面向移动社交网络的位置隐私保护方法

许志凯，张宏莉，史建焘，田志宏
（哈尔滨工业大学 计算机科学与技术学院，黑龙江 哈尔滨 150001）

摘　要：移动社交网络为人们的生活带来了极大的便利，但用户在享受这些服务带来便利的同时，个人位置隐私受到了严重威胁。首先对用户位置隐私保护需求进行了形式化描述，继而针对用户的敏感兴趣点泄露问题，提出了一种情景感知的隐私保护方法。该方法将位置信息、社交关系、个人信息引入到知识构建算法中以计算兴趣点间的相关性，并利用该相关性及时空情景实时判断发布当前位置是否会泄露用户隐私，进而实现了隐私保护与服务可用性间的平衡。最后通过仿真实验验证了该方法的有效性。

关键词：社交网络；移动社交网络；位置隐私；推理攻击
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.2096-109x.2015.00007

Location privacy protection for mobile social network

XU Zhi-kai, ZHANG Hong-li, SHI Jian-tao, TIAN Zhi-hong
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Given its high utility value, mobile social network services(MSNS), however, has raised serious concerns about users' location privacy. The location privacy requirements of users in MSNS are personal and dynamic, thus a metric called confidence was proposed to quantify the privacy risks. To avoid the adversary inferring users' privacy, a method of legation privacy protection was designed to calculate the correlation between the locations through location information, social relation and personal information. Then the correlation and the space-time background were used to evaluate whether the users' published geo-content meet the user's privacy requirement. Eventually, our experimental results demonstrate the validity and practicality of the proposed strategy.

Key words: social network; mobile social network; location privacy; inference attack

1　引言

随着智能终端的普及和 3G、Wi-Fi 等通信技术的发展，社交网络与移动互联网逐步融合，形成了移动社交网络服务（MSNS，mobile social network services）。用户不仅可以实时地获取位置信息，获得相应的服务（如查找附近最近的咖啡馆、酒店等），还可以与朋友分享带有位置标签的内容（如在某个语义位置签到、发现邻近好友等）。移动社交网络的位置服务将虚拟社区与现实社会相结合，丰富了人们的社交方式，为人们的生活带来了极大的便利；然而，人们在使用这些服务的同时，也面临隐私泄露的威胁。例如，攻击者通过关联分析、多次查询等方式可推测出用户的

收稿日期：2015-09-12；修回日期：2015-10-08。通信作者：许志凯，zhikaixu@foxmail.com
基金项目：国家自然科学基金资助项目（61202457, 61173144, 61402137, 61402149）；国家重点基础研究发展计划（“973”计划）基金资助项目（2011CB302605, 2013CB329602）
Foundation Items: The National Natural Science Foundation of China(61202457, 61173144, 61402137, 61402149); The National Basic Research Program of China (973 Program) (2011CB302605, 2013CB329602)

生活习惯、健康状态等信息。位置隐私能否得到妥善保护已成为制约移动社交网络发展的瓶颈问题。

当前移动社交网络中的位置隐私保护还处于初级阶段，最新的研究试图将基于位置的服务中已有的隐私保护方法应用到移动社交网络中，如 k-匿名[1]、空间匿名[2]等，但这些研究均未考虑位置数据与个人信息的关联问题。由于社交网络真实性、交互性的特点，移动社交网络中位置数据与个人信息（如社交关系、个人兴趣等）是直接关联的，且用户间的位置隐私不是孤立的，在这种情况下如何度量用户隐私的泄露程度，继而平衡服务的可用性与用户的隐私需求已成为移动社交网络位置隐私保护研究中亟需解决的难点问题。

位置隐私可界定为个人、机构等主体不愿意被外部知晓的位置信息，本文将这些位置称为敏感兴趣点。位置隐私保护旨在防止用户的位置隐私被非授权者访问。在移动社交网络中，用户自行设置位置信息的访问权限，用户的隐私需求具有个性化、动态化的特点。如一些用户将教堂视为敏感兴趣点，认为这可能泄露他的宗教信仰；但另一些用户却愿意在教堂共享自己的位置，以便于与好友聚会交流。由此，本文引入置信度来形式化地描述用户的隐私需求，度量隐私泄露的风险，提出一种情景感知的位置隐私保护框架，将位置信息、社交关系、个人信息引入到知识构建算法中以计算兴趣点之间的相关性，并利用该相关性及时空情景实时判断发布当前位置是否满足用户的隐私需求，实现隐私保护与服务质量间的平衡。

2　相关工作

当前移动社交网络中关于位置隐私保护的研究刚刚起步。最新的研究试图将基于位置服务（LBS，location based services）中的隐私保护方法，如空间匿名[1-3]（spatial cloaking）、隐私信息检索[4-6]（PIR, private information retrieval）等应用到移动社交网络中，具体方法如下。

1) 基于空间匿名的位置隐私保护

该方法通过增加用户位置信息的不确定性以保护用户的位置隐私。文献[1]最早将关系数据库中的 k-匿名引入到位置隐私保护中，其基本思想是：在发布位置信息时，以一块覆盖其他 k-1 个用户的空间区域代替用户的精确位置，从而使攻击者无法从 k 个用户中鉴别出某个用户。文献[2]在保证 k 匿名的同时，引入了位置 l-多样化的概念,即在 k 个用户中保证至少 l 个用户的查询请求是不一样的。文献[3]将该空间匿名应用到移动社交网络中的邻近好友发现服务中，通过匿名区域之间的距离推测用户真实的距离，但方法会导致服务质量下降，不适合一些需要精确位置的服务。

2) 基于加密的位置隐私保护

文献[4]将密文检索理论应用到位置隐私保护中，用户可以在不泄漏位置信息的情况下检索服务器上的任意数据项。文献[5]提出一种基于可查询加密（searchable encryption）的隐私保护框架，在数据加密的情况下实现好友之间签到位置的查询。文献[6]提出一种基于同态加密（homomorphic encryption）和 CP-ABE 的细粒度的位置查询协议，根据用户与查询者的信任关系，确定位置精度（如方圆 10 m、20 m、100 m 的位置范围），用户信任查询者的程度越高，查询者获得的位置精度越高。但文献[5,6]中的加密运算对于移动用户的预计算开销和运行时计算开销较大。

此外，部分研究者结合位置之间的时空关联性来探索位置隐私保护的途径。文献[7]定量分析了零星的位置泄露引起的运动轨迹泄露问题。文献[8]提出攻击者在获得用户最大移动速度的情况下，可推理出匿名区域中的查询用户。文献[9,10]通过位置模糊和时间模糊的方法保证连续提交的匿名区域间在速度上是可达的。这些研究主要关注移动速度信息对位置隐私泄露的影响，对于其他深入的关联分析结果导致的隐私泄露问题未有涉及。针对以上研究的不足，本文提出的面向移动社交网络的位置隐私保护方法具有以下特点：可满足移动社交网络中用户个性化和动态的隐私需求；可解决对移动社交网络中的背景知识进行关联分析导致的隐私泄露问题。

3　位置隐私保护需求的形式化描述

本小节首先介绍了移动社交网络的拓扑模

型，然后给出了移动社交网络中用户隐私需求的形式化定义，分析了保证用户隐私安全必须满足的条件。

3.1 移动社交网络拓扑

移动社交网络的拓扑可以概括为人与人之间关系、人与位置之间关系、位置与位置之间关系，如图 1 所示。人与人之间的关系是指社交网络中的信任关系；人与位置的关系是指人与地理位置的交互；位置与位置之间的关系是指地理位置之间的相关性。用户不仅可以通过位置信息获取服务，还可以从用户产生的位置信息中发现新知识并加以利用，如通过用户移动轨迹的相似性为用户推荐好友；但同时用户的身份和位置信息相互关联，攻击者可通过社交网络中的社交关系、个人信息推理用户的位置隐私。

3.2 位置隐私保护需求形式化描述

定义 1 兴趣点及敏感兴趣点。兴趣点（POI, point of interest）是电子地图上某个地标，用以标示该地所代表的政府部门、商业机构、旅游景点、交通设施等。如图 1 所示，本文利用兴趣点标示用户位置，将兴趣点记为 l_i，兴趣点集合记为 L；将用户不愿发布的兴趣点记为敏感兴趣点 s_i，这种兴趣点的集合记为 S，其中，$S \subseteq L$。

定义 2 移动轨迹 H_u。给定用户 u，其在时刻 t_1 和 t_n 之间的移动轨迹可以表示为一个按时间排列的序列 $H_u=\{(l_1, t_1), (s_2, t_2), (l_3, t_3), \cdots, (s_{i-1}, t_{i-1}), (l_i, t_i), \cdots, (l_n, t_n)\}$, (l_i, t_i)是序列 H_u 中的一个元素，表示用户 u 在时刻 t_i 访问过兴趣点 l_i，(s_{i-1}, t_{i-1})表示用户 u 在时刻 t_{i-1} 访问过敏感兴趣点 s_{i-1}。

定义 3 用户的位置保护隐私需求 SC_u^t。给定用户 u, 其在时刻 t 的对位置隐私的需求可以表示为一个集合 $SC_u^t=\{(s_1, c_1),\cdots, (s_i, c_i), \cdots, (s_n, c_n)\}$，其中，$(s_i, c_i)$表示用户 u 对敏感兴趣点 s_i 的置信度需求为 c_i。令 P_i 表示攻击者推理出的用户 u 已访问或将访问兴趣点 s_i 的置信度，当 $P_i<c_i$ 对于任意一个敏感兴趣点 $s_i \in S$ 均成立时, 用户的位置隐私受到了良好保护。

在移动社交网络中，由于文化背景不同，用户对同一信息往往具有差异化的价值取向和界定标准，因而位置隐私具有个体性，即对于不同的用户 u_i 和 $u_j(i \neq j)$, $SC_{ui} \neq SC_{uj}$。即使用户主体确定，随着用户所处时空情景不同，用户对位置隐私的界定也会发生变化，如一些用户在周末对位置隐私有更高的需求，另一些用户与家人在一起时更关注自己的位置隐私。因而位置隐私具有动态性,即使用户主体确定，在不同的时刻 t_1 和 t_2, $SC_u^{t_1} \neq SC_u^{t_2}$。

3.3 攻击模型

敏感兴趣点推理攻击模型如图 2 所示。$H_u=\{(l_i, t_i), (s_j, t_j), (l_{i+1}, t_{i+1})\}$。设用户 u 在 t 时刻发布兴趣点 l_i, 在 t_{i+1} 时刻发布兴趣点 l_{i+1}。令 $\lambda=|t_{i+1}-t_i|$。如果 λ 远大于从 l_i 移动到 l_{i+1} 所需的正常时间间隔, 则攻击者可推理出用户在 t_i 到 t_{i+1} 间访问过其他兴趣点。当攻击者具有一定的背景知识时，如大部分用户遵循$(l_i \rightarrow s_j \rightarrow l_{i+1})$的移动规律, 即使用户未发布兴趣点 s_j，攻击者也可推理出用户 u 很有可能访问过 s_j。

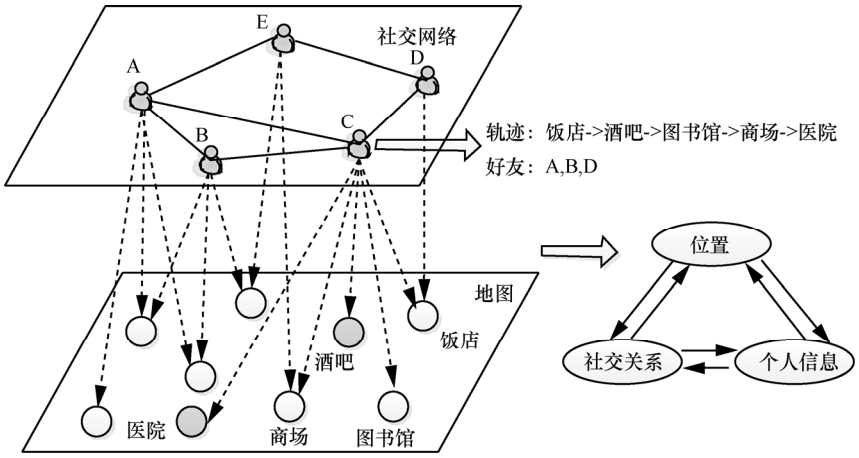


图 1 移动社交网络拓扑

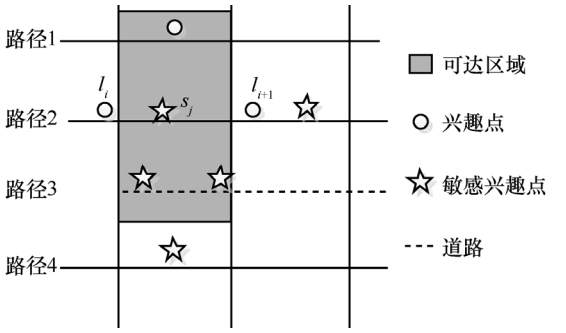


图 2　敏感兴趣点推理攻击模型

如第 3.1 节所述，社交网络中用户的身份和位置信息相互关联，位置信息可沿着好友关系在社交网络中传播，因此本文假设攻击者具有以下背景知识：①用户的移动轨迹 H；②社交网络中的其他信息 F，F 包括用户的社交关系、个人信息等。攻击者可利用上述背景知识推理用户隐藏的敏感兴趣点 s_i，如果存在一个位置点 $s_i \in S$，当攻击者推理出的用户已访问或将访问兴趣点 s_i 的置信度 $P(s_i|F,H,\lambda) > c_i$ 时，用户的隐私被破坏。

在移动社交网络中，用户自行决定是否发布位置信息，共享位置信息可获得更优质的服务，但面临隐私泄露带来的安全威胁，需要通过合理的计算来平衡服务的可用性与用户的隐私需求。

4　情景感知的位置隐私保护框架

本小节首先提出一种位置隐私保护框架，描述了位置隐私保护的基本过程，然后给出了计算兴趣点之间相关性的两种知识构建算法，进而提出了一种情景感知的隐私安全保护算法。该算法利用兴趣点之间的相关性及时空情景，实时判断发布当前位置是否会泄露用户的隐私，进而实现了隐私保护与服务可用性的平衡。

4.1　位置隐私保护基本过程

位置隐私保护框架如图 3 所示。本文中的位置隐私保护框架包含 3 个角色：用户、隐私保护服务器（PPS, privacy preserving server）和社交网络数据存储服务器（SNDSS, social network data store server）。其中，用户是移动社交网络中信息的发布者与查询者；PPS 利用 SNDSS 提供的数据构建知识，并计算用户的位置隐私保护需求是否得到满足；SNDSS 存储移动社交网络中用户已发布的信息，并提供信息查询与信息发布两种服务。

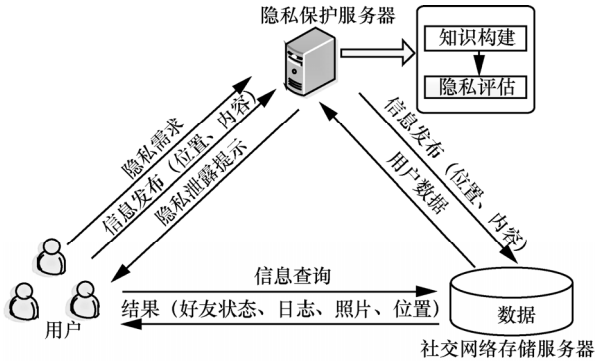


图 3　位置隐私保护框架

1) 信息查询

用户直接向 SNDSS 发起信息查询请求，由 SNDSS 返回用户请求的信息，如带有位置标签的日志、状态、照片等。

2) 信息发布

用户将信息发布请求<l, t, SC_u^t, *geo-content*>发送到 PPS，其中，l 表示用户的当前位置，t 表示当前时间，*geo-content* 表示带有位置标签的信息。PPS 收到用户的信息发布请求后，实时判断发布当前位置 l 是否满足用户对隐私保护的要求，如果是，则将该信息转发到 SNDSS；反之则提醒用户可能泄露的敏感兴趣点 s_i 及其置信度，由用户决定是否继续发布当前位置。

4.2　知识构建

PPS 基于 SNDSS 提供的移动轨迹信息 H 及社交网络信息 F 计算兴趣点之间的相关性。本文首先提出一种基本的知识构建算法，并在此基础上提出了一种基于协同过滤的启发式算法。

4.2.1　基本的知识构建算法

用户的移动行为在特定区域内遵循一定的规律性[11]，这些生活模式和社会规律可以通过分析大量用户的数据集合得出。频繁模式是指频繁出现在数据集中的模式。如频繁地同时出现在用户移动轨迹集 H 中的兴趣点的集合是频繁模式。本文利用频繁模式树[12]挖掘用户移动轨迹 H 中的频繁模式，进而得出兴趣点之间的关联规则(l_i, $l_{i+1} \Rightarrow l_j$)。关联规则的置信度，即兴趣点之间相关性的计算式为

$$P(l_j \mid l_i, l_{i+1}) = \frac{support(l_i, l_j, l_{i+1})}{support(l_i, l_{i+1})} \tag{1}$$

其中，*support* (l_i, l_j, l_{i+1})表示关联规则的支持度计数，即依次包含兴趣点 l_i、l_j 和 l_{i+1} 的移动轨迹数。

由于移动轨迹 H 带有方向性, 因此 $support(l_i, l_{i+1}) \neq support(l_{i+1}, l_i)$。

4.2.2 基于协同过滤的知识构建算法

基本方法未考虑用户对同一兴趣点差异化的价值取向。对于不同用户，同一关联规则的置信度是不同的。针对这一问题，本文提出一种基于协同过滤的知识构建算法，如图 4 所示。该方法通过分析用户移动轨迹 H 得到用户对已访问兴趣点的隐式评价，进而发现与其具有相似兴趣的用户，并基于最近邻居计算用户对未访问兴趣点的评分，进而得出对于该用户关联规则$(l_i, l_{i+1} \Rightarrow l_j)$的置信度。

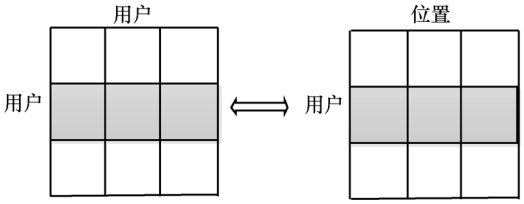


图 4 基于用户的协同过滤

1) 用户兴趣模型

设移动社交网络中存在两个集合：一个是用户的集合 U，总数是 m；一个是兴趣点的集合 L，总数是 n。现将这两个集合合并起来，排成一个 $m \times n$ 阶矩阵 R，行向量是用户集合 U，列向量是兴趣点集合 L,矩阵中的元素数是 $m \times n$。给定用户 u_k，其兴趣向量可以表示为 $R_{u_k}=<(l_1,r_{k1}), \cdots ,(l_i, r_{ki},) \cdots , (l_n, r_{kn})>$，$r_{ki}$ 表示兴趣点 l_i 的权重，即用户 u_k 对兴趣点 l_i 的评分。r_{ki} 可以通过统计用户对兴趣点 l_i 的重复访问次数得出。受信息检索中加权技术 TF-IDF 的启发，本文将兴趣点 l_i 视为“单词”，将用户的移动轨迹 H_{u_k} 视为“文档”，r_{ki} 的计算式为

$$r_{ki} = \frac{u_k.count_i}{\sum_{l_j \in H_{uk}} u_k.count_j} \cdot \lg \frac{|U|}{|\{u|u.count_i \geq 1, u \in U\}|} \quad (2)$$

其中，$count_i$ 表示用户访问兴趣点 l_i 的次数，U 表示所有用户的集合。

2) 关联规则置信度计算

矩阵 R 建立初期可能是一个稀疏矩阵。用户对未访问兴趣点的评分可由用户的最近邻居得出。本文用 Jaccard 系数计算用户社交关系之间的相似性，用余弦相似度计算用户兴趣的相似性。用户相似度 $sim(u_i, u_j)$的计算式为

$$\mathrm{sim}(u_i, u_j) = \alpha \frac{R_{ui} \cdot R_{uj}}{|R_{ui}|^2 |R_{uj}|^2} + (1-\alpha) \frac{|F_i \cap F_j|}{|F_i \cup F_j|} \quad (3)$$

其中，F_i 和 F_j 分别表示 u_i 和 u_j 好友的集合，α 是一个常数，$0<\alpha<1$。寻找最近邻的目标就是对每个用户 u，在用户空间中查找用户集合 $U'=\{u_1,u_2,\cdots,u_k\}$，使得 $u \notin U'$，并且 u_1 与 u 的相似性 $\mathrm{sim}(u_1, u)$最高，u_2 与 u 的相似性 $\mathrm{sim}(u_2, u)$次之，依此类推。令 l_j 是用户 u_k 未访问的兴趣点。r_{kj} 可由 u_k 的近邻用户对兴趣点 l_j 的评分得出，表示为

$$r_{kj} = \overline{r_k} + \frac{\sum_{u_i \in U'} \mathrm{sim}(u_i, u_k)}{|U'|} \cdot \sum_{u_i \in U'} (r_{ij} - \overline{r_i}) \cdot \mathrm{sim}(u_i, u_k) \quad (4)$$

其中，$\overline{r_k}$ 表示用户 u_k 对兴趣点的平均评分。用户对兴趣点的评分直接影响到用户的行为模式[11]，因此可将用户 u_k 对兴趣点的评分与关联规则的置信度线性结合得出关联规则对于该用户的置信度。对于用户 u_k，关联规则$(l_i, l_{i+1} \Rightarrow l_j)$的置信度计算式为

$$P(l_j | l_i, l_{i+1}) = \frac{(1+\beta \cdot r_{kj}) support(l_i, l_j, l_{i+1})}{(1+\beta \cdot \overline{r_k}) support(l_i, l_{i+1})} \quad (5)$$

其中，$support(l_i, l_j, l_{i+1})$与式(1)中的定义相同，$\beta$ 是一个常数，$0<\beta<1$。

4.3 隐私安全保护算法

PPS 收到用户 u_k 的信息发布请求$< l_{i+1}, t_{i+1}, SC_{uk}^{t}, geo\text{-}content>$后的处理流程如算法 1 所示。PPS 利用兴趣点之间的相关性及时空背景实时判断：①发布当前位置 l_{i+1} 是否会泄露 u_k 将要访问的敏感兴趣点 s_i（算法第 1)~5)行，其中 $P(s_i | l_{i+1})$可由式(1)或式(5)得出）；②发布当前位置 l_{i+1} 是否会泄露 u_k 已访问的敏感兴趣点 s_i（算法第 6)~20)行）。当且仅当任意一个敏感兴趣点 $s_i \in S$ 泄露的概率 P_i 均小于用户在该点的置信度要求 c_i 时，发布当前位置 l_{i+1} 满足用户的隐私需求。

算法 1 隐私评估算法

INPUT：$SC_{u_k}^{t}=\{(s_1, c_1),\cdots, (s_i, c_i),\cdots, (s_n, c_n)\}$

$H_{u_k}= <(l_1, t_1),(l_2, t_2),\cdots, (l_i, t_i)>$

l_{i+1} // 用户的当前位置

t_{i+1} // 当前时间

OUTPUT: G // 可能泄露的敏感兴趣点集合

$G \leftarrow \phi$

1) **for** each s_i in $S(k)$ //$S(k)$: 用户 k 的敏感兴趣点的集合
2) **if** $P(s_i \mid l_{i+1}) > c_i$
3) $G = G \cup < s_i, P(s_i \mid l_{i+1}) >$
4) **end if**
5) **end for**
6) $\triangle t = t_{i+1} - t_i$
7) **if** $\triangle t \le dis(l_i, l_{i+1}) / v_{max}$
8) **return** G
9) **else**
10) $R=\{l_m \mid dis(l_i,l_m)+dis(l_m,l_{i+1}) \le \Delta t \cdot v_{\max}\}$
11) **end if**
12) **if** $R \cap S(k) = \phi$
13) **return** G
14) **else**
15) **for** each s_i in $R \cap S(k)$
16) **if** $P(s_i \mid l_i, l_{i+1}, R) > c_i$
17) $G = G \cup < s_i, P(s_i \ l_i, l_{i+1}, R) >$
18) **end if**
19) **end for**
20) **end if**
21) **return** G

如图 2 所示，令 l_i 表示用户 u_k 在时刻 t_i 发布过的兴趣点，PPS 首先计算两次信息发布请求间的时间间隔 $\Delta t = t_{i+1} - t_i$，如果 $\Delta t < dis(l_i, l_{i+1}) / v_{\max}$，则说明在两次信息发布请求之间，$u_k$ 没有访问过其他兴趣点。其中，$dis(l_i, l_{i+1})$表示 l_i 与 l_{i+1} 间的曼哈顿距离，$v_{\max}$ 表示 u_k 的最大移动速度，PPS 将该信息发布请求转发到到 SNDSS（算法 6)~8)行）；反之如图 2 所示，PPS 将计算在 Δt 内 u_k 可能访问的兴趣集合 $R=\{l_m \mid dis(l_i, l_m)+dis(l_m, l_{i+1}) \le \Delta t \cdot v_{\max}\}$（算法第 10)行）。针对 $R \cap S(k)$中的每个兴趣点 s_i，PPS 利用后验概率 $P(s_i \mid l_i, l_{i+1}, R)$计算 u_k 访问过 s_i 的置信度 P_i，进而判断用户对 s_i 的置信度要求 c_i 是否得到满足（算法 15)~19)行）。其中，后验概率 $P(s_i \mid l_i, l_{i+1}, R)$ 可由贝叶斯公式计算得出

$$p(s_i \mid l_i, l_{i+1}, R) = \frac{P(l_i, s_i, l_{i+1})}{P(l_i, R, l_{i+1})} = \frac{P(s_i \mid l_i, l_{i+1})}{\sum_{l \in R} P(l \mid l_i, l_{i+1})} \quad (6)$$

其中，$P(s_i \mid l_i, l_{i+1})$可由式(1)或式(5)得出。当 $P_i < c_i$ 对于任意一个敏感兴趣点 $s_i \in S$ 均成立时，共享位置 l_{i+1} 满足用户的隐私要求，PPS 将该信息发布请求转发到 SNDSS；反之，PPS 则通知用户可能泄露的敏感兴趣点 s_i 及其置信度 $P(s_i \mid l_i, l_{i+1}, R)$，最终由用户决定是否继续发布当前位置 l_{i+1}。

5　仿真实验

为测试本文提出的位置隐私保护方法，本次实验所处的硬件环境为：Inter(R) Celeron G530, 500 GB 硬盘，4 GB 内存，Win7 32 位系统。实验数据集为文献[13]公开的 Foursquare 的签到数据集。该数据集包含 18 107 个用户在 2010 年 3 月到 2011 年 1 月间的签到位置、签到时间及好友关系。实验从两个方面对位置隐私保护算法进行评估：①位置隐私保护方法的有效性；②位置服务的可用性。

5.1　位置隐私保护方法的有效性

如第 3.2 节所述，移动社交网络中的位置信息是由用户主动共享的，用户通常不会在敏感兴趣点签到，因此本文所用数据集中的用户签到轨迹没有包含用户真正的敏感兴趣点。针对这一问题，本文在实验中假设当且仅当用户 u_k 签到轨迹中包含兴趣点 l_i 时，用户 u_k 访问过兴趣点 l_i。基于上述假设，本文为每个用户生成了两个敏感兴趣点集合：①从用户的签到轨迹中随机删除一些兴趣点作为敏感兴趣点集合 A；②随机添加一些与用户已签到位置（地理上）临近的兴趣点作为敏感兴趣点集合 B。本文用以下两个指标评价本文中位置隐私保护方法的有效性：①正确率（true positive），用户访问过集合 A 中兴趣点的平均置信度；②误报率（false positive），用户访问过集合 B 中兴趣点的平均置信度。

在实验过程中, 从数据集中随机选取了 1 000 名用户，从每个用户的签到轨迹中选取两个兴趣点 l_a 和 l_b（用户在 l_a 和 l_b 的签到间隔时间应大于平均的签到间隔时间）。针对每个用户，实验中分别删除 l_a 和 l_b 间的 1、3、5 和 10 个兴趣点作为兴趣点集合 A，添加与 l_a 和 l_b 在地理位置上临近的 5、10、15 和 20 个兴趣点作为兴趣点集合 B。基本算法为基于第 4.2.1 节知识构建算法的隐私保护算法, 协同过滤算法为基于第 4.2.2 节中知识构

建算法的隐私保护算法。

准确率实验和误报率实验的实验结果如图 5 和图 6 所示。两种算法的准确率均未随删除的兴趣点的增加而明显降低，当被删除的兴趣点数为 1 时，基于协同过滤算法的准确率为 82%；当被删除的兴趣点数为 10 时，算法的准确率为 67%。算法的误报率始终较低，当增加的兴趣点数为 20 时，算法的误报率为 10%。实验结果表明本文中的隐私保护算法可以准确地计算出用户敏感兴趣点隐私泄露的概率，进而保护用户的位置隐私。实验同时证明了当兴趣点之间存在较强的相关性时，只在敏感兴趣点关闭移动社交服务不能有效地保护用户的隐私。

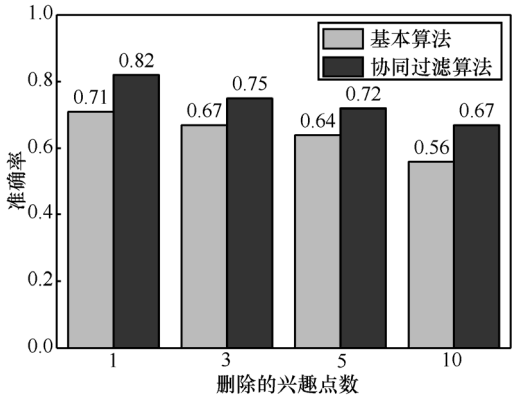


图 5 准确率实验

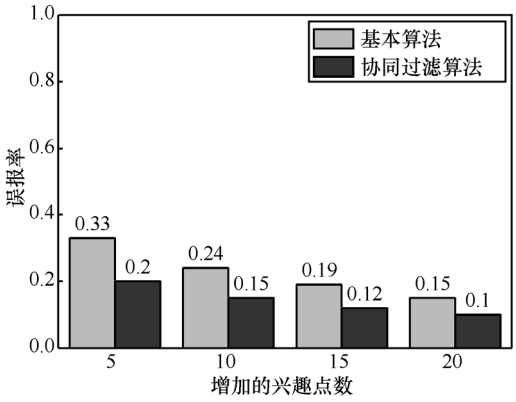


图 6 误报率实验

从图 5 和图 6 中可以看出，协同过滤算法在各种条件下准确率均高于基本算法，且误报率低于基本算法。如被删除兴趣点的数为 5 时，基本算法的准确率为 64%，协同过滤算法的准确率为 72%。这是因为基本算法只利用了大众的历史数据，当用户的行为模式具有自己的个性化特征时，基本算法难以判断该用户是否访问了特定兴趣点。

5.2 隐私保护度与服务可用性分析

在移动社交网络中，位置隐私保护需要建立在服务可用性的基础上，因此本文从可用性的角度对所提出的位置隐私保护算法进行了评估。可用性指标 γ 定义为

$$\gamma=\frac{\eta'}{\eta} \tag{7}$$

其中，η 表示用户想发布的位置信息数，η' 表示用户在保证隐私安全的基础上可发布的位置数。实验中随机从数据集中选取 5%和 10%的兴趣点作为敏感兴趣点，为计算方便，实验中假设用户在敏感兴趣点的隐私需求是相同的（本文中的方法支持用户个性化、动态的隐私需求）。通过改变用户在敏感兴趣点的置信度要求以测试在用户不同的隐私需求下的服务的可用性。

敏感兴趣点比例为 5%和 10%的服务可用性实验结果如图 7 和图 8 所示。当敏感兴趣点比例为 5%、用户的置信度要求为 0.5 时，协同过滤算法的服务可用性为 0.93；当用户的置信度要求提高到 0.1 时，该算法的服务可用性为 0.78。当敏感兴趣点的比例为 10%、用户的置信度要求为 0.5 时，协同过滤算法的服务的可用性为 0.9；当用户的置信度要求提高到 0.1 时，该算法的服务可用性为 0.64。由此可以得出本文中算法的服务可用性较高，且并未随用户置信度要求的提高而快速下降。这主要因为本文方法的误报率较低（其中协同过滤算法的误报率在 0.2 以下），只有与敏感点相关性较高的兴趣点的泄露概率才会超过用户的置信度要求。实验结果说明本文中的位置保护方法可以在满足用户隐私需求的基础上，保证服务的可用性。

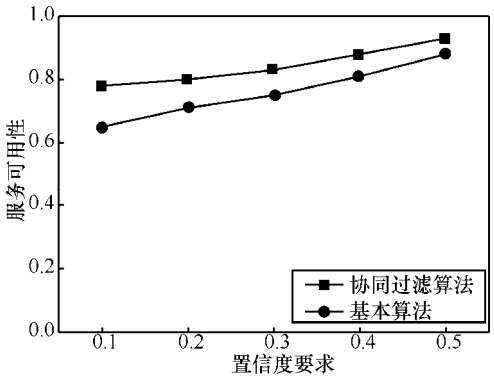


图 7 服务可用性实验（敏感兴趣点比例为 5%）

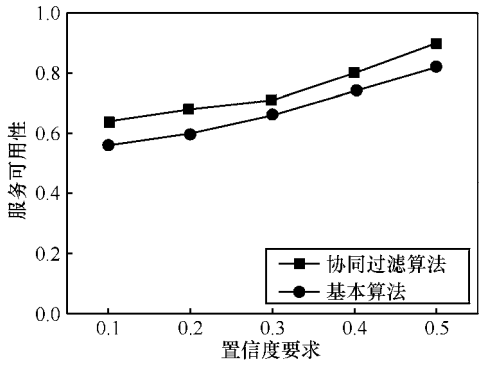


图 8　服务可用性实验（敏感兴趣点比例为 10%）

6　结束语

本文研究了移动社交网络中用户敏感兴趣点泄露问题，利用置信度将用户个性化、动态的隐私需求进行了形式化描述，在此基础上提出了一种情景感知的隐私保护方法。该方法利用兴趣点间的相关性及时空情景计算共享当前位置能否满足用户的隐私需求。为实现实时预测，本文提出两种知识构建算法离线的计算兴趣点间的相关性。基本算法基于大众的历史轨迹，基于协同过滤算法进一步引入了用户的社交关系及个人信息。本文在真实数据集上进行了充分实验，结果证明该方法可以在保护用户隐私的前提下，保证服务可用性。

参考文献：

[1] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, San Francisco, CA. 2003: 31-42.

[2] BAMBA B, LIU L, PESTI P, et al. Supporting anonymous location queries in mobile environments with privacygrid[C]//Proceedings of the 17th International World Wide Web Conference, Beijing. 2008:237-246.

[3] LI H P, HU H, XU J. Nearby friend alert: location anonymity in mobile geosocial networks[J]. Pervasive Computing, 2013,12(4): 62-70.

[4] HENGARTNER U. Hiding location information from location-based services[C]//Proceedings of International Conference on Mobile Data Management, Mannheim. 2007:268-272.

[5] ZHAO X, LI L, XUE G. Checking in without worries: location privacy in location based social networks[C]//Proceedings of IEEE INFOCOM, Turin.2013: 3003-3011.

[6] LI X, JUNG T. Search me if you can: privacy-preserving location query service[C]//Proceedings of IEEE INFOCOM, Turin. 2013: 2760-2768.

[7] SHOKRI R, THEODORAKOPOULOS G, DANEZIS G, et al. Quantifying location privacy: the case of sporadic locate on exposure[C]//Proceedings of the 11st International Conference on Privacy Enhancing Technologies Symposium (PETS), Waterloo, ON. 2011:57-76.

[8] CHENG R, ZHANG Y, BERTINO E, et al. Preserving user location privacy in mobile data management infrastructures[C]//Proceedings of the 6th Privacy Enhancing Technologies (PETS), Cambridge. 2006:393-412.

[9] WANG Y, XU D, HE X, et al. L2P2: location-aware location privacy protection for location-based services[C]//Proceedings of IEEE INFOCOM, Orlando, FL. 2012: 1996-2004.

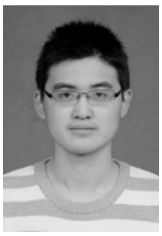
[10] GHINITA G, GABRIEL M L, SILVESTRI C, et al. Preventing velocity-based linkage attacks in location-aware applications[C]//Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information System, Seattle, WA. 2009:246-255.

[11] NOULAS S, SCELLATO C, MASCOLO C, et al. An empirical study of geographic user activity patterns in foursquare[C]//Proceedings of 5th International Conference on Weblogs and Social Media, Barcelona, Catalonia. 2011:17-21.

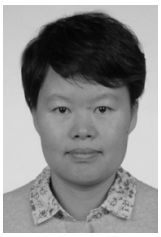
[12] HAN J, PEI J, YIN Y, et al. Mining frequent patterns without candidate generation: a frequent-pattern tree approach[J]. Data Mining & Knowledge Discovery, 2004, 8(1): 53-87.

[13] GAO H, TANG J, LIU H. Exploring social-historical ties on location-based social networks[C]//Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, Dublin. 2012: 114-121.

作者简介：



许志凯（1988-），男，山东潍坊人，哈尔滨工业大学博士生，主要研究方向为云计算和信息安全。



张宏莉（1973-），女，吉林榆树人，哈尔滨工业大学教授、博士生导师，主要研究方向为网络与信息安全、网络测量与建模、网络计算、并行处理等。

史建焘（1980-），男，黑龙江哈尔滨人，哈尔滨工业大学讲师，主要研究方向为 P2P 和网络安全。

田志宏（1978-），男，黑龙江哈尔滨人，哈尔滨工业大学副教授、博士生导师，主要研究方向为网格通信和信息安全。