An Efficient Privacy-Preserving Fingerprint-Based Localization Scheme Employing Oblivious Transfer

Mengxuan Sun(✉), Xiaoju Dong, Fan Wu, and Guihai Chen

Shanghai Key Laboratory of Scalable Computing and Systems,
Shanghai Jiao Tong University, Shanghai, China
sunmengxuan@sjtu.edu.cn, {dong-xj,fwu,gchen}@cs.sjtu.edu.cn

Abstract. The tremendous growth of WiFi fingerprint-based localization techniques has significantly facilitated localization services. The traditional techniques pose a threat to both client's and server's privacies, because it is likely to reveal sensitive information about the client and the server during providing localization services. Many existing works have proposed privacy preserving localization schemes based on homomorphic cryptographic systems. However, the state of the art homomorphic cryptographic systems turn out to bear a time-consuming process for recourse-constrained devices. Hence, preserving location privacy while guaranteeing efficiency and usability is still a challenging problem. In this paper, we propose a privacy preserving indoor localization scheme employing oblivious transfer, called OTPri, to preserve the privacy of both clients and server in the process of localization in an efficient way. Our method enables a client to efficiently compute her location locally at client side with a small amount of additional overhead compared with the non-privacy-preserving scheme. Meanwhile, we conduct comprehensive experiments, including single-floor and multi-floor scenarios in our office building. The evaluation results demonstrate the efficiency improvement and overhead reduction of our proposed scheme compared with a classical privacy-preserving indoor localization scheme.

1 Introduction

The explosive popularity of portable mobile devices such as smartphones and tablets are fostering the emergence of location-aware applications and services for mobile users. Exemplary applications [19] include sounding the security alert when entering dangerous areas, posting location-based advertisements, locating a friend, etc. Due to the lack of GPS signals for indoor localization, a large body of research has come up with numerous techniques. A prevalent method is to measure received signal strength as a fingerprint, and match it with the sampled fingerprints in the database, including radio frequency [2,5,33], acoustic signals [12,23,24], infrared ray [28], GSM [21], the combination of ambience features [1,30], etc. A common idea among all these techniques is to reduce the site survey

L. Zhu and S. Zhong (Eds.): MSN 2017, CCIS 747, pp. 110–132, 2018.
https://doi.org/10.1007/978-981-10-8890-2_9

effort and distance error to realize indoor localization. Nevertheless, the majority of clients are reluctant to disclose their privacy while asking for an accurate service. Meanwhile, the server has to protect its database from unauthorized acquisition. Therefore, protecting privacy while guaranteeing the usability of the WiFi fingerprint-based localization system is an important foundation to ensure practicability.

Typically, the WiFi fingerprint-based localization system consists of two phases [29], including offline training and online operating phase. In the offline training phase, the server acquires the received signal strengths and the corresponding coordinate information of sampled locations and stores them in the database for future reference. During the online operating phase, a client who needs a localization service first measures the signal strength at current location and then submits it to the server for matching. Finally, the server employs an algorithm to determine client's location.

Although regarded as a promising approach for indoor localization, there are still considerable potential privacy leakages in such a paradigm of localization service. From client's perspective, the client has to expose its fingerprint and location directly to the server when requesting services, which will enable the server to trace client's location, and give third-party an opportunity to breach the client's privacy. Existing works indicate that the adversary can steal the individuals habits, activities, and relationships by location traces [18,25]. Therefore, the loss of privacy can lead to bad consequences [8], including location-based spams, damage of reputation or economic and physical violences. From server's perspective, although the transmission of complete database to the client may protect client's privacy [15], the database may be disclosed and utilized for commercial profit. Meanwhile, the continuous transmission of massive amounts of data will consume the device's resource, extend query process and compromise the network health. Moreover, the service provider has a strong demand for the protection for its WiFi fingerprint database from the unauthorized reveal. Hence, a privacy preserving scheme should be carefully designed to ensure confidentiality and usability.

According to the potential privacy leakage and the corresponding requirements, there are several challenges in designing a privacy-preserving localization scheme [7]. First, the scheme should meet both client's and server's requirements for data privacy, which means keeping their data safe from each other and malicious third party while acquiring all necessary information to achieve accurate localization, thus it makes the design much more complicated than the localization itself. Second, considering the resource-constrained characteristic of portable devices, the scheme should avoid complicated computation and large amount of communication overhead to ensure quick response and low cost, however, existing privacy-preserving localization schemes employing homomorphic cryptographic turn out to be a time-consuming process for portable devices and the performance degrades in larger scenarios. Third, since precision is a core objective of localization system, it directly influences the usability and user experience of the system. Nevertheless, the introduction of privacy-preserving function certainly will influence the accuracy of the localization scheme, thus we should improve

the accuracy to the greatest extent. Therefore, it is a challenging problem to achieve an overhead-performance balanced system while guaranteeing the data privacy of both sides.

To protect clients' location privacy in location-based services, some approaches including k-anonymity [17,31] and mix zones [4]) have been proposed. However, WiFi fingerprint-based localization lacks trusted third parties to apply them. Furthermore, these works submit the users' location information with the requests to protect the location privacy of the users requesting location-based services, which assumes that each client has obtained service without any privacy concern.

Considering the challenges as stated, toward this end, we are motivated to design a privacy-preserving localization system based on oblivious transfer [6] named OTPri.

To avoid the exposure of client's fingerprint, the localization scheme should protect any side information that may lead to a coarse estimate of the location in addition to the protection of the exact location of the client, meanwhile, the client has to provide as few information as possible to acquire the information that meets the requirement of accurate localization. Therefore, whenever a client needs to be localized, she sends an AP id from her vicinity to the server, then the server will choose the corresponding data entries which are stored at its side. Even though the id of the vicinity AP is exposed, the server can only confirm a wider area of the client's location, which meets the demand for client privacy. Moreover, due to the characteristic of oblivious transfer, the server cannot figure out which data entries the client has chosen, and the client cannot get any other information except her choices, besides, the server can put constraint on the number of data entries that the client can obtain during one localization process and the total number of requests a single client can achieve, thus preserving the server's data security to the greatest extent (Fig. 1).

The major contributions of this paper are summarized as two-fold:

- We present and formulate a privacy-preserving indoor localization scheme employing oblivious transfer to achieve a privacy-overhead-balanced construction to solve the privacy issues during the localization process. Meanwhile, we reduce the computation and communication overhead to the utmost extent, which has a decrease of nearly 40% considering the computation and communication overhead. And the localization process is conducted locally at the client side.
- We elaborate the privacy property for the proposed scheme and evaluate its performance by comprehensive experiments in both single-floor and multiple-floor scenarios in our office buildings. By comparing with existing privacy-preserving localization solutions, we verify the efficiency improvement and overhead reduction of the proposed scheme compared with PriWFL algorithm [6].

The remainder of this paper is organized as follows. In Sect. 2, we define the system model and present the threat model and technical preliminaries. In Sect. 3, we present the design motivation of the proposed privacy-preserving localization

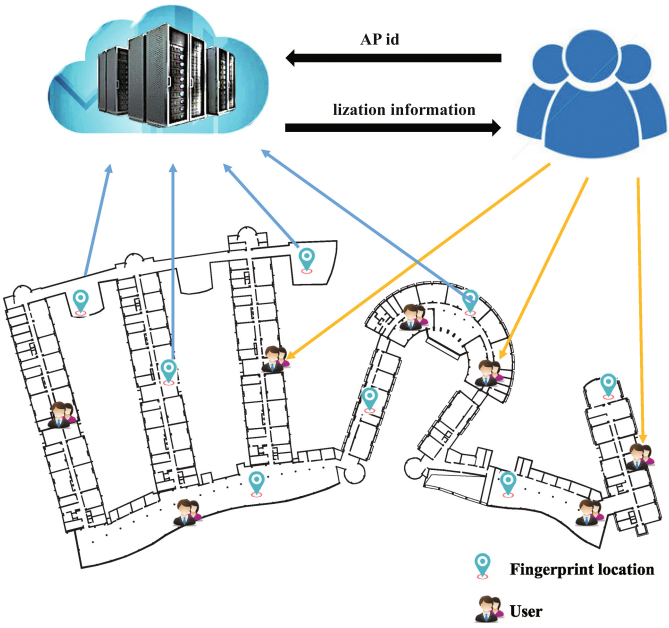

Fig. 1. OTPri system design

scheme, its details and privacy analysis. Section 4 reports our extensive experiments on this scheme. Section 5 briefly discusses the related work. Finally, in Sect. 6, we conclude our work.

2 Background and Attack Model

2.1 Overview of WiFi Fingerprint-Based Localization

WiFi fingerprint-based localization uses the WiFi signal strength to infer the location of a user. WiFi fingerprint-based localization is mainly composed of two phases [2], including offline phase and online phase. In the offline phase, the server selects N WiFi access points to represent fingerprints, then the service provider measures the average WiFi signal strength of the WiFi access points at M locations in the interested area, denoted as $V_i = \{v_1, v_2, ..., v_j, ..., v_N\}$, $i \in [1, M]$, where v_j is the average WiFi signal strength at (x_i, y_i) from the jth access point AP_j, and N is the totality of access points. Then the service provider stores the sampled fingerprints and their corresponding coordinates $(i, (x_i, y_i), V_i)$ in the WiFi fingerprint database D.

In the online phase, a client who intends to locate herself first measures the signal strengths from N access points, indicated as $V' = (v'_1, v'_2, ..., v'_j, ..., v'_N)$, and

sends its fingerprint to the server. Then, the server computes the Euclidean distances between V' and all the M sampled fingerprints as $d_i = ||V' - V_i||^2, i \in [1, M]$.

$$\begin{aligned} d_i &= \| V' - V_i \|^2 = \sum_{j=1}^{N} (v_{i,j} - v'_j)^2 \\ &= \sum_{j=1}^{N} v_{i,j}^2 + \sum_{j=1}^{N} (-2v_{i,j} * v'_j) + \sum_{j=1}^{N} v_j'^2 \end{aligned} \tag{1}$$

In the last step, the server selects k smallest values of d_i and finds out the corresponding coordinates of these $d_i s$, then estimates the client's location by computing the centroid of these locations.

2.2 Threat Model

The clients and service providers act in a semi-honest manner [10], in which they independently follow the protocol during localization process, but will try to extract useful information from the communication. Besides, we also assume that the third-party cannot steal privacy through eavesdropping the communication because of the encryption of the message sent between the client and server. Thus, in this paper, the prevention of privacy leakage in a normal localization operation is considered.

Our study considers both the client's location privacy and the service provider's data privacy. From the client's perspective, the client intends to acquire the localization service without compromising its location privacy. The location information can be theft by a curious service provider who collects the locations of the customers to make marketing and sales strategies or sells them for profit, namely *client privacy attack*. Therefore, the proposed scheme should prevent the attackers from stealing client's information, including the client's location and her sampled WiFi RSS signals, while providing accurate localization service.

From the service provider's perspective, the fingerprint database should be protected from unauthorized reveal. The database of the server may be downloaded or simulated by a malicious client to make profit, namely *database privacy attack*. Consequently, the server needs to protect its collected fingerprint database from learning or simulating by others in the process of localization.

2.3 Security Model

In privacy-preserving indoor localization scheme, we use the standard security model [9,11] in presence of semi-honest participants, in which the client and the server will follow the scheduled protocol, but might try to compute additional information by received messages. We use simulation argument to define security in this setting: if no additional information is revealed to the participants during protocol execution, which means no party can compute the view of protocol

execution using that party's input and output only, the protocol is unconditionally secure. The notion of privacy-preserving for semi-honest participants is formalized using the definition below:

Definition 1. The client and the server engage in a protocol t, in which they cooperatively compute function $f(in_1, in_2) = (out_1, out_2)$, where in_i and out_i respectively represent input and output of client and server. During the execution of protocol t, we use $VIEW_t(P_i)$ to denote the view of a participant. More precisely, the participants' input, random coin tosses r_i and messages $m_1, ..., m_t$ passed between the parties during protocol execution form $P_i's$ view: $VIEW_t(P_i) = (in_i, r_i, m_1, ..., m_t)$. We define time simulator S_i such that:

$$S_i(in_i, f(in_1, in_2)) \equiv VIEW_p(P_i), out_i \tag{2}$$

where "$\equiv$" denotes computational indistinguishability. If for each party P_i, such a probabilistic polynomial time simulator exists, the protocol t is unconditionally secure.

Indistinguishability. Two probability ensembles X_i and Y_i, indexed by i, are (computationally) indistinguishable if for any PPTM D, polynomial $p(n)$ and sufficiently large i, it holds that

$$|Pr[D(X_i) = 1] - Pr[D(Y_i) = 1]| \leq 1/p(i) \tag{3}$$

2.4 Security Assumptions

For our privacy-preserving scheme against semi-honest client, we assume the hardness of Decisional Diffie-Hellman (DDH) problem [6].

Decisional Diffie-Hellman (DDH). Let $p = 2q + 1$ where p, q are two primes, and G_q be the subgroup of Z_p^* with order q. The following two distribution ensembles are computationally indistinguishable:

- $Y_1 = (g, g^a, g^b, g^{ab})_{G_q}$, where g is a generator of G_q, and $a, b \in {}_R Z_q$.
- $Y_2 = (g, g^a, g^b, g^c)_{G_q}$, where g is a generator of G_q, and $a, b, c \in {}_R Z_q$.

2.5 k-out-of-n Oblivious Transfer

In this paper, we adopt k-out-of-n Oblivious Transfer [6] to protect both the client's privacy and the server's fingerprint database. Therefore, we briefly review the fundamental of k-out-of-n oblivious transfer.

Oblivious transfer (OT) is an important primitive used in many cryptographic protocols. An oblivious transfer protocol involves two parties, the sender S and the receiver R. S has some messages and R wants to obtain some of them via interaction with S. The security requirement is that S wants R to obtain the message of her choice only and R does not want S to know what she chooses. A k-out-of-n OT (OT_n^k) scheme is an OT scheme in which R chooses k messages at the same time, where $k < n$.

The sender S has n secret messages $m_1, m_2, ..., m_n$ from message space G_q, and the semi-honest receiver R wants to get k of them.

In our scheme, there is no need for trapdoor specification or initialization, which means the system parameters can be repeatedly used by all senders and receivers, and each pair of sender and receiver does not need to hold any secret key.

3 Design of OTPri

In this section, we propose the construction of our privacy preserving indoor localization scheme employing oblivious transfer (OTPri), a WiFi fingerprint-based indoor localization employing oblivious transfer to preserve privacy.

3.1 Preliminary Design

Our proposed scheme protects both client's and server's information, and achieves high efficiency in the process of indoor localization. The key idea of our scheme is to mask the query by oblivious transfer [6], thus the server cannot know the client's choice. Meanwhile, the client only obtains the information of her choice. We demonstrate the challenges and our corresponding solutions in this subsection.

Privacy Preservation in Indoor Localization. To avoid the exposure of client's fingerprint, the localization scheme should protect any side information that may lead to a coarse estimate of the location in addition to the protection of the exact location of the client. In the query process of other fingerprint-based localization methods, the server has access to client's fingerprint and estimated location. Therefore, the privacy of client's location is leaked to the server. To avoid this kind of threats, we integrate oblivious transfer with traditional fingerprint-based localization scheme, which allows the client to choose the necessary information for localization on her own and keep her choices from the others. Meanwhile, due to the characteristic of oblivious transfer, the client cannot learn anything other than her choices. And the server can put constraint on the number of data entries that the client can obtain during one localization process and the total number of requests a single client can achieve, thus preserving the server's data security to the greatest extent.

Time Efficiency. Privacy-preserving indoor localization has been researched very extensively in the last few years. Many schemes use homomorphic encryption. However, it requires computationally expensive public-key operations that scale very inefficiently for larger security parameters, which is a time-consuming process for resource-constrained devices such as smartphones. Moreover, the user has to generate a pair of keys each time when it needs to be localized. To shorten the process of query, our scheme employs oblivious transfer in which the parameters can be used repeatedly by all possible clients and servers without any initialization.

3.2 Scheme Details

Our scheme involves three phases, including Pre-Process Phase, Oblivious Transfer Phase and Location Determination Phase.

Pre-process Phase. After collecting the fingerprints from a building $(i, (x_i, y_i), V_i = ((v_{ij})|_{j=1}^{N})_{i=1}^{M})$, where i is an index, M is the total number of sampled locations, N is the totality of APs, (x_i, y_i) is the coordinate of the specific location, V_i represents the WiFi fingerprint at the specific location (x_i, y_i), the server stores the results in a 2-D matrix $MATRIX[N][M]$, which records the RSS value of N APs at M geo-locations. Moreover, the server stores the table $T = (i, (x_i, y_i))_{i=1}^{M}$ which records the indices and their corresponding coordinates of sampled locations. The Radiomap MATRIX can be of the following format:

$$\begin{gathered}\text{Radiomap(MATRIX)}\\ AP_{1,1}, AP_{1,2}, ..., AP_{1,M} \Rightarrow x_1, y_1\\ AP_{2,1}, AP_{2,2}, ..., AP_{2,M} \Rightarrow x_2, y_2\\ AP_{3,1}, AP_{3,2}, ..., AP_{3,M} \Rightarrow x_3, y_3\\ ...\\ AP_{N,1}, AP_{N,2}, ..., AP_{N,M} \Rightarrow x_N, y_N\end{gathered}$$

Each row in this radiomap represents a data entry. This process can be executed before a client uses the localization service, and only needs to be performed once. Whenever a client needs a localization service, she first measures its WiFi RSS value of each AP at current location, denoted as $V' = (v'_1, v'_2, ..., v'_N)$. Then she chooses one AP-id, named j, from her vicinity and sends this AP-id (j) to the server. After receiving the port number of the AP, the server searches its MATRIX and finds out all the data entries that have nonzero signal value at this AP, where $AP_{i,j} \neq 0$, $i \in (\alpha_1, \alpha_2, ..., \alpha_l)$, then the server renumbers the indices of those data entries and forms a map $(i, signal)_{i=1}^{l}$ between the renumbered indices and their corresponding signal values at this AP, after that the server sends the map to the client. Meanwhile, the server finds out the union set of APs that have nonzero signal value at these data entries, $C = \{\beta_1, \beta_2, ..., \beta_n\}$ where $AP_{\alpha_1,\beta_1}, AP\alpha_2, \beta_2, ..., AP\alpha_l, \beta_n \neq 0$, and sends the set of port numbers of these APs $(C = \{\beta_1, \beta_2, ..., \beta_n\})$ to the client, which indicates the delivery order of signal values at Oblivious Transfer Phase. The sever will send these signal values and their corresponding coordinates by column. For example, if 2 is included in $\{\beta_1, \beta_2, ..., \beta_n\}$, then the server will send the signal value at AP_2 of these chosen data entries in a transfer process in Oblivious Transfer Phase.

Oblivious Transfer Phase. Considering the map $(i, signal)_{i=1}^{l}$ received by the client, first, the client finds out the k nearest data entries based on their RSS values, then masks her choices $C = \{\sigma_1, \sigma_2, ..., \sigma_k\}$ by oblivious transfer. Eventually, the server sends other APs' signal values in accordance with the order in $C = \{\beta_1, \beta_2, ..., \beta_n\}$ and their corresponding coordinates one by one.

For system parameters, g, h is two generators of G_q, and $\log_g h$ is not revealed to any party. (g, h, G_q) are universal parameters, which means they can be used repeatedly by all possible clients and the server if $\log_g h$ is not revealed.

During each transfer, the server sends the signal value of l data entries at AP_j, $j \in \{\beta_1, \beta_2, ..., \beta_n\}$, denoted as $m_1, m_2, ..., m_l$. The procedure of each transfer of signal value is as follows:

- System parameters: (g, h, G_q);
- Server has messages: $m_1, m_2, ..., m_l$;
- Client's choices: $\sigma_1, \sigma_2, ..., \sigma_k$;

1. Client chooses two polynomials:

$$f(x) = a_0 + a_1 x + ... + a_{k-1} x^{k-1} + x^k \tag{4}$$

$$f'(x) = b_0 + b_1 x + ... + b_{k-1} x^{k-1} + x^k \tag{5}$$

where $a_0, a_1, ..., a_{k-1} \in Z_q$ and $b_0 + b_1 x + ... + b_{k-1} x^{k-1} + x^k = (x - \sigma_1)(x - \sigma_2)...(x - \sigma_k)$ mod q.

2. Client to Server:

$$\begin{aligned} A_0 &= g^{a_0} h^{b_0} \\ A_1 &= g^{a_1} h^{b_1} \\ &... \\ A_{k-1} &= g^{a_{k-1}} h^{b_{k-1}} \end{aligned} \tag{6}$$

3. Server computes

$$d_i = (g^{k_i}, m_i B_i^{k_i}) \tag{7}$$

where $k_i \in Z_q^*$ and

$$\begin{aligned} B_i &= g^{f(i)} h^{f'(i)} \\ &= A_0 A_1^i ... A_{k-1}^{i^{k-1}} (gh)^{i^k} \; mod \; p \end{aligned} \tag{8}$$

for $i = 1, 2, ..., l$.

4. Server to Client: $d_1, d_2, ..., d_l$.
5. Let $d_i = (U_i, V_i)$, the client computes $m_{\sigma_i} = V_{\sigma_i} / U_{\sigma_i}^{f(\sigma_i)}$ mod p for each σ_i.

First, the client constructs a k-degree polynomial $f'(x)$, which satisfies $f'(i) = 0$ if and only if $i \in \{\sigma_1, \sigma_2, ..., \sigma_k\}$. Next, another random k-degree polynomial $f(x)$ is selected to mask the chosen polynomial $f'(x)$. Then, the client sends the masked choices $A_0, A_1, ..., A_{k-1}$ to the server.

After the server receives these requests, he first computes $B_i = g^{f(i)} h^{f'(i)}$ by computing $A_0 A_1^i ... A_{k-1}^{i^{k-1}} (gh)^{i^k}$ mod p. The server has no idea of which $f'(i)$ is equal to zero, for $i = 1, 2, ..., n$ because of the random polynomial $f(x)$. Next, the server encrypts each message m_i by public key B_i. Then, the server sends the encrypted messages $d_1, d_2, ..., d_k$ to the client.

For each d_i, $i \in \{\sigma_1, \sigma_2, ..., \sigma_k\}$, since $B_i = g^{f(i)}h^{f'(i)} = g^{f(i)}h^0 = g^{f(i)}$, the client can get these messages with secret key $f(i)$. If $i \notin \{\sigma_1, \sigma_2, ..., \sigma_k\}$, since the client cannot compute $(g^{f(i)}h^{f'(i)})^{k_i}$ with the knowledge of g^{k_i} and $f(i), f'(i)$ only, the client gets no access to the message m_i.

Correctness. For each message received by the client $d_i = (U_i, V_i)$, the chosen messages m_{σ_i}, $i = 1, 2, ..., k$, are computed as

$$\begin{aligned} V_{\sigma_i}/U_{\sigma_i}^{f(\sigma_i)} &= m_{\sigma_i} * (g^{f(\sigma_i)}h^{f'(\sigma_i)})^{k_{\sigma_i}}/g^{k_{\sigma_i}f(\sigma_i)} \\ &= m_{\sigma_i} * (g^{f(\sigma_i)} * 1)^{k_{\sigma_i}}/g^{k_{\sigma_i}f(\sigma_i)} \\ &= m_{\sigma_i} \end{aligned} \tag{9}$$

Location Determination Phase. In this phase, the user computes the squared Euclidean distance d_i between V' and V_i, $i = 1, 2, ..., k$. Then, it sorts the distances and determines the q smallest distances $d_{I_1}, d_{I_2}, ..., d_{I_q}$. These q nearest neighbors form C. Finally, the client estimates her location by computing the centroid of the q neighbors.

$$\begin{cases} \parallel V'_h - V_1 \parallel = \sum_{j=1}^{N}(v'_{h,j} - v_{1,j})^2 = d_{h,1} \\ \parallel V'_h - V_2 \parallel = \sum_{j=1}^{N}(v'_{h,j} - v_{2,j})^2 = d_{h,2} \\ ... \\ \parallel V'_h - V_k \parallel = \sum_{j=1}^{N}(v'_{h,j} - v_{k,j})^2 = d_{h,k} \end{cases} \tag{10}$$

$$\begin{cases} x = \frac{\sum_{j \in C} x_j}{q} \\ y = \frac{\sum_{j \in C} y_j}{q} \end{cases} \tag{11}$$

3.3 Parameter Set

Considering the accuracy-overhead balanced construction and the restriction of oblivious transfer protocol, we must carefully choose parameters which are involved in the localization process.

Locations and Access Points. As the previous descriptions have stated, N access points are selected to measure WiFi signal strengths to represent a specific indoor location, and these access points should be chosen to efficiently differentiate each location. For example, an AP that has very low RSS values on all locations should be eliminated due to its neglectable effect on localization process. Moreover, the M sampled locations should be distinct from each other and evenly distributed to represent the building's floor plan as detailed as possible.

Number of Data Entries in the Map. After receiving the port number of the AP from the client, the server searches its MATRIX and finds out all the data entries that have nonzero signal value at this AP, then the server forms a

map $(i, signal)$, $i \in [1, l]$ and sends it to the client. Note that the number of data entries in the map is the totality of choices in Oblivious Transfer Phase, which remains as a constant during localization process, denoted as l. Different queries may have different l because the AP id chosen by the client may vary when her current location varies, and the number of data entries that have nonzero signal value at this chosen AP varies as the AP id changes. According to the system process displayed before, l has a deep influence on system performance.

Size of *k*. In Oblivious Transfer Phase, the client finds out the k nearest data entries based on their RSS values, then masks her choices by oblivious transfer. An oversize k will lead to the unnecessary exposure of server's database and extend the process of query. However, if the size of k is too small, the alternative set will be too small that the client will be unable to compute her location accurately. Thus, an appropriate k is needed to balance precision, privacy and overhead.

3.4 Communication Cost

The communication cost in localization scheme mainly centers on Oblivious Transfer Phase which uses two rounds, $O(k)$ messages are sent in the first round, where the client asks for k data entries from the server, then $O(l)$ messages are sent in the second round, where the server responds for the request. As for computation, the client computes $3k + 2$ and the server computes $(k + 2)l$ modular exponentiations. The complexity analysis results are summarized in Table 1.

Table 1. Complexity analysis results

Phases	Communication	Computation (Exp)
Client to Server	$O(k)$	$3k + 2$
Server to Client	$O(l)$	$(k + 2)l$

3.5 Security Analysis

In Sect. 2, we presented two attack models related to the privacy of user and database. The security analysis of these models is represented in this section.

Theorem 1. OTPri is resistent to user privacy attack.

Proof. Pre-process Phase: From the client's perspective, the client only provides the server with the id of AP from her vicinity. Therefore, the server can only confirm a wider area of the client's location.

Oblivious Transfer Phase: Client's privacy-indistinguishability-If there is x in C, but not in C', or vice versa, we say two sets C and C' are diverse. In oblivious transfer phase, the transcript of the choice $C = \{\sigma_1, \sigma_2, ..., \sigma_k\}$ received

by the server is indistinguishable from the transcript of $C' = \{\sigma'_1, \sigma'_2, ..., \sigma'_k\}$. Which means if the received messages of the server for C and C' are identically distributed, the choices of the client are unconditionally secure.

For choices $C = \{\sigma'_1, \sigma'_2, ..., \sigma'_k\}$, every tuple $(b'_0, b'_1, ..., b'_{k-1})$ that represents the choices corresponds to a tuple $(a'_0, a'_1, ..., a'_{k-1})$ that satisfies $A_i = g^{a'_i} h^{b'_i}$ for $i = 0, 1, ..., k-1$. Therefore, the client's choices are unconditionally secure.

Location Determination Phase: Since Location Determination Phase is conducted locally at the client's side, neither server nor any other third party can acquire the client's location or her sampled WiFi RSS signals, thus the client's location privacy is naturally preserved.

Theorem 2. OTPri is resistent to database privacy attack.

Proof. Pre-Process Phase: Since the index of map $(i, signal)$, $i \in [1, l]$ sent from the server to the client is renumbered before sending it to the client, the client is unable to realize the original mapping between index and RSS signal value of the MATRIX.

Oblivious Transfer Phase: Server's security-indistinguishability-For any choices that don't belong to set $C = \{\sigma_1, \sigma_2, ..., \sigma_k\}$, they should be indistinguishable from the random ones. Which means the client gets no information about messages m_i if she is semi-honest, $i \notin \{\sigma_1, \sigma_2, ..., \sigma_k\}$.

We prove that m_is look random if the DDH assumption holds, $i \notin \{\sigma_1, \sigma_2, ..., \sigma_k\}$. First, the random variable for the unchosen messages is defined below:

$$C = (g, h, (g^{k_{i_1}}, m_{i_1}(g^{f(i_1)} h^{f'(i_1)})^{k_{i_1}}), ..., \\ (g^{k_{i_{i-k}}}, m_{i_{n-k}}(g^{f(i_{n-k})} h^{f'(i_{n-k})})^{k_{i_{n-k}}}))$$

where $k_{i_1}, k_{i_2}, ..., k_{i_{n-k}} \in {}_R Z_q^*$. Since the polynomial $f(x)$ and $f'(x)$ are selected by the client, besides, $f'(i_1), ..., f'(i_{n-k}) \neq 0$, C can be simplified as below:

$$C' = (g, h, (g^{k_{i_1}}, h^{k_{i_1}}), ..., (g^{k_{i_{n-k}}}, h^{k_{i_{n-k}}}))$$

In multiple samples, the indistinguishability is preserved. Therefore, the prove of the following two distributions

- $C = (g, h, g^r, h^r)$, where $h \neq 1$, $r \in {}_R Z_q^*$
- $X = (g, h, x_1, x_2)$, where $h \neq 1$, $x_1, x_2 \in {}_R G_q$

are distinguishable by a polynomial-time distinguisher D is necessary. To solve the DDH problem, we can construct another polynomial-time machine D', whose sub-routine is D.

Machine D'
Input : (g, u, v, w) (eitherfrom Y_1 or Y_2 in DDH)
Output : $D(g, u, v, w)$

If D distinguishes C and X with non-negligible advantage ε, D' distinguishes Y_1, Y_2 in the DDH problem with at least non-negligible advantage $\varepsilon - 2/q$, where $dist(C, Y_1) = 1/q$ and $dist(X, Y_2) = 1/q$.

Therefore, the indistinguishability is naturally proved.

Even if the communication between the server and client is intercepted by a third-party, due to the characteristic of oblivious transfer, the third-party cannot obtain the data because the message between the server and client is encrypted in Oblivious Transfer Phase, and the decryption requires the knowledge of client's choices, which are only grasped by the client.

4 Experiment Results

In this section, several experiments are conducted to demonstrate the performance of OTPri scheme. And We compare the performance of OTPri system with another privacy-preserving fingerprint-based localization system PriWFL [15].

4.1 Experiment Setup

To simulate the real circumstances to the greatest extent and thoroughly evaluate the performance of OTPri, we employ two scenarios in our experiment. One is a single-floor scenario, and the other one is a two-floor scenario. The experiment is carried out in our department's compound buildings, consisting of five buildings connected by two corridors, including laboratory rooms with different sizes, long narrow corridors and arc spaces. Its purpose is to evaluate the performance of our scheme in the context of a large and complicated scenario.

In the data collection process of our experiment, at each sampled location, we measure the values of received signal strength of 425 WiFi APs and sustain the measurement for 30 s to record the change of the RSS values during this period, then take the average as the averaged RSS values. We use DELL Vostro 2420 Laptop with Linux system and an IEEE 802.11 Atheros Communications AR9485 Wireless Network Adapter to receive signals from each access point. Here we regard the various APs as the same because the type of them cannot be controlled.

4.2 Performance in a Single Floor

First, our experiment is conducted in the first floor of our office building. We choose 111 points to build the database and 40 points as queries. The queries are selected to represent as many typical places as possible. This data set contains 353 APs. Our experiments in this section mainly discuss how variables influence the precision and time performance of our proposed scheme. We use localization precision and time cost for a query as two metrics to evaluate the performance of our scheme.

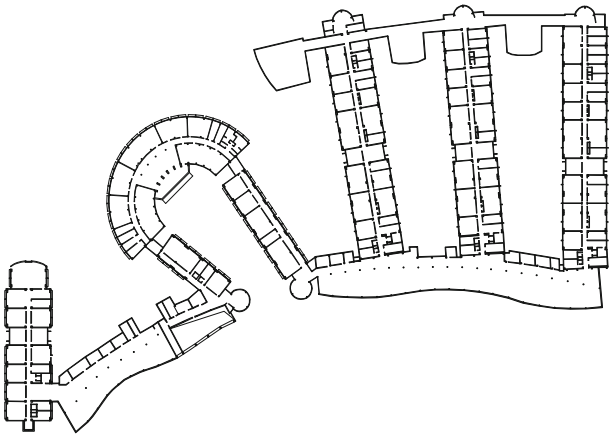

Fig. 2. Floor plan of department building

Precision vs. *k*. In this section, we vary k to explore the relationship between the precision of our scheme and the variable k. As analyzed before, an appropriate k is needed to balance precision, privacy and overhead. Figure 3 shows the cumulative distribution function (CDF) of localization errors both in the baseline algorithm PriWFL and the proposed scheme OTPri in this paper. As depicted in the figure, OTPri provides a 40% error of 3.6 m and a 80% error of 6.6 m when $k = 5$, then the precision remains almost the same when k continues to grow, which achieves a similar accuracy with PriWFL. In order to obtain the best performance, at least 5 candidates are needed to determine the client's location (Fig. 2).

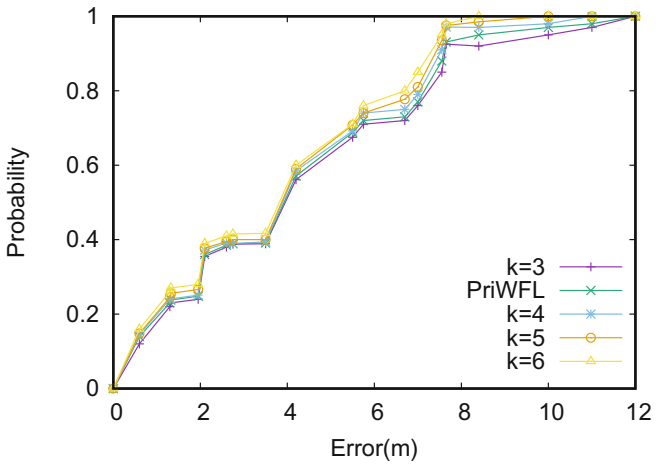

Fig. 3. Precision in a single floor when k varies

Time vs. *l*. As analyzed in Sect. 3, the major time cost is from Oblivious Transfer Phase. Since it is k-out-of-l oblivious transfer, the scheme performance will be influenced as l varies. As experiment setup, we set k as 3 and configure l as 6, 8, 10, 12 and 14 to evaluate the average run time of the scheme. Figure 4 depicts the relationship between the time cost and l. From this figure we can observe that the time cost increases from 3.003 s to 7.018 s as l increases. Compared with other privacy-preserving scheme that costs at least 12 s each query when 15 access points are considered [15], our scheme is more practical and efficient, thus showing enormous potential in practical utility.

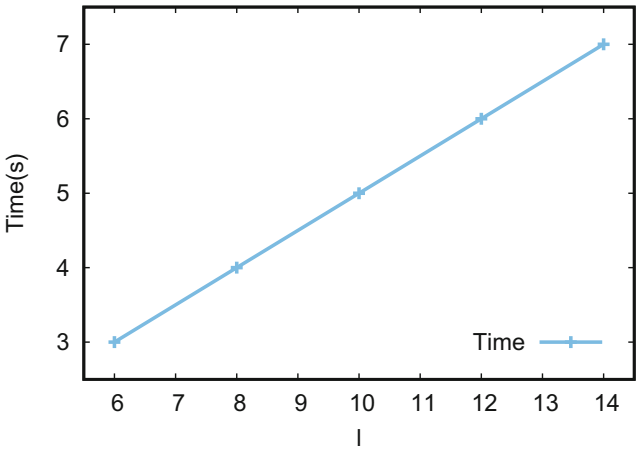

Fig. 4. Time for query in a single floor when l varies

Time vs. *k*. In this section, we evaluate the time performance of the scheme when k varies, we set l as 8 by default and configure k as 2, 3, 4, 5 and 6, then measure the time cost under different k settings. As depicted in Fig. 5, the time cost increases from 3.153 s to 7.250 s as k increases.

4.3 Performance in Multiple Floors

We test the performance of our scheme in a multi-floor scenario. In this scenario, there are as many as 425 APs. The fingerprint database consists of 222 locations and 84 queries, separately 40 and 44 for the first and second floor. The arrangement of fingerprint database locations and queries are mostly symmetric in the two floors. Apart from (x, y) coordinates to locate the points, we add the third dimension z to form a 3D scenario. Separately we assume the points in the first and second floor have $z = 8$ and 12. When the estimated z for each query is less than 10, we believe it is in the first floor and when z is larger than 10, it is in the second floor. We analyze the precision achieve and time cost reduction by our scheme in this section.

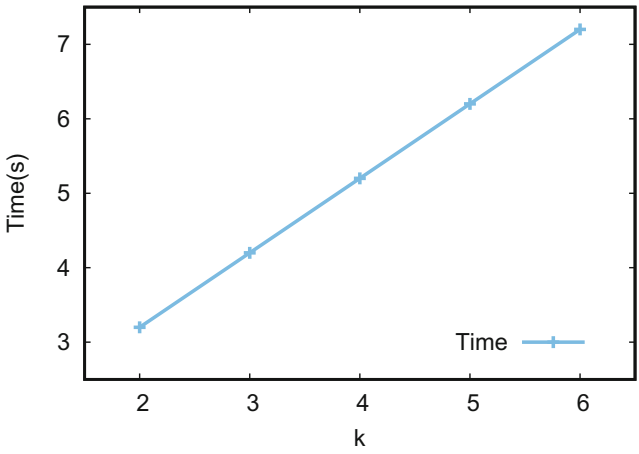

Fig. 5. Time for query in a single floor when k varies

Precision in Multiple Floors. First we calculate the cumulative distribution function (CDF) of the error. The default configuration is $k = 5$. As shown in Fig. 6, our OTPri scheme performs with a 25% error of 1.9 m and 60% error of 4.7 m, while PriWFL has a 25% error of 2.3 m and 60% error of 5.0m. It is obvious that our scheme works better in this case. We can see from this result that turning the 2D floor plan into 3D does not influence much of the precision. In fact, of all queries, 100% of them are located correctly to their own floors. The maximum estimated height of the first-floor points is 8.8 m and the minimum estimated height of the second-floor points is 10.7 m.

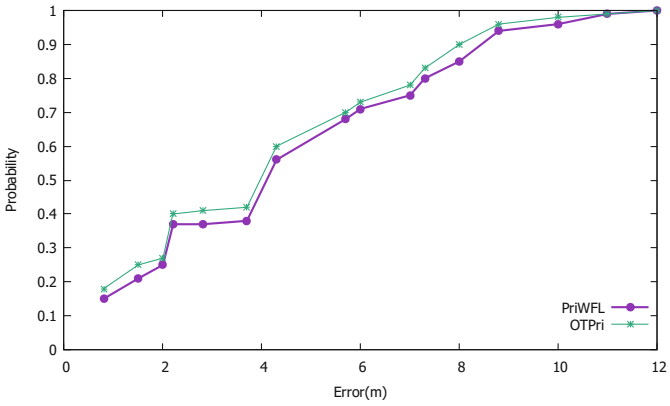

Fig. 6. Precision in multi-floor scenario

To further look into how floors are identified for the queries, we record the top 5, 10, 20, 50 and 100 candidates with the nearest signal values at a specific AP and count their occurrences. The histogram is shown in Fig. 7. Averagely 92.5% of the top 5 candidates for queries in floor 1 are in the same floor. For queries in floor 2, the number is 90.9%. This is enough to identify a query's floor since most choices of the client are within 5. Naturally, as c grows, there's a tendency that this percentage decreases because more candidates that have a near distance in other floors are added. As c increases, the percentage of candidates being in the same floor with queries approximates 0.5.

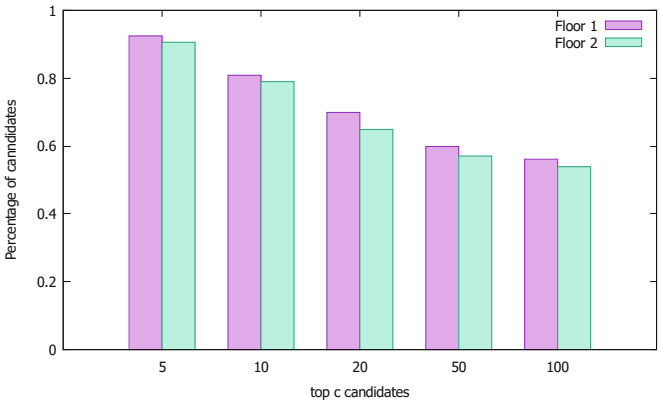

Fig. 7. Percent of candidates

Time Cost in Multiple Floors

Time vs. l. As experiment setup, we set k as 5 and configure l as 6, 8, 10, 12 and 14 to evaluate the average run time of the scheme. Figure 8 depicts the relationship between the time cost and l. From this figure we can observe that the time cost increases from 3.013 s to 6.961 s as l increases. Comparatively, our scheme shows a great advantage in practical uses even in larger scenario. Moreover, the query time grows linearly with l, showing a predictable upper bound for any l.

Time vs. k. We evaluate the time performance of the scheme in multi-floor scenario when k varies, we set l as 8 by default and configure k as 2, 3, 4, 5 and 6, then measure the time cost under different k settings. As depicted in Fig. 9, the time cost increases from 3.253 s to 7.321 s as k increases.

Analysis on Wi-Fi Access Point Number and Database Size. So far in all of our experiments, the full database is used, consisting of 425 Wi-Fi access points and 222 database locations in the two floors. However, in this section, we perform a sensitivity analysis to see how many APs and locations in the database

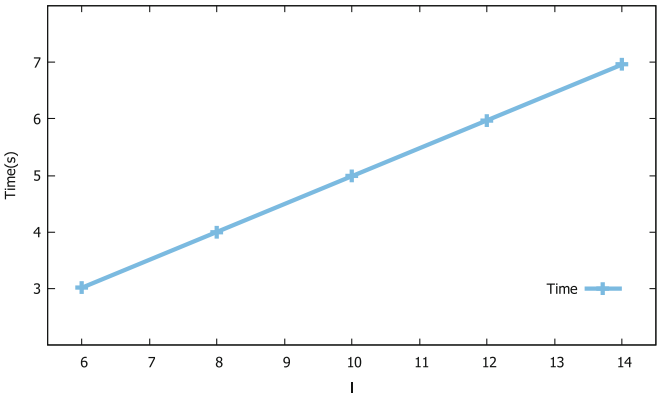

Fig. 8. Time for query in multi floors when l varies

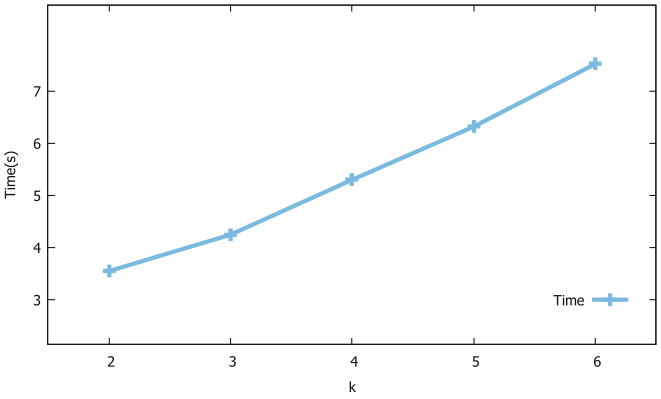

Fig. 9. Time for query in multi floors when k varies

are actually needed to calculate an accurate coordinate of a query. We keep certain portion of APs and database points, and run our OTPri scheme to see the median value of errors. Specifically, the APs and points are chosen evenly with a certain stride. For instance, we select APs with order number 1, 4, 7, 10, ...

Figures 10 and 11 shows the median error when different number of Wi-Fi access points n and location points m are used. It is obvious that larger m and n both lead to smaller errors. In order to obtain the best performance, at least 50 Wi-Fi APs and 80% of the fingerprint database are needed. In the case of Wi-Fi APs, the median error decreases sharply from around 20 m to 3 m until AP number reaches 50, then it remains almost the same when AP number continues to grow. So we can safely use an AP number of 50 in this scenario. As for the points in the fingerprint database, though the median error also declines very fast till 60% of the database is used, it continues to decline at a much lower rate when more than 60% is used, indicating that a larger data set still leads to a better performance.

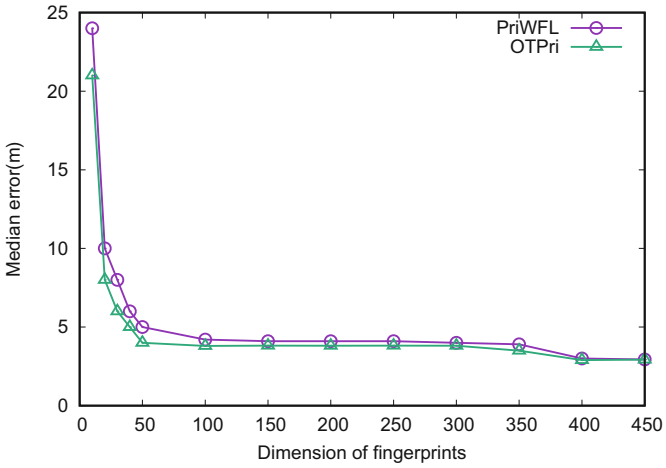

Fig. 10. Median error with different AP numbers

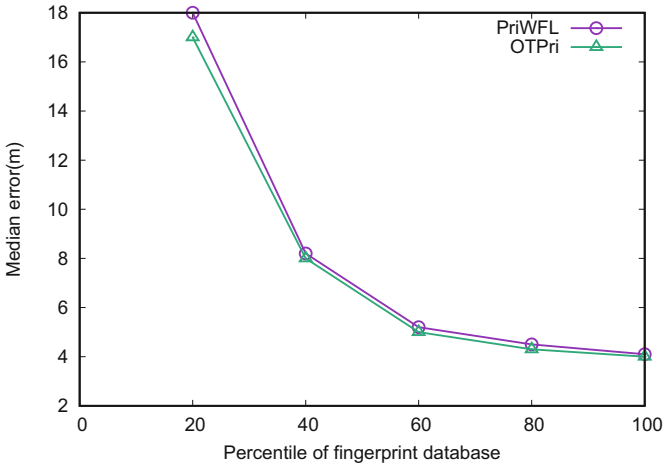

Fig. 11. Median error with different database size

Bandwidth Cost in Oblivious Transfer Phase. In this section, we configure k as 5 and the number of WiFi access points as 425, then investigate the impact of database size M on the bandwidth cost of Oblivious Transfer Phase and compare the results with PriWFL [15]. We observe from Fig. 12 that as M increases from 50 to 400, the bandwidth cost increases from 6.018 KB to 50.009 KB, which is much less than PriWFL [15].

Discussion. The experimental results show that our scheme achieves a similar performance approach in terms of precision but significantly lower online computation overhead and thus total protocol execution latency and energy

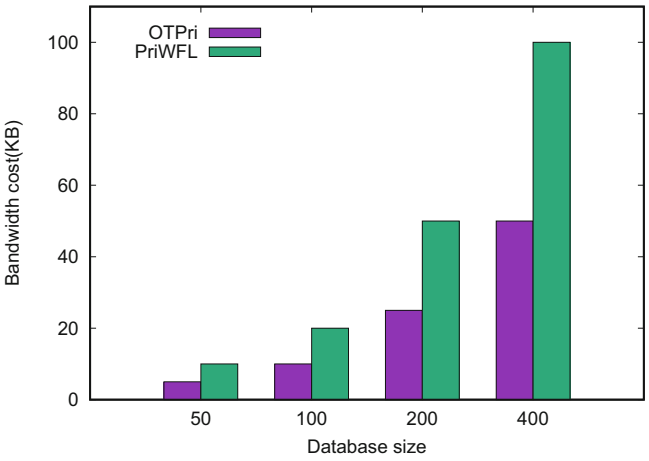

Fig. 12. Bandwidth cost

consumption compared with PriWFL, making it more practical than PriWFL to realize privacy-preserving indoor localization.

Two directions need to be investigated to further reduce the execution time delay of our scheme. First, we can employ advanced network with higher transmission speed to reduce the transmission time [32]. Since a growing number of existing mobile devices support advanced network, the transmission time will be substantially reduced. Second, we can further optimize the Java implementation of Oblivious Transfer [13] to reduce the online operation time.

5 Related Work

In this section, we discuss two related research works, including indoor localization based on signal-fingerprint and privacy-preserving indoor localization.

5.1 Fingerprint-Based Indoor Localization

Among all of the fingerprint-based localization schemes, there are other methods apart from harnessing WiFi signal strengths as fingerprints for locating. Classical works including the famous RADAR [2,3], Horus [33], OIL [22], PlaceLab [14], LANDMARC [20] employ RF signals as identification for each location. And other sources of signatures have also been explored, including geo-magnetism, FM radio [5], background acoustic noise, etc. SurroundSense [1] uses ambience features like sound, light, color as fingerprints, thus the options for fingerprinting techniques are largely diversified. Through abstracting these signatures into fingerprints, a location in indoor area can be represented by these fingerprints, and a nearer fingerprint usually indicates a nearer location, thus we can utilize these fingerprints to estimate user's location. However, these approaches demand a thorough site survey to build up the fingerprint database. If these fingerprints can be abstracted into numbers, our proposed scheme can be applied.

5.2 Privacy-Preserving Indoor Localization

To protect the client's location privacy and the server's data security during localization process, several techniques have been proposed to achieve privacy-preserving indoor localization. A common approach is to encrypt the communication between users and servers with cryptosystems. Li et al. proposed Pri-WFL scheme [15] to encrypt the localization process with Paillier cryptosystem. Though secure enough, the performance decreases in larger scenarios. Li and Jung [16] designed a suite of privacy-preserving location query protocols to balance the required privacy guarantee and computation overhead. In another similar work, Shu et al. [26] employed information hiding and homomorphic encryption techniques to design multi-lateral privacy-preserving localization protocols for three privacy levels. Other schemes, including k-anonymity [11,17,31] and mix zones [4], employ pseudonyms to prevent server from tracking of user's real location or its long-term movements. Vu et al. [27] used locality-sensitive hashing to partition the users into k-anonymous groups. There are four categories in protection strategy of location privacy: (1) regulatory approaches, (2) privacy policy based approaches, (3) anonymity based approaches, and (4) obfuscation based approaches.

Most of these techniques require communication through a trusted third intermediary, which may not be not practical in some real-life settings.

6 Conclusion and Future Work

In this work, we have proposed a privacy-preserving WiFi fingerprint-based localization scheme employing oblivious transfer called OTPri. By employing oblivious transfer in this scheme, the client can locally compute her location with no need for transmission of the whole database. Meanwhile, the client only has to expose a single AP id from its vicinity, moreover, the client cannot learn anything other than her choices, thus preserving the client's and server's data privacy. Through analysis, we have proved that this scheme guarantees fast localization and small overhead while preserving both the privacy of server and clients via oblivious transfer. Finally, experiments based on comprehensive dataset are conducted to prove the effectiveness of the scheme, and the results show that OTPri achieves a much better time performance and much less overhead compared with PriWFL approach. Meanwhile, the experimental studies have shown that OTPri achieves a similar performance compared with the PriWFL approach in terms of precision.

We can investigate a few directions for our future work. First, there are other efficient schemes for privacy-preserving indoor localization. Second, it would be interesting if we expand the experiments to additional public facilities with WiFi coverage, such as coffee shops, libraries, or grocery stores, to characterize and measure the performance of the proposed scheme, and use different measurements for location privacy, such as the one in [25].

References

1. Azizyan, M., Constandache, I., Choudhury, R.R.: Surroundsense: mobile phone localization via ambience fingerprinting. In: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, pp. 261–272. ACM (2009)
2. Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user location and tracking system. In: Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000), vol. 2, pp. 775–784. IEEE (2000)
3. Bahl, P., Padmanabhan, V.N., Balachandran, A.: Enhancements to the radar user location and tracking system. Microsoft Research, 2(MSR-TR-2000-12), pp. 775–784 (2000)
4. Beresford, A.R., Stajano, F.: Mix zones: user privacy in location-aware services. In: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, pp. 127–131. IEEE (2004)
5. Chen, Y., Lymberopoulos, D., Liu, J., Priyantha, B.: FM-based indoor localization. In: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, pp. 169–182. ACM (2012)
6. Chu, C.-K., Tzeng, W.-G.: Efficient k-out-of-n oblivious transfer schemes with adaptive and non-adaptive queries. In: Vaudenay, S. (ed.) PKC 2005. LNCS, vol. 3386, pp. 172–183. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-30580-4_12
7. Du, X., Chen, H.H.: Security in wireless sensor networks. IEEE Wireless Commun. **15**(4), 60–66 (2008)
8. Du, X., Guizani, M., Xiao, Y., Chen, H.-H.: Secure and efficient time synchronization in heterogeneous sensor networks. IEEE Trans. Veh. Technol. **57**(4), 2387–2394 (2008)
9. Du, X., Xiao, Y., Chen, H.-H., Wu, Q.: Secure cell relay routing protocol for sensor networks. Wirel. Commun. Mob. Comput. **6**(3), 375–391 (2006)
10. Du, X., Xiao, Y., Guizani, M., Chen, H.-H.: An effective key management scheme for heterogeneous sensor networks. Ad Hoc Netw. **5**(1), 24–34 (2007)
11. Hei, X., Du, X., Wu, J., Hu, F.: Defending resource depletion attacks on implantable medical devices. In: Global Telecommunications Conference (GLOBECOM 2010), pp. 1–5. IEEE (2010)
12. Huang, W., Xiong, Y., Li, X.-Y., Lin, H., Mao, X., Yang, P., Liu, Y.: Shake and walk: acoustic direction finding and fine-grained indoor localization using smartphones. In: Proceedings of IEEE INFOCOM, pp. 370–378. IEEE (2014)
13. Ishai, Y., Kilian, J., Nissim, K., Petrank, E.: Extending oblivious transfers efficiently. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 145–161. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45146-4_9
14. LaMarca, A., et al.: Place lab: device positioning using radio beacons in the wild. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) Pervasive 2005. LNCS, vol. 3468, pp. 116–133. Springer, Heidelberg (2005). https://doi.org/10.1007/11428572_8
15. Li, H., Sun, L., Zhu, H., Lu, X., Cheng, X.: Achieving privacy preservation in wifi fingerprint-based localization. In: Proceedings of IEEE INFOCOM, pp. 2337–2345. IEEE (2014)
16. Li, X.-Y., Jung, T.: Search me if you can: privacy-preserving location query service. In: Proceedings of IEEE INFOCOM, pp. 2760–2768. IEEE (2013)

17. Liu, X., Liu, K., Guo, L., Li, X., Fang, Y.: A game-theoretic approach for achieving k-anonymity in location based services. In: Proceedings of IEEE INFOCOM, pp. 2985–2993. IEEE (2013)
18. Ma, L., Teymorian, A.Y., Cheng, X.: A hybrid rogue access point protection framework for commodity wi-fi networks. In: The 27th Conference on Computer Communications (INFOCOM 2008), pp. 1220–1228. IEEE (2008)
19. Mohapatra, D., Suma, S.: Survey of location based wireless services. In: IEEE International Conference on Personal Wireless Communications (ICPWC 2005), pp. 358–362. IEEE (2005)
20. Ni, L.M., Liu, Y., Lau, Y.C., Patil, A.P.: LANDMARC: indoor location sensing using active RFID. Wirel. Netw. **10**(6), 701–710 (2004)
21. Otsason, V., Varshavsky, A., LaMarca, A., de Lara, E.: Accurate GSM indoor localization. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 141–158. Springer, Heidelberg (2005). https://doi.org/10.1007/11551201_9
22. Park, J.-G., Charrow, B., Curtis, D., Battat, J., Minkov, E., Hicks, J., Teller, S., Ledlie, J.: Growing an organic indoor location system. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 271–284. ACM (2010)
23. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket location-support system. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 32–43. ACM (2000)
24. Priyantha, N.B., Miu, A.K., Balakrishnan, H., Teller, S.: The cricket compass for context-aware mobile applications. In: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, pp. 1–14. ACM (2001)
25. Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., Hubaux, J.-P.: Quantifying location privacy. In: IEEE Symposium on Security and Privacy (SP), pp. 247–262. IEEE (2011)
26. Shu, T., Chen, Y., Yang, J., Williams, A.: Multi-lateral privacy-preserving localization in pervasive environments. In: Proceedings of IEEE INFOCOM, pp. 2319–2327. IEEE (2014)
27. Vu, K., Zheng, R., Gao, J.: Efficient algorithms for k-anonymous location privacy in participatory sensing. In: Proceedings of IEEE INFOCOM, pp. 2399–2407. IEEE (2012)
28. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. ACM Trans. Inf. Syst. (TOIS) **10**(1), 91–102 (1992)
29. Xiao, Y., Du, X., Zhang, J., et al.: Internet Protocol Television (IPTV): the killer application for the next-generation Internet. IEEE Commun. Mag. **45**(11), 126–134 (2007)
30. Xiao, Y., Rayi, V.K., Sun, B., Du, X., Hu, F., Galloway, M.: A survey of key management schemes in wireless sensor networks. Comput. Commun. **30**(11), 2314–2341 (2007)
31. Yang, D., Fang, X., Xue, G.: Truthful incentive mechanisms for k-anonymity location privacy. In: Proceedings of IEEE INFOCOM, pp. 2994–3002. IEEE (2013)
32. Yao, X., Han, X., Du, X., Zhou, X.: A lightweight multicast authentication mechanism for small scale IoT applications. IEEE Sens. J. **13**(10), 3693–3701 (2013)
33. Youssef, M., Agrawala, A.: The Horus WLAN location determination system. In: Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services, pp. 205–218. ACM (2005)