this Markov chain, such that (i, j) is in E if there exists an edge from i to j with probability $p\prime(i, j) > 0$.

We assume that this Markov structure chain gives the movement pattern of each user and what differentiates between users is their transition probabilities. That is, for fixed locations i and j, two different users could have two different transition probabilities. For simplicity, let us assume that all users start at location (state) 1, i.e., $X_u(1) = 1$ for all $u = 1, 2, \cdots$. This condition is not necessary and can be easily relaxed; however, we assume it here for the clarity of exposition. We now state and prove the theorem that gives the condition for perfect location privacy for a user in the above setting.

Theorem 3. For an irreducible, aperiodic Markov chain with r states and $|E|$ edges, if $m = cn^{\frac{2}{|E|-r}-\alpha}$, where $c > 0$ and $\alpha > 0$ are constants, then

$$\lim_{n\to\infty} I(X_1(k); \mathbf{Y}^{(m)}) = 0, \quad \forall k \in \mathbb{N}, \tag{3.6}$$

i.e., user 1 has perfect location privacy.

Proof. Let $M_u(i,j)$ be the number of observed transitions from state i to state j for user u. We first show that $M_{\Pi(u)}(i,j)$'s provide a sufficient statistic for the adversary when the adversary's goal is to obtain the permutation $\Pi^{(n)}$. To make this statement precise, let us define $\mathbf{M}_u^{(m)}$ as the matrix containing $M_u(i,j)$'s for user u:

$$\mathbf{M}_u^{(m)} = \begin{bmatrix} M_u(1,1) & M_u(1,2) & \cdots & M_u(1,r) \\ M_u(2,1) & M_u(2,2) & \cdots & M_u(2,r) \\ \cdots & \cdots & \cdots & \cdots \\ M_u(r,1) & M_u(r,2) & \cdots & M_u(r,r) \end{bmatrix}$$

Also, let $\mathbf{M}^{(m)}$ be the ordered collection of $\mathbf{M}_u^{(m)}$'s. Specifically,

$$\mathbf{M}^{(m)} = \left(\mathbf{M}_1^{(m)}, \mathbf{M}_2^{(m)}, \cdots, \mathbf{M}_n^{(m)}\right)$$

The adversary can obtain $\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big)$, a permuted version of $\mathbf{M}^{(m)}$. In particular, we can write

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

We now state a lemma that confirms $\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big)$ is a sufficient statistic for the adversary, when the adversary's goal is to recover $\Pi^{(n)}$. Here, $\mathbf{Y}^{(m)}$ is the collection of anonymized observations of users' locations available to the adversary.

Lemma 6. Given $\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big)$, the random matrix $\mathbf{Y}^{(m)}$ and the random permutation $\Pi^{(n)}$ are conditionally independent. That is

$$P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}, \text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) = P\left(\Pi^{(n)} = \pi \;\middle|\; \text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) \tag{3.7}$$

Lemma 6 is proved in Section 3.4.1.

We assumed the Markov chain to be irreducible and aperiodic so that when we are determining $p(i, j)$'s, there are d degrees of freedom, where d is equal to $|E| - r$. This is because for each state i, we must have

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

Thus, the Markov chain of the user u is completely determined by d values of $p(i, j)$'s which we show as

$$\mathbf{P}_u = (p_u(1), p_u(2), \cdots, p_u(d))$$

and $\mathbf{P}_u$'s are known to the adversary for all users. Note that the choice of $\mathbf{P}_u$ is not unique; nevertheless, as long as we fix a specific $\mathbf{P}_u$, we can proceed with the proof. We define E_d as the set of d edges whose $p(i, j)$'s belong to $\mathbf{P}_u$. Let $R_{\mathbf{p}} \subset \mathbb{R}^d$ be the range of acceptable values for $\mathbf{P}_u$. For example, in Figure 3.4 we have $|E| = 6$ and $r = 3$, so we have three independent transitions probabilities. If we choose p_1, p_2, and p_3 according to the figure, we obtain the following region

$$R_{\mathbf{p}} = \{(p_1, p_2, p_3) \in \mathbb{R}^3 : 0 \leq p_i \leq 1 \text{ for } i = 1, 2, 3 \text{ and } p_1 + p_2 \leq 1\}.$$

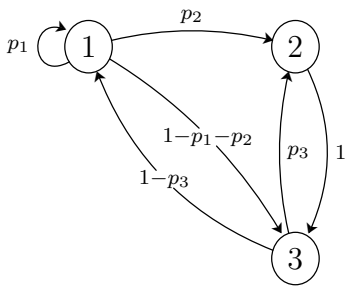

Figure 3.4: Three states Markov chain example

The statistical properties of each user are completely known to the adversary since she knows the Markov chain of each user. The adversary wants to be able to distinguish between users by having m observations per user and also knowing $\mathbf{P}_u$'s for all users.

In this model, we assume that $\mathbf{P}_u$ for each user u is drawn independently from a d-dimensional continuous density function, $f_{\mathbf{P}}(\mathbf{p})$. As before, we assume there exist positive constants $\delta_1, \delta_2 > 0$, such that

$$\begin{cases} \delta_1 < f_{\mathbf{P}}(\mathbf{p}) < \delta_2 & \mathbf{p} \in R_{\mathbf{p}} \\ f_{\mathbf{P}}(\mathbf{p}) = 0 & \mathbf{p} \notin R_{\mathbf{p}} \end{cases}$$

We now claim that the adversary's position in this problem is mathematically equivalent to the the i.i.d model where the number of locations r is equal to $d+1$ where $d = |E| - r$. First, note that since the Markov chain is irreducible and aperiodic, it has a unique stationary distribution which is equal to the limiting distribution. Next, define $\mathbf{Q}_u$ to be the vector consisting of all the transition probabilities of user u. In particular, based on the above argument, we can represent $\mathbf{Q}_u$ in the following way:

$$\mathbf{Q}_u = [\mathbf{P}_u, \mathbf{P}_u\mathbf{B}],$$

where $\mathbf{B}$ is a non-random d by $|E| - d$ matrix. Now, note that $\mathbf{P}_u\mathbf{B}$ is a non-random function of $\mathbf{P}_u$. In particular, if $M_u(i, j)$ shows the observed number transitions from state i to state j for user

u, then we only need to know $M_u(i,j)$ for the edges in E_d, as the rest will be determined by the linear transform defined by $\mathbf{B}$. This implies that the decision problem for the adversary is reduced to the decision problem on transition probabilities in $\mathbf{P}_u$ and the adversary only needs to look at the $M_u(i,j)$'s for the edges in E_d. Now, this problem has exactly the same structure as the i.i.d model where the number of locations r is equal to $d+1$ where $d = |E| - r$. In particular, $M_u(i,j)$'s have multinomial distributions and the statement of Theorem 3 follows by applying Theorem 2.

□

3.4.1 Proof of Lemma 6

Here, we provide a formal proof for Lemma 6 which we restate as follows. In the Markov chain setting of Section 3.4, we have the following: Given $\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big)$, the random matrix $\mathbf{Y}^{(m)}$ and the random permutation $\Pi^{(n)}$ are conditionally independent. That is

$$P\left(\Pi^{(n)} = \pi \;\middle|\; \mathbf{Y}^{(m)} = \mathbf{y}, \text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) = P\left(\Pi^{(n)} = \pi \;\middle|\; \text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) \tag{3.8}$$

Proof. Remember

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

Note that $\mathbf{Y}^{(m)}$ (and therefore $\mathbf{y}$) is an m by n matrix, so we can write

$$\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n),$$

where for $u = 1, 2, \cdots, n$, we have

$$\mathbf{y}_u = \begin{bmatrix} y_u(1) \\ y_u(2) \\ \vdots \\ y_u(m) \end{bmatrix}.$$

Also, $\mathbf{m}$ is a collection of n matrices so we can write

$$\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_n).$$

For an $r \times r$ matrix $\mathbf{m} = [m(i,j)]$, let us define $D(\mathbf{m})$ as the set of sequences $(x_1, x_2, \cdots, x_m) \in \{1, 2, \cdots, r\}^m$ that satisfy the following properties:

1. $x_0 = 1$;
2. The number of transitions from i to j in $(x_1, x_2, \cdots, x_m)$ is equal to m_{ij} for all i and j. That is, the number of indices k for which we have $x_k = i$ and $x_{k+1} = j$ is equal to $m(i,j)$.

We now show that the two sides of Equation 3.8 are equal. The right hand side probability can be written as

$$\begin{aligned} P\left(\Pi^{(n)} = \pi \;\middle|\; \text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) &= \frac{P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m} \;\middle|\; \Pi^{(n)} = \pi\right) P\left(\Pi^{(n)} = \pi\right)}{P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right)} \\ &= \frac{P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \pi\big) = \mathbf{m} \;\middle|\; \Pi^{(n)} = \pi\right)}{n! P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right)} \\ &= \frac{P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \pi\big) = \mathbf{m}\right)}{n! P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right)}. \end{aligned}$$

Now note that

$$
\begin{aligned}
P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \pi\big) = \mathbf{m}\right) &= P\left(\bigcap_{j=1}^{n}\left(\mathbf{M}^{(m)}_{\pi^{-1}(j)} = \mathbf{m}_j\right)\right) \\
&= P\left(\bigcap_{u=1}^{n}\left(\mathbf{M}^{(m)}_{u} = \mathbf{m}_{\pi(u)}\right)\right) \\
&= \prod_{u=1}^{n} P\left(\mathbf{M}^{(m)}_{u} = \mathbf{m}_{\pi(u)}\right) \\
&= \prod_{u=1}^{n} \sum_{(x_1,x_2,\cdots,x_m)\in D(\mathbf{m}_{\pi(u)})} P(X_u(1)=x_1, X_u(2)=x_2, \cdots, X_u(m)=x_m) \\
&= \prod_{u=1}^{n} \sum_{(x_1,x_2,\cdots,x_m)\in D(\mathbf{m}_{\pi(u)})} \prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)} \\
&= \prod_{u=1}^{n} \left(|D(\mathbf{m}_{\pi(u)})| \prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)}\right) \\
&= \left(\prod_{k=1}^{n} |D(\mathbf{m}_k)|\right)\left(\prod_{u=1}^{n}\prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)}\right)
\end{aligned}
$$

Similarly, we obtain

$$
\begin{aligned}
P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \Pi^{(n)}\big) = \mathbf{m}\right) &= \sum_{\text{all permutations } \pi'} P\left(\text{Perm}\big(\mathbf{M}^{(m)}, \pi'\big) = \mathbf{m} \;\middle|\; \Pi^{(n)} = \pi'\right) P\left(\Pi^{(n)} = \pi'\right) \\
&= \frac{1}{n!} \sum_{\text{all permutations } \pi'} \left(\prod_{k=1}^{n} |D(\mathbf{m}_k)|\right)\left(\prod_{u=1}^{n}\prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi'(u)}(i,j)}\right) \\
&= \frac{1}{n!} \left(\prod_{k=1}^{n} |D(\mathbf{m}_k)|\right) \sum_{\text{all permutations } \pi'} \left(\prod_{u=1}^{n}\prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi'(u)}(i,j)}\right).
\end{aligned}
$$

Thus, we conclude that the right hand side of Equation 3.8 is equal to

$$
\frac{\prod_{u=1}^{n}\prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)}}{\sum_{\text{all permutations } \pi'}\left(\prod_{u=1}^{n}\prod_{i,j} p_u(i,j)^{\mathbf{m}_{\pi'(u)}(i,j)}\right)}.
$$

Now let us look at the left hand side of Equation 3.8. We can write

$$P\left(\Pi^{(n)}=\pi \;\middle|\; \mathbf{Y}^{(m)}=\mathbf{y}, \text{Perm}\big(\mathbf{M}^{(m)},\Pi^{(n)}\big)=\mathbf{m}\right)=P\left(\Pi^{(n)}=\pi \;\middle|\; \mathbf{Y}^{(m)}=\mathbf{y}\right).$$

This is because $\text{Perm}\big(\mathbf{M}^{(m)},\Pi^{(n)}\big)$ is a function of $\mathbf{Y}^{(m)}$. We have

$$P\left(\Pi^{(n)}=\pi \;\middle|\; \mathbf{Y}^{(m)}=\mathbf{y}\right)=\frac{P\left(\mathbf{Y}^{(m)}=\mathbf{y} \;\middle|\; \Pi^{(n)}=\pi\right)P\left(\Pi^{(n)}=\pi\right)}{P\left(\mathbf{Y}^{(m)}=\mathbf{y}\right)}$$

We have

$$\begin{aligned}P\left(\mathbf{Y}^{(m)}=\mathbf{y} \;\middle|\; \Pi^{(n)}=\pi\right)&=\prod_{u=1}^{n}P(X_u(1)=y_{\pi(u)}(1),X_u(2)=y_{\pi(u)}(2),\cdots,X_u(m)=y_{\pi(u)}(m))\\&=\prod_{u=1}^{n}\prod_{i,j}p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)}.\end{aligned}$$

Similarly, we obtain

$$P\left(\mathbf{Y}^{(m)}=\mathbf{y}\right)=\frac{1}{n!}\sum_{\text{all permutations }\pi'}\prod_{u=1}^{n}\prod_{i,j}p_u(i,j)^{\mathbf{m}_{\pi'(u)}(i,j)}$$

Thus, we conclude that the left hand side of Equation 3.8 is equal to

$$\frac{\prod_{u=1}^{n}\prod_{i,j}p_u(i,j)^{\mathbf{m}_{\pi(u)}(i,j)}}{\sum_{\text{all permutations }\pi'}\left(\prod_{u=1}^{n}\prod_{i,j}p_u(i,j)^{\mathbf{m}_{\pi'(u)}(i,j)}\right)},$$

which completes the proof.

□

CHAPTER 4

SIMULATION

4.1 I.I.D. Model

In the two states model with n number of users, let p_u be the probability of user u being at state 1 and the observation vector $\mathbf{Y}^{(m)}$ consists of m observation for each user.

If the following holds,

1. $m = cn^{2-\alpha}$, which $c, \alpha > 0$ and are constant
2. $p_1 \in (0, 1)$
3. $(p_2, p_3, \cdots, p_n) \sim f_P$, $0 < \delta_1 < f_P < \delta_2$
4. $P = (p_1, p_2, \cdots, p_n)$ be known to the adversary

then, we have perfect location privacy for all the users, e.g. for the first user we have,

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} I\left(X_1(k); \mathbf{Y}^{(m)}\right) = 0.$$

4.1.1 Threat Model

In this framework, we assume that the adversary knows the exact profile of each of the users. Here, by observing $\mathbf{Y}^{(m)}$, the adversary aims to break the permutation function and map each anonymized collection of m observations to a user.

In this framework we assume that the adversary breaks the permutation function using the maximum likelihood method. Suppose the number of observations that the adversary collects for

each user in the two state model is $m = \alpha n^2$ and she performs maximum likelihood to recover the mapping $\Pi^{(n)}$. For anonymized user u, she counts the number of times that the user was at state 1 during m observations and then she compares that with probability of all n users being at state 1 and matches the anonymized user to the user with the closest probability of being at state 1.

4.1.2 Error Probability of The Adversary

Using maximum likelihood, the adversary finds the permutation function $\tilde{\Pi}^{(n)}$. She obtains the probability of each anonymized user being at state 1 from her observations, compares them with her prior knowledge and matches the closest ones and de-anonymizes the users.

We define the error probability of the adversary to be

$$P_e(\alpha, n) = E\left[\frac{|\{i : \tilde{\Pi}(i) \neq \Pi(i)\}|}{n}\right]. \tag{4.1}$$

Theorem 4. In the above setting, define

$$P_e(\alpha) = \lim_{n\to\infty} P_e(\alpha, n). \tag{4.2}$$

Then, the following statements hold:

1. The function $P_e(\alpha)$ is well defined (the limit exists) for all $\alpha \in \mathbb{R}^+$.
2. $0 < P_e(\alpha) < 1$ for all $\alpha \in \mathbb{R}^+$.
3. $P_e(\alpha)$ is a decreasing function of α.
4. $\lim_{\alpha\to 0} P_e(\alpha) = 1$ and $\lim_{\alpha\to\infty} P_e(\alpha) = 0$.

Figure 4.1 shows $P_e(\alpha)$ as a function of α when users' probabilities (i.e., p_i's) are uniformly distributed in $(0, 1)$. In this simulation, we generated n number of p_us as the adversary's prior knowledge. Then we generated an observation vector for each user based on the probability of

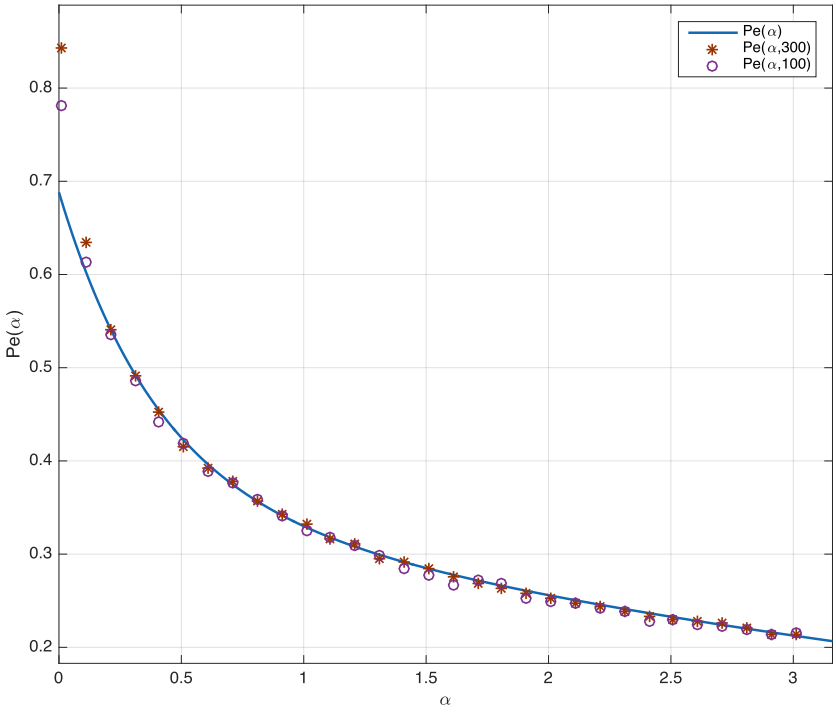

Figure 4.1: Error probability of the adversary $P_e(\alpha)$ vs. α. $P_e(\alpha)$ and $P_e(\alpha, N)$ when p has uniform distribution in $(0, 1)$ for $n = \{100, 300\}$.

that user being at state 1. Having the observation vector, we tried to match users and observation vectors using maximum likelihood. By obtaining $\tilde{\Pi}^{(n)}$, we computed the error probability that is shown in Fig. 4.1 for different number of users.

Simulation results in Figure 4.1 shows that as α grows, the adversary's error probability goes to zero which shows that the adversary maps users with low error probability. On the other hand, as α becomes smaller, the error probability approaches 1. These results are consistent with the our main result that users have perfect privacy if the adversary obtains less than $O(n^{\frac{2}{r-1}})$ observations per user.

4.2 Markov Chain Model

With n number of users, observation vector $\mathbf{Y}^{(m)}$ consists of m observations for each user.

For an irreducible, aperiodic Markov chain with r states and $|E|$ edges, if $m = cn^{\frac{2}{|E|-r}-\beta}$, where $c > 0$ and $\beta > 0$ are constants, then

$$\lim_{n\to\infty} I(X_1(k); \mathbf{Y}^{(m)}) = 0, \quad \forall k \in \mathbb{N}, \tag{4.3}$$

i.e., user 1 has perfect location privacy.

4.2.1 Threat Model

In this framework, we assume an adversary who is the strongest adversary and has all the statistical knowledge of the users' movements. Here, by observing $\mathbf{Y}^{(m)}$, the adversary aims to break the permutation function and map each anonymized collection of m observations to a user.

In this framework we assume that adversary breaks the permutation function using the maximum likelihood method. Suppose the number of observations that the adversary collects for each user is $m = \alpha n^{\frac{2}{|E|-r}}$ and she performs maximum likelihood to recover the mapping $\Pi^{(n)}$.

4.2.2 Error Probability of the Adversary

Here, we provide some simulation results that verify the result in Theorem 3. We consider a network with n users and r locations. The possible path of each user can be modeled as an irreducible, aperiodic Markov chain with r states and $|E|$ number of edges. After obtaining m observations per user, the adversary estimates transition probabilities $\tilde{p}_u(i, j)$. If we consider the number of transition from state i to j as $m_{(i,j)}$ and the number of times the user was at state i as m_i then $\frac{m_{(i,j)}}{m_i}$ gives the transition probability from state i to j for that user. By using nearest neighbor decoding in $\mathbb{R}^d$, the adversary matches a user with the closest transition probabilities to the observed paths.

In our simulations we consider $r = 3$ and $m = \alpha n^{\frac{2}{|E|-r}}$. We used four different Markov chains as users' movement models and tried to de-anonymize the users by observing them m times.

We define the error probability of the adversary to be

$$P_e(\alpha) = E\left[\frac{|\{i : \tilde{\Pi}(i) \neq \Pi(i)\}|}{n}\right]. \tag{4.4}$$

We see that if the adversary's number of observations, m, is more than $O(n^{\frac{2}{|E|-r}})$, then the adversary's error probability goes to zero. On the other hand, if the number of observations is much smaller, then the error probability goes to one, suggesting that users might have perfect location privacy.

First, we model each user's path as a Markov chain $MC1$ shown in figure 4.2. Since in this model $|E| = 4$ we can write $m = \alpha n^2$.

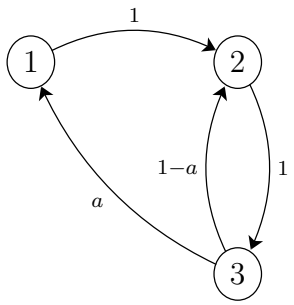

Figure 4.2: The Markov chain MC1 which models of users' path.

For n users, we create $a \in (0, 1)$ which are i.i.d. We then generate m observations for each user. Here, m_1 is the number of times that a user was at state 1. m_2 and m_3 are defined in the same manner. $m_{(3,1)}$ was the number of jumps from state 3 to state 1. By calculating $\tilde{a} = \frac{m_{(3,1)}}{m_3}$ for all users and comparing $\tilde{a}$ for each path to all the users' probability a , we matched the closest a and $\tilde{a}$. Figure 4.3 shows the error probability that we obtain in simulating users moving in MC1 model.

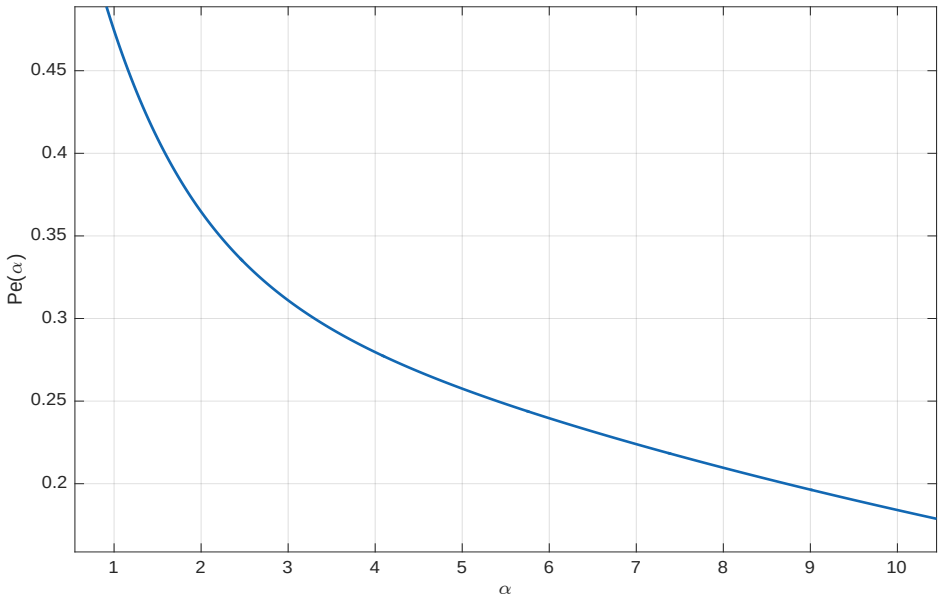

Figure 4.3: $Pe(\alpha)$ vs. α for Markov chain MC1 with $n = 500$.

Second, we model each user's path as a Markov chain $MC2$ shown in Figure 4.4. Since in this model $|E| = 5$ we can write $m = \alpha n$.

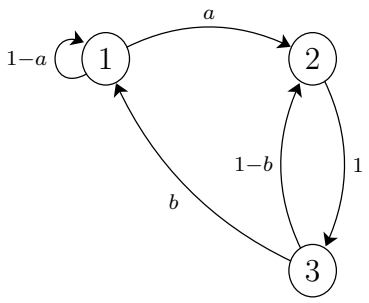

Figure 4.4: The Markov chain MC2 which models of users' path.

For n users, we create $(a, b) \in (0, 1)$ which are i.i.d. We then generate m observations for each user. By calculating $\tilde{a} = \frac{m_{(1,2)}}{m_1}$ and $\tilde{b} = \frac{m_{(3,1)}}{m_3}$ for all users and comparing $\tilde{a}, \tilde{b}$ for each path to all

the users' probabilities a and b , we matched the closest (a, b) and $(\tilde{a}, \tilde{b})$ in $\mathbb{R}^2$. Figure 4.5 shows the error probability that we obtain in simulating users moving in MC2 model.

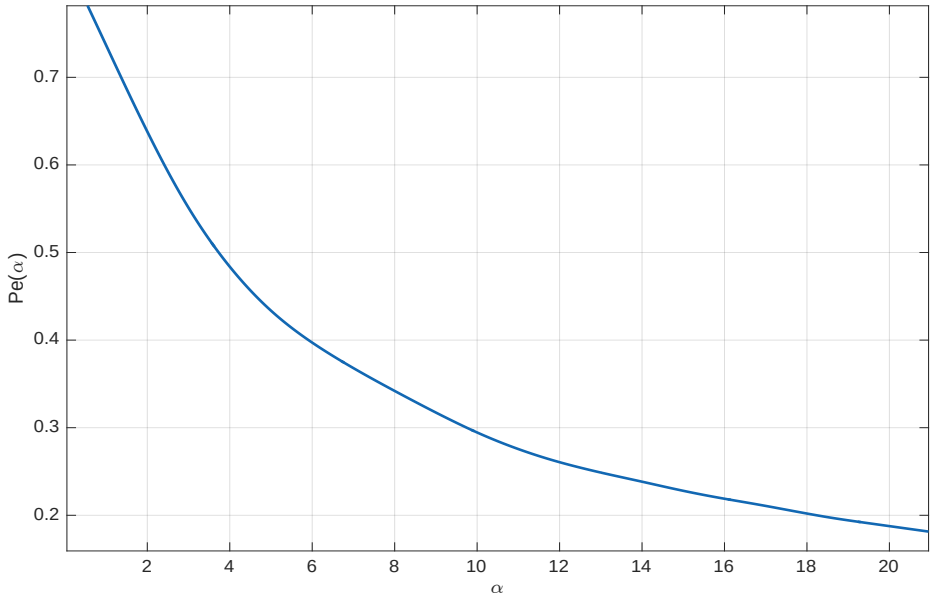

Figure 4.5: $Pe(\alpha)$ vs. α for Markov chain MC2 with $n = 500$.

Third, We model each user's path as a Markov chain $MC3$ shown in Figure 4.6. Since in this model $|E| = 6$ we can write $m = \alpha n^{\frac{2}{3}}$.

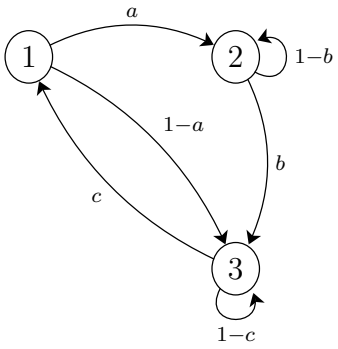

Figure 4.6: The Markov chain MC3 which models of users' path.

For n users, we create $(a, b, c) \in (0, 1)$ which are i.i.d and assumed them to be known to the adversary. We then generate m observations for each user. By calculating $\tilde{a} = \frac{m_{(1,2)}}{m_1}$ and $\tilde{b} = \frac{m_{(2,3)}}{m_2}$ and $\tilde{c} = \frac{m_{(3,2)}}{m_3}$ for all users and comparing $\tilde{a}, \tilde{c}$ for each path to all the users' probabilities a, b and c, we matched the closest (a, b, c) and $(\tilde{a}, \tilde{b}, \tilde{c})$ in $\mathbb{R}^3$. Figure 4.7 shows the error probability that we obtain in simulating users moving in MC3 model.

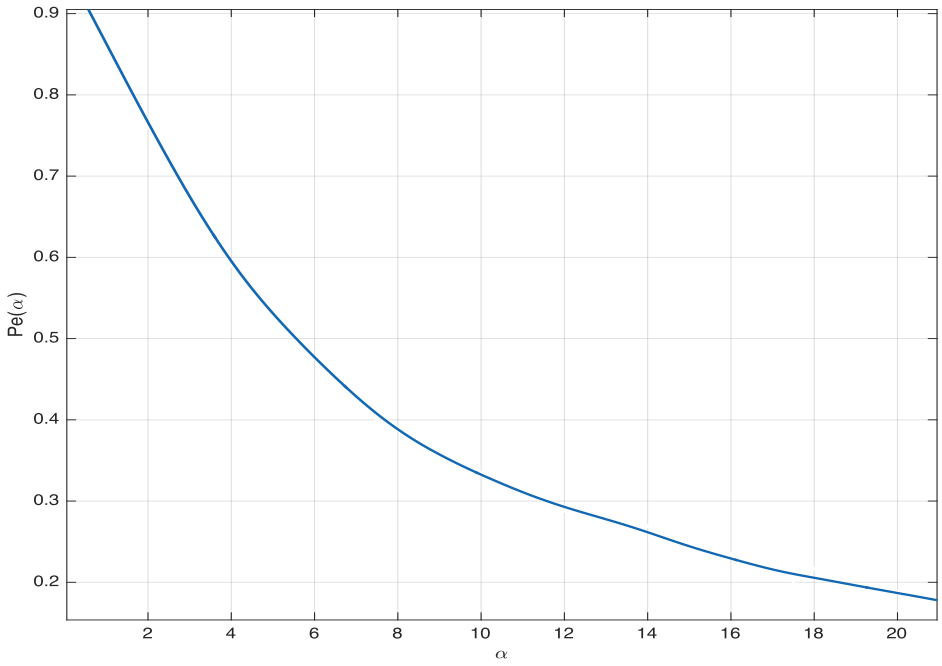

Figure 4.7: $Pe(\alpha)$ vs. α for Markov chain MC3 with $n = 500$.

For a fixed $\alpha = 5$, Figure 4.8 shows that as n increases, the error probability of the adversary converges to a fix positive value. We have repeated this for different values of α and have observed the same effect. This is consistent with our result that $m = O(n^{\frac{2}{|E|-r}})$ is the threshold for perfect privacy.

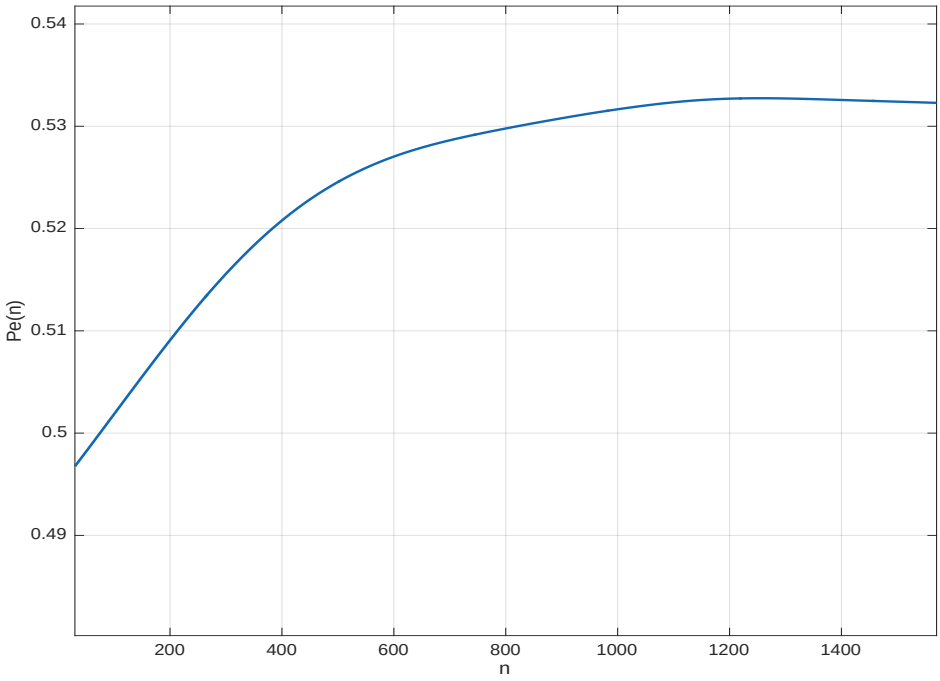

Figure 4.8: $P_e(n)$ vs. n for Markov chain MC3 with $\alpha = 5$.

Then, we model each user's path as a Markov chain $MC4$ shown in Figure 4.9. Since in this model $|E| = 7$ we can write $m = \alpha\sqrt{n}$.

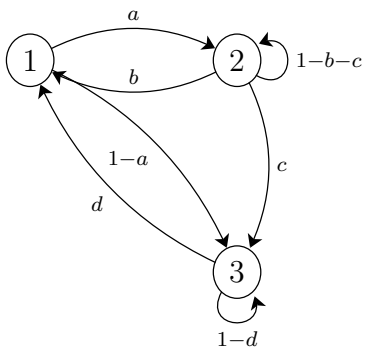

Figure 4.9: The Markov chain MC4 which models of users' path.

For n users, we create $(a, b, c, d) \in (0, 1)$ which are i.i.d and assumed them to be known to the adversary. We then generate m observations for each user. By calculating $\tilde{a} = \frac{m_{(1,2)}}{m_1}$ and $\tilde{b} = \frac{m_{(2,1)}}{m_2}$ and $\tilde{c} = \frac{m_{(2,3)}}{m_2}$ and $\tilde{d} = \frac{m_{(3,1)}}{m_3}$ for all users and comparing $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ for each path to all the users' probabilities a, b, c, d, we matched the closest (a, b, c, d) and $(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d})$ in $\mathbb{R}^4$. Figure 4.10 shows the error probability that we obtain in simulating users moving in MC3 model.

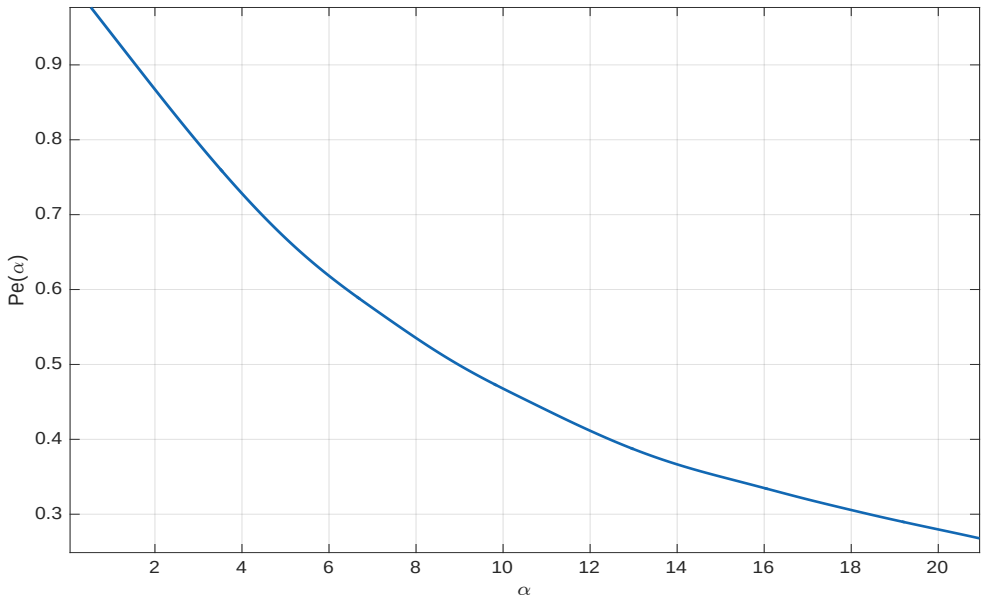

Figure 4.10: $Pe(\alpha)$ vs. α for Markov chain MC4 with $n = 500$.

Simulation results in Figures 4.3, 4.5, 4.7, 4.10 show that as α grows, the adversary's error probability goes to zero which shows that the adversary maps users with low error probability. On the other hand, as α becomes smaller, the error probability approaches 1. These results are consistent with the our main result that users have perfect privacy if the adversary obtains less than $O(n^{\frac{2}{|E|-r}})$ observations per user.

CHAPTER 5

CONCLUSION AND FUTURE WORK

We provided an information theoretic definition for perfect location privacy using the mutual information between users actual location and the anonymized observation that the adversary collects. Here, we assumed the strongest adversary who has all the statistical knowledge of the users' movements. The anonymization technique creates a pseudonym for each user at any time using a random permutation on the set of all users, $\{1, 2, \cdots, n\}$.

First, we model users movements independent from their previous locations. In this model, we have n users and r locations. We prove that if the number of observations that the adversary collects, m, is less than $O(n^{\frac{2}{r-1}})$, then users will have perfect location privacy. So, if the anonymization method changes the pseudonyms of the users before m observations made by the adversary for each user, then the adversary cannot distinguish between users.

We assumed that the location of a user is independent from his previous locations and also independent from other users' location. This assumption will fail immediately using this framework for real world data. Markov chain models are known to be more realistic models in terms of modeling users' movement rather than independent patterns. In Markov chain models, users' next location depends on the current location. Then, we extended our framework by using Markov chain model, a more realistic model, to model users' movements. By using the same notion of perfect location privacy we show the feasibility of achieving perfect location privacy using Markov chain.

Using Markov chains we prove that perfect location privacy is achievable if the pseudonym of the user is changed before $O(n^{\frac{2}{|E|-r}})$ observations is made by the adversary. If the anonymization

method changes the pseudonyms of the users before m observations made by the adversary for each user, then all the users have perfect location privacy.

Several issues may arise in such a framework. Achieving perfect location privacy is dependent on how unique the Markov chain of a user is. In the best case, all the users have a same Markov chain model of movements with similar transition probabilities. In this case, adversary cannot distinguish between them by observing even for large amount of time. On the other hand, some users may have a very unique Markov model for their movements in which case, the adversary is able to find the user with very limited number of observations. Users can be classified in to two groups: (1) users who have perfect location privacy if the number of observations collected by the adversary, m, is below some threshold, (2) users who will never achieve perfect location privacy when only anonymization is used. That is, a finite number of observations is enough to give the adversary a strictly positive probability of identifying a user correctly. The key to the above analysis seems to be in defining a uniqueness measure for the user's Markov chain. That is, users who have too unique transition graphs are insecure in terms of location privacy and other privacy protecting mechanisms have to get involved.

Extending this work using other location privacy protecting mechanisms can help users to protect their location information. Other LPPMs such as location obfuscation LPPMs allow users to report their location less precisely. Users are able to add noise or hide their location for certain amounts of time. By adding this method to our framework, users are able to both change their pseudonyms over time and also slightly change their location before reporting it. This may result to achieve more common Markov models for all the users' movements and since this change may decrease the uniqueness of users' Markov chain, they are more likely to achieve perfect location privacy.

BIBLIOGRAPHY

[1] Statistics and facts about Google. http://www.statista.com/topics/1001/google/.

[2] TensorFlow machine learning software. http://tensorflow.org/.

[3] "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/.

[4] 50,000 Uber driver names, license numbers exposed in a data breach, February 2015. http://arstechnica.com/business/2015/02/50000-uber-driver-names-license-plate-numbers-exposed-in-a-data-breach/.

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

[6] Andrés, Miguel E, Bordenabe, Nicolás E, Chatzikokolakis, Konstantinos, and Palamidessi, Catuscia. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (2013), ACM, pp. 901–914.

[7] Bamba, Bhuvan, Liu, Ling, Pesti, Peter, and Wang, Ting. Supporting anonymous location queries in mobile environments with privacygrid. In *Proceedings of the 17th international conference on World Wide Web* (2008), ACM, pp. 237–246.

[8] Beresford, Alastair R, and Stajano, Frank. Location privacy in pervasive computing. *IEEE Pervasive computing*, 1 (2003), 46–55.

[9] Bordenabe, Nicolás E, Chatzikokolakis, Konstantinos, and Palamidessi, Catuscia. Optimal geo-indistinguishable mechanisms for location privacy. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (2014), ACM, pp. 251–262.

[10] Cai, Y., and Xu, G. Cloaking with footprints to provide location privacy protection in location-based services, Jan. 1 2015. US Patent App. 14/472,462.

[11] Calmon, Flavio P, Makhdoumi, Ali, and Médard, Muriel. Fundamental limits of perfect privacy. In *Information Theory (ISIT), 2015 IEEE International Symposium on* (2015), IEEE, pp. 1796–1800.

[12] Chang, Shan, Qi, Yong, Zhu, Hongzi, Zhao, Jizhong, and Shen, Xuemin Sherman. Footprint: Detecting sybil attacks in urban vehicular networks. *Parallel and Distributed Systems, IEEE Transactions on 23*, 6 (2012), 1103–1114.

[13] Chatzikokolakis, Konstantinos, Andrés, Miguel E, Bordenabe, Nicolás Emilio, and Palamidessi, Catuscia. Broadening the scope of differential privacy using metrics. In *Privacy Enhancing Technologies* (2013), Springer, pp. 82–102.

[14] Chatzikokolakis, Konstantinos, Palamidessi, Catuscia, and Stronati, Marco. Geo-indistinguishability: A principled approach to location privacy. In *Distributed Computing and Internet Technology*. Springer, 2015, pp. 49–72.

[15] Chatzikokolakis, Konstantinos, Palamidessi, Catuscia, and Stronati, Marco. Location privacy via geo-indistinguishability. *ACM SIGLOG News 2*, 3 (2015), 46–69.

[16] Cheng, Reynold, Zhang, Yu, Bertino, Elisa, and Prabhakar, Sunil. Preserving user location privacy in mobile data management infrastructures. In *Privacy Enhancing Technologies* (2006), Springer, pp. 393–412.

[17] Chow, Chi-Yin, Mokbel, Mohamed F, and Liu, Xuan. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *GeoInformatica 15*, 2 (2011), 351–380.

[18] Chow, Richard, and Golle, Philippe. Faking contextual data for fun, profit, and privacy. In *Proceedings of the 8th ACM workshop on Privacy in the electronic society* (2009), ACM, pp. 105–108.

[19] Corser, George P, Fu, Huirong, and Banihani, Abdelnasser. Evaluating location privacy in vehicular communications and applications. *IEEE Transactions on Intelligent Transportation Systems 17*, 9 (2016), 2658–2667.

[20] Csiszár, Imre. Almost independence and secrecy capacity. *Problemy Peredachi Informatsii 32*, 1 (1996), 48–57.

[21] Dano, Mike. 55% of U.S. iOS users with Google Maps use it weekly, 2013. `http://www.fiercemobileit.com/story/55-us-ios-users-google-maps-use-it-weekly/2013-08-27`.

[22] Dewri, Rinku. Local differential perturbations: Location privacy under approximate knowledge attackers. *Mobile Computing, IEEE Transactions on 12*, 12 (2013), 2360–2372.

[23] Dewri, Rinku, and Thurimella, Ramakrisha. Exploiting service similarity for privacy in location-based search queries. *IEEE Transactions on Parallel and Distributed Systems 25*, 2 (2014), 374–383.

[24] Dritsas, Stelios, Gritzalis, Dimitris, and Lambrinoudakis, Costas. Protecting privacy and anonymity in pervasive computing: trends and perspectives. *Telematics and informatics 23*, 3 (2006), 196–210.

[25] Duckham, Matt, and Kulik, Lars. A formal model of obfuscation and negotiation for location privacy. In *Pervasive computing*. Springer, 2005, pp. 152–170.

[26] Duckham, Matt, Kulik, Lars, and Birtley, Athol. A spatiotemporal model of strategies and counter strategies for location privacy protection. In *Geographic Information Science*. Springer, 2006, pp. 47–64.

[27] Freudiger, Julien, Manshaei, Mohammad Hossein, Hubaux, Jean-Pierre, and Parkes, David C. Non-cooperative location privacy. *IEEE Transactions on Dependable and Secure Computing 10*, 2 (2013), 84–98.

[28] Freudiger, Julien, Raya, Maxim, Félegyházi, Márk, Papadimitratos, Panos, and Hubaux, Jean-Pierre. Mix-zones for location privacy in vehicular networks.

[29] Freudiger, Julien, Shokri, Reza, and Hubaux, Jean-Pierre. On the optimal placement of mix zones. In *Privacy enhancing technologies* (2009), Springer, pp. 216–234.

[30] Gedik, Buğra, and Liu, Ling. Location privacy in mobile systems: A personalized anonymization model. In *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on* (2005), IEEE, pp. 620–629.

[31] Gedik, Buğra, and Liu, Ling. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *Mobile Computing, IEEE Transactions on 7*, 1 (2008), 1–18.

[32] Ghinita, Gabriel, Kalnis, Panos, Khoshgozaran, Ali, Shahabi, Cyrus, and Tan, Kian-Lee. Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (2008), ACM, pp. 121–132.

[33] Gruteser, Marco, and Grunwald, Dirk. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services* (2003), ACM, pp. 31–42.

[34] Guynn, Jessica. Google open sources TensorFlow machine learning software, November 2015. `http://www.usatoday.com/story/tech/2015/11/09/google-open-sources-tensorflow-machine-learning-software/75456822/`.

[35] Ho, Shen-Shyang, and Ruan, Shuhua. Differential privacy for location pattern mining. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS* (2011), ACM, pp. 17–24.

[36] Hoh, Baik, and Gruteser, Marco. Protecting location privacy through path confusion. In *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on* (2005), IEEE, pp. 194–205.

[37] Hoh, Baik, Gruteser, Marco, Xiong, Hui, and Alrabady, Ansaf. Preserving privacy in gps traces via uncertainty-aware path cloaking. In *Proceedings of the 14th ACM conference on Computer and communications security* (2007), ACM, pp. 161–171.

[38] Kalnis, Panos, Ghinita, Gabriel, Mouratidis, Kyriakos, and Papadias, Dimitris. Preserving anonymity in location based services.

[39] Kalnis, Panos, Ghinita, Gabriel, Mouratidis, Kyriakos, and Papadias, Dimitris. Preventing location-based identity inference in anonymous spatial queries. *Knowledge and Data Engineering, IEEE Transactions on 19*, 12 (2007), 1719–1733.

[40] Khoshgozaran, Ali, and Shahabi, Cyrus. 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.

[41] Khoshgozaran, Ali, Shahabi, Cyrus, and Shirani-Mehr, Houtan. Location privacy: going beyond k-anonymity, cloaking and anonymizers. *Knowledge and Information Systems 26*, 3 (2011), 435–465.

[42] Kido, Hidetoshi, Yanagisawa, Yutaka, and Satoh, Tetsuji. An anonymous communication technique using dummies for location-based services. In *Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on* (2005), IEEE, pp. 88–97.

[43] Kido, Hidetoshi, Yanagisawa, Yutaka, and Satoh, Tetsuji. Protection of location privacy using dummies for location-based services. In *Data Engineering Workshops, 2005. 21st International Conference on* (2005), IEEE, pp. 1248–1248.

[44] Krumm, John. A survey of computational location privacy. *Personal and Ubiquitous Computing 13*, 6 (2009), 391–399.

[45] Lee, Jaewoo, and Clifton, Chris. Differential identifiability. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), ACM, pp. 1041–1049.

[46] Li, Tiancheng, and Li, Ninghui. 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* (2009), ACM, pp. 517–526.

[47] Liu, Xinxin, Liu, Kaikai, Guo, Linke, Li, Xiaolin, and Fang, Yuguang. A game-theoretic approach for achieving k-anonymity in location based services. In *INFOCOM, 2013 Proceedings IEEE* (2013), IEEE, pp. 2985–2993.

[48] Lu, Hua, Jensen, Christian S, and Yiu, Man Lung. Pad: privacy-area aware, dummy-based location privacy in mobile services. In *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access* (2008), ACM, pp. 16–23.

[49] Ma, Zhendong, Kargl, Frank, and Weber, Michael. A location privacy metric for v2x communication systems. In *Sarnoff Symposium, 2009. SARNOFF'09. IEEE* (2009), IEEE, pp. 1–6.

[50] Machanavajjhala, Ashwin, Kifer, Daniel, Abowd, John, Gehrke, Johannes, and Vilhuber, Lars. Privacy: Theory meets practice on the map. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on* (2008), IEEE, pp. 277–286.

[51] Malandrino, Francesco, Borgiattino, Carlo, Casetti, Claudio, Chiasserini, Carla-Fabiana, Fiore, Marco, and Sadao, Roberto. Verification and inference of positions in vehicular networks through anonymous beaconing. *IEEE Transactions on Mobile Computing 13*, 10 (2014), 2415–2428.

[52] Manshaei, Mohammad Hossein, Zhu, Quanyan, Alpcan, Tansu, Bacşar, Tamer, and Hubaux, Jean-Pierre. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR) 45*, 3 (2013), 25.

[53] Mokbel, Mohamed F, Chow, Chi-Yin, and Aref, Walid G. The new casper: query processing for location services without compromising privacy. In *Proceedings of the 32nd international conference on Very large data bases* (2006), VLDB Endowment, pp. 763–774.

[54] Nguyen, Hiep H, Kim, Jong, and Kim, Yoonho. Differential privacy in practice. *Journal of Computing Science and Engineering 7*, 3 (2013), 177–186.

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

[56] Paulet, Russell, Kaosar, Md Golam, Yi, Xun, and Bertino, Elisa. Privacy-preserving and content-protecting location based queries. *Knowledge and Data Engineering, IEEE Transactions on 26*, 5 (2014), 1200–1210.

[57] Salamatian, Salman, Zhang, Amy, du Pin Calmon, Flavio, Bhamidipati, Sandilya, Fawaz, Nadia, Kveton, Branislav, Oliveira, Pedro, and Taft, Nina. 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.

[58] Sankar, Lalitha, Rajagopalan, S Raj, and Poor, H Vincent. Utility-privacy tradeoffs in databases: An information-theoretic approach. *Information Forensics and Security, IEEE Transactions on 8*, 6 (2013), 838–852.

[59] Shankar, Pravin, Ganapathy, Vinod, and Iftode, Liviu. Privately querying location-based services with sybilquery. In *Proceedings of the 11th international conference on Ubiquitous computing* (2009), ACM, pp. 31–40.

[60] Shokri, Reza. *Quantifying and protecting location privacy*. PhD thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2013.

[61] Shokri, Reza. Optimal user-centric data obfuscation. *arXiv preprint arXiv:1402.3426* (2014).

[62] Shokri, Reza, Theodorakopoulos, George, Danezis, George, Hubaux, Jean-Pierre, and Le Boudec, Jean-Yves. Quantifying location privacy: the case of sporadic location exposure. In *Privacy Enhancing Technologies* (2011), Springer, pp. 57–76.

[63] Shokri, Reza, Theodorakopoulos, George, Le Boudec, Jean-Yves, and Hubaux, Jean-Pierre. Quantifying location privacy. In *Security and Privacy (SP), 2011 IEEE Symposium on* (2011), IEEE, pp. 247–262.

[64] Shokri, Reza, Theodorakopoulos, George, Papadimitratos, Panos, Kazemi, Ehsan, and Hubaux, Jean-Pierre. Hiding in the mobile crowd: Locationprivacy through collaboration. *IEEE transactions on dependable and secure computing 11*, 3 (2014), 266–279.

[65] Shokri, Reza, Theodorakopoulos, George, Troncoso, Carmela, Hubaux, Jean-Pierre, and Le Boudec, Jean-Yves. Protecting location privacy: optimal strategy against localization attacks. In *Proceedings of the 2012 ACM conference on Computer and communications security* (2012), ACM, pp. 617–627.

[66] Smith, Craig. By The Numbers 24 Amazing Uber Statistics, September 2015. `http://expandedramblings.com/index.php/uber-statistics/`.

[67] Smith, Craig. By The Numbers: 45 Amazing Yelp Statistics, May 2015. `http://expandedramblings.com/index.php/yelp-statistics/`.

[68] Sweeney, Latanya. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10*, 05 (2002), 571–588.

[69] Sweeney, Latanya. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10*, 05 (2002), 557–570.

[70] Timberg, Craig. 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/`.

[71] Um, Jung-Ho, Kim, Hee-Dae, and Chang, Jae-Woo. An advanced cloaking algorithm using hilbert curves for anonymous location based service. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on* (2010), IEEE, pp. 1093–1098.

[72] Wang, Wei, and Zhang, Qian. Privacy-preserving collaborative spectrum sensing with multiple service providers. *IEEE Transactions on Wireless Communications 14*, 2 (2015), 1011–1019.

[73] Wernke, Marius, Skvortsov, Pavel, Dürr, Frank, and Rothermel, Kurt. A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing 18*, 1 (2014), 163–175.

[74] Xue, Mingqiang, Kalnis, Panos, and Pung, Hung Keng. Location diversity: Enhanced privacy protection in location based services. In *Location and Context Awareness*. Springer, 2009, pp. 70–87.

[75] Yamamoto, Hirosuke. A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.). *IEEE Transactions on Information Theory 29*, 6 (1983), 918–923.

[76] Ying, Bidi, Makrakis, Dimitrios, and Mouftah, Hussein T. Dynamic mix-zone for location privacy in vehicular networks. *Communications Letters, IEEE 17*, 8 (2013), 1524–1527.

[77] Zhang, Yuan, Tong, Wei, and Zhong, Sheng. On designing satisfaction-ratio-aware truthful incentive mechanisms for¡ inline-formula¿¡ tex-math notation=. *IEEE Transactions on Information Forensics and Security 11*, 11 (2016), 2528–2541.

[78] Zhangwei, Huang, and Mingjun, Xin. A distributed spatial cloaking protocol for location privacy. In *Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference on* (2010), vol. 2, IEEE, pp. 468–471.

[79] Zhong, Ge, and Hengartner, Urs. A distributed k-anonymity protocol for location privacy. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on* (2009), IEEE, pp. 1–10.

[80] Zhou, Tong, Choudhury, Romit Roy, Ning, Peng, and Chakrabarty, Krishnendu. 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* (2007), IEEE, pp. 1–8.

[81] Zurbaran, Mayra Alejandra, Avila, Karen, Wightman, Pedro, and Fernandez, Michael. Nearrand: Noise-based location obfuscation based on random neighboring points. *IEEE Latin America Transactions 13*, 11 (2015), 3661–3667.