移动环境下 LBS 位置隐私保护

彭志宇* 李善平
(浙江大学计算机科学与技术学院 杭州 310027)

摘 要：用 k 匿名模型对基于位置信息的服务(LBS)中的位置隐私进行保护是近年来研究的热点。在移动用户不断发出查询的场景下，该文提出了移动模式攻击(MPA)，使得传统的针对孤立查询的隐私保护算法均失效。基于熵理论，提出了熵匿名度度量，并以此为基础提出了移动环境下的模糊化算法 Mclique，实验证明其有效地抵御了 MPA 攻击。通过简化 Mclique 算法中熵的计算，提出了快速模糊化算法 Fclique，实验证明 Fclique 不仅仍具有较强的 MPA 抵御能力，且极大提高了时间效率。

关键词：移动模式攻击；隐私保护；k 匿名；基于位置信息的服务
中图分类号：TP309 **文献标识码**：A **文章编号**：1009-5896(2011)05-1211-06
DOI: 10.3724/SP.J.1146.2010.01050

Protecting Location Privacy in Location-based Services in Mobile Environments

Peng Zhi-yu Li Shan-ping
(*College of Computer Science and Technology, Zhejiang University, Hangzhou* 310027, *China*)

Abstract: The k-anonymity model is employed to protect the location privacy in Location-Based Services (LBS) in recent years. A Moving-Pattern Attack (MPA) is presented, in a scenario that the moving user keeps delivering queries. It is shown that traditional algorithms fail in this attack. In order to protect against MPA, a novel anonymity measurement based on entropy is proposed, which leads to a cloaking algorithm under mobile environments, i.e. Mclique. Experiments show that Mclique protects effectively user privacy against MPA. By simplifying the computing of entropy in Mclique, a fast cloaking algorithm, Fclique, is proposed. Experiments show that Fclique is capable of surviving MPA, and reduces greatly the time complexity as well.

Key words: Moving Pattern Attack (MPA); Privacy protection; k-anonymity; Location-Based Service (LBS)

1 引言

在基于位置信息的服务(Location Based Services, LBS)中，用户通过向服务端提供自己的位置信息，得到相应的查询结果。如查询离自己最近的医院，周边的饭店有哪些等。然而，用户位置隐私在 LBS 中存在着严重的威胁[1]。

LBS 位置匿名研究是解决这一隐私威胁的途径。现有的位置匿名研究借鉴了数据库中的 k 匿名机制[2]，将用户的精确位置模糊化为一个足够大的区域(简称模糊区域)发给 SP(Service Provider)，而这个区域中包含了其它至少 k–1 个用户，致使服务端无法得知究竟是 k 个用户中的哪一个发来的查询。如文献[3]提出了一种基于四分树(Quard Tree)结构的算法来进行位置信息的 k 匿名模糊化。文献[4]提出的 CliqueCloak 算法则允许用户设置个性化的隐私和服务质量参数，通过找到一些可以组成匿名团(clique)的用户，来确定这些用户共同的模糊化区域。文献[5]提出了一种不依赖系统可信度的模糊化算法。然而，这些研究考虑的都仅是孤立的服务，即把每一次的模糊化都看成一个独立的事件进行处理。而当用户在移动环境中不断发出 LBS 申请时，如果攻击者得到用户的历史模糊化区域，这些算法的匿名效果则将遭受严峻的考验，连续查询攻击[6,7]和最大速度攻击[8,9]是这一类攻击的典型。其中前者是通过对用户发送同一查询的概率进行假设和建模，从而猜测出查询所属的用户。而后者通过得到用户的最大移动速度，从而破坏用户在当前模糊化区域中的位置匿名。

与这两种攻击类似，本文提出一种移动模式攻击(Moving-Pattern Attack, MPA)，同样是利用历史模糊区域对位置隐私进行攻击，但 MPA 通过对用户移动模式信息的预测能够更高效和准确地定位用户。图 1 是一个 MPA 攻击的示例。假设用户 A, B, C 在 t_i 时刻被成功匿名产生了一个模糊区域 R_i，而攻击者获知用户 A 行驶在限速 v (km/h)的公路

2010-09-29 收到，2011-03-09 改回
国家自然科学基金(60773180)资助课题
*通信作者：彭志宇 pzy202@163.com

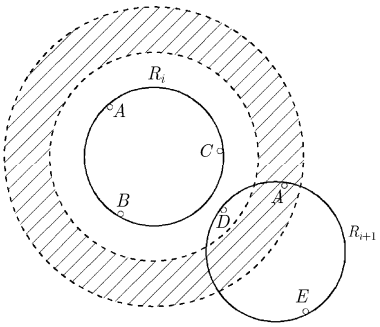

图 1 MPA 攻击示例

上，则其可以假定用户 A 的运行速度将在$[v-e,v+e]$(km/h)的范围内，通过计算得知用户 A 在 t_{i+1}时刻必出现在阴影区域中。若在 t_{i+1}时刻 A, D, E 3个用户产生了模糊区域 R_{i+1}，则攻击者可以肯定 A 在 t_{i+1}时刻将出现在阴影区域与 R_{i+1}的交集处，而此区域中只有 A 一个用户，即匿名失败。

一般地，攻击者通过观察或者计算得知了某用户 A 的运动模式，利用 A 在 T_i时刻的模糊区域 R_i 计算出 T_{i+1}时刻 A 的位置概率密度分布函数 $u(x)$，$u(x)$满足以下条件

$$\int_0^R u(x)\mathrm{d}x = 1 \tag{1}$$

其中 x 为平面上一点与 R_i的圆心的距离，R 为用户在最大速度下此段时间可到达的距离。一旦攻击者获取了 A 在 T_{i+1}时刻的模糊区域 R_{i+1}，$u(x)$就可以用来有效降低用户 A 在 R_{i+1}中的位置分布概率，使得 k 匿名的匿名度量 k 变得毫无意义，甚至在极端情况下，攻击者能准确定位出 A 的位置。

2 系统模型

整个 LBS 匿名系统包括 3 个部分，用户、可信的匿名器以及服务提供商。用户将明码的 LBS 查询(id,loc,query)发给匿名器，其中 id 是用户的真实 ID，loc 是用户的位置，query 是用户所要查询的内容。匿名器将 id 用一个假名 id′替代，再采用 LBS 匿名算法得到一个模糊区域 R，将用户的位置 loc 用 R 替代，于是将新的查询(id′,R,query)发送给服务提供商。服务提供商根据匿名查询的内容将结果返回给匿名器，最后由匿名器将此结果返回给用户。

定义 1(用户的熵) 若在一个匿名区域 R 中的 n 个用户的位置点分别是 $L_1, L_2, \cdots, L_n$，则对 R 中的任一用户 v，把它的熵 $E(v)$定义为

$$E(v) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

其中 p_i是用户 v 出现在位置 L_i上的概率。

定义 2(熵匿名度) 把用户 v 的熵匿名度定义为

$$K(v) = 2^{E(v)} \tag{3}$$

为满足用户的个性化需求，系统允许用户指定如下参数：

(1)k 代表用户对匿名程度的要求。即用户至少处于与其他 k-1 个人处于不可区分的状态。

(2)$R_{\max}$ 代表用户可以接受的最大模糊区域的半径。模糊区域越大，其返回的查询结果的质量就越差，因此 $R_{\max}$用以控制 QoS(Quality of Service)。

(3)dt 代表最迟响应的时间间隔。即从用户发出查询到接收结果之间的最大时间间隔，在移动环境中一般时间间隔越大，用户偏离原位置的距离越远，查询结果的效果也越差，因此这也是控制 QoS 的参数。

定义 3(无向图模型) 定义 $G(V,E)$为一无向图，V 是图中的节点集合，每一个节点代表一个向匿名器提交了 LBS 查询的用户。E 是图中的边的集合，图中的两个点 v 和 w 之间存在一条边当且仅当下列条件满足：

$$\mathrm{Distance}(v,w) \leqslant \min(v.R_{\max}, w.R_{\max}) \tag{4}$$

其中 Distance(v,w)表示点 v和点 w之间的直线距离。

定义 4(匿名集) 在一个图 $G(V,E)$中，一个用户集合 U 成为一个匿名集，当且仅当下列两个条件均满足：

$$\forall v,\ w \in U,\quad \{v,w\} \in E \tag{5}$$

$$\forall v \in U,\quad K(v) \geqslant v.k \tag{6}$$

3 移动环境下的 LBS 匿名算法

3.1 寻找极大完全图

一个点集成为匿名集需要满足式(5)和式(6)两个条件。本节探讨怎样寻找满足式(5)的点集。给出下面的定义。

定义 5(完全图) 在无向图 $G(V,E)$中，如果存在一个子图 C，其中 C 中的节点是两两相连的，则 C 叫做完全图，且 C 可以叫做 G 的一个完全子图。

定义 6(极大完全子图) 对一个完全图 C 来说，如果不存在满足以下条件的完全图 C'：$C \neq C'$ 且 $C \subset C'$，则称 C 为一个极大完全子图。

式(5)的目标即在一个图 G 中找到所有的极大完全子图。本文采取了如下的累进式维护极大完全子图的方法来解决这一问题。

定义 7 (动态图) 对一个无向图 $G(V,E)$，它的动态图 G_t定义为 $G_t = (V,E_t)$，其中 $t = 0,1,\cdots,|E|$; $E_{t-1} \subset E_t$; $E_0 = \varnothing$。

把 C_t 记为 G_t 中所有极大完全子图的集合。特别地，当 $t = 0$ 时，G_0 是一个没有边存在的图，$C_0 =$

$\{\{1\},\{2\},\cdots,\{n\}\}$。

定义 8　在动态图 G_t 中，若 A 为一个点集，定义 $[A]_t$ 为 G_t 中所有与 A 的交集不为空的极大完全子图的集合。

根据以上定义，可得定理：

定理 1　对动态图 G_t，若 $\{v,w\}$ 是从 t 到 $t+1$ 时刻所加入的新边，那么 C_{t+1} 与 C_t 有如下的关系：

(1)所有在 C_t 中而不在 $[\{v,w\}]_t$ 中的元素，都在 C_{t+1} 中；

(2)对所有 $(A,B)\in[\{v\}]_t\times[\{w\}]_t$ 有如下论断：

(a)令 $L=(A\cap B)\cup\{v,w\}$，则 L 是一个完全图，若 L 是极大完全图，则 $L\in C_{t+1}$

(b)若 $|A-B|=1$，则 $A\notin C_{t+1}$；否则，若 $|A-B|>1$ 且 A 仍是极大完全图，则 $A\in C_{t+1}$

(c)若 $|B-A|=1$，则 $B\notin C_{t+1}$；否则，若 $|B-A|>1$ 且 B 仍是极大完全图，则 $B\in C_{t+1}$

(3) C_{t+1} 中所有的元素都经由上述两条规则得来。

证明　对(1)，因为所有在 C_t 中而不在 $[\{v,w\}]_t$ 中的元素都不包含点 v 和点 w，故新边 $\{v,w\}$ 的加入不改变这些极大完全图。

对(2)(a)，因为 A 和 B 本身都是极大完全图，故其交集显然也是极大完全图，又由定义知 A 中的点都与 v 有边，故 $A\cap B$ 中的点都与 v 有边，同理 $A\cap B$ 中的点都与 w 有边，所以 $(A\cap B)\cup\{v,w\}$ 是完全图。

对(2)(b)，在 G_t 中 $\{v,w\}$ 是不存在的，故 $v\notin[\{w\}]_t$，所以 $|A-B|=1$ 隐含 $A-B=\{v\}$，可知 $A\subset L$，故 A 不是极大完全图，$A\notin C_{t+1}$。

同理(2)(c)可证。

对(3)，假设有 $M\in C_{t+1}$ 则需证明 M 必由(1)或(2)得来。分情况讨论，若 $M\cap\{v,w\}=\varnothing$，则显然 $M\in C_t$ 成立，故 M 是由(1)得来；若 $\{v,w\}\subseteq M$，则可知 $M-\{v\}\in C_t$，$M-\{w\}\in C_t$，则 M 是由(2)(a)得来的；若 $v\in M$ 而 $w\notin M$，则 $M\in[\{v\}]_t$，这时若存在一个 B 满足 $B\in[\{w\}]_t$ 且 $|A-B|=1$，那么根据(2)(a)知 $L=(A\cap B)\cup\{v,w\}\in C_{t+1}$ 且 $M\subset L$，这与 $M\in C_{t+1}$ 矛盾，故不存在这样的 B，所以 M 是由(2)(b) 得来的；同理若 $w\in M$ 而 $v\notin M$ 时也可证。于是(3)得证。　证毕

定义 9　对集合 M 和 N，定义 $\mathrm{Max\_Join}(M\times N)$ 如下：$\mathrm{Max\_Join}(M\times N)=\mathrm{Maxi}(\mathrm{Join}(M\times N))$，其中 $\mathrm{Join}(M\times N)=\{A\cap B|A\in M,B\in N\}$，$\mathrm{Maxi}(M)$ 定义为若 $A\in M$，且对所有 $B\in M$ 都有 $A\not\subset B$，则 $A\in\mathrm{Maxi}(M)$。

推论 1　对动态图 G_t，若 $\{v,w\}$ 是从 t 到 $t+1$ 时刻所加入的新边，那么 C_{t+1} 与 C_t 有如下的关系：

(1)所有在 C_t 中而不在 $[\{v,w\}]_t$ 中的元素，都在 C_{t+1} 中；

(2)所有 $L\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$，$L\cup\{v,w\}\in C_{t+1}$；

(3)对所有 $A\in[\{v\}]_t$，如果存在 $C\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$，且 $A-\{v\}\subseteq C$，则 $A\notin C_{t+1}$，否则 $A\in C_{t+1}$；

(4)对所有 $B\in[\{w\}]_t$，如果存在 $C\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$，且 $B-\{w\}\subseteq C$，则 $B\notin C_{t+1}$，否则 $B\in C_{t+1}$；

(5) C_{t+1} 中所有的元素都经由上述 4 条规则得来。

证明　(1)同定理 1 中的证明。

对(2)，由 Max_Join 的定义结合定理 1 可知 $L\cup\{v,w\}$ 是完全图，则只需证明它是极大完全图。假设存在 $M\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$ 且 $M\cup\{v,w\}\notin C_{t+1}$，则一定存在 N 使得 $M\subset N$ 且 $N\cup\{v,w\}\in C_{t+1}$，由定理 1(3)的证明知 N 是由定理 1(2)(a)得来，于是 $N\in\mathrm{Join}([\{v\}]_t\times[\{w\}]_t)$，而 $M\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$，可知 $N\subset M$，这与前面 $M\subset N$ 矛盾。

对(3)，若存在 $C\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$ 且 $A-\{v\}\subseteq C$，则 $C\cup\{v,w\}$ 是完全图，而显然 $A\subset(C\cup\{v,w\})$，于是 A 不可能是极大完全图，$A\notin C_{t+1}$。若对所有的 $C\in\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$ 都有 $A-\{v\}\supset C$，则说明 $A\cup\{w\}$ 不是完全图，故由于 $A\in C_t$ 知在加了 $\{v,w\}$ 后 A 仍是极大完全图，所以 $A\in C_{t+1}$。

(4)同(3)可证。

对(5)，因为推论 1 的 4 条规则与定理 1 中的规则是分别对应的，故可得。　证毕

根据推论 1，给出寻找所有极大完全图的算法。

算法 1

(1)初始时，令 $C_0=\{\{1\},\{2\},\cdots,\{n\}\}$ 以及 $t=0$；

(2)令 $C_{t+1}=C_t$；

(3)若 t 时刻加入的边为 (v,w)，令 $M=\mathrm{Max\_Join}([\{v\}]_t\times[\{w\}]_t)$；

(4)对所有的 $L\in M$，将 $L\cup\{v,w\}$ 加入到 C_{t+1} 中；

(5)对所有的 $A\in[\{v\}]_t$ 做如下动作：对所有的 $L\in M$，若有 $A-\{v\}\subseteq L$ 则将 A 从 C_{t+1} 中删除；

(6)对所有的 $B\in[\{w\}]_t$ 做如下动作：对所有的 $L\in M$，若有 $B-\{w\}\subseteq L$ 则将 B 从 C_{t+1} 中删除；

(7)令 $t=t+1$，若 t 等于 n 则算法结束，否则继续执行(2)-(7)的指令。

3.2 熵匿名度测试

本节探讨匿名集定义中式(6)的满足。把完全图中的点集称为候选集，把式(6)称为熵匿名测试。本

节探讨的即在候选集中通过熵匿名测试来寻找匿名集。

若一个点集 U 含有 n 个点分别为 $N_1,N_2,\cdots,N_n$，其中 N_i 的位置坐标为 (x_i,y_i)，节点 N_i 在 T_k 时刻的模糊化区域 R_i 的圆心坐标为 (x_{o_i},y_{o_i})，且其在 T_{k+1} 时刻的位置概率密度分布函数为 $u_i(r)$，其中 r 为平面上任一点与 R_i 的圆心的距离。现在来计算点 N_i 的熵匿名度。

点 N_j 与 R_i 的圆心的距离 l_{ij} 为

$$l_{ij}=\sqrt{(x_{o\_i}-x_j)^2+(y_{o\_i}-y_j)^2} \tag{7}$$

于是 t_{k+1} 时刻点 N_i 位于 N_j 点位置处的概率 p_{ij} 为

$$p_{ij}=u(l_{ij})\Big/\sum_{k=1}^{n}u(l_{ik}) \tag{8}$$

可以得出此刻 N_i 点的熵为

$$E(N_i)=-\sum_{k=1}^{n}p_{ik}\log_2 p_{ik} \tag{9}$$

其熵匿名度为

$$K(N_i)=2^{E(N_i)} \tag{10}$$

于是，当 $K(N_i)\geq N_ik$ 时，则 N_i 的熵匿名度测试通过。一个候选集中所有点都通过了匿名度测试，则这个匿名集就成为匿名集。

3.3 节点加入

LBS 匿名系统在每次查询到来时做出匿名集查找的动作。一个查询对应着动态图中的一个节点，所以新的查询到来意味着动态图中新节点的加入。

定理 2 在动态图 $G_t(V,E_t)$ 中加入节点 v 以及 n 条与 v 相关的边 $\{v,w_1\},\{v,w_2\},\cdots,\{v,w_n\}$ 构成的 G_{t+n}，若 G_t 中不存在匿名集，则 G_{t+n} 中的匿名集一定包含节点 v。

证明 假设 G_{t+n} 中存在一个匿名集 $\{N_1,N_2,\cdots,N_n\}$，且 $v\neq N_i$，$0\leqslant i\leqslant n$。可知此匿名集中任一条边 $\{N_i,N_j\}$ 都与 v 无关，即 $\{N_i,N_j\}\in E_t$，又因 $N_i\in V$，故匿名集 $\{N_1,N_2,\cdots,N_n\}$ 完全在 G_t 中，这与 G_t 不存在匿名集的假设矛盾，故知 G_{t+n} 中不存在不包含 v 的匿名集。 证毕

根据定理 2，给出新节点进入时的匿名算法。

算法 2

(1)为新加入的查询创建一个新的节点 v;

(2)找到 v 的所有邻居，并将它们放入堆栈 $v_$neighbor 中;

(3)从堆栈 $v_$neighbor 中弹出一个节点 w;

(4)将一条边 (v,w) 加入到图中，并按照算法 1 更新极大完全图;

(5)如果堆栈 $v_$neighbor 不为空，则继续执行指令(3)-指令(5)，否则执行指令(6);

(6)对所有 $[\{v\}]_t$ 中的极大完全图 C，做如下动作：

(a)令 $D=C-\{v\}$;

(b)对所有 D 的子集 E，对集合 $E+\{v\}$ 进行熵匿名度测试，如果测试成功则将其作为匿名集返回，且算法结束。

3.4 节点离开

当匿名集被找到或者有节点到达最迟响应时间限度还没有匿名成功，就有节点从动态图离开。节点离开时，需要相应地更新整个动态图的极大完全图。

算法 3

(1)对 $[v]_t$ 中的所有元素 c 做如下动作：

(a)将 v 从 c 中删除;

(b)将 c 加入集合 TmpSet 中;

(2)设 MCSet 为当前所有极大完全图的集合，对所有 d 属于 MCSet ，做如下动作：

如果 $v\in d$ 则将 v 从 d 中删除;

(3)对所有 TmpSet 中的元素 c 做如下动作：对所有 MCSet 中的元素 d，如果 $c\subseteq d$ 则将 c 从 MCSet 中删除。

4 模拟实验

4.1 实验设置

实验采用著名的 Network-based Generator of Moving Objects[10]模拟器对市区的交通网络图中各用户的运动轨迹进行模拟。定义产生了 10000 个用户，其速度设定为 3 个档次，其中 2000 个低速用户的最大速度在 5–15 km/h 之间，6000 个中速用户的速度在 30–50 km/h 之间，2000 个高速用户的最大速度在 80–120 km/h 之间。每个用户不断发出查询，每个查询都在 2–7 之间随机选一个数做为其匿名要求 k。初始查询在 0–100 s 之间，其查询间隔固定为 50 s，其最大延迟时间 dt 设定为 3 s，这样保证了下一个查询发出之前，上一个查询已经成功返回或者超时停止。每个用户的最大模糊区域的半径 $R_{\max}$ 设定为其最大时速的 1/12。

4.2 防范移动模式攻击

移动模式攻击(MPA)假设攻击者可以得到所有用户产生的模糊区域以及这些区域中的用户 ID。攻击者利用某一用户过去最近的两个模糊区域的圆心距离以及其相应的间隔时间计算出用户的平均速度，并以此作为用户下一段时间的运行速度，从而在下一时刻模糊区域产生时判断用户的大概位置。作为示例，假设用户 A 在 T_i 时刻产生了以 O_1 为圆心的模糊区域 R_1，在 T_{i+1} 时刻产生了以 O_2 为圆心

的模糊区域 R_2，若 O_1 与 O_2 之间的距离为 r，预测用户在 T_{i+2} 时刻出现的位置与 O_2 距离 l 将为

$$l = r(T_{i+2} - T_{i+1})/(T_{i+1} - T_i) \tag{11}$$

攻击者把 A 用户在 T_{i+2} 时刻产生的模糊区域中与 O_2 的距离最接近 l 的点识别成 A 点。

为了防范 MPA 攻击，Mclique 算法需要预测用户每一次匿名成功后的位置概率密度函数。有研究表明，在乡村公路和高速公路上，运行车速一般成正态分布[11]。即运行车速的概率密度分布满足：

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{12}$$

为了得到数学期望 μ 和标准差 σ，本文用 4.1 节所描述的模拟器和实验设置进行了模拟，统计所有用户在每个时间单元里的运动速度与其实际最大速度的关系，以用户本身的最大速度为 1，综合所有用户的数据进行回归得到：μ=0.926，σ=0.136。若 T_i 时刻用户的模糊化区域圆心为 O，T_{i+1} 时刻的查询与 T_i 时刻的时间间隔为 Δt，用户的最大速度为 v，则用户在 T_{i+1} 时刻的位置概率密度分布函数为

$$u(x) = v\Delta t\frac{1}{\sqrt{2\pi}\times 0.136}\exp\left(-\frac{(x-0.926)^2}{0.136^2}\right) \tag{13}$$

其中 x 代表用户在 T_{i+1} 时刻的位置与 O 的距离。

本文考察了 MPA 攻击下，Mclique 与 Clique Cloak 的匿名效果对比。对具有不同匿名要求 k 的用户分别进行统计，考察匿名要求分别从 2 到 7 时，Mclique 的被识别率，CliqueCloak 的被识别率以及理论识别率的对比。从图 2 的模拟结果可见随着匿名要求 k 的上升，Mclique 的识别率迅速下降，并十分接近理论识别率，而 CliqueCloak 算法的识别率则下降缓慢，甚至在理论匿名度为 6，7 时，攻击者还有 30%以上的机率能够识别出用户。需要说明的是，当理论匿名度为 2 时，两种算法的识别率都低于理论识别率，这是因为这些匿名要求为 2 的用户往往与其他匿名要求更高的用户组成了匿名集，故其匿名效果反而优于理论值。

5　时间效率及性能改进

模糊化处理时间是从一个新查询进入开始计时，直到其匿名成功所耗费的时间。Mclique 算法需要枚举候选集的子集进行熵计算，故其模糊化处理时间有指数级的复杂度。

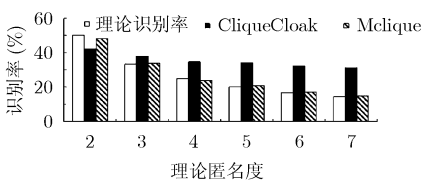

图 2 防范 MPA 攻击的效果

快速算法 Fclique 基于两个事实来改进算法。首先式(2)所示的熵函数 $E(v)$，在所有的 p_i 相等或者很接近时取到最大或较大值，即寻找匿名集时要尽量避免概率密度的“奇异点”。其次，若用户在上一时刻的模糊区域中各点出现的概率基本相等，则即使预测出它的速度，其下一位置也将呈现一个大致等概率的可能区域。

如图 3 所示，R_i 是某用户 A 在 T_i 时刻的模糊区域，半径为 r，假设用户的速度为 v，且 A 在 R_i 中的位置概率均匀分布，则 Δt 时间后其位置概率密度将均匀地分布在阴影部分所示的环形区域上，我们将这一区域称为均匀分布环形区。Fclique 的做法是，用户 A 在 T_{i+1} 时刻只在均匀分布环形区内寻找匿名伙伴。

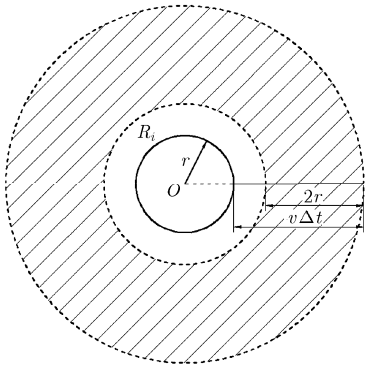

图 3 Fclique 的概率密度分布预测

体现到算法上，仅须将定义 3 中的 v 和 w 有边的条件除式(4)外再增加一条：双方互在对方的均匀分布环形区内。若所有匿名伙伴的位置概率密度均相等，则其熵匿名度就等于匿名集内节点的数目。这样，对 Fclique 算法来说，定义 4 中的式(6)需要改为

$$\forall v \in U,\ |U| \geq v.k \tag{14}$$

Fclique 在寻找极大完全图的算法与 Mclique 完全一致，而在随后的匿名集查找中无需进行熵匿名度测试，代之以线性的测试。算法如下。

算法 4

(1)令 $s = |U|$；

(2)若 s 等于 0，则查找匿名集失败，算法结束；

(3)对 U 中所有的元素 v，做如下动作：如果 $vk > s$，则将 v 从 U 中删除；

(4)如果 $|U|$ 等于 s 则将 U 做为匿名集返回，且算法结束；否则继续执行指令(1)–指令(4)。

图 4 是匿名算法的时间效率分析图。Mclique

的模糊化处理时间随匿名度的增大而迅速增加，势头甚至超过 CliqueCloak 算法，这归结于其枚举候选集的子集时带来的指数化时间复杂度。Fcliuqe 算法的处理时间则相当稳定，一是因为避免了 Mcliuqe 的熵匿名度测试，二是在查找极大完全图时使用了累进式算法，从而速度大大优于 CliqueCloak 算法。

图 5 是加入了 Fcliuqe 算法的受 MPA 攻击效果图。Fclique 的被识别率稍高于 Mclique，但与 CliqueCloak 算法比起来，Fclique 随匿名度增加而识别率递减的趋势非常明显。且 Fclique 的被识别率比理论识别率最大也仅高出 1%–2%，是一个可以接受的结果。

匿名成功率是在最迟响应时间之前匿名成功的查询数和总查询数的比值，是反应匿名算法性能的重要指标。图 6 对 3 种算法的匿名成功率进行了比较。Mcliuqe 比 CliuqeCloak 有 1%–4%的下降，这是因为相比 CliuqeCloak，Mcliuqe 的匿名条件更加苛刻，且这是为增加匿名效果而无法避免的。Fclique 比 Mcliuqe 的成功率又有 1%–3%的下降，这是因为 Fcliuqe 算法只在均匀分布环形区内找匿名伙伴，其匿名条件又苛刻了一层。然而，从数据上看，Fcliuqe 的成功率也不过比最高的 Cliuqe Cloak 下降了 1%–7%，这并非一个不可接受的范围。

6 结论

本文对 LBS 中的传统隐私保护算法提出了一种基于移动模式的攻击算法，实验证明此攻击算法将对 LBS 的隐私产生巨大威胁。通过将用户上一次的模糊化区域与用户移动模式的预测加入匿名算法中，提出了移动环境下的 LBS 匿名算法 Mclique，以及其快速版本 Fclique。模拟实验证明了它们有效减低了攻击者对用户位置的识别率，加强了对用户隐私的保护。而且 Fclique 通过略微牺牲匿名成功率和攻击者识别率，却在时间性能上有显著的改进，是一个切实可行的算法。

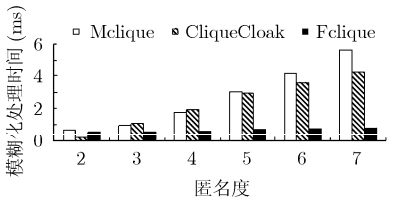

图 4 匿名算法的时间效率

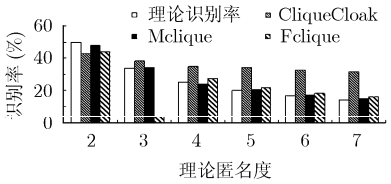

图 5 Fclique 防 MPA 攻击的能力

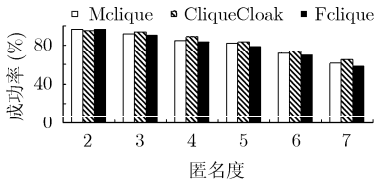

图 6 匿名成功率比较

参 考 文 献

[1] Mokbel M F, Chow C Y, and Aref W G. The new Casper: query processing for location services without compromising privacy[C]. Proceedings of the International Conference on Very Large Data Bases. New York, 2006: 763–774.

[2] Sweeney L. K-anonymity: a model for protecting privacy[J]. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 2002, 10(5): 557–570.

[3] Gruteser M and Grunwald D. Anonymous usage of location based services through spatial and temporal cloaking[C]. ACM/USENIX MobiSys, New York, 2003: 69–78.

[4] Gedik B and Liu L. Location privacy in mobile systems: a personalized anonymization model[C]. Proceedings of International Conference on Distributed Computing Systems, Columbus, 2005: 620–629.

[5] Hu H and Xu J. Non-exposure location anonymity[C]. Proceedings of the IEEE International Conference on Data Engineering, Shanghai, 2009: 1120–1131.

[6] 林欣, 李善平, 杨朝晖. LBS 中连续查询攻击算法及匿名性度量[J]. 软件学报, 2009, 20(4): 1058–1068.
Lin X, Li S P, and Yang Z H. Attacking algorithms against continous queries in LBS and anonymity measurement[J]. *Journal of Software*, 2009, 20(4): 1058–1068.

[7] 潘晓, 郝兴, 孟小峰. 基于位置服务中的连续查询隐私保护研究[J]. 计算机研究与发展, 2010, 47(1): 121–129.
Pan X, Hao X, and Meng X F. Privacy preserving towards continuous query in location-based services[J]. *Journal of Computer Research and Development*, 2010, 47(1): 121–129.

[8] Xu J, Tang X, Hu H, and Du J. Privacy-conscious location-based queries in mobile environments[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2010, 21(3): 313–326.

[9] Pan X, Xu J, and Meng X. Protecting location privacy against location-dependent attack in mobile services[C]. Proceeding of the ACM Conference on Information and Knowledge Management, Napa Valley, 2008: 1475–1476.

[10] Brinkhoff T. A framework for generating network-based moving objects[J]. *Geoinformatica*, 2002, 6(2): 153–180.

[11] Speed prediction for two lane rural highways. Washington, DC, Federal Highway Administration, 2000.

彭志宇： 男，1982 年生，博士生，研究方向为信任管理、隐私保护.

李善平： 男，1963 年生，教授，博士生导师，研究方向为分布式计算、信息集成技术.