

基于马尔可夫模型的同态加密位置隐私保护方案

周凯^{1,2,3}, 彭长根^{2,3}, 朱义杰^{3,4}, 何建琼^{1,2,3}

- (1. 贵州大学计算机科学与技术学院, 贵州 贵阳 550025;
2. 贵州大学贵州省公共大数据重点实验室, 贵州 贵阳 550025;
3. 贵州大学密码学与数据安全研究所, 贵州 贵阳 550025;
4. 贵州省网络数据保密工程技术研究中心, 贵州 贵阳 550025)

摘要: 针对基于位置服务的位置隐私与查询隐私保护问题, 提出一种基于马尔可夫模型的同态加密位置隐私保护方案。首先, 随机置换匿名用户真实身份, 结合用户的历史查询内容, 构建马尔可夫状态转移矩阵; 其次, 预查询用户的历史高频内容及其马尔可夫链下的预测内容, 并且存储相应结果集; 最后, 对该方案双预测系统的安全性进行了分析。该方案使服务器满足 $k+1$ 个查询内容, 并使恶意服务器或攻击者无法判定查询用户的真实身份与查询内容之间的对应关系, 实现了用户位置隐私与查询隐私的保护。同时, 利用同态加密密文的可计算性和保密性, 实现了面向密文数据的统计分析和隐私数据的安全存储。

关键词: 基于位置的服务; 查询隐私; 马尔可夫链; 同态加密; 匿名性

中图分类号: TP302

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00137

Homomorphic encryption location privacy-preserving scheme based on Markov model

ZHOU Kai^{1,2,3}, PENG Chang-gen^{1,2,3}, ZHU Yi-jie^{3,4}, HE Jian-qiong^{1,2,3}

- (1. College of Computer Science and Technology, Guizhou University, Guiyang 550025, China;
2. Guizhou Provincial Key Laboratory of Public Big Data, GuiZhou University, Guiyang 550025, China;
3. Institute of Cryptography & Data Security, Guizhou University, Guiyang 550025, China;
4. Guizhou Provincial Engineering and Technology Research Center of Cyber Data Security, Guiyang 550025, China)

Abstract: Homomorphic encryption location privacy-preserving scheme based on Markov mode was proposed to solve the problem of location privacy and query privacy protection in location-based service systems. Firstly, the anonymous user's identity were permuted randomly and the Markov state transition matrix combining with the user's historical query content was constructed. Secondly, system previously queries the user's high frequency content and the prediction content under Markov chain, then store the corresponding result sets. Finally, the security of the scheme's double prediction system was analyzed. The scheme makes the LBS receives $k+1$ query contents which let malicious server or attacker can't determine the corresponding relation between queried user's real identity and queried content. So the user's location privacy and query privacy can be protected. Meanwhile, the computability and confidentiality of homomorphic encryption ciphertext were used to realize the statistical analysis of ciphertext-oriented data and the secure storage of private data.

Key words: location-based services, inquiry privacy, Markov chain, homomorphic encryption, anonymity

收稿日期: 2016-11-22; 修回日期: 2016-12-27。通信作者: 彭长根, peng_stud@163.com

基金项目: 国家自然科学基金资助项目 (No. 61262073, No. 61662009); 贵州省哲学社会科学规划青年课题基金资助项目 (No.16GZQN06); 贵州省科技基金计划基金资助项目 (No.黔科合基础[2016]1023); 贵州大学研究生创新基金资助项目 (No.2016050)

Foundation Items: The National Natural Science Foundation of China (No.61262073, No.61662009), The Philosophy and Social Sciences Planning Project of Guizhou Province (No.16GZQN06), The Science and Technology Foundation of Guizhou Province (No.Gzuihou-Science-Contact [2016]1023), The Graduate Innovation Foundation of Guizhou University (No.2016050)

1 引言

移动互联网可移动化、可定位化以及个性化服务的发展,促进了基于位置服务(LBS, location-based service)的快速发展。基于位置服务将设备定位技术、无线通信技术以及地理信息管理等多种技术相互集成应用,从而为用户提供与其当前空间位置相关的个性化服务^[1]。当前,LBS已经在社会、经济、生活等领域产生了深远的影响,基于位置服务在给人民生活带来各种便利的同时,也面临个人隐私被泄露的风险。文献[2,3]报道了某人利用GPS跟踪前女友事件,文献[4]研究了用户轨迹开始和结束的地点,可以推测出用户的家庭住址等信息。

近几年,关于基于位置服务隐私保护的研究从未中断过,也取得了丰厚的成果,大部分的位置隐私保护研究工作是基于中心服务器结构和P2P网络结构^[5]。文献[6]介绍了位置大数据的相关概念以及位置大数据的隐私威胁定义。目前,位置隐私保护技术主要有:位置 k -匿名技术^[7]、假名技术^[8]、假位置泛化技术^[9]、位置区域的模糊化技术^[10]、混合区域技术^[11]、加密技术^[12,13]等。基于隐马尔可夫模型进行概率推测的位置隐私保护的研究也有很多,文献[14]提出一种MaskIt方法,主要思想为用隐马尔可夫模型形式化攻击者获取用户连续提交位置信息的过程。文献[15]使用Markov模型来进行位置的预测,但由于模型仅考虑当前时刻位置的影响,使预测精度较低。文献[16]提出了一种用户能判断待提交位置信息的模糊性对候选集精确性影响的位置服务框架,从而权衡预先查询的隐私强度和实用性。

针对大部分方案,用户在使用LBS时,必须要在服务质量和隐私保护之间做出一个折中的权衡。安全的位置隐私保护和高质量的基于位置服务就像鱼和熊掌,两者不可兼得。一方面,位置隐私保护技术最大程度地将用户精确的位置信息隐藏起来;另一方面,享用位置服务的用户希望能得到更精确的位置信息,从而享受更高质量的服务。

针对以上问题,本文提出一种基于中心服务器结构的位置隐私保护方法。首先,通过一个置

换表将用户的身份信息匿名。其次,引入马尔可夫模型、混淆查询信息以及预查询概念,实现基于位置查询隐私保护。同时,混淆用户的查询信息,保护用户的身份信息和查询内容的对应关系。最后,通过分析用户的历史查询内容,预查询用户的历史高频率内容及马尔可夫链下的预测内容,并且存储相应结果集。期间,利用同态加密密文的可计算性和保密性,实现面向密文数据的统计分析和隐私数据的安全存储。

2 准备知识

本节介绍需用的准备知识,包括同态加密、马尔可夫链等基本概念和性质。

定义1 同态加密。

同态加密算法由一个四元组($KeyGen, Enc, Dec, Eval$)和3个集合(M, Π, C)构成,其中, $M = (m_1, \dots, m_n)$ 是明文集合, $\Pi = (\pi_1, \dots, \pi_n)$ 是相应的密文集合, C 是可行的电路集合。四元组中算法描述如下。

$KeyGen(1^k, \rho)$ 密钥生成算法:输入安全参数 k 和随机数 ρ ,输出公钥、私钥(pk, sk)。

$Enc(pk, m)$ 加密算法:输入公钥 pk 、明文 M ,加密生成密文 Π 。

$Dec(sk, \Pi)$ 解密算法:输入私钥 sk 和密文集 Π ,解密输出明文 M 。

$Eval(pk, \pi_1, \dots, \pi_n, f_c)$ 评估算法:输入公钥 pk 、密文集 $(\pi_1, \dots, \pi_n)$,其中 $\pi_i \in \Pi, 1 \leq i \leq n$ 和可行函数的电路描述 f_c ,输出密文 $\pi \leftarrow f_c(\pi_1, \dots, \pi_n)$,使

$$\Pr \left[\begin{array}{l} \forall f_c \in M \wedge \pi_i \leftarrow Enc(pk, m_i) \pi \\ \leftarrow f(\pi_1, \dots, \pi_n) : Dec(sk, \pi) \neq f(m_1, \dots, m_n) \end{array} \right] \approx 0$$

同态加密的形式化定义如下。

定义2 设 m_0 和 m_1 是明文集合 M 中的元素, f 是 M 上的运算, E 是 M 上的加密算法,如果存在一个有效的操作函数 F 使

$$F(E(m_0), E(m_1)) = E(f(m_0, m_1)) \quad (1)$$

则称加密算法 E 对运算 f 是同态的。

1) 加法同态

设 m_0 和 m_1 是明文集合 M 中的元素, E 是 M 上的加密算法, D 是 M 上的解密算法,若存在一

个有效的操作函数 $\oplus$ 使

$$D(E(m_0) \oplus E(m_1)) = D(E(m_0 + m_1)) \quad (2)$$

则称加密算法 E 满足加法同态。

2) 乘法同态

设 m_0 和 m_1 是明文集合 M 中的元素, E 是 M 上的加密算法, D 是 M 上的解密算法,若存在一个有效的操作函数 $\otimes$ 使

$$D(E(m_0) \otimes E(m_1)) = D(E(m_0 \times m_1)) \quad (3)$$

则称加密算法 E 满足乘法同态。

定义3 马尔可夫链。

随机过程 $\{X_n, n \in T\}$,若对于任意的整数 $n \in T$ 和离散状态空间 $I, i_0, i_1, \dots, i_{n+1} \in I$,满足条件概率

$$\begin{aligned} P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} \\ = P\{X_{n+1} = i_{n+1} | X_n = i_n\} \end{aligned} \quad (4)$$

则称 $\{X_n, n \in T\}$ 为马尔可夫链,简称马氏链。

定义4 状态转移概率矩阵。

马尔可夫链 $\{X_n, n \in T\}$ 在时刻 n 的转移概率为

$$p_{ij}(n) = P\{X_{n+1} = j | X_n = i\} \quad (5)$$

其中, $i, j \in I$ 。若 $p_{ij}(n)$ 与时刻 n 无关,则称马尔可夫链是齐次的,并记 $p_{ij}(n)$ 为 p_{ij} 。对于由 n 个状态组成的状态空间,转移概率可表示为矩阵形式,记为

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (6)$$

其中, $p_{ij} \geq 0, i, j \in I; \sum_{j \in I} p_{ij} = 1, i \in I$ 。

3 查询隐私保护方案

3.1 用户身份匿名方案

当用户发生一个查询行为,位置匿名系统会随机地生成一个“假名”,即用户的第二重身份。用户发送查询信息时,会发送加密生成的“假名”以及注册时发放的唯一公钥。该系统主要保护用户的身份信息,每次查询时,用户的“假名”都可能不同,以达到混淆用户身份信息的效果。

3.1.1 系统初始化

当用户初次使用匿名系统时,需向可信中心

提供个人的身份信息(姓名、电话号码等),用 U_{id_i} 表示用户 u_i 的身份信息。

具体构造方案如下。

1) $KeyGen(1^k, \rho)$ 密钥生成算法: 输入安全参数 k 和随机数 ρ ,生成公钥集 $PK = (pk_1, \dots, pk_n)$ 、私钥集 $SK = (sk_1, \dots, sk_n)$,中心服务器公钥为 pk_T ,私钥 sk_T 。每个用户 $u_i (1 \leq i \leq n)$ 唯一的公钥为 pk_i 、私钥为 sk_i 。

2) 用户 u_i 在时刻 t 随机生成一个随机数 $r_{i,t}$ 作为临时密钥。其中, $r_{i,t}$ 可以向中心服务器随时更改,每个用户的 $r_{i,t}$ 都不一致。

3.1.2 假名生成

通过已经生成的映射表将用户的真实ID进行混淆,其中 U_{id_i} 都是日常用户的真实ID。 U_{id_i}' 是系统预设的一些“假名”,且在现实中真实存在。通过身份匿名表达达到以假乱真的效果,其中,用户 U_{id_i} 的身份信息具有如图1所示的多种关系。

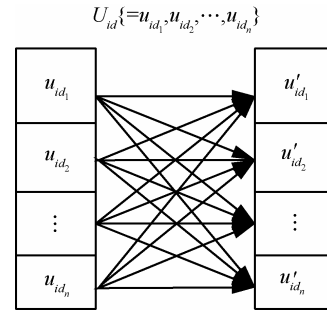


图1 用户身份匿名表

用户 u_i 每次发送查询消息,假名ID都会不同。在这种映射关系下,即使攻击方(敌手)获得位置信息或者身份ID,也很难确定是哪个用户 U_{id_i} 提出的位置服务请求。因此可以很好地起到混淆用户身份,保护用户的身份隐私不泄露。

3.1.3 服务请求提交

在用户与LBS服务器的通信中,本文采用一次一密的加密方式。具体方案如下。

1) 用户 u_i 在时刻 t 随机生成一个随机数 $r_{i,t}$ 作为临时密钥,并生成查询信息 $Q_i = \langle loc, q_i, r_{i,t}, U_{id_i}' \rangle$,其中, loc 是位置信息, q_i 是查询内容, U_{id_i}' 是用户生成的假名。

2) 用户 u_i 用可信中心的公钥 pk_T 将 Q_i 加密并发送给可信中心。

3) 可信中心用私钥 sk_T 解密,获得用户 u_i 的

loc 、 q_i 、 $r_{i,t}$ 、 U_{id} 。根据用户的 $r_{i,t}$ 对用户进行身份核对。

4) 可信中心发送 loc 、 q_i 、 U_{id} 给 LBS 服务器。

3.2 查询隐私保护方案

本文方案根据用户的历史查询信息建模，用频率替代概率，将历史查询信息进行聚类分析。在分析中引入同态加密，在密文状态下进行统计计算，然后利用马尔可夫预测模型，预查询可能的查询信息，使查询隐私满足 k -匿名，且提高查询效率。第三方服务器中的用户集合 $U = \{u_1, \dots, u_n\}$ ， $u_i, 1 \leq i \leq n$ 表示集合中的第 i 个用户， u_i 用户的历史查询信息集合 $Q_i = \{q_{i_1}, q_{i_2}, \dots, q_{i_m}\}$ 。查询隐私保护方案分为 2 个模型： k -匿名混淆模型和马尔可夫预测模型。

3.2.1 k -匿名模型

KeyGen($1^k, \rho$) 密钥生成算法：输入安全参数 k 和随机数 ρ ，生成公钥集 $PK = (pk_1, \dots, pk_n)$ 、私钥集 $SK = (sk_1, \dots, sk_n)$ ，分配给每个用户 $u_i (1 \leq i \leq n)$ 唯一的公钥 pk_i 、私钥 sk_i 。

Enc(pk, m)：输入公钥 pk_i 、 u_i 用户的历史查询信息集合 $Q_i = \{q_{i_1}, q_{i_2}, \dots, q_{i_m}\}$ ，加密生成密文 $Q_i^* = \{q_{i_1}^*, q_{i_2}^*, \dots, q_{i_m}^*\}$ 。

利用同态加密的密文可计算性，将用户 u_i 历史查询事件集合用二元组 $\langle A_i, B_i \rangle = \{\langle a_{i_1}, b_{i_1} \rangle, \langle a_{i_2}, b_{i_2} \rangle, \dots, \langle a_{i_k}, b_{i_k} \rangle\}$ 表示， a 表示不同的查询事件， b 表示每种查询事件的总数， A 是密文状态。求出每种查询事件的概率 $P = \{p_{i_1}, \dots, p_{i_k}\}$ ，其中

$$p_{i_j} = \frac{b_{i_j}}{\sum_{l=1}^k b_{i_l}} (j=1, \dots, k) \quad (7)$$

将 $p_{i_1}, \dots, p_{i_k}$ 从大到小进行排序，得到排序后的 $P' = (p'_{i_1}, \dots, p'_{i_k})$ 以及对应概率的查询事件集合 $A'_i = \{a'_{i_1}, a'_{i_2}, \dots, a'_{i_k}\}$ ，集合 A'_i 是密文。当检测到用户上传位置信息时，经匿名技术处理，选择概率排名前 n 个 $\{a'_{i_1}, a'_{i_2}, \dots, a'_{i_n}\}$ 查询内容，其中 $n (n < k)$ 的大小由用户设置。用户 u_i 历史查询事件也可能来自多个用户上传的数据。

Dec(sk, Π) 解密算法：输入私钥 sk_i 和密文集 $\{a'_{i_1}, a'_{i_2}, \dots, a'_{i_n}\}$ ，解密输出明文状态的查询内容 $\{D(a'_{i_1}), D(a'_{i_2}), \dots, D(a'_{i_n})\}$ 。提前发送给服务提供商

(SP) 并存放加密的候选结果集 $\{R(D(a'_{i_1})), R(D(a'_{i_2})), \dots, R(D(a'_{i_n}))\}$ 。将用户发送的查询内容 q 与预查询内容 $\{D(a'_{i_1}), D(a'_{i_2}), \dots, D(a'_{i_n})\}$ 进行匹配。实现查询的伪代码，见 Query 算法。

此时分为 2 种情况：1) 当 q 在 $\{a'_{i_1}, a'_{i_2}, \dots, a'_{i_n}\}$ 中时，服务器将对应的加密结果直接发送给用户，并更新历史查询事件集合和查询数量；2) 当 q 不在 $\{a'_{i_1}, a'_{i_2}, \dots, a'_{i_n}\}$ 中时， q 与 $\{a'_{i_{n+1}}, a'_{i_2}, \dots, a'_{i_k}\}$ 进行匹配，更新查询事件和查询数量。

Query 算法如下。

输入：查询 q

输出： $R(D(a'_i))$ or 0

- 1) for $t \leftarrow 1$ to k do
- 2) if ($q == D(a'_t)$) // 判断查询与预查询内容是否相同
- 3) Return ($R(D(a'_t))$)
- 4) end
- 5) else
- 6) $t = t + 1$
- 7) repeat
- 8) if ($t > k$)
- 9) return 0
- 10) end

3.2.2 马尔可夫预测模型

该模型是以 u_i 用户 ($i=1, \dots, n$) 的查询信息密文集合 $Q_i^* = \{q_{i_1}^*, q_{i_2}^*, \dots, q_{i_m}^*\}$ 作为源数据。在此模块中，主要工作是使用马尔可夫链预测模型，求出状态转移矩阵。具体方案如下。

1) 根据 k -匿名混淆模型分析得到二元组

$$\langle A_i, B_i \rangle = \{\langle a_{i_1}, b_{i_1} \rangle, \langle a_{i_2}, b_{i_2} \rangle, \dots, \langle a_{i_k}, b_{i_k} \rangle\}$$

用户 u_i 共有 k 种查询事件。

2) 将连续的查询内容集合 $Q_i^* = \{q_{i_1}^*, q_{i_2}^*, \dots, q_{i_m}^*\}$ ，分成以下二元组集合 $\langle q_{i_1}^*, q_{i_2}^* \rangle, \langle q_{i_2}^*, q_{i_3}^* \rangle, \dots, \langle q_{i_{m-1}}^*, q_{i_m}^* \rangle$ 。通过统计计算求得概率转移矩阵 C_i 。

$$C_i = \begin{bmatrix} p_{i_{11}} & p_{i_{12}} & \dots & p_{i_{1k}} \\ p_{i_{21}} & p_{i_{22}} & \dots & p_{i_{2k}} \\ \dots & \dots & \dots & \dots \\ p_{i_{k1}} & p_{i_{k2}} & \dots & p_{i_{kk}} \end{bmatrix} \quad (8)$$

3) 加密 C_i 矩阵。

将矩阵 C_i 按列分组, 其中 $A[i] = (p_{1i}, p_{2i}, \dots, p_{ki})$ 。将源矩阵按随机生成的映射表 Map 进行转换, 矩阵 C_i 的列混淆, 输出混淆矩阵 $B = \{B[1], \dots, B[k]\}$, 矩阵 B 存放在数据库中。如图2所示。

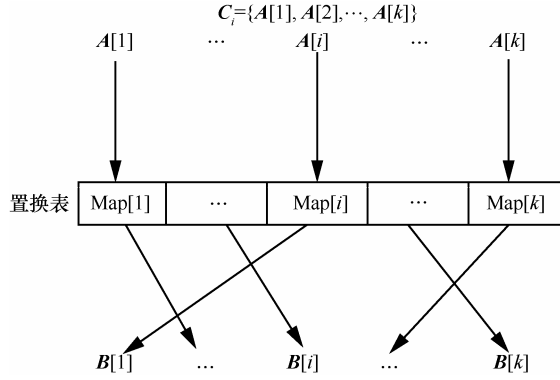


图2 加密 C_i 矩阵

4) 根据用户 u_i 的最近一次查询内容 $q_{i_m}^*$, 利用求得的概率转移矩阵 C_i , 预测出下一个查询事件 $q_{i_{m+1}}^*$ 。

5) 用同态解密算法 D , 解密得到 $D(q_{i_{m+1}}^*)$, $D(q_{i_{m+1}}^*)$ 与 $(a'_{i_1}, a'_{i_2}, \dots, a'_{i_n})$ 不相等。根据用户 u_i 发送的地理信息 loc 和匿名后的随机假名 $u_i d'_i$, 提前查询 $D(q_{i_{m+1}}^*)$, 并存储服务器返回的候选答案集 R_i 。

6) 将 $D(q_{i_{m+1}}^*)$ 与用户上传的查询内容 Q 进行配对, 返回结果。更新数据库中的查询事件和查询数量。

但是, 也可能出现预测内容与用户查询内容不同的情况。那么, 系统先匿名用户的身份, 然后将查询 Q 发送给服务提供商, 相应的结果直接返回给用户。

在基于位置服务中, k -匿名模型和马尔可夫预测模型同时工作。利用 k -匿名模型提前查询出高频率查询信息的候选答案集。使用马尔可夫预测模型预查询预测信息。每当系统检测到用户打开位置服务, 就自动触发预测, 发送 k -匿名模型和马尔可夫预测模型的查询信息。此时, 位置服务器至少要接收到 $k+1$ 个互不相同的查询内容, 使服务器无法辨别真实的查询内容。另外, 通过预查询和存储结果, 提高用户的查询效率。

4 方案分析

LBS 系统中的攻击者分为内部攻击者和外部攻击者。本方案的安全性分析如下。

4.1 内部攻击下的安全性

内部攻击者是指 LBS 服务器拥有所有用户的一切查询请求消息, 可以利用所掌握的查询信息和查询用户身份信息进行推断攻击, 其攻击目的是获得用户身份信息与位置数据的对应关系。

攻击者假设: 攻击者是一个全局攻击者, 攻击者可以获取 LBS 服务器数据, 拥有一定的边信息^[17]辅助的推测攻击。攻击者以获取查询用户的身份信息、查询内容为目标, 攻破位置数据与身份信息的对应关系。

攻击环境假设: 1) 所有用户身份集合 $U_{id} = \{u_{id_1}, u_{id_2}, \dots, u_{id_n}\}$ 。2) 在隐私保护系统中, 注册的用户没有退出。用户可能发生连续查询行为, 而且用户所需的查询内容恰好在集合 $\{a'_i, a'_{i_2}, \dots, a'_{i_n}\} \cup \{q_{i_{m+1}}^*\}$ 中。

4.1.1 身份匿名

身份匿名是指 LBS 服务器既不能区分用户的真实身份, 也不能判断不同的查询信息是否来自同一用户。在本文方案中, 用户 u_i 每次查询的假名均来自于身份匿名表的随机选择。任何一个假名其实都是一个现实中真实存在的。由于身份匿名表中的假名与用户 u_i 的 ID 无直接关系, 即使敌手得到了所有假名集合 U_{id_i} , LBS 服务器也无法从用户的任意一个假名识别真实身份。不仅如此, 即使用户 u_i 发送连续查询信息, 敌手获得一串假名也无法判定出用户的真实身份。

4.1.2 位置隐私

假设用户 u_i 在一段时间 ΔT 内发生了 n 次服务请求查询。显然, 攻击者访问 LBS 数据库, 在时刻 t_j , LBS 服务器获悉用户真实查询内容的概率为 $p_{q_{t_j}} = \frac{1}{k+1}$, 从用户 u_i 假名识别其真实身份

的概率为 $p_{id_{t_j}} = \frac{1}{n}$, 从用户提交的 loc_j 个位置中

识别真实的位置概率为 $p_{loc_{t_j}} = \frac{1}{|loc_j|}$ 。这样获悉

一次用户真实身份、真实位置信息、查询内容的

对应关系的概率为 $p_{id_j, loc_j, qur_j} = \frac{1}{n(k+1)|loc_j|}$ ，获

悉真实身份和位置关系的概率为 $p_{id_j, loc_j} = p_{id_j} p_{loc_j}$ 。

对于用户 u_i 在一个时间段 ΔT 内的整个位置轨迹，本文定义在 n 次查询中查询内容与真实身份泄露的平均概率为

$$\overline{p_{id_j, loc_j, qur_j}} = \frac{\sum_{j=1}^n p_{id_j} p_{loc_j} p_{qur_j}}{n} \quad (9)$$

定义用户 u_i 在 n 次查询中泄露位置和身份信息

$$\overline{p_{id_j, loc_j}} = \frac{\sum_{j=1}^n p_{id_j} p_{loc_j}}{n} = \frac{\sum_{j=1}^n p_{id_j} p_{loc_j}}{n} \quad (10)$$

接下来，对基于 k -匿名的方案^[7]和本文方案的隐私程度做简要的分析。

按照文献[7]的 k -匿名构造方案，用户 u_i 在时间段 ΔT 内使用相同的假名，假设敌手获悉其假名所对应真实身份的概率为 $p_{name_{i_j}}$ ， k 的值设为

$k_1 = |loc_j|$ ，则 $\overline{p_{id_j, loc_j}}$ 修改为

$$\overline{p_{id_j, loc_j}} = \frac{(\max_{j=1}^n p_{name_{i_j}}) \sum_{j=1}^n p_{loc_j}}{n} \quad (11)$$

假设其中的 m 次查询收到了边信息辅助的推测攻击。当 LBS 服务器在时刻 tp 拥有边信息时，其掌握了用户的真实身份和位置，既 $\max_{j=1}^n p_{name_{i_j}} = 1, p_{loc} = 1$ ，而当在所有查询时刻没有掌握边信息时， $\max_{j=1}^n p_{name_{i_j}} = 0, p_{loc} = \frac{1}{k_1}$ 。将这些值代入，

可以得到

$$\overline{P}_1 = \begin{cases} 0, & m = 0 \\ \frac{m + (n-m) \frac{1}{k_1}}{n}, & m > 0 \end{cases} \quad (12)$$

m 次查询收到了边信息辅助的推测攻击。当 LBS 服务器在时刻 tp 拥有边信息时，其掌握了用户的真实身份和位置。此时的 $p_{id_j, loc_j, qur_j} = p_{id_j} p_{loc_j} p_{qur_j} = 1$ ，而当在所有查询时刻没有掌握边信息时， $\max_{j=1}^n p_{name_{i_j}} = 0, p_{loc} = \frac{1}{k_1}$ 。将这些值

代入，可以得到

$$\overline{P}_2 = \begin{cases} 0, & m = 0 \\ \frac{m + (n-m) \frac{1}{k_1}}{n}, & m > 0 \end{cases} \quad (13)$$

对于本文方案，由于采用了假名的方法， p_{id_j} 在不同的时刻 t_j 是相互独立的。所以， p_{id_j} 的值是 1 或 0，而 $p_{loc_j} = 1$ 。即使攻击者获悉用户的确切位置，匿名身份，此时攻击者获悉用户查询内容的概率为 $p_{qur} = \frac{1}{k+1}$ 。则

$$\overline{p}_1 = \overline{p_{id_j, loc_j}} = \frac{\sum_{j=1}^n p_{id_j} p_{loc_j}}{n} = \frac{m}{n} \quad (14)$$

$$\overline{p}_2 = \overline{p_{id_j, loc_j, qur_j}} = \frac{\sum_{j=1}^n p_{id_j} p_{loc_j} p_{qur_j}}{n} = \frac{m}{n(k+1)} \quad (15)$$

显然，当存在边信息时，文献[7]方案中用户泄露的平均概率远大于本文方案，即

$$\overline{P}_1 - \overline{P}_1' = \frac{m + (n-m) \frac{1}{k_1}}{n} - \frac{m}{n} = \frac{n-m}{nk_1} \quad (16)$$

$$\begin{aligned} \overline{P}_2 - \overline{P}_2' &= \frac{m + (n-m) \frac{1}{k_1}}{n} - \frac{m}{n(k+1)} \\ &= \frac{(n-m) \frac{k+1}{k_1} + mk}{n(k+1)} > 0 \end{aligned} \quad (17)$$

当中心服务器中存放的历史查询信息达到一定程度时，用户的 k 值越大，用户的查询隐私保护程度越高，攻击者攻破位置信息与身份数据的对应关系越难。

4.2 外部攻击下的安全性

外部攻击分为 2 种：被动攻击和主动攻击。本文方案中，用户 u_i 与可信中心 T 之间的通信均是密文状态，没有相应的密钥是无法解密得到相应明文的。

在发放用户 u_i 的公、私钥对阶段。用户 u_i 向可信中心发送一个临时随机数密钥 $r_{i,t}$ ，然后可信中心为其发放 $r_{i,t}$ 加密的公钥 pk_i 、私钥 sk_i 。由于攻击者无法得到可信中心的密钥 sk_T ，他无法伪

造可信机构的消息来欺骗用户 u_i ，也无法得到用户 u_i 的临时随机数密钥 $r_{i,t}$ ，不能解密消息得到用户的公钥 pk_i 、私钥 sk_i 。在服务请求提交阶段，攻击者无法伪装成用户 u_i 与 LBS 服务器通信，这样用户 u_i 的临时随机数密钥 $r_{i,t}$ 以及中心服务器的私钥 sk_T 就无法被可信中心服务器验证，也无法伪装成 LBS 服务器。

5 结束语

本文研究了基于位置服务中的位置隐私、查询隐私保护问题，提出了一种基于马尔可夫预测模型下的隐私保护模型。该模型基于保护用户身份信息与位置数据对应关系的思想，通过一个置换表匿名用户的身份信息，利用 k -匿名混淆模型、马尔可夫预测模型，用 k 个高频率的查询信息混淆用户的真实查询信息，并通过系统的预查询、存储，提高中心服务器结构下的查询效率，最后证明了该方案的安全性。

参考文献：

- [1] JENSEN C S. 5-Database aspects of location-based services[J]. Journal of Location Based Services, 2004: 115-147.
- [2] Man accused of stalking ex-girlfriend with GPS[EB/OL]. <http://www.foxnews.com/story/0,2933,131487,00.Html>.
- [3] Authorities: GPS system used to stalk woman[EB/OL]. http://www.usatoday.com/tech/news/2002-12--gps-stalker_x.html.
- [4] WICKER S B. The loss of location privacy in the cellular age[J]. Communications of the ACM, 2012, 55(8): 60-68.
- [5] 王宇航, 张宏莉, 余翔湛. 移动互联网中的位置隐私保护研究[J]. 通信学报, 2015, 36(9): 230-243.
WANG Y H, ZHANG H L, YU X Z. Research on location privacy in mobile internet [J]. Journal of Communication, 2015, 36(9): 230-243.
- [6] 王璐, 孟小峰. 位置大数据隐私保护研究综述[J]. 软件学报, 2014, 25(4): 693-712.
WANG L, MENG X F. Location privacy preservation in big data era: a survey[J]. Journal of Software, 2014, 25(4): 693-712.
- [7] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(05): 557-570.
- [8] PIRAMUTHU, On existence proofs for multiple RFID tags[C]// The International Conference on Pervasive Services, IEEE. 2006: 317-320.
- [9] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C]// The International Conference on Pervasive Services(ICPS '05). 2005:88-97.
- [10] ARDAGNA CA, CREMONINI M, DAMIANI E, et al. Location privacy protection through obfuscation-based techniques[M]//Data and Applications Security XXI. Berlin Heidelberg :Springer. 2007: 47-60.
- [11] PALANISAMY B, LIU L. MobiMix: protecting location privacy with mix-zones over road networks[C]//The 2011 IEEE 27th International Conference on Data Engineering. 2011: 494-505.
- [12] MASCETTI S, FRENI D, BETTINI C, et al. Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies [J]. VLDB Journal, 2010, 20(4): 541-566
- [13] LI X Y, JUNG T. Search me if you can: privacy-preserving location query service[C]//Infocom, IEEE. 2012: 2760-2768.
- [14] TZ M, NATH S, GEHRKE J. MaskIt: privately releasing user context streams for personalized mobile applications[C]//The ACM Sigmod International Conference on Management of Data. 2012: 289-300.
- [15] CHEN M, LIU Y, YU X. NLPMM: a next location predictor with Markov modeling[C]//Advances in Knowledge Discovery and Data Mining. 2014:186-197.
- [16] DEWRI R, THURIMELLA R. Exploiting service similarity for privacy in location-based search queries[J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 25(2): 374-383.
- [17] YOU T H, PENG W C, LEE W C. Protecting moving trajectories with dummies[C]//The International Conference on Mobile Data Management. 2007: 278-282.

作者简介：



周凯 (1991-), 男, 浙江衢州人, 贵州大学硕士生, 主要研究方向为密码学与可信计算。



彭长根 (1963-), 男, 贵州锦屏人, 博士, 贵州大学教授、博士生导师, 主要研究方向为密码学、信息安全。

朱义杰 (1989-), 男, 山东临沂人, 贵州大学硕士生, 主要研究方向为密码学与可信计算。

何建琼 (1991-), 贵州遵义人, 贵州大学硕士生, 主要研究方向为密码学与安全协议。