P^3-LOC: A Privacy-Preserving Paradigm-Driven Framework for Indoor Localization

Ping Zhao, Hongbo Jiang, *Senior Member, IEEE*, John C. S. Lui, *Fellow, IEEE, ACM*, Chen Wang, *Member, IEEE*, Fanzi Zeng, Fu Xiao, and Zhetao Li

***Abstract*— Indoor localization plays an important role as the basis for a variety of mobile applications, such as navigating, tracking, and monitoring in indoor environments. However, many such systems cause potential privacy leakage in data transmission between mobile users and the localization server (LS). Unfortunately, there has been little research done on privacy issue, and the existing privacy-preserving solutions are *algorithm-driven*, each designed for specific localization algorithms, which hinders their wide-scale adoption. Furthermore, they mainly focus on users' location privacy, while the LS's data privacy cannot be guaranteed. In this paper, we propose a Privacy-Preserving Paradigm-driven framework for indoor LOCalization (P^3-LOC). P^3-LOC takes the advantage that most indoor localization systems share a common two-stage localization paradigm: information measurement and location estimation. Based on this, P^3-LOC carefully perturbs and cloaks the transmitted data in these two stages and employs specially designed "k-anonymity" and "differential privacy" techniques to achieve the provable privacy preservation. The key advantage is that P^3-LOC does not rely on any prior knowledge of the underlying localization algorithms, and it guarantees both users' location privacy and the LS's data privacy. Our extensive experiments from the measured data have validated that P^3-LOC provides privacy preservation for general indoor localization techniques. In addition, P^3-LOC is comparable with the state-of-the-art algorithm-driven techniques in terms of localization error, computation, and communication overhead.**

Manuscript received January 4, 2018; revised April 17, 2018, June 24, 2018, and August 26, 2018; accepted October 29, 2018; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor X.-Y. Li. Date of publication November 21, 2018; date of current version December 14, 2018. This work was supported in part by the National Natural Science Foundation of China under under Grant 61872416, Grant 61572219, Grant 61872416, Grant 61502192, Grant 61732017, Grant 61671216, Grant 61471408, Grant 51479159, Grant 91538203, and Grant 41701479, in part by the Initial Research Funds for Young Teachers of Donghua University, and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Hongbo Jiang.)*

P. Zhao is with the College of Information Science and Technology, Donghua University, Shanghai 201620, China (e-mail: pingzhao2014ph@gmail.com).

H. Jiang and F. Zeng are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: hongbojiang2004@gmail.com; zengfanzi@hnu.edu.cn).

J. C. S. Lui is with the Computer Science and Engineering Department, The Chinese University of Hong Kong, Hong Kong (e-mail: cslui@cse.cuhk.edu.hk).

C. Wang is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: cwangwhu@gmail.com).

F. Xiao is with the College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210042 China, and also with the Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing 210042 China (e-mail: xiaof@njupt.edu.cn).

Z. Li is with the College of Information Engineering, Xiangtan University, Xiangtan 411105, China (e-mail: liztchina@hotmail.com).

Digital Object Identifier 10.1109/TNET.2018.2879967

***Index Terms*— Privacy, indoor localization, k-anonymity, differential privacy, paradigm-driven.**

I. INTRODUCTION

AS AN enabling technology for the future Internet of Things, indoor localization has a variety of applications in indoor environments, such as navigating, tracking, and monitoring. As such it has attracted growing commercial and academic interest [1]–[3]. Unfortunately, when this services are offered, either users' locations may be exposed to the untrusted localization server (LS), or a trusted LS's localization-related information could be breached by malicious users [4]–[7].

A. Privacy Concerns in Indoor Localization

Existing indoor localization systems largely fall into fingerprint-based [8], [9], model-based [10]–[12], and dead-reckoning-based indoor localization [13], [14]. In fact, each of them has security concerns with respect to users' location privacy and LS's data privacy [15], [16].

Fingerprint-based indoor localization estimates users' locations by mapping users' measured signal against the pre-built fingerprint database. Model-based indoor localization calculates users' locations based on geometrical models that characterize the relationship between signal transmitters and receivers. Dead-reckoning-based indoor localization utilizes inertial sensors to estimate the position change since the last update. Overall, to provide localization services, in these three kinds of indoor localization techniques, the LS needs to provide users localization-related information, e.g., fingerprint database, radio transmission parameters, radio map, and floor plan, etc. (cf. Steps (1) and (3) in Fig. 1). Likewise, to be localized, users are supposed to send LS the measured signal (hereafter geo-information), e.g., the Wi-Fi received signal strength (RSS), time of arrival (ToA), angular velocity, etc (cf. Steps (2) and (4) in Fig. 1). However, the LS may be untrusted and disclose users' locations deduced from the geo-information to attackers or advertisers, and therefore users' *location privacy* is breached. Similarly, users may be malicious, tampering the estimated locations of other users using the localization-related information, or disclosing the localization-related information to attackers. As a result, the LS's *data privacy* is disclosed. In summary, location privacy and data privacy could be disclosed in the above three kinds of indoor localization techniques.

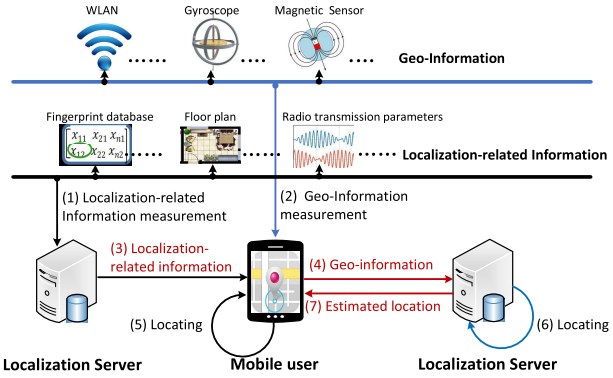


Fig. 1. A two-stage indoor localization paradigm. Stage 1 (information measurement) includes Steps (1) and (2); Stage 2 (location estimation) includes Steps (3)-(5), or (4)-(6)-(7). In Stage 2, (4)-(6)-(7), geo-information is provided to the LS; in Stage 2, (3)-(5), the localization-related information is transmitted to users.

To make matters worse, users and the LS are vulnerable to serious attacks [4]–[6], [17], in the event of location/data privacy disclosure. Specifically, when the location privacy of users is breached by an untrusted LS, sensitive personal information such as one's life style, social relationship, and political beliefs, etc., can be easily revealed, thereby exposing users to spams, or even blackmails and physical violence [18], etc. Likewise, when the database of the LS (e.g., fingerprint, radio map, and floor plan, etc.) is breached by malicious users, malicious users can (i) infer the estimated locations of other users, incurring location privacy risks, and then breach users' habits, relationships, and interests, etc., leading to more serious privacy leakage [19]; (ii) attack the localization infrastructure, e.g., wireless routers, sensor nodes, etc., according to the locations of infrastructure recorded in the database [20], [21].

B. Existing Work

There are only few studies proposed to address the privacy concerns for indoor localization. Li *et al.* conducted a pioneering work, where users encrypted the measured RSS using homomorphic encryption, and the LS randomly chose several APs to locate users. In [17], a similar privacy-preserving technique via homomorphic encryption and fuzzy logic was proposed. Similarly, Armengo *et al.* [6] considered an indoor environment where privacy can be preserved using an encryption algorithm. Higuchi *et al.* [5] focused on extending the existing crowd-tracking system to preserve privacy.

The problems of existing studies are two-fold. First, they are algorithm-driven: solutions are designed for a specific indoor localization algorithm, and thus cannot be applied to others. Specifically, [6], [17], and [22] are designed for a specific RSS or CSI fingerprint-based localization algorithm which searches for k closest matches. Therefore they cannot be applied to model-based [10]–[12], dead-reckoning-based techniques [13], [14], or other fingerprint-based techniques [8], [9]. Furthermore, the work in [5] heavily relies on special hardware (crow tracking system), inevitably resulting in the limited applicability to other localization algorithms.

Second, existing efforts [5], [6], [17], [22] mainly focus on users' location privacy, while the localization server (LS)'s data privacy cannot be guaranteed. Specifically, the fingerprint locations were disclosed and thus the fingerprint data can be disclosed in [6] and [22]. The study in [5] ignored the important issue of LS's data privacy. It is noted that while Wang *et al.* [17] claimed that the data privacy threat on the LS can be addressed in common cases, malicious users could disclose LS's fingerprint database via purpose-designed operations. Attackers, by sending localization queries at both the reference locations and users' locations at the same time, purposely construct dependent equations whose solutions are exactly the fingerprint data.

C. Our Approach

To address the above-mentioned two problems, we propose P^3-LOC, a Privacy-Preserving *Paradigm-driven* framework for indoor LOCalization. Our key observation is that most if not all of indoor localization systems share a two-stage localization paradigm [23]: Stage 1, information measurement (i.e., Steps (1), (2) in Fig. 1), and Stage 2, location estimation (i.e., Steps (3), (4), (5), (6), and (7) in Fig. 1). To concrete, the localization-related information (resp. geo-information) measured in Stage 1 needs to be transmitted to users (resp. the LS) in Stage 2, to perform the localization process. Therefore, we investigate a general framework for indoor localization where the data is transmitted in an indistinguishable manner.

On this basis, P^3-LOC perturbs and cloaks the data which is transmitted in Stage 1 and Stage 2, i.e., Steps (3), (4), and (7) in Fig. 1, without reliance on any prior knowledge of the underlying localization algorithms. Specifically, to preserve users' location privacy against LS in Step (4), P^3-LOC first segments the geo-information into pieces, and then utilizes specially designed "k-anonymity" and "differential privacy" techniques to perturb and cloak the pieces. The perturbed and cloaked pieces, i.e., the inputs of location estimation, can prevent users from getting access to the localization-related information in Step (3) and other users' estimated locations in Step (7). By doing so, LS's data privacy is preserved against users, and an individual user's location privacy is preserved against the LS and other users.

Contributions: To the best of our knowledge, P^3-LOC is the first work with provably guaranteeing both users' location privacy and the LS's data privacy, and meanwhile can be applied to any localization system that complies with the two-stage localization paradigm. P^3-LOC also offers several salient features. First, it is largely distributed, incurring negligible overhead to the server. Each user locally segments, perturbs, and cloaks his own geo-information, and identifies his estimated location (i.e., the localization outcome). Second, it is user-friendly, as users can predefine their desired privacy requirements. Lastly, it is scalable, only requiring a small amount of computation and communication overhead. Our extensive experiments from measured data have demonstrated the effectiveness and efficiency of P^3-LOC compared to other state-of-the-art solutions.

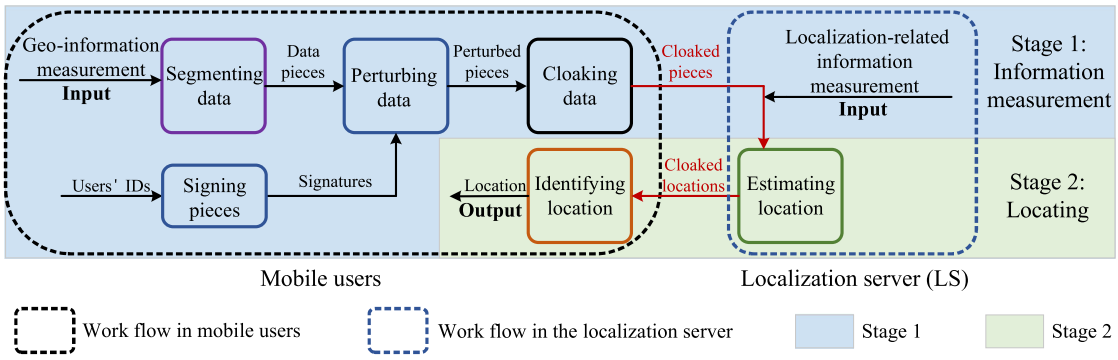


Fig. 2. The work flow of P^3-LOC.

The remainder of this paper is organized as follows. Section II presents some preliminary knowledge of our approach. Section III describes our solution in detail, followed by the performance evaluation in Section IV. Finally, Section V concludes the paper.

II. Preliminary

A. Attacker Model

In this paper, we consider the potential privacy disclosure at two kinds of entities: users and the LS.

(a) **LS's attacks**. Like [17], [22], and [24], we consider the LS is semi-honest. That is, the LS will strictly execute the indoor localization algorithm, i.e., honestly estimate and return users' locations. But it may attempt to disclose users' location privacy in the localization process. Specifically, in the building block (cf. Fig. 2), Estimating location, the LS tries to identify each user's geo-information and then takes one step further getting the user's estimated location with the help of the localization algorithm.

(b) **Users' attacks**. Like [17], [22], and [25], we consider the users are curious-but-honest. A specific user may intend to disclose other users' locations in the building blocks (cf. Fig. 2), Cloaking data and Identifying location, via identifying other users' geo-information and estimated locations respectively. Furthermore, it also tries to get access to the LS's localization-related information in the building block, Identifying location. However, it will honestly perform localization operations (i.e., honestly send their geo-information).

Note that users and the LS may send tampered geo-information and tampered locations respectively. But such a case does not exist in the scenario we considered, where users and the LS honestly send exact geo-information and estimated locations respectively. Moreover, attacks in such a case are always named location injection attacks or spoofing attacks, etc. The corresponding privacy preserving algorithms against these attacks is another topic. What's more, a larger number of existing studies, e.g., [7], [26], [27], have designed effective algorithms to resist these attacks. In addition, it is out of the scope of this paper to address how to encourage users and LS to honestly execute algorithms. We refer interested users to [25] and [29] for additional information.

B. Design Goals

P^3-LOC is designed to preserve users' location privacy and the LS's data privacy in indoor localization systems as long as they comply with the two-stage paradigm. In particular, to preserve location privacy, each user's geo-information and localization outcome should be only identified by himself, and be indistinguishable by other users as well as the LS. To preserve data privacy, the localization-related information should be only available to the LS.

C. k-Anonymity

k-Anonymity is first proposed in [29] to protect the information leakage in data release. The definition is as follows:

Definition 1: The "*attributes*" of table $T(A_1, \ldots, A_n)$ are $\{A_1, \ldots, A_n\}$, and a *quasi-identifier* of the table denoted by Q_T is a set of sensitive attributes $\{A_i, \ldots, A_j\} \subseteq (A_1, \ldots, A_n)$ [29].

Note that sensitive attributes are recognized by data holder and regarded as the quasi-identifier.

Definition 2: When each sequence of values in $T[Q_T]$ has no less than k duplicates in $T[Q_T]$, $T(A_1, \ldots, A_n)$ meets "*k-anonymity*". The duplicates of each sequence of values in $T[Q_T]$ constitute a "*cloaking set*"(CS), and each of the duplicates is cloaked in this CS [29].

For example, the attributes in Table I are {Gender, ZIP, Problem}, the quasi-identifier recognized by the data holder is Q_T ={Gender, ZIP}, and each sequence of values in $T[Q_T]$ (e.g., T[Gender="F"], T[ZIP="0214$*$"]) appears with $k = 2$ occurrences. Therefore u_1 and u_3 cannot be distinguished. That is, u_1 and u_3's problems are cloaked in a CS ($k = 2$).

However, k-anonymity cannot be directly applied for P^3-LOC, as a specific user's geo-information is disclosed to other users who are cloaked in the same CS. Therefore, we propose to segment the geo-information into pieces and cloak every piece so that each of the other users can only obtain one piece of the specific user's geo-information (cf. Sections III-B, III-C).

D. ε-Differential Privacy

The formal definition of ε-differential privacy is as follows:

TABLE I
AN EXAMPLE OF k-ANONYMITY; Q_T ={GENDER, ZIP}, $k = 2$

	Gender	ZIP	Problem
u_1	M	0214*	Short breath
u_2	F	0218*	Chest pain
u_3	M	0214*	Hypertension
u_4	F	0218*	Obesity

Definition 3: A randomized algorithm Alg gives ε-*differential* privacy if for any dataset D, any tuple $t \in D$, and any $S \in Range(Alg)$ [30], we have

$$e^{-\varepsilon} \leq \frac{Pr[Alg(D) = S]}{Pr[Alg(D_{-t}) = S]} \leq e^{\varepsilon}, \tag{1}$$

where $Range(Alg)$ is the output range of Alg; D_{-t} is the dataset where the tuple t is removed from D; and $\varepsilon \in (0, 1)$.

ε-Differential privacy can prevent the information disclosure caused by the absence of any *single* tuple in D. However, it cannot be directly applied to P^3-LOC, as we aim to prevent information disclosure caused by the absence of any λ tuples in D, where λ is the number of missing tuples in D, and $\lambda \geq 1$. Thus, we propose the (λ, ε)-differential privacy (cf. Definition 5) to prevent information disclosure caused by any λ tuples.

ε-Differential privacy is composable as stated below.

Theorem 4 (Sequential Composition [30]): Assume each of mechanisms $\mathcal{M}_i$, $i = (1, \ldots, r)$, provides ε_i-differential privacy. $\mathcal{M}$ performs $\mathcal{M}_i$, $i = (1, \ldots, r)$ with independent randomness. Then $\mathcal{M}$ satisfies $\sum_{i=1}^{r} \varepsilon_i$-differential privacy.

On the basis of Theorem 4, we can obtain the composition property of (λ, ε)-differential privacy (cf. Theorem 7). By incorporating this composition property, we then propose a novel data perturbation mechanism to achieve the (λ, ε)-differential privacy (cf. Section III-C).

III. DESIGN OF P^3-LOC

A. Overview

P^3-LOC includes six building blocks, as shown in Fig. 2.

(1) *Segmenting data.* Each user segments his geo-information measurements into data pieces.

(2) *Perturbing data.* A data perturbation mechanism based on the ε-differential privacy technique is conducted, in order to inject noise into each piece.

(3) *Signing Pieces.* Each user labels his pieces that will be exchanged with other users' data, through adapting the cryptographic hash function.

(4) *Cloaking data.* P^3-LOC performs a distributed data transfer strategy where every user can cloak each of his pieces into a cloaking set (CS). Then the user sends cloaked pieces to the LS.

(5) *Estimating location.* Upon receiving the pieces from users, the LS conducts the location estimation and returns the data tuples containing the cloaked locations.

(6) *Identifying location.* Since each user's geo-information is available to himself, his real location can be identified when he receives the data tuples from the LS.

B. Segmenting Data

Assume there are N_U users u_1, u_2,..., u_{N_U}, and each user u_i ($i \in \{1, \ldots, N_U\}$) has a geo-information x_i and privacy parameters (k_i, λ) predefined by himself. x_i is the RSS or CSI (channel signal information), etc., in e.g., WiFi fingerprint-based indoor localization. Moreover, in model-based indoor localization, x_i may be ToA, TDoA (time difference of arrival), etc. Furthermore, in dead-reckoning-based indoor localization, x_i may be the angular velocity, acceleration, azimuth angle, etc. The privacy parameter k_i means that his geo-information (resp. his estimated location) should be cloaked in a CS containing no less than $(k_i - 1)$ other users' geo-information (resp. estimated locations). Parameter λ is the length of event sequence to be defined below.

To protect users' geo-information against the LS, a straightforward method is to cloak every user u_i's geo-information into a CS with no less than $(k_i - 1)$ other users' geo-information using k-anonymity techniques, so that the LS cannot distinguish u_i's geo-information from others cloaked in the CS. Unfortunately, this simple method cannot preserve a specific user's location privacy against other users whose geo-information is cloaked in the same CS with the user, as the user has to share geo-information with others to complete the k-anonymity operation.

In contrast, in P^3-LOC, each user (say, the user u_i) first breaks his owned geo-information x_i into κ_i (κ_i is set to be k_i) random pieces $x_{i\gamma}$ ($\gamma \in \{1, \ldots, \kappa_i\}$) such that $x_i = \sum_{\gamma=1}^{\kappa_i} x_{i\gamma}$. To concrete, we assign random to pieces $x_{i\gamma}$ ($\gamma \in \{1, \ldots, \kappa_i - 1\}$), and let $x_{i\kappa_i} = x_i - \sum_{j=1}^{\kappa_i - 1} x_{ij}$. Then, we cloak each of these pieces with the pieces from neighbors. That is, u_i's geo-information x_i is cloaked with $(\kappa_i - 1)$ geo-information from other users. As a result, each user in same CS with u_i has the knowledge of only one of u_i' pieces, and the LS cannot distinguish x_i from no less than $(\kappa_i - 1)$ geo-information from other users. It is important to note that segmenting x_i into more pieces is still workable, but at the expense of increasing the computation and communication cost.

After that, the geo-information blocks of all users are

$$\begin{bmatrix} X_1 & X_2 & \cdots & X_i & \cdots & X_{N_U} \\ O_1 & O_2 & \cdots & O_i & \cdots & O_{N_U} \end{bmatrix} \tag{2}$$

where $X_i = [x_{i1}, x_{i2}, \cdots, x_{i\kappa_i}]^T$, $O_i = [0, 0, \cdots, 0]^T_{\kappa_i' \times 1}$, and $\kappa_i' = \max\{\kappa_1, \kappa_2, \cdots, \kappa_{N_U}\} - \kappa_i$.

It is noted that for the purpose of locally cloaking data (the details can be found in Section III-E), similar to [31], P^3-LOC demands that nearby users can directly communicate with each other (say via Bluetooth or WiFi interface). As a result, the neighbors whose data pieces cloaked with the user u_i may collude with each other to infer the user u_i's geo-information, and then further breach other personal information, resulting in serious privacy risks. To address such privacy concern, before locally cloaking data, we propose a data perturbation mechanism based on differential privacy in the next section.

C. Perturbing Data

The goal of the data perturbation mechanism is to inject noise into pieces $x_{i\gamma}$ so that neighbors whose data pieces

cloaked with the user u_i cannot infer the geo-information of user u_i when no more than λ neighbors collude with each other.

Motivated by the properties of ε-differential privacy (cf. Definition 3) that it can prevent the information disclosure caused by the absence of any single tuple in D, we try to propose a relaxed version of ε-differential privacy, (λ, ε)-differential privacy. (λ, ε)-differential privacy can prevent information disclosure caused by the absence of no more than λ tuples in D. That is, when less than λ malicious neighbors collude with each other, they cannot infer the user u_i's geo-information. Note that when the number of colluded users increases to be larger than λ, P^3-LOC cannot protect users' location privacy against the above collusion attacks.

Inherited from ε-differential privacy, a formally defined (λ, ε)-differential privacy and the event sequence are given by

Definition 5: A randomized algorithm $\mathcal{M}$ satisfies (λ, ε)-differential privacy if for any input dataset S_n, any tuple $D_i \in S_n$, and any $R \in Range(\mathcal{M})$, it holds that:

$$e^{-\varepsilon} \leq \frac{Pr[\mathcal{M}(S_n) = R]}{Pr[\mathcal{M}(S'_n) = R)]} \leq e^{\varepsilon}, \tag{3}$$

where $S_n = (D_1, D_2, \ldots, D_n)$, $S_n[i] = D_i$, D_i is a dataset; for each $S_n[i_1]$, $S_n[i_2]$, $S'_n[i_1]$, $S'_n[i_2]$, $i_1 < i_2$, $S_n[i_1] \neq S'_n[i_1]$, and $S_n[i_2] \neq S'_n[i_2]$, it holds $i_2 - i_1 + 1 \leq \lambda$; $S'_n[i_1]$ is obtained by removing or adding a row in $S_n[i_1]$; $\mathcal{M}(S_n[i]) = R[i]$.

Definition 6: An event sequence of length λ' is $\{S'_n[i_1], \ldots, S'_n[i_{\lambda'}]\}$, where $S_n[i_1] \neq S'_n[i_1]$, $S_n[i_2] \neq S'_n[i_2]$, $\ldots$, $S_n[i_{\lambda'}] \neq S'_n[i_{\lambda'}]$.

According to Definitions 5 and 6, (λ, ε)-differential privacy can prevent information disclosure caused by an event sequence of length λ' ($\lambda' \leq \lambda$). That is, (λ, ε)-differential privacy can preserve the user's location privacy when less than λ malicious users collude with each other.

Combing Theorem 4 with Definition 5, we have:

Theorem 7: The randomized algorithm $\mathcal{M}$ can be decomposed into n algorithms $\mathcal{M}_1$, $\mathcal{M}_2$, $\ldots$, $\mathcal{M}_n$, and each $\mathcal{M}_i$ generates independent randomness. Then $\mathcal{M}$ meets (λ, ε)-differential privacy, if

$$\begin{cases} \mathcal{M}_i(D[i]) = R[i], \\ e^{-\varepsilon} \leq \dfrac{Pr[\mathcal{M}_i(D_i) = R[i]]}{Pr[\mathcal{M}_i(D'_i) = R[i]]} \leq e^{\varepsilon} \\ \sum_{\gamma=i}^{i+\lambda-1} \varepsilon_\gamma \leq \varepsilon, \quad i \in (1, \ldots, n) \end{cases}$$

Proof: See Appendix A. ∎

Theorem 7 implies that the privacy budget ε can be viewed as the sum of the privacy budget in every sub-sequence of length λ in S_n (e.g., $S_n[1], \ldots, S_n[\lambda]$). Namely, any sub-sequence of length λ do meet the ε-differential privacy. Accordingly, we propose the following data perturbation mechanism $\mathcal{M}$ to inject noise into pieces as follows.

Denote the set of user u_i's data pieces as $\mathcal{D}_i = \{D_{i1}, D_{i2}, \ldots, D_{i1}, \ldots, D_{i\kappa_i}\}$, where $D_{il} = [x_{il}]_{d\times 1}$. The inputs of the data perturbation mechanism $\mathcal{M}$ are users' pieces $\mathcal{D}_i$, and $\mathcal{M}$ can be decomposed into mechanisms $\mathcal{M}_1$, $\ldots$, $\mathcal{M}_l$, $\ldots$, $\mathcal{M}_{\kappa_i}$ such that $\mathcal{M}_l(D_{i1}) = \mathbf{o}_{il} \in Range(\mathcal{M})$. Each mechanism $\mathcal{M}_l$ generates independent randomness and meets ε_l-differential privacy.

Algorithm 1 DATA PERTURBATION MECHANISM $\mathcal{M}$

Require: D_{il}, $\mathbf{o}_{i1}$, $\mathbf{o}_{i2}$, $\ldots$, $\mathbf{o}_{i(l-1)}$, ε_1,ε_2, $\ldots$, and ε_{l-1}
Ensure: $\mathbf{o}_{il}$, ε_l
1: // Sub mechanism $\mathcal{M}_{l1}$
2: Compute $\mathbf{c}_{il} = Q(D_{il})$
3: Set $\omega_{l,1} = 2\lambda x_i/(\varepsilon d)$, $dis = 1/d \sum_{j=1}^{d} \mid \mathbf{o}_{i(l-1)}[j] - \mathbf{c}_{il}[j] \mid + Lap(\omega_{l,1})$
4: $\varepsilon_{l,1} = \varepsilon/(2\lambda)$
5: // Sub mechanism $\mathcal{M}_{l2}$
6: Set remaining privacy budget $\varepsilon_{l,2} = \varepsilon/2 - \sum_{\tau=l-\lambda+1}^{l-1} \varepsilon_{\tau,2}$
7: Set $\omega_{l,2} = 2x_i/\varepsilon_{l,2}$
8: **if** $dis > \omega_{l,2}$ **then** $\mathbf{o}_{il} = \mathbf{c}_{il} + \langle Lap(\omega_{l,2}) \rangle^d$
9: **else** $\mathbf{o}_{il} = \mathbf{o}_{i(l-1)}$
10: $\varepsilon_l = \varepsilon_{l,1} + \varepsilon_{l,2}$

Next, we need to determine the output $\mathbf{o}_{il}$ and privacy budget ε_l. First of all, we define the following symbols. Let an statistic function be Q: $D_{il} \longrightarrow R^d$; $Q(D_{il}) = \mathbf{c}_{il}$; $\mathbf{c}_{il}[j]$ is the count of column j of D_{il}; $\Delta(Q) = \max_{D_{il}, D'_{il} \in \mathcal{D}} \| Q(D_{il}) - Q(D'_{il}) \|_1 = x_i$. We adapt the budget distribution in study [32] to our settings, and design the algorithm of data perturbation mechanism $\mathcal{M}$ as follows.

As shown in Algorithm 1, mechanism $\mathcal{M}_l$ is decomposed into two sub mechanisms $\mathcal{M}_{l1}$ and $\mathcal{M}_{l2}$. $\mathcal{M}_{l1}$ computes the dissimilarity, according to D_{i1}, $\mathbf{o}_{i1}$, $\mathbf{o}_{i2}$, $\ldots$, $\mathbf{o}_{i(l-1)}$, ε_1,ε_2, $\ldots$, and ε_{l-1}. According to the dissimilarity, $\mathcal{M}_{l2}$ computes $\mathbf{o}_{il}$ and ε_l. Sub mechanism $\mathcal{M}_{l1}$ computes the statistical result $\mathbf{c}_{il}$ on D_{il} (Line 2). Then it computes the dissimilarity dis and injects Laplace noise with scale $\omega_{l,1}$ into dis (Line 3). $\varepsilon_{l,1}$ is set to be $\varepsilon/(2\lambda)$ to perturb the dis (Line 4). $\mathcal{M}_{l2}$ first computes the remaining privacy budget $\varepsilon_{l,2}$ and Laplace noise with scale $\omega_{l,2}$ (Lines 6 and 7). Then it outputs $\mathbf{o}_{il}$ and ε_l (Lines 8, 9 and 10).

On top of the data perturbation mechanism, we have

Theorem 8: The proposed data perturbation mechanism $\mathcal{M}$ satisfies the requirement of (λ, ε)-differential privacy.

Proof: See Appendix B. ∎

This shows that the data perturbation mechanism can preserve users' location privacy. The users' pieces are processed via our proposed data perturbation mechanism listed above, and the following geo-information blocks of all users are obtained.

$$\begin{bmatrix} \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_i & \cdots & \hat{X}_{N_U} \\ O_1 & O_2 & \cdots & O_i & \cdots & O_{N_U} \end{bmatrix}, \tag{4}$$

where $\hat{X}_i = [\hat{x}_{i1}, \hat{x}_{i2}, \cdots, \hat{x}_{i\kappa_i}]^T$. Thereafter, with the geo-information blocks, users need to sign their pieces such that the LS knows which pieces are coming from the same user end for the purpose of location estimation in Section III-F.

D. Signing Pieces

A user u_i signs his pieces $\hat{X}_i$ with the signature tag_i. Since the signature composed of u_i's ID ID_i will reveal the identity

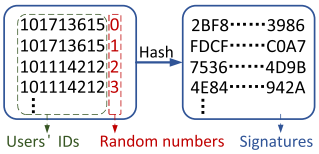


Fig. 3. An example of generating signatures.

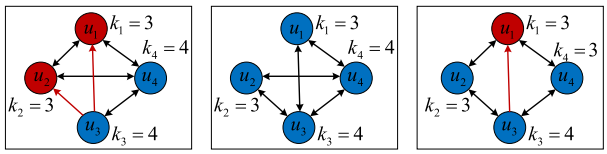


Fig. 4. (a) Non-effective transfer strategy. (b) The effective transfer strategy, where N is an even. (c) The effective transfer strategy, where N is an odd.

privacy of u_i, we propose to generate the signature using the following cryptographic hash function:

$$tag_i = H(ID_i \parallel random_i), \tag{5}$$

where $random_i$ is an random number, H is a cryptographic hash function, and $\parallel$ is string concatenation operation. For example, in Fig. 3, u_1's ID "101713615" is appended to random number "0", and then is hashed into the signature "$2BF8\ldots\ldots3986$".

Next, users exchange pieces with each other to cloak the pieces according to the data transfer strategy explained in Section III-E.

E. Cloaking Data

To cloak users' geo-information using k-anonymity technique, the number of users should be no less than the maximum of the number of each user's pieces (cf. Definition 2). We assume $(N_U - 1)$ users denoted by $u_1, u_2, \ldots, u_{N_U-1}$ are waiting for being located since $N_U - 1 < \max(\kappa_1, \kappa_2 \ldots, \kappa_{N_U-1})$, and a new user u_{N_U} requests for localization service. u_{N_U} broadcasts a hello message consisting of the number of his pieces κ_{N_U} via Bluetooth or WiFi interface [31]. Upon receiving the hello message, every user (say, the user u_i) sends back an acknowledgement consisting of κ_i. Thereafter u_{N_U} determines whether N_U is no less than $\max(\kappa_1, \kappa_2 \ldots, \kappa_{N_U})$. If so, u_{N_U} sends back an acknowledgement. Each user performs his transfer strategy in a distributed manner. He keeps one piece to himself and sends all the remaining pieces to other users according to his transfer strategy, so as to cloak his pieces using the k-anonymity technique. Note that existing work [33], [34] can be directly applied to motivate users to exchange their pieces with others to complete k-anonymity, which is out of the scope of this work.

A straightforward method to exchange pieces is to randomly transfer pieces to other users. If so, some selfish users may free-ride on others' efforts. That is, these selfish users send less pieces to the LS than their pieces, but in fact each user should send the same number of pieces as his own pieces to the LS. Conversely, these users offering free rides send more pieces to the LS than their pieces. For example, in Fig. 4(a), u_3 free-rides on efforts of u_1 and u_2, since u_3 sends three pieces to u_1, u_2, u_4 and only receives one piece from u_4. Therefore u_3 only needs to send two pieces to the LS ($2 < \kappa_3 = 4$), and u_1 (resp. u_2) has to send four pieces to the LS ($4 > \kappa_1 = \kappa_2 = 3$). To this end, we propose the following effective transfer strategy.

Definition 9: The effective transfer strategy among N_U users meets: $\arg\min \sum_{\gamma=1}^{\gamma=N_U} f_3(\kappa_{\gamma\gamma} - (\kappa_\gamma - 1))$, where $\kappa_{\gamma\gamma}$ is the number of pieces the user u_γ receives from others, and $f_3(x) = x$ when $0 \le x$, otherwise $f_3(x) = 0$.

For example, in Fig. 4(a), $\sum_{\gamma=1}^{\gamma=4} f_3(\kappa_{\gamma\gamma} - (\kappa_\gamma - 1)) = 2$ which means u_1 and u_2 have to send one more piece to the LS, compared to κ_1 and κ_2. Namely, u_1 and u_2 offer free rides for other users. But in Fig. 4(b), each user u_i ($i \in (1, \ldots, 4)$) only sends κ_i pieces to the LS, as $\sum_{\gamma=1}^{\gamma=4} f_3(\kappa_{\gamma\gamma} - \kappa_\gamma) = 0$. Namely, no user offers free ride. Therefore, the transfer strategy in Fig. 4(b) is the effective transfer strategy.

Based on the definition of the effective transfer strategy, we have the following result.

Theorem 10: If the number of pieces to be exchanged (denoted by N) is even (resp. odd), there must be even (resp. odd) users sending one more piece to the LS than their pieces in the effective transfer strategy.

Proof: See Appendix C. ■

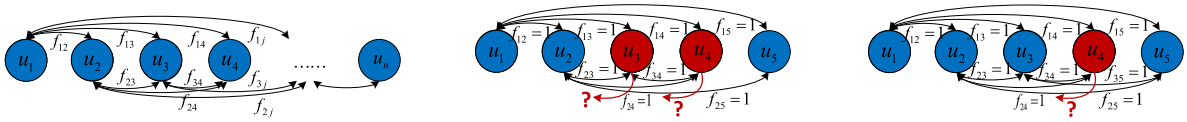
For instance, in Fig. 4(c), one user (i.e., u_1) needs to send one more piece to the LS due to $N = 9$.

Motivated by Theorem 10 and Definition 9, the proposed data transfer strategy is as follows. When $\max(\kappa_1, \kappa_2 \ldots, \kappa_{N_U}) \le N_U$, each user performs the following steps to compute his transfer strategy to cloak his pieces with k-anonymity technique. Without loss of generality, we assume $\kappa_{N_U} \le \cdots \le \kappa_2 \le \kappa_1$. As shown in Fig. 5(a), we define $f_{\gamma_1\gamma_2} = 1$ when u_{γ_1} and u_{γ_2} exchange pieces, and $f_{\gamma_1\gamma_2} = f_{\gamma_2\gamma_1}$ ($\gamma_1 \ne \gamma_2, \gamma_1, \gamma_2 \in (1, \ldots, N_U)$). Then we get

$$\begin{cases} f_{12} = f(\kappa_1 - 1 - 1) = f(\kappa_1 - 2) \\ f_{13} = f(\kappa_1 - 1 - f_{12} - 1) \\ \vdots \\ f_{1\gamma} = f(\kappa_1 - 1 - \sum_{\tau=2}^{\gamma-1} f_{1\tau} - 1), \end{cases} \tag{6}$$

where $\gamma \in (2, \ldots, N_U)$; $f(x) = 1$ when $0 \le x$, and otherwise, $f(x) = 0$.

$$\begin{cases} f_{23} = f(\kappa_2 - 1 - f_{12} - 1) \times f_2(f_{13} - (\kappa_3 - 1)) \\ \vdots \\ f_{2\gamma} = f(\kappa_2 - 1 - f_{12} - \sum_{\tau=3}^{\gamma-1} f_{2\tau} - 1) \times f_2(f_{1\gamma} - (\kappa_\gamma - 1)), \end{cases} \tag{7}$$



(a) (b) (c)

Fig. 5. (a) The illustration of computing the effective transfer strategy. (b) (c) Examples of computing effective transfer strategy. (b) N is an even, and $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 5$, $\kappa_5 = 3$. (c) N is an odd, and $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 5$, $\kappa_5 = 4$.

where $\gamma \in (3, \ldots, N_U)$; $f_2(x) = 1$ when $1 \leq x$, and otherwise $f_2(x) = 0$. We can generalize Eqs. (6) and (7) as

$$\begin{cases} f_{\gamma(\gamma+1)} = f(\kappa_\gamma - 1 - \sum_{\tau=1}^{\gamma-1} f_{\tau\gamma} - 1) \\ \qquad \times f_2(\sum_{\tau=1}^{\gamma-1} f_{\tau(\gamma+1)} - (\kappa_{\gamma+1} - 1)) \\ f_{\gamma(\gamma+z)} = f(\kappa_\gamma - 1 - \sum_{\tau=1}^{\gamma-1} f_{\tau\gamma} - \sum_{\tau=\gamma+1}^{\gamma+z-1} f_{\gamma\tau} - 1) \\ \qquad \times f_2(\sum_{\tau=1}^{\gamma-1} f_{\tau(\gamma+z)} - (\kappa_{\gamma+z} - 1)), \end{cases} \tag{8}$$

where $z \in (2, \ldots, N_U - \gamma)$, $\gamma \in (1, \ldots, N_U)$. Thereafter, each user u_i only needs to compute the following equation to determine his transfer strategy $f_{i\tau}$:

$$\sum_{\tau=1, \tau \neq i}^{m} f_{\tau i} = \kappa_i - 1, \quad 1 \leq m \leq N_U. \tag{9}$$

It is important to note that, according to Theorem 10, some users cannot compute their transfer strategy using Eqs. (8) and (9). Taking Fig. 5(b) as an example, N is 18; u_1, u_2, and u_5 can get their transfer strategy respectively according to Eqs. (8) and (9), but obviously one piece of u_3 and one piece of u_4 are not cloaked. In summary, the above data transfer strategy needs to be modified.

Before diving into the details of how to modify the above data transfer strategy, we present the following result.

Theorem 11: When N pieces of N_U users need to be exchanged, there will be $N - 2\sum_{\gamma_1=1}^{N_U-1} \sum_{\gamma_2=\gamma_1+1}^{N_U} f_{\gamma_1\gamma_2}$ pieces which cannot be transferred to other users, according to Eqs. (8) and (9).

Proof: See Appendix D. ■

For example, in Fig. 5(c), one piece of u_4 cannot be cloaked, as $(N - 2\sum_{\gamma_1=1}^{4} \sum_{\gamma_2=\gamma_1+1}^{5} f_{\gamma_1\gamma_2}) = 1$.

According to Theorem 11, we propose to adjust the transfer strategy as follows. First, all users compute transfer strategy as Eqs. (8) and (9). We denote these users of which some pieces are not transferred to other users (recall Theorem 11) as $u_\nu, u_{\nu+1}, \ldots, u_{\nu+\eta}$, $\nu \in (1, \ldots, N_U)$, $\eta \in (0, \ldots, N_U - \nu)$. Then, every user u_i $(i \in (\nu, \ldots, \nu + \eta))$ randomly sends the remaining pieces to $\kappa_i - (\sum_{\tau=1, \tau \neq i}^{N_U} f_{\tau i} + 1)$ other users each of which (e.g., u_γ) meets $f_{i\gamma} = 0$.

In summary, we design a distributed transfer strategy in this section. Next, each user sends cloaked pieces to the LS for location estimation.

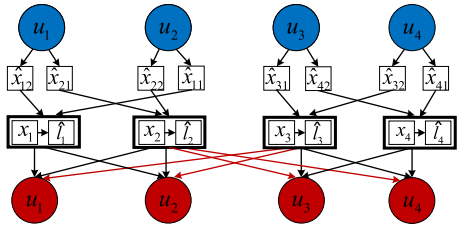


Fig. 6. Illustration of localization process, $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 2$.

F. Estimating Location

Although users' pieces are cloaked, the localization outcome of a specific user cannot be protected against other users who have received the user's pieces. For example, in Fig. 6, u_1 sends one of his piece $\hat{x}_{12}$ and one of u_2's piece $\hat{x}_{21}$ to the LS. Since the LS cannot distinguish u_1 and u_2, both locations of u_1 and u_2 are returned to u_1. As u_1 can identify his location, the possibility that u_1 identifies u_2's location is 1, which is larger than u_2's privacy requirement $1/k_2 = 1/2$. As a result, u_2's location privacy is disclosed to u_1.

To that end, we design a privacy-preserving localization strategy performed at the server side. Specifically, upon receiving pieces from the N_U users, the LS first computes the geo-information of each user u_i, according to the signatures, i.e., $x_i = \sum_{\gamma=1}^{\gamma=\kappa_i} \hat{x}_{i\gamma}$. Then the LS carries out the underlying localization algorithm to locate each user u_i using x_i. Assume a specific user u_j sends pieces $\hat{x}_{j\gamma}, \hat{x}_{(j+1)\gamma}, \ldots, \hat{x}_{(j+r)\gamma}$ to the LS. Denote the corresponding localization outcomes as $\hat{l}_j, \hat{l}_{j+1}, \ldots, \hat{l}_{j+r}$, and the set of tuples consisting of the $(r+1)$ geo-information and $(r+1)$ localization outcomes as $\mathcal{L}_1 = \{(x_j, \hat{l}_j), (x_{j+1}, \hat{l}_{j+1}), \ldots, (x_{j+r}, \hat{l}_{j+r})\}$. To meet different privacy requirements (i.e., $k_{j+1}, \ldots, k_{j+r}$) of users $u_{j+1}, \ldots, u_{j+r}$, the LS randomly selects $[\max(k_{j+1}, \ldots, k_{j+r}) - r]$ other tuples (denote the set of these tuples as $\mathcal{L}_2$) from $\mathcal{L}$, which meet the following constraints: $\mathcal{L}_2 \cap \mathcal{L}_1 = \varnothing$, $\mathcal{L}_2 \subset \mathcal{L}$, where $\mathcal{L} = \{(x_1, \hat{l}_1), \ldots, (x_{N_U}, \hat{l}_{N_U})\}$ is the set of tuples consisting of these N_U users' geo-information and localization outcomes. Lastly, the LS returns tuples (that is, the cloaked location) in $\mathcal{L}_2 \cup \mathcal{L}_1$ to u_j.

Taking Fig. 6 as an illustrative example, except for tuples $(x_1, \hat{l}_1)$ and $(x_2, \hat{l}_2)$ that should be returned to u_1, tuple $(x_3, \hat{l}_3)$ is also sent to u_1 as $k_2 = 2$.

G. Identifying Location

Since each user's geo-information is revealed to himself (and only to himself), his real location can be identified through checking whether his geo-information is contained in

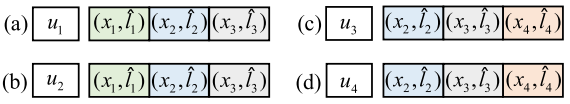


Fig. 7. Illustration of identifying locations.

the tuples returned by the LS. For example, in Fig. 7(a), upon receiving the tuples $(x_1, \hat{l}_1)$, $(x_2, \hat{l}_2)$, and $(x_3, \hat{l}_3)$ returned by the LS, u_1 can identify his estimated location $\hat{l}_1$ through detecting whether the geo-information contained in these tuples is the same as his geo-information x_1. Likewise, u_2, u_3, and u_4 in Figs. 7(b-d) can identify their estimated locations.

H. Discussions

P^3-LOC provides several salient features, as presented in the following theorems.

Theorem 12: P^3-LOC guarantees each user's location privacy against other users and the LS at the expected level $(1/k)$ when no users collude. Moreover, when less than λ users collude with each other, P^3-LOC provides (λ, ε)-differential privacy on the user's location privacy against other users and protects the user's location privacy against the LS at the expected level $(1/k)$. Furthermore, P^3-LOC protects the LS's data privacy.

Proof: See Appendix E. ■

Theorem 13: P^3-LOC guarantees the data utility with the average error $\frac{4x_i(2^\lambda-1)}{\varepsilon\lambda}+\frac{2x_i\lambda}{\varepsilon d}$ in data perturbation mechanism $\mathcal{M}$, for a specific user u_i.

Proof: See Appendix F. ■

Theorem 14: Both the computation cost and the communication cost introduced by P^3-LOC are at most $O(N_U)$ where N_U is the total number of mobile users.

Proof: See Appendix G. ■

Theorem 12 characterizes the degree of privacy preservation, Theorem 13 implies the guarantee on data utility, and Theorem 14 shows the scalability of P^3-LOC.

IV. Performance Evaluation

We conduct extensive experiments via measured data, and compare P^3-LOC with PriWFL [22] and MCA [6], privacy-preserving algorithms for fingerprint–based indoor localization. As there is no prior work on privacy preservation in model based and dead-reckoning based indoor localization, we simply compare P^3-LOC with a model-based indoor localization algorithm [10] (dubbed as ZERO) and a dead-reckoning based indoor localization algorithm [35] (dubbed as IPLOS). Note that ZERO and IPLOS do not provide any privacy guarantees.

In the fingerprint-based indoor localization, each of the WiFi fingerprints is the average value of 100 RSS values measured at each sampling location labeled by red spots (cf. Fig. 8(a)). In model-based indoor localization, we periodically measure the RSS from all APs (cf. Fig. 8(b)), map RSS to geographical distances, and calibrate signal-distance mapping parameters. In dead-reckoning based indoor localization, users' locations are estimated by an accelerometer. The movement speed is 2 steps/sec, and the stride length is set as 70cm.

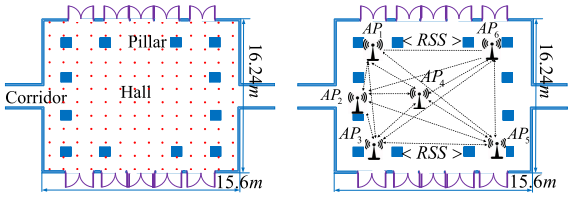


Fig. 8. The layout of the office building for (a) fingerprinting-based and (b) model-based indoor localization.

Other default parameter settings are as follows: $N_U = 80$, $\varepsilon = 1$; $\kappa_i = k_i \in [2, 5]$, $\lambda \in [1, 4]$; every real number is represented by 24 bits; the tolerable delay $d_t = 0.01$s; the query interval is 15s. The RSS data and sensor data are measured by Samsung Galaxy S5; all the experiments are implemented in $C{+}{+}$ and conducted on a desktop PC with an Intel Core $i5$ 3.30 GHz processor and $8G$ memory.

A. Privacy Preservation

We propose the following metrics: the degree of location privacy preservation (LPP) and the degree of data privacy preservation (DPP) to quantify the privacy preservation. LPP refers to the proportion of users whose location is identified with a probability less than $1/k$. When all users cannot get access to the localization-related information except for their estimated locations, the data privacy of LS is protected, and DPP equals 1. That is, when users get partial information about localization-related information with a probability 0, DPP is set to 1; otherwise 0. For example, in the WiFi fingerprint-based indoor localization [22], the locations of fingerprints recorded in fingerprint database are disclosed to users, and malicious users can get a similar fingerprint database. Namely, users have a certain belief about partial information of localization-related information. In such a case, LS's data privacy is not protected, and DPP is set to 0.

1) Privacy Preservation in Fingerprint Based Indoor Localization: Fig. 9(a) shows the privacy preservation in default settings. It shows that P^3-LOC can protect data privacy and provide more than 93% LPP in existing three kinds of localization techniques. In contrast, PriWFL and MCA are only applicable to the fingerprint-based localization algorithm which searches for nearest matches, and can only protect 60% and 58% users' location privacy without any protection for data privacy. That is because (i) no user can access the database in the LS in P^3-LOC (proved in Theorem 12), while the locations of fingerprints are disclosed to users in PriWFL and MCA; (ii) P^3-LOC is a paradigm driven general framework, and therefore is applicable to all underlying localization systems; (iii) PriWFL and MCA employs homomorphic encryption resulting in many expired queries, and thus less users' location privacy can be protected.

Fig. 9(b) shows the impact of N_U on the privacy preservation. First, the LPP in PriWFL, MCA, and P^3-LOC decrease with the increase of N_U, as more users to be processed result in more expired queries, and thus less users' location privacy can be protected. Second, the LPP in P^3-LOC is more robust

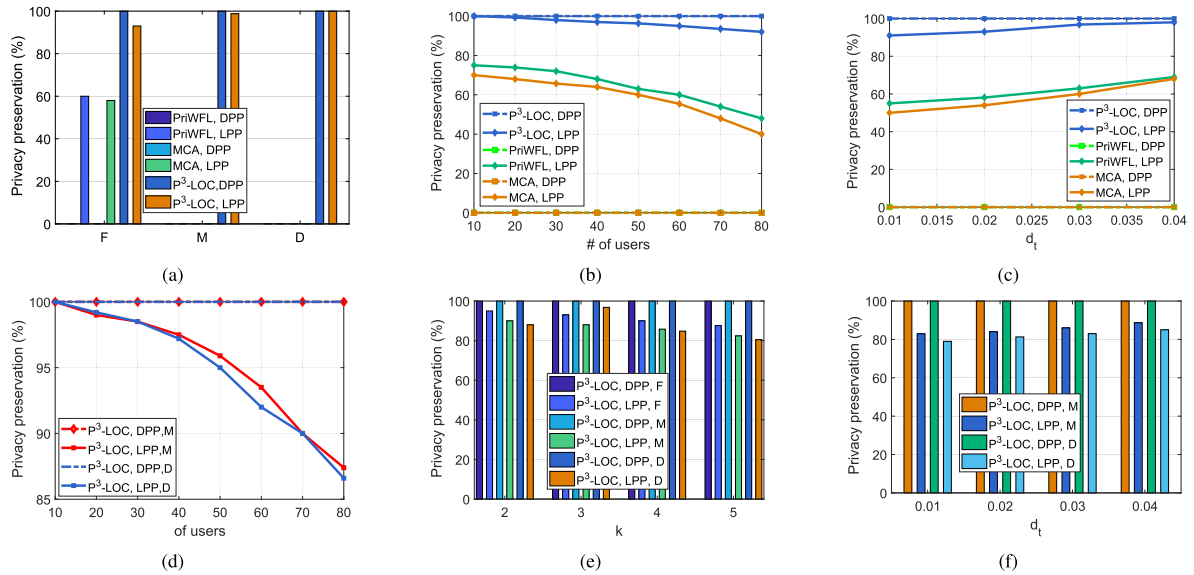


Fig. 9. (a) The privacy preservation in default setting; "F", "M", and "D" refer to fingerprinting-based, model-based, and dead-reckoning based localization, respectively. (b) The impact of N_U on privacy preservation in F; (c) The privacy preservation varying with d_t in F. (d) The privacy preservation varying with N_U in M and D. (e) The privacy preservation varying with k. (f) The privacy preservation varying with d_t in M and D.

to N_U, as it takes less time to process users' queries in P^3-LOC than PriWFL. Lastly, the DPP in the three algorithms do not vary with N_U. The reason is that PriWFL and MCA cannot protect data privacy while P^3-LOC does. Also, P^3-LOC outperforms PriWFL and MCA in terms of both LPP and DPP.

Fig. 9(c) shows the impact of d_t. It shows that the privacy preservation increases with d_t, as a larger d_t enables algorithms to process more queries. In addition, the LPP in P^3-LOC is more robust to d_t, since more expired queries in PriWFL and MCA can be processed when we prolong the d_t.

Since PriWFL does not use k-anonymity, we only focus on the impact of k on the performance of P^3-LOC. A larger k results in more pieces to be processed and more expired queries. As such, in Fig. 9(e), the LPP in P^3-LOC decreases when we increase k. Lastly, the DPP in P^3-LOC is also not affected by k.

2) Privacy Preservation in Model-Based Indoor Localization: Since ZERO and IPLOS do not protect privacy, we only focus on the privacy preservation in P^3-LOC.

In model-based indoor localization, the LPP in P^3-LOC decreases with N_U (cf. Fig. 9(d)) and k (cf. Fig. 9(e)), and increases with d_t (cf. Fig. 9(f)). The reason is that increasing N_U and k result in more expired queries, and thus less users' location privacy can be protected. Moreover, a larger d_t enables more queries to be processed.

3) Privacy Preservation in Dead-Reckoning-Based Indoor Localization: More users and larger k deteriorate the LPP in P^3-LOC, which is shown in Figs. 9(d) and 9(e). In addition, in Fig. 9(f), P^3-LOC benefits from a larger d_t.

B. Data Utility Guarantee

The data utility loss is resulted from the data perturbation mechanism $\mathcal{M}$ that meets (λ, ε)-differential privacy and injects Laplace noise into pieces. Then, the geo-information containing noise further affects the localization accuracy. That is, the settings of parameters λ and ε affect the data utility of users' geo-information (i.e., the localization accuracy), as the scale of Laplace noise is determined by λ and ε. Therefore, we propose the metric, cumulative distribution function (CDF) of the localization errors, to quantify the data utility loss. The localization error of a specific user u_i is measured as $|\hat{l}_i - l_i|$, where $\hat{l}_i$ is the estimated location of u_i, and l_i is the exact location (i.e., ground truth).

The cumulative distribution function (CDF) of localization errors in fingerprint-based localization is shown in Figs. 10(a) and 10(b). We observe from Fig. 10(a) that the performance of P^3-LOC is comparable to that of PriWFL and MCA, as 50% of the localization errors in P^3-LOC and PriWFL and MCA are less than 2.2m, the remaining 50% of the localization errors are around 3.1m, and the average localization errors are about 2.17m, 2.16m, and 2.09m. That is because, PriWFL uses the homomorphic encryption, and P^3-LOC provides differential privacy. Moreover, in Fig. 10(a), the localization errors in P^3-LOC increase with the increasing parameter λ, because a larger parameter λ enlarges the noise. Furthermore, Fig. 10(b) shows that the localization errors in P^3-LOC decrease with the increasing parameter ε, as increasing ε decreases the scale of noise.

In model-based localization, ZERO does not outperform P^3-LOC, as the average localization errors in P^3-LOC and ZERO are about 2.36m and 2.24m (cf. Fig. 10(c)). In addition, the localization errors in P^3-LOC increase with the increasing parameter λ, and decrease with the parameter ε, which is shown in Figs. 10(c) and 10(d). The reasons are analyzed as above. Likewise, in dead-reckoning-based localization, P^3-LOC is still preferable, since the average localization errors in P^3-LOC and IPLOS are 1.19m and 1.13m (cf. Fig. 10(e)). Furthermore, the localization errors in P^3-LOC vary with

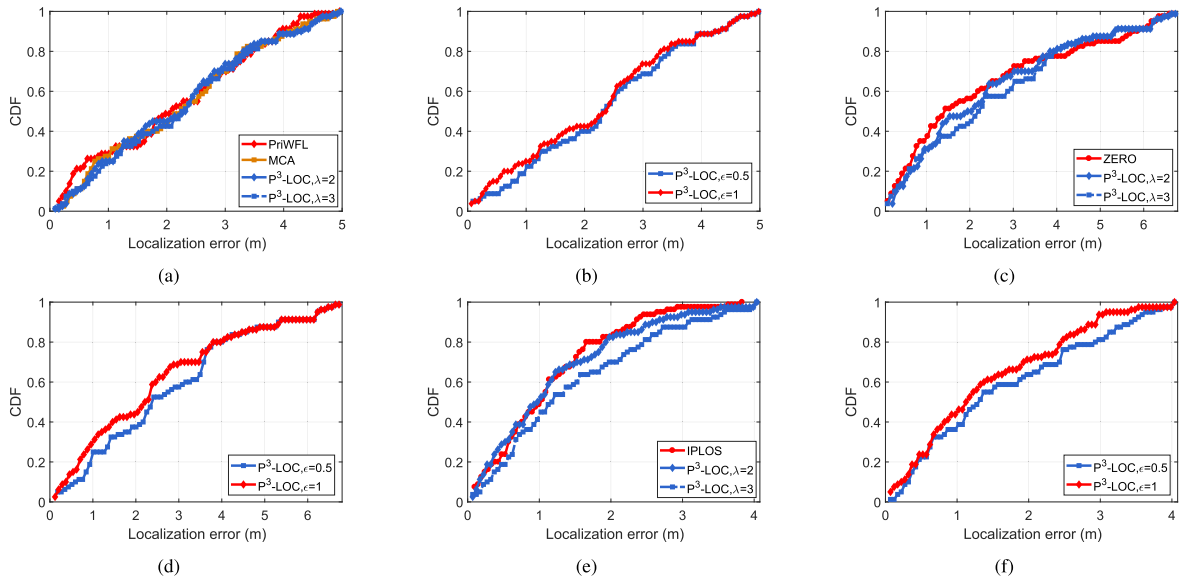


Fig. 10. The localization error in (a) (b) fingerprint-based, (c) (d) model-based, and (e) (f) dead-reckoning-based localization.

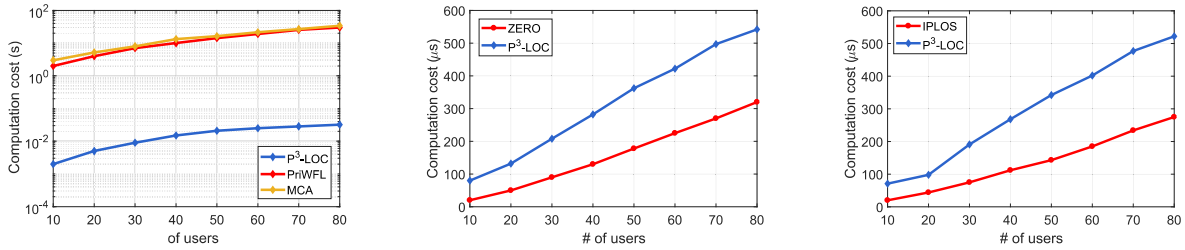


Fig. 11. The computation cost in (a) fingerprint-based, (b) model-based and (c) dead-reckoning-based localization.

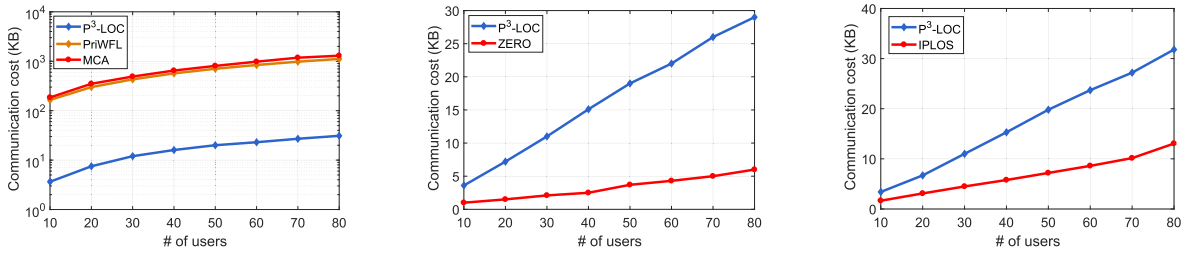


Fig. 12. The communication cost in (a) fingerprint-based, (b) model-based and (c) dead-reckoning-based localization.

parameters λ and ε (cf. Figs. 10(e) and 10(f)), which is same to the trends in Figs. 10(a) and 10(b).

C. Robustness

We propose the metrics, computation overhead (CPO) and communication overhead (CMO) to quantify the scalability of P^3-LOC.

1) Computation Cost: In fingerprint-based localization, the CPO in P^3-LOC, PriWFL, and MCA indeed increases with N_U, because more users result in more computation operations (cf. Fig. 11(a)). In addition, the CPO in PriWFL and MCA are about 10^3 times larger than that in P^3-LOC, which is attributed to the homomorphic encryption employed in PriWFL and MCA. Lastly, the CPO in P^3-LOC is more robust to N_U than that in PriWFL and MCA, which is resulted from the heavy computation overhead in PriWFL and MCA.

In model-based and dead-reckoning-based localization, the CPO in all algorithms increases with N_U, as algorithms have to locate more users (cf. Figs. 11(b) and 11(c)). Furthermore, the CPO in P^3-LOC is a bit larger than that in ZERO and IPLOS, because ZERO and IPLOS ignore the privacy preservation. Lastly, P^3-LOC is still preferable, as it only incurs microsecond-level computation cost.

2) Communication Cost: In fingerprint-based indoor localization, the CMO in PriWFL is about 150 times larger than that in P^3-LOC (cf. Fig. 12(a)). Because the length of the ciphertext

is 1024 bits, which is quite larger than the 24 bits in P^3-LOC. Second, the CMO in P^3-LOC and PriWFL increases with N_U, as more users lead to more pieces. Lastly, the CMO in P^3-LOC is more robust than that in PriWFL, which is attributed to the homomorphic encryption in PriWFL.

In model-based and dead-reckoning-based localization, as shown in Figs. 12(b) and 12(c), the CMO in all algorithms is proportional to N_U. In addition, the CMO in P^3-LOC is a bit larger than that in ZERO and IPLOS.

V. Conclusion

In this paper, we have presented P^3-LOC, a privacy-preserving paradigm-driven framework for indoor localization. It guarantees both users' location privacy and the LS's data privacy. Moreover, it is applicable to any localization system which complies with the aforementioned two-stage indoor localization paradigm. Extensive experiments via measured data have demonstrated the effectiveness and efficiency of P^3-LOC in terms of both privacy preservation and QoS guarantees.

The temporal-spatial correlation of the location dataset enables adversary to launch spatio-temporal correlation attacks [36], location-depend attacks [24], inference attacks [19], etc. It is another topic, protecting location privacy against these attacks, but the corresponding privacy preserving solutions [24], [37], [38] can be combined with our scheme. So in terms of the future work, we plan to explore the temporal-spatial correlation among locations to enhance our scheme.

Appendix

A. Proof of Theorem 7

Proof: Since each $\mathcal{M}_i$ generates independent randomness, so

$$Pr[\mathcal{M}(S_n) = R] = \prod_{\gamma=1}^{n} Pr[\mathcal{M}_\gamma(S_n[\gamma]) = R[\gamma]] \quad (10)$$

Similarly, for S'_n of S_n, the following holds:

$$Pr[\mathcal{M}(S'_n) = R] = \prod_{\gamma=1}^{n} Pr[\mathcal{M}_\gamma(S'_n[\gamma]) = R[\gamma]] \quad (11)$$

According to Definition 5, for $1 \le \gamma \le i-1$ and $i+\lambda \le \gamma \le n$, $S_n[\gamma] = S'_n[\gamma]$. Thus, we get,

$$\frac{Pr[\mathcal{M}(S_n) = R]}{Pr[\mathcal{M}(S'_n) = R]} = \prod_{\gamma=i}^{i+\lambda-1} \frac{Pr[\mathcal{M}_\gamma(S_n[\gamma]) = R[\gamma]]}{Pr[\mathcal{M}_\gamma(S'_n[\gamma]) = R[\gamma]]} \quad (12)$$

As $\mathcal{M}_i$ satisfies ε_i-differential privacy, and $S'_n[\gamma]$ and $S_n[\gamma]$ differ on at most one record, thus $\frac{Pr[\mathcal{M}_\gamma(S_n[\gamma])=R[\gamma]]}{Pr[\mathcal{M}_\gamma(S'_n[\gamma])=R[\gamma]]} \le e^{\varepsilon_\gamma}$. We further get,

$$\frac{Pr[\mathcal{M}(S_n) = R]}{Pr[\mathcal{M}(S'_n) = R]} \le \prod_{\gamma=i}^{i+\lambda-1} e^{\varepsilon_\gamma} = e^{\sum_{\gamma=i}^{i+\lambda-1} \varepsilon_\gamma} \quad (13)$$

Since $\sum_{\gamma=i}^{i+\lambda-1} \varepsilon_\gamma \le \varepsilon$, therefore $\frac{Pr[\mathcal{M}(S_n)=R]}{Pr[\mathcal{M}(S'_n)=R]} \le e^{\varepsilon}$. According to Definition 5, $\mathcal{M}$ meets (λ, ε)-differential privacy.

In summary, Theorem 7 holds. ■

B. Proof of Theorem 8

Proof: We first need to prove that $\mathcal{M}_{l1}$ meets $\varepsilon_{l,1}$-differential privacy with $\varepsilon_{l,1} = \varepsilon/(2\lambda)$, and $\mathcal{M}_{l2}$ is $\varepsilon_{l,2}$-differentially private for $\varepsilon_{l,2} = \varepsilon/2 - \sum_{\tau=l-\lambda+1}^{l-1} \varepsilon_{\tau,2}$.

To begin with, we give the following lemma in existing work [39].

Lemma 15: For all f: $D \longrightarrow \mathbb{R}^d$, the following mechanism is ε-differentially private [39]: $\mathcal{M}_f(x) = f(x) + \langle Lap(\Delta(f)/\varepsilon)\rangle^d$, where $\langle Lap(\Delta(f)/\varepsilon)\rangle^d$ means injecting a zero-mean Laplace distribution with scale $\omega = \Delta(f)/\varepsilon$ to each of the d output values of $f(D)$ [32].

$\mathcal{M}_{l1}$ outputs $Q'(D_i) = 1/d \sum_{j=1}^{d} \mid \mathbf{o}_{i(1-1)}[j] - \mathbf{c}_{il}[j] \mid + Lap(\omega_{l,1})$. So the absence of a row in D_i changes the above result by $\Delta(Q') = \frac{x_i}{d}$. Then $\mathcal{M}_{l1}$ injects Laplace noise with scale $2\lambda x_i/(\varepsilon d)$. According to Lemma 15, $\mathcal{M}_{l1}$ is $\varepsilon_{l,1}$-differential privacy for $\varepsilon_{l,1} = \varepsilon/(2\lambda)$. $\mathcal{M}_{l2}$ outputs $Q(D_{i1}) = \mathbf{c}_{il}$. The sensitivity of Q is $\Delta(Q) = x_i$, and it injects Laplace noise with scale $2x_i/(\varepsilon/2 - \sum_{\tau=l-\lambda+1}^{l-1} \varepsilon_{\tau,2})$. Thus, according to Lemma 15, $\mathcal{M}_{l2}$ is $\varepsilon_{l,2}$-differentially private, where $\varepsilon_{l,2} = \varepsilon/2 - \sum_{\tau=l-\lambda+1}^{l-1} \varepsilon_{\tau,2}$.

Then we need to prove that in any sliding window of length λ, $\sum_{\tau=l}^{l+\lambda-1} \varepsilon_\tau \le \varepsilon, l \in (1, \ldots, \kappa_i)$, according to Theorem 7. In $\mathcal{M}_l$, $\varepsilon_\tau = \varepsilon_{\tau,1} + \varepsilon_{\tau,2}$. So we need to prove $\sum_{\tau=l}^{l+\lambda-1} \varepsilon_{\tau,1} + \sum_{\tau=l}^{l+\lambda-1} \varepsilon_{\tau,2} \le \varepsilon$. Since $\sum_{\tau=l}^{l+\lambda-1} \varepsilon_{\tau,1} = \sum_{\tau=l}^{l+\lambda-1} \varepsilon/(2\lambda) = \varepsilon/2$, we only need to prove $\sum_{\tau=l}^{l+\lambda-1} \varepsilon_{\tau,2} \le \varepsilon/2$. As $\mathcal{M}_{l2}$ uses half of the available privacy budget, $\sum_{\tau=l}^{l+\lambda-1} \varepsilon_{\tau,2} \le \varepsilon/2$ holds.

In summary, Theorem 8 holds. ■

C. Proof of Theorem 10

Proof: $2C_{N_U}^2 = N_U(N_U - 1)$ pieces at most can be exchanged among N_U users. It is obviously $N_U(N_U - 1)$ is an even. Thus there must be even pieces that can be pairwise exchanged among the N_U users. Furthermore, when N is an even, there must be even pieces that cannot be transferred to others and therefore even users have to send one more piece to the LS. Likewise, there are odd users sending one more piece to LS than their pieces, when N is an odd.

Thus, Theorem 10 holds. ■

D. Proof of Theorem 11

Proof: First, users compute their transfer strategy according to Eqn. (8) and (9). Then $\sum_{\gamma_1=1}^{N_U} \sum_{\gamma_2=\gamma_1+1}^{N_U} f_{\gamma_1\gamma_2}$ pairs of users can pairwise exchange pieces. Thus, the number of pieces exchanged among $\sum_{\gamma_1=1}^{N_U} \sum_{\gamma_2=\gamma_1+1}^{N_U} f_{\gamma_1\gamma_2}$ pairs of users is $2\sum_{\gamma_1=1}^{N_U} \sum_{\gamma_2=\gamma_1+1}^{N_U} f_{\gamma_1\gamma_2}$, and therefore, $(N - 2\sum_{\gamma_1=1}^{N_U} \sum_{\gamma_2=\gamma_1+1}^{N_U} f_{\gamma_1\gamma_2})$ pieces are left and cannot be transferred to other users.

In summary, Theorem 11 holds. ■

E. Proof of Theorem 12

Proof: In first three building blocks, every user can only access his pieces. In fourth block, each user can only access the perturbed pieces of other users, and thus cannot get the exact geo-information of others, incurring protection for location privacy of users. In summary, the location privacy is preserved

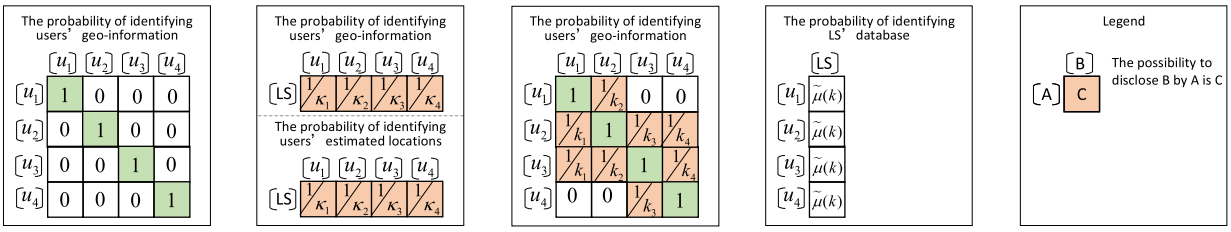


Fig. 13. The illustration of privacy preservation for the users in Fig. 6. (a) Location privacy preservation in the first four building blocks; (b) Location privacy preservation in the fifth building block; (c) Location privacy preservation in sixth building block; (d) Data privacy preservation in P^3-LOC, and $\widetilde{\mu}(k)$ is a negligible function with respect to the privacy parameter k; (e) Legend.

in first four building blocks, which is shown in Fig. 13(a). In the fifth block, i.e., estimating data, the LS cannot identify the geo-information (resp. the estimated location) of very user u_i from $(\kappa_i - 1)$ geo-information (resp. estimated locations) of others, which is illustrated in Fig. 13(b). Lastly, when the LS sends back the localization outcomes, each user can only identify his own location, and cannot distinguish others' locations. As shown in Fig. 13(c), each user u_i distinguish the geo-information of u_j ($i \in (1, \ldots, 4)$, $j \neq i$) with the possibility $1/\kappa_j$ or 0. In summary, since $\kappa_i \geq k_i$, the location privacy of each user u_i is preserved at the pre-specified level $1/k_i$.

In addition, it is proved in Theorem 8 that the proposed data perturbation mechanism satisfies (λ, ε)-differential privacy. Thus P^3-LOC can provide ε-differential privacy on users' location privacy when less than λ users collude with each other.

Lastly, each user only get his own location, and cannot get any information about database in the LS (cf. Fig. 13(d)). Specifically, as shown in Fig. 6, the user u_1 sends pieces $\hat{x}_{12}$ and $\hat{x}_{21}$ containing noise to the LS. Then LS returns user u_1 tuples $(x_1, \hat{l}_1)$, $(x_2, \hat{l}_2)$, and $(x_3, \hat{l}_3)$ as shown in Fig. 7(a). Assume LS performs the localization algorithm [22] that treats the centroid of the locations of nearest fingerprints as the user u_1's location (recall that P^3-LOC is paradigm-driven and thus is applicable to all underlying localization algorithms). In such a case, u_1 can get:

$$\begin{cases} \|x_1 - V_{1,1}\|_2 = dis_{1,1} \\ \vdots \\ \|x_1 - V_{1,n}\|_2 = dis_{1,n} \\ \dfrac{L_{1,1} + \cdots + L_{1,n}}{n} = \hat{l}_1 \\ \|x_2 - V_{2,1}\|_2 = dis_{2,1} \\ \vdots \\ \|x_2 - V_{2,n}\|_2 = dis_{2,n} \\ \dfrac{L_{2,1} + \cdots + L_{2,n}}{n} = \hat{l}_2 \\ \|x_3 - V_{3,1}\|_2 = dis_{3,1} \\ \vdots \\ \|x_3 - V_{3,n}\|_2 = dis_{3,n} \\ \dfrac{L_{3,1} + \cdots + L_{3,n}}{n} = \hat{l}_3 \end{cases} \tag{14}$$

where $V_{i,1}, V_{i,2}, \ldots, V_{i,n}$ $(i = 1, 2, 3)$ are the fingerprints that are nearest to u_i, $L_{i,\tau}$ $(\tau = 1, 2, \ldots, n)$ is the location of fingerprint $V_{i,\tau}$, and $dis_{i,\tau}$ is the distance between fingerprint $V_{i,\tau}$ and geo-information x_i in signal space. Obviously, $V_{i,\tau}$, $L_{i,\tau}$, and $dis_{i,\tau}$ are unknown variables as user u_1 only receives tuples $(x_1, \hat{l}_1)$, $(x_2, \hat{l}_2)$, and $(x_3, \hat{l}_3)$ from LS, and the number of unknown variables is larger than the number of sub-equations in Eq. (14). Thus, there are infinite solutions to Eq. (14). Note that disclosing data privacy of LS means getting the values of $V_{i,\tau}$ and $L_{i,\tau}$. Denote the corresponding infinite solutions by $\Gamma(k) = \{so_1, so_2, \ldots\}$, and the variable $Se(k)$ is the solution selected by malicious user u_1. Both $\Gamma(k)$ and $Se(k)$ are related to privacy parameter k, as sub-equations in Eq. (14) are constrained by privacy parameter k (cf. Section III-F). Furthermore, variable $Se(k)$ is uniformly distributed among the infinite solutions, as malicious u_1 can always make random guesses. Therefore, the probability to disclose the LS's data privacy can be denoted by $\widetilde{\mu}(k) = \frac{1}{No(\Gamma(k))}$, where $No(\Gamma(k))$ is the cardinality of $\Gamma(k)$. Furthermore, we can conclude that $\widetilde{\mu}(k)$ is a negligible function with respect to privacy parameter k: given $k > 0$ and any a positive polynomial $Poly(k)$, $\widetilde{\mu}(k) \ll \frac{1}{Poly(k)}$. That is, the probability to disclose the LS's data privacy is negligibly small with respect to the privacy parameter k. In summary, the data privacy of LS is also protected.

To sum up, Theorem 12 holds. ■

F. Proof of Theorem 13

Proof: Consider the worst case, where no privacy budget is recycled from the datasets that is out of the window of size λ. In this case, the sub mechanism $\mathcal{M}_{l,2}$ exponentially distributes the privacy budget, i.e., $\varepsilon/4, \varepsilon/8, \ldots, \varepsilon/2^{\lambda+1}$. Therefore, the error in the window of size λ is $\frac{1}{\lambda}(4x_i/\varepsilon + 8x_i/\varepsilon + \cdots + x_i 2^{\lambda+1}/\varepsilon) = \frac{4x_i(2^\lambda - 1)}{\varepsilon\lambda}$ at most. Sub mechanism $\mathcal{M}_{l,1}$ outputs the dissimilarity with the Laplace noise of scale $\lambda_{l,1}$. The expect estimation of the dissimilarity is $2\lambda x_i/\varepsilon d$, due to the noise of $\mathcal{M}_{l,1}$. So the error induced by $\mathcal{M}_{l,1}$ is $2\lambda x_i/\varepsilon d$. In summary, the error in P^3-LOC is at most $\frac{4x_i(2^\lambda - 1)}{\varepsilon\lambda} + 2x_i\lambda/\varepsilon d$.

Overall, Theorem 13 holds. ■

G. Proof of Theorem 14

Proof: Let us see the computation cost at first. In first building block, segmenting data, each user u_i segments κ_i

TABLE II
COMPLEXITY ANALYSIS

Building block	Computation	Communication
Segmenting data	$\sum_{\gamma=1}^{N_U} \kappa_\gamma$	none
Perturbing data	$\sum_{\gamma=1}^{N_U} \kappa_\gamma$	none
Cloaking data	$\sum_{\gamma=1}^{N_U} (\kappa_\gamma - 1)$	$\sum_{\gamma=1}^{N_U} (\kappa_\gamma - 1)$
Estimating location	$\sum_{\gamma=1}^{N_U} k_\gamma$	$\sum_{\gamma=1}^{N_U} k_\gamma$
Identifying location	$\sum_{\gamma=1}^{N_U} \max(k)$	$\sum_{\gamma=1}^{N_U} \max(k)$

pieces, and thus the computation complexity is $T(N_U) = \kappa_i$ (i.e., $O(1)$) for each user u_i, and $T(N_U) = \sum_{\gamma=1}^{N_U} \kappa_\gamma$ (i.e., $O(N_U)$) for the N_U users. In perturbing data, N_U privacy budget ε_γ ($\gamma \in (1, \ldots, N_U)$) are distributed for N_U users incurring the computation complexity $T(N_U) = N_U$, and each user processes his pieces with (λ, ε)-differential privacy resulting in the computation complexity $T(N_U) = \sum_{\gamma=1}^{N_U} \kappa_\gamma$. In third building block, cloaking data, each user u_i has to compute $(\kappa_i - 1)$ times to search for $(\kappa_i - 1)$ other users, incurring the computation complexity $T(N_U) = \sum_{\gamma=1}^{N_U} (\kappa_\gamma - 1)$ (i.e., $O(N_U)$) for the N_U users. In estimating location, the LS has to locate all N_U users, and thus the computation complexity is $\sum_{\gamma=1}^{N_U} k_\gamma$. Since each user receives $\max(k)$ localization output at most, the computation complexity in identifying location is $T(N_U) = \sum_{\gamma=1}^{N_U} max(k)$ (i.e., $O(N_U)$). In summary, the computation complexity in P^3-LOC is at most $O(N_U)$.

Next, we turn to the communication cost. First, in third building block, cloaking data, each user u_i sends $(\kappa_i - 1)$ pieces to other users. Thus, the communication cost in exchanging pieces is at most $T(N_U) = \sum_{\gamma=1}^{N_U} (\kappa_\gamma - 1)$ (i.e., $O(N_U)$). In fourth building block, estimating location, each user u_i sends κ_i pieces to the LS, and thus the communication cost is at most $T(N_U) = \sum_{\gamma=1}^{N_U} \kappa_\gamma$ (i.e., $O(N_U)$). In identifying location, each user receives $\max(k)$ localization output at most. So the communication cost is at most $T(N_U) = \sum_{\gamma=1}^{N_U} \max(k)$ (i.e., $O(N_U)$). In summary, the communication cost in P^3-LOC is at most $O(N_U)$.

We depict the total complexity in Table II.

Overall, Theorem 14 holds. ■

REFERENCES

[1] G. Fei, J. Niu, and L. Duan, "WAIPO: A fusion-based collaborative indoor localization system on smartphones," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2267–2280, Apr. 2017.

[2] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv.*, vol. 49, no. 2, p. 25, 2016.

[3] C. Wang, H. Lin, and H. Jiang, "CANS: Towards congestion-adaptive and small stretch emergency navigation with wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1077–1089, May 2016.

[4] T. Shu, Y. Chen, and J. Yang, "Protecting multi-lateral localization privacy in pervasive environments," *IEEE/ACM Trans. Netw.*, vol. 23, no. 5, pp. 1688–1701, Oct. 2015.

[5] T. Higuchi, P. Martin, S. Chakraborty, and M. Srivastava, "AnonyCast: Privacy-preserving location distribution for anonymous crowd tracking systems," in *Proc. ACM UbiComp*, 2015, pp. 1119–1130.

[6] P. Armengol, R. Tobkes, K. Akkaya, B. S. Çiftler, and I. Güvenç, "Efficient privacy-preserving fingerprint-based indoor localization using crowdsourcing," in *Proc. IEEE MASS*, Oct. 2015, pp. 549–554.

[7] P. Zhao *et al.*, "ILLIA: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous LBS queries," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1033–1042, Apr. 2018.

[8] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 2479–2487.

[9] M. Wang, Z. Zhang, X. Tian, and X. Wang, "Temporal correlation of the RSS improves accuracy of fingerprinting localization," in *Proc. IEEE INFOCOM*, Apr. 2016, pp. 1–9.

[10] H. Lim, L.-C. Kung, J. C. Hou, and H. Luo, "Zero-configuration, robust indoor localization: Theory and experimentation," in *Proc. IEEE INFOCOM*, Apr. 2006, pp. 1–12.

[11] M. Youssef, A. Youssef, C. Rieger, U. Shankar, and A. Agrawala, "PinPoint: An asynchronous time-based location determination system," in *Proc. ACM MobiSys*, 2006, pp. 165–176.

[12] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," in *Proc. IEEE INFOCOM*, Mar./Apr. 2003, pp. 1734–1743.

[13] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1281–1293, Jul. 2013.

[14] Z. Yang *et al.*, "Mobility increases localizability: A survey on wireless indoor localization using inertial sensors," *ACM Comput. Surv.*, vol. 47, no. 3, p. 54, Apr. 2015.

[15] H. Li, H. Zhu, and D. Ma, "Demographic information inference through meta-data analysis of Wi-Fi traffic," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1033–1047, May 2018.

[16] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 4, pp. 646–660, Jul./Aug. 2018.

[17] X. Wang, Y. Liu, Z. Shi, X. Lu, and L. Sun, "A privacy-preserving fuzzy localization scheme with CSI fingerprint," in *Proc. IEEE GLOBECOM*, Dec. 2015, pp. 1–6.

[18] V. Primault, S. B. Mokhtar, and L. Brunie, "Privacy-preserving publication of mobility data with high utility," in *Proc. IEEE ICDCS*, Jun./Jul. 2015, pp. 802–803.

[19] M. Backes, M. Humbert, J. Pang, and Y. Zhang, "walk2friends: Inferring social links from mobility profiles," in *Proc. ACM CCS*, 2017, pp. 1943–1957.

[20] T. Shu, Y. Chen, J. Yang, and A. Williams, "Multi-lateral privacy-preserving localization in pervasive environments," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 2319–2327.

[21] T. Wang and Y. Yang, "Analysis on perfect location spoofing attacks using beamforming," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2778–2786.

[22] H. Li, L. Sun, H. Zhu, X. Lu, and X. Cheng, "Achieving privacy preservation in WiFi fingerprint-based localization," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 2337–2345.

[23] Y. Liu, Z. Yang, X. Wang, and L. Jian, "Location, localization, and localizability," *J. Comput. Sci. Technol.*, vol. 25, no. 2, pp. 274–297, 2010.

[24] H. Jiang, P. Zhao, and C. Wang, "RobLoP: Towards robust privacy preserving against location dependent attacks in continuous LBS queries," *IEEE/ACM Trans. Netw.*, vol. 26, no. 2, pp. 1018–1032, Apr. 2018.

[25] E. D. Karim and J. Lampkins, "Disincentivizing/incentivizing malicious/honest behavior on the Internet via privacy-preserving appcoins," in *Proc. IEEE ICNP*, Oct. 2014, pp. 630–635.

[26] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proc. ACM CCS*, 2012, pp. 617–627.

[27] L. Wang, J.-H. Cho, R. Chen, and J. Chen, "PDGM: Percolation-based directed graph matching in social networks," in *Proc. IEEE ICC*, May 2017, pp. 1–7.

[28] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2399–2407.

[29] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness, Knowl. Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[30] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proc. ACM CCS*, 2015, pp. 1298–1309.

[31] R. Zhang, J. Shi, Y. Zhang, and C. Zhang, "Verifiable privacy-preserving aggregation in people-centric urban sensing systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 268–278, Sep. 2013.

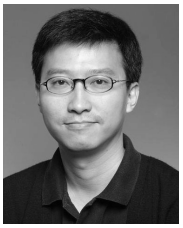
[32] G. Kellaris, S. Papadopoulos, X. Xiao, and D. Papadias, "Differentially private event sequences over infinite streams," *Proc. VLDB Endowment*, vol. 7, no. 12, pp. 1155–1166, 2014.

[33] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k-anonymity location privacy," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2994–3002.

[34] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "INCEPTION: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proc. ACM MobiHoc*, 2016, pp. 341–350.
[35] S. Lee, B. Kim, H. Kim, R. Ha, and H. Cha, "Inertial sensor-based indoor pedestrian localization with minimum 802.15.4a configuration," *IEEE Trans. Ind. Inform.*, vol. 7, no. 3, pp. 455–466, Aug. 2011.
[36] R. Shokri, G. Theodorakopoulos, J.-Y. L. Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE S&P*, May 2011, pp. 247–262.
[37] H. Liu, X. Li, H. Li, J. Ma, and X. Ma, "Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services," in *Proc. IEEE INFOCOM*, May 2017, pp. 1–9.
[38] X. Pan, J. Xu, and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1506–1519, Aug. 2012.
[39] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*. Berlin, Germany: Springer, 2006, pp. 265–284.

Ping Zhao received the B.E. degree from the Tianjin University of Science and Technology, China, in 2013, and the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, in 2018. In 2018, she joined Donghua University as a Faculty Member, where she is currently an Assistant Professor. Her research of interests is in the areas of mobile computing, information security, and Internet of Things.

Hongbo Jiang (M'08–SM'14) received the Ph.D. degree from Case Western Reserve University in 2008. He was a Professor with the Huazhong University of Science and Technology. He is currently a Full Professor with the College of Computer Science and Electronic Engeering, Hunan University. His research concerns computer networking, especially algorithms and protocols for wireless and mobile networks. He is serving as an Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING, and an Associate Technical Editor for the *IEEE Communications Magazine*.



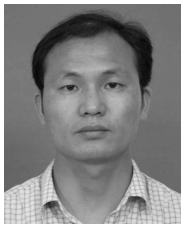
John C. S. Lui (M'93–SM'02–F'10) was born in Hong Kong. He received the Ph.D. degree in computer science from the University of California at Los Angeles in 1992. From 2005 to 2011, he was the Chairman of the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, where he is currently the Choh-Ming Li Professor. His current research interests are in communication networks, network/system security (such as cloud security and mobile security), network economics, network sciences (such as online social networks and information spreading), cloud computing, large-scale distributed systems, and performance evaluation theory.

Dr. Lui is an elected member of the IFIP WG 7.3, a Fellow of the ACM, and a Senior Research Fellow of the Croucher Foundation. He was the Chair of the ACM SIGMETRICS from 2011 to 2015. He received various departmental teaching awards, the CUHK Vice-Chancellor's Exemplary Teaching Award, and the CUHK Faculty of Engineering Research Excellence Award (2011–2012). He was a co-recipient of the Best Paper Award in the IFIP WG 7.3 Performance 2005, the IEEE/IFIP NOMS 2006, SIMPLEX 2013, and ACM RecSys 2017. He has been serving on the Editorial Board of the *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON COMPUTERS, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, *Journal of Performance Evaluation*, *Journal of Network Science*, and the *International Journal of Network Security*. He is currently a Senior Editor of the IEEE/ACM TRANSACTIONS ON NETWORKING.

Chen Wang (S'10–M'13) received the B.S. and Ph.D. degrees from the Department of Automation, Wuhan University, China, in 2008 and 2013, respectively. From 2013 to 2017, he was a Postdoctoral Research Fellow with the Networked and Communication Systems Research Laboratory, Huazhong University of Science and Technology, China. In 2017, he joined the Huazhong University of Science and Technology as a Faculty Member, where he is currently an Associate Professor. His research interests are in the broad areas of wireless networking, Internet of Things, and mobile computing, with a recent focus on privacy issues in wireless and mobile systems.

Fanzi Zeng received the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2005. Since 2005, he has been with the School of information science and engineering, Hunan University, Changsha, China, where he is currently a Professor. His general interests are in the areas of signal processing for wireless communications, estimation, and detection theory. His current research focuses on cognitive radio technology.

Fu Xiao received the Ph.D. degree in computer science and technology from the Nanjing University of Science and Technology, China. He is currently a Professor and a Ph.D. Supervisor with the School of Computer, Nanjing University of Posts and Telecommunications, China. His main research interests include wireless networking and sensor networks.



Zhetao Li received the B.Eng. degree in electrical information engineering from Xiangtan University in 2002, the M.Eng. degree in pattern recognition and intelligent system from Beihang University in 2005, and the Ph.D. degree in computer application technology from Hunan University in 2010. He was a Visiting Researcher with Ajou University in 2012. In 2013, he was a Clerk Fellow with the Department of Science and Technology, Ministry of Education, China. He is currently a Professor with the College of Information Engineering, Xiangtan University. His research interests include wireless networks and Internet of Things.