Research on Location Privacy Protection Mechanism Based on No-Trusted User Cooperation

Kunfu Wang¹*, Wei Feng¹, Wanfeng Mao¹ and Hui Wang²

¹ System Engineering Research Institute of China State Shipbuilding Corporation, Beijing, China. 100036
² School of computer science and technology, Henan Polytechnic University, Jiaozuo, 454000
* Corresponding Author Email: wangkunfu1166@163.com

Abstract. With the development of Internet and wireless location technology, location-based service makes people's life more convenient, and the threat to user privacy is growing. After analyzing the advantages and disadvantages of the existing location privacy protection models, this paper proposes a new MHT model based on untrusted users. Through the multi-hop transmission of the request information, the user's location information can be rid of the correlation with the user, and the query request can be desensitized. It can provide precise location service and resist the collusion attack and reasoning attack, which effectively protects the user Location privacy.

1. Introduction

In recent years, with the development of wireless communication technology and location technology[1], location-based service (LBS) has become increasingly close to people. In the mobile Internet, there are a large number of applications based on location-based services, and people's life is more convenient [2].

By sending the user's point of interest and location information to LBS server, users can obtain a variety of location-based services. However, if there is no scientific protection measures in the process of generating, storing and using location information, sensitive information such as user's health status, financial information, social relations, interests and emotional state may be leaked, and attackers can spread malicious advertisements and steal users' property according to these information. Therefore, protecting users' privacy information from being leaked has become an urgent problem to be solved. In recent years, researchers at home and abroad have proposed a variety of location privacy protection methods, among which k-anonymity model [3] is one of the most commonly used methods. The basic idea [4] is to replace the user's real location with an anonymous region of $\hat{k}$ users including the target user, so that the attacker can't distinguish the real location of the target user.

In the client server architecture [5], users use historical information to anonymously process real query requests through algorithms, so as to achieve the purpose of protecting users' location privacy. However, the system with this structure has higher requirements for clients, and requires certain storage capacity and computing power. The advantage is that it does not need other entities to participate.

The centralized structure [6] generally requires the introduction of a third-party anonymous server, which is responsible for processing the user's query request, sending the request information instead of the real user, and sending the server's returned results to the user. The disadvantage is easy to produce
single point of failure and bottleneck problems, the advantage is to reduce the performance requirements of the client.

The distributed structure [7,8], each mobile user cooperates with each other, shares part of the location information, and forms an anonymous space to achieve the k-anonymity effect. After processing, it achieves the purpose of protecting location privacy. The disadvantages are easy to generate communication delay and vulnerable to collusion attacks, and the advantages are to effectively avoid single point of failure and bottleneck problems.

With the development of data mining technology [9], the algorithm of k-anonymity model is vulnerable to the semantic attack of attackers, and the ability to resist collusion attack and reasoning attack is weak. To solve these problems, this paper proposes a multi hop transmission (MHT), The model is not only suitable for snapshot queries, but also for continuous queries, and can provide accurate location services against collusion attacks, reasoning attacks and semantic attacks. Under the same resource, the privacy protection strength of this model is stronger than that of k-anonymity model.

2. System Structure and Related Definitions

2.1. System Structure

MHT mechanism system consists of three parts: user, trusted cloud computing center and LBS server.

1. Users are usually intelligent terminals (such as mobile phones, tablets, etc.), which need to have the following functions:
   ① Positioning function (such as GPS, Beidou, etc.) can obtain the user's current position;
   ② Communication with location service providers can transmit information through cellular networks (such as 4G, 5g, etc.);
   ③ It can communicate with other users in short distance (such as WiFi, Bluetooth, etc.). When users send their own query requests, they are target users, and when they participate in other users' query requests, they are neighbor users.

2. The trusted Cloud Computing Center (TC3) selects the next hop neighbor users according to the privacy protection requirements of each user who enters the network or requests that have been sent in the user cycle of the network, and returns the set of optional next hop neighbor users to the corresponding users.

3. LBS server can retrieve the request information and package the result to the user who sent the request. It has the ability of encryption and decryption.

2.2. Related Definitions

Define 1 information transmission track: the information transmission track that the target user sends the request to the LBS server. It is expressed as:

\[ M_0 = \{ u_0, u_1, u_2, \ldots, u_h, S \} \]

Where \( h (h \in N^+) \) is the user privacy protection strength, that is, the number of hops; \( u_0 \) represents the target user, \( u_1 \ldots u_h \) is the neighbor user, respectively called one hop neighbor user, two hop neighbor user, \( h \) hop neighbor user; \( S \) is LBS server.

Define 2 The product of function vector: defines the operational relationship between \( \otimes \) and \( \oplus \) functions and vectors.
1

\[ f(x) \otimes \begin{bmatrix} p_{\min} \\
p_{\max} \\ r \end{bmatrix} = \begin{cases} \frac{1}{1-r} \left(1 - \frac{x}{p_{\min}}\right) & x \in [p_{\min}, p_{\min}] \\ 1 & x \in [p_{\max}, p_{\max}] \\ 1 - \frac{1}{r} \left(\frac{x}{p_{\max}} - 1\right) & x \in (p_{\max}, (1+r)p_{\max}] \\ 0 & x \in ELSE \end{cases} \] (1)

Where \([p_{\min}, p_{\max}]\) is the optimal value range, the range extension tolerance of \(r\) value evaluation makes \(rp_{\min}\) the lower limit and \((1+r)p_{\max}\) the upper limit.

\[ g \otimes \begin{bmatrix} p_{\min} \\
p_{\max} \\ r \end{bmatrix} = g(p_{\min}, p_{\max}, r) \] (2)

**Define 3 Reserve price**: When the target user looks for neighbor users, the quantitative value of the condition that the user who wants to participate in the request process of the target user can provide.

1. The bidding reserve price of any user \(i\) satisfies the following formula:

\[
Price_i(dist, \theta, Cnt) = \left[ f(dist) \ f(\theta) \ f(Cnt) \right] \otimes M \cdot W_i = \left[ f(dist) \otimes M_{i1} \ f(\theta) \otimes M_{i2} \ f(Cnt) \otimes M_{i3} \right] \cdot W_i
\] (3)

Among them:

a. \(M_{i1}(i=1,2,3)\) is the \(i\)-th column vector of matrix \(M\);

b. \(W_i\) is the weight vector and is the column vector of \(3 \times 1\);

c. \(dist_i\) is the Euclidean distance between user \(i(x_1, x_2)\) and \(u(y_1, y_2)\), calculated with \(\sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}\);

d. \(\theta_i(\theta \in [-\pi, \pi])\) represents the radian of the connecting line between \(i\) and user \(u\) and the due east direction (positive direction). The radian \(\theta\) between \(l_i(x_1, x_2)\) and \(l_j(y_1, y_2)\) is calculated according to the following formula:

\[
\theta = \arccos\left(\frac{x_1y_1 + x_2y_2}{\sqrt{(x_1^2 + x_2^2)(y_1^2 + y_2^2)}}\right)
\]

e. Where \(Cnt_i\) is the number of users \(i\) participating in the information transmission track at the current time;

f. \(M\) is the default attribute parameter matrix:

\[
M \otimes \begin{bmatrix} R_{\min} & \theta_{\min} & Cnt_{\min} \\
R_{\max} & \theta_{\max} & Cnt_{\max} \\
\lambda_R & \lambda_\theta & \lambda_{Cnt} \end{bmatrix}
\]
Where $R_{\text{min}}, R_{\text{max}}, \theta_{\text{min}}, \theta_{\text{max}}, \text{Cnt}_{\text{min}}, \text{Cnt}_{\text{max}}$ represents the minimum and maximum value of the optimal value range of distance, radian and number of auxiliary trajectories respectively; $\lambda_{r}, \lambda_{\theta}, \lambda_{\text{Cnt}}$ represents the range extension tolerance of different value evaluation.

2. For the $3 \times n$ matrix $U$ composed of competitive reserve price of $n$ users, the bidding reserve price satisfies the following formula:

$$\phi = \text{Price} \oplus U = [\text{Price} \oplus U_{1} \text{ Price} \oplus U_{2} \cdots \text{ Price} \oplus U_{n}]$$

Where $U_{i}(i = 1, 2, \cdots, n)$ is the $i$-th column vector of matrix $U$, and $U_{i} = [\text{dist}_{i}, \theta_{i}, \text{Cnt}_{i}]^{T}$.

**Definition 4 Cloud computing information** $MSG_{U2C}$: The data format sent by the user to TC3 meets the following formula:

$$MSG_{U2C} = \{\text{LOC}_{u}, W_{u}, M_{u}, ADR_{u}, NUM_{Q}\}$$

Where $\text{LOC}_{u}$ represents the location of user $u$; $W_{u}$ is the user’s weight vector; $M_{u}$ represents the default attribute parameter matrix of user $u$; $ADR_{u}$ represents the communication address of user $u$; $NUM_{Q}$ is the number of users returned by TC3.

**Definition 5 Request information** $MSG_{U2S}$: The request information format of the target user meets the following formula:

$$MSG_{U2S} = \{h, N_{\text{rdm}}, E_{\text{PK}_{S}} (POI, \text{Loc}, r, \text{PK}_{U})\}$$

Where $N_{\text{rdm}}$ is a random value and $E$ is an asymmetric encryption function; $POI$ represents the interest point information of the target user; $\text{Loc}$ represents the longitude and latitude coordinates of the target user; $r$ is the query radius; $\text{PK}_{S}, \text{PK}_{U}$ represents the public key of LBS server and target user respectively.

**Definition 6 Request return information** $MSG_{S2U}$: The information returned by the LBS server after processing the request information satisfies the following formula:

$$MSG_{S2U} = \{N_{\text{rdm}}, E_{\text{PK}_{U}} (\text{Result})\}$$

Where $\text{Result}$ is the result of processing user request information by LBS server.

**Definition 7 Routing table** $L_{i}$: Record the last hop route of user $I$ in the information transmission track, which satisfies the following formula:

$$L_{i} = \{N_{\text{rdm}}, ADR\}$$
When user $i$ participates in multiple information transmission trajectories, $N_{rdm}$ distinguishes different information transmission trajectories.

3. MHT Mechanism

3.1. Auction Mechanism

When different users participate in the query request of the target user, the privacy protection help provided by different users is also different. The more users participate in the information transmission trajectory, the higher the security of their own query requests. Therefore, users will compete to participate in the query request process of other users, assuming that they will not reject the auxiliary requests of other users. Therefore, when the user selects the next hop neighbor user, the auction mechanism is introduced to quantify the user's attribute information, and then weighted. Finally, the quantized value is the bidding reserve price of the user, which is used to compete with other users to participate in the query request of the target user.

3.2. Find Next Hop Neighbor User Set

When all users join the network or are already in the network, they send $MSG_{U2C}$ to TC3 every certain time. TC3 calculates the bidding reserve price of other users when they participate in the auction to assist the user to query. And return the set $U_{next}^i$ of the communication address $ADR_u$ of $NUM_Q$ users with the highest bidding price to the user. This process should be completed before making a query.

3.3. Query Implementation Process

The target user constructs the information transmission track according to the algorithm, and transmits the query information forward along the information transmission track. When the LBS server receives the query request, it searches the qualified records in the database, and reverses the query results along the information transmission track until it is returned to the target user.

3.4. Performance Analysis

For communication overhead. The bandwidth consumption of MHT mechanism is not greater than that of k-anonymity mechanism. The k-anonymity mechanism needs to regenerate the anonymous region every time it queries, fuzzizes the location information, and filters the query results. MHT mechanism only needs to select users from $U_{next}^i$ each time, and the query results are accurate.

4. Safety Analysis

4.1. Resistance to Collusion Attacks

Assume that the attacker set is $Adv = \{v_1, v_2, \ldots, v_m\}$, and explain it in four cases.

1. $\exists u_i (i \in [1, h])$ makes $u_i \in Adv$, that is, there are malicious neighbor users. $u_i$ cannot analyze the encrypted part based on the received $MSG_{U2S}$. $u_i$ only knows the information of the neighbor users in the previous hop and the neighbor users in the next hop. It is impossible for any user other than the target user to obtain a complete $M_Q$. 

2. TC3 $\in \text{Adv}$, The cloud computing center is not trusted. When $MSG_{U_{2C}}$ is sent to TC3, the user does not disclose his ID to TC3. Even if TC3 corresponds $MSG_{U_{2C}}$ to the real user, TC3 cannot determine the next hop neighbor user of the user because the user randomly selects the user in $U_{next}$ when selecting the next hop neighbor user. Because the secret key of $MSG_{U_{2S}}$ encryption part is the public key of LBS server, TC3 has no corresponding private key, so it can't resolve its encryption part.

3. LBS $\in \text{Adv}$, That is, the LBS server is not trusted. Lbs server can not get the information transmission track $M_Q$ completely, and $MSG_{U_{2S}}$ does not contain any information about the target user.

4. $\exists u_i (i \in [1, h])$ and TC3,LBS $\in \text{Adv}$. Because the hop number $h$ is only known by the target user, even the first hop neighbor user only knows the hop number of the previous hop user when the request information is obtained, and it is not sure that the last hop neighbor user is the target user. The attacker cannot determine all users in the information transmission path.

4.2. Resist Reasoning Attack

Suppose $N_s$ represents the whole network set and $P_E(\text{Event})$ represents the probability of success of $\text{Event}$.

$$P_E(u_i \in M_Q | M_Q \in N_s) = P_E(u_j \in M_Q | M_Q \in N_s)$$

(5)

$$P_E(u_i \in M_Q | M_Q \in N_s) = \frac{P_E(u_i \in M_Q \cap M_Q \in N_s)}{P_E(M_Q \in N_s)} = \frac{P_E(u_i \in N_s)}{P_E(M_Q \in N_s)}$$

Because

$$P_E(u_j \in M_Q | M_Q \in N_s) = \frac{P_E(u_j \in N_s)}{P_E(M_Q \in N_s)}$$. No matter what the value of $P_E(M_Q \in N_s)$ is, the simplification result of the above formula is

$$P_E(u_i \in N_s) = P_E(u_j \in N_s)$$, which is obviously true. Therefore, formula (5) is proved.

5. Conclusion

This paper analyzes the advantages and disadvantages of the common system structure model of LBS, encrypts the user's query request when the untrusted entity participates in the real user's location service, and sends the query request to the LBS server with the assistance of $h$ neighbor users. After the LBS server processes the query result, it encrypts the query result and feeds back to the target user along the information transmission path. Performance analysis and security analysis are carried out to highlight the advantages of MHT mechanism.

6. Reference

[1] Lv Shufang , Research review on wireless sensor networks node localization[J] Transducer and Microsystems Technologies, 2016.35(05):p.1-3+8.

[2] Paulet Russell,Kaosar Md.Golam,Yi Xun,et al.Privacy-Preserving and Content-Protecting Location Based Queries[C].in 2012 IEEE 28th International Conference on Data Engineering.2012.

[3] Zhang Shaobo,Choo Kim-Kwang Raymond,Liu Qin,et al.Enhancing privacy through uniform grid and caching in location-based services[J].Future Generation Computer Systems, 2018.86:p.881-892.
[4] Zhang Shaobo, Li Xiong, Tan Zhiyuan, et al. A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services [J]. Future Generation Computer Systems, 2019, 94: p. 40-50.

[5] Wang Jie, Wang Chunru, Ma Jianfeng, Li Hongtao. Dummy location selection algorithm based on location semantics and query probability [J], Journal on Communications, 2020 (3): p. 53-61.

[6] Pan Xiao, Xu Jianliang, and Meng Xiaofeng. Protecting Location Privacy against Location-Dependent Attack in Mobile Services [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 24: p. 1506-1519.

[7] Wu Dan-dan, Lu Xin. Location Anonymous Algorithm Based on User Collaboration under Distributed Structure [J]. Computer Science, 2019, 46(4): p. 158-163.

[8] ZHOU Hong-ya, ZHANG Qian. Location Privacy Preservation in Mobile P2P Encryption Based on Non-trusted Users [J]. Computer Technology and Development, 2016, 26 (05): 104-109+119.

[9] LI XueLong, GONG HaiGang. A survey on big data systems [J]. Science in China (Information Sciences), 2015, 45(01): 1-44.