doi:10.11959/j.issn.1000-436x.2015320

高效的轨迹隐私保护方案

李凤华[1]，张翠[1]，牛犇[1]，李晖[2]，华佳烽[2]，史国振[3]

(1. 中国科学院 信息工程研究所 信息安全国家重点实验室，北京 100195;
2. 西安电子科技大学 综合业务网理论及关键技术国家重点实验室，陕西 西安 710071; 3. 北京电子科技学院 信息安全系，北京 100070)

摘 要：作为基于位置服务中的一种重要信息，智能终端用户的轨迹隐私保护问题日益受到广大研究者的重视。为解决这一问题，综合考虑了用户所处区域的背景信息、用户行动模式和轨迹相似性等特征，构建了(k−1)条难以被拥有背景信息的敌手所区分的虚假轨迹，从而为移动用户提供 k-匿名级别的轨迹隐私保护。相对于现有技术，该方案不依赖于任何可信第三方，能够在保证虚假轨迹与真实轨迹相似性的基础上有效抵御拥有背景信息的敌手的攻击。实验结果表明了方案的有效性和高效性。

关键词：无线网络安全；基于位置服务；位置隐私；轨迹隐私

中图分类号：TN 929.5 **文献标识码**：A

Efficient scheme for user's trajectory privacy

LI Feng-hua[1], ZHANG Cui[1], NIU Ben[1], LI Hui[2], HUA Jia-feng[2], SHI Guo-zhen[3]

(1. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100195, China;
2. State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China;
3. Department of Information Security, Beijing Electronic Science and Technology Institute, Beijing 100070, China)

Abstract: As one of the most important information in location-based services (LBS), the trajectory privacy for smart devices users has gained increasingly popularity over recent years. To address this problem, (k−1) dummy trajectories to achieve trajectory k-anonymity with comprehensively considering side information were generated, user's mobility pattern and the trajectory similarity etc. Without relying on the trusted third party, the scheme could provide trajectory k-anonymity against adversaries with side information by generating (k−1) realistic dummy trajectories. The evaluation results indicate its effectiveness and efficiency.

Key words: wireless network security; location-based services; location privacy; trajectory privacy

1 引言

随着智能移动终端的迅猛发展和移动互联网的不断成熟，基于位置服务(LBS, location-based service)已经得到广泛应用。用户可以在应用商店里下载并安装各种基于位置服务的应用软件，利用各种网络定位技术或者由自身移动设备的 GPS 模块获取位置及其他相关信息，向提供基于位置服务的服务器(LBS 服务器)发送基于该位置信息的查询，以获得周边的餐饮服务(大众点评)、天气情况、导航信息和交通拥堵状况(百度地图)等，从而享受基于位置服务带来的便利。

然而，用户通过连续不断的提交基于位置的服务请求以享受基于位置服务带来便利的同时，个人隐私也可能已经暴露给了不可信的第三方，例如为用户提供服务的 LBS 服务器。LBS 服务器收集用户所提交的位置信息，并以此为基础采用数据挖掘[1]等技术获取涉及用户位置隐私的相关信息。例如，

收稿日期：2015-05-30；**修回日期**：2015-09-09

基金项目：国家自然科学基金—广东联合基金资助项目（U1401251）；国家高技术研究发展计划（“863”计划）基金资助项目（2012AA013102）；教育部重点基金资助项目（209156）

Foundation Items: The National Natural Science Foundation of China—Guangdong Union Foundation (U1401251); The National High Technology Research and Development Program of China (863 Program) (2012AA013102); The Key Program of Scientific and Technology Research of Ministry of Education (209156)

提供导航服务的服务器通过分析用户的导航轨迹，可以推断出用户的健康状况(例如该用户经常出入于医院)，社会地位(例如该用户经常出入于高级社交场所)等。但是这种方法不适应于数据稀疏的场合。文献[2]总结了基于位置服务中的各种安全及隐私泄漏问题，指出了来自 LBS 服务器的威胁，尤其当该服务器不可信或被恶意第三方收买、控制以后，可能会将对用户相关隐私信息的长期监测和分析结果泄漏给一些恶意机构以谋取更多的利益，比如将用户信息贩卖给恶意广告商等，从而严重威胁用户财产甚至生命安全。因此，对用户轨迹隐私的保护刻不容缓。

目前，除了传统基于策略和基于密码学工具的解决方案外，考虑到移动终端的资源受限性，较多研究方案致力于避免传统解决方案中存在的算法性能等问题，设计轻量级的解决方案。其中，最为广泛采用的是 *k*-匿名[3~5]技术，其基本思路是通过生成（*k*−1）条虚假信息，并将其与用户真实位置或轨迹一起发送给 LBS 服务器，以达到匿名真实位置或轨迹的目的。该技术一般通过匿名服务器(anonymizer)[6~9]来收集这（*k*−1）条虚假信息。虽然这类方法可以降低移动终端的计算负担，但是该方法要求每个用户将自己的真实服务请求发送给匿名服务器，所以匿名服务器的性能将影响该类隐私保护方法的效果，一旦该匿名服务器出现任何状况，将导致隐私保护系统崩溃。基于移动客户端的解决方案[10~12]可以有效避免此类问题，通过近距离通信技术，用户相互协作以实现隐私保护。但是此类方案要求用户长时间打开能量消耗相对较大的设备或接口（例如，WiFi 或者 Bluetooth 等），从而引起资源的大幅消耗和用户参与度的降低。为了解决上述问题，研究者提出了一种独立式结构[13, 14]，用户可以根据自身的隐私需求由移动终端独立完成位置隐私保护。通过在本地生成多条虚假信息，并将用户真实信息隐匿于其中，从而实现不依赖可信第三方的隐私保护。此类方案大都忽略了敌手可能掌握的背景信息，导致用户的真实信息会以较高概率被主动攻击者从虚假信息中猜测出。例如，一些随机生成的虚假位置会落在一些不经常发出查询的位置，如河面上、大洋中、险峰顶等。显然，敌手通过地图信息或者特定区域的服务请求发送概率等可以非常容易地排除这些位置，从而影响隐私保护效果。

本文在充分考虑背景信息、用户行动模式和轨迹相似性等特征可能被敌手获取的前提下，设计了一种不依赖于任何可信第三方的、高效的、基于虚假轨迹的隐私保护算法。本方案以用户真实轨迹为基础，通过轨迹旋转保证用户行动轨迹的相似性，并通过基于背景信息的虚假轨迹调整策略来确保生成的虚假轨迹能够应对各种来自拥有背景信息敌手的攻击。最终实现对真实用户轨迹 *k*-匿名级别的隐私保护目标。

2　相关工作

2.1　轨迹隐私保护技术

针对智能移动终端的轨迹隐私保护问题，学术界涌现了大量的解决方案，大致可以划分为两类：基于可信第三方服务器的方法和基于移动终端的分布式方法。这些方案主要通过空间隐匿(spatial cloaking)、混合区域(mix-zone)、路径混淆(path confusion)和虚假轨迹(dummy trajectories)等技术手段实现。

在基于可信第三方服务器的一类方案中，空间隐匿[5]是一种常用的隐私保护方法。众多用户将自身相关信息暴露给匿名服务器(anonymizer)，匿名服务器获取到用户的位置信息、运动方向和速度，构造一个包括至少 *k* 个用户的最小空间隐匿区域作为服务请求内容发送给 LBS 服务器，从而实现轨迹 *k*-匿名来保护用户的轨迹隐私。然而，此类方案均难以避免安全和性能上的瓶颈，容易导致单点失效攻击。随后，Hwang 等[8]在现有空间隐匿方案的基础上，提出了 *r*-匿名时间模糊算法。然而该方案同样依赖于可信第三方服务器，同时在匿名过程中存在其他相关问题，例如，若用户忽略了敌手可以获取到的背景信息，敌手便可利用背景信息区分出空间隐匿区域中不合理的区域，缩小猜测范围，并以较大猜测概率猜中用户的真实轨迹，对用户隐私造成严重威胁。为此，Xu 等[7]提出了一种使用用户历史轨迹来充当虚假轨迹实现轨迹 *k*-匿名的方法。该方法确保参与 *k*-匿名的虚假轨迹真实且有效，从而增加敌手猜测出用户真实轨迹的难度。然而，由于未能充分考虑背景信息对用户隐私保护策略的影响，如用户在历史轨迹上每一个位置发送请求的概率问题，使敌手仍会以较大概率猜测出用户的真实轨迹。基于混合区域的方案[9, 15]往往通过周期性变换假名来保护用户轨迹隐私。然而，此类方案中所

采用到的混合路由器(mix-router)同样可被视为一个可信第三方，单点失效问题仍然存在。最后，在路径混淆方案中，Hoh 等[16]采用路径混淆方法，利用预期的距离误差去量化敌手猜测出用户真实轨迹的可能性。但是该方案的安全级别取决于用户轨迹的一些基本特性，一旦敌手能够预先获取地图信息或者通过长期观测获得用户常见的运动模式等信息，那么用户的轨迹隐私问题将面临着巨大的威胁。为了避免依赖可信第三方，Dong 等[14]提出了基于点对点的 Jointcache 路径混淆方法，该方法通过混淆附近用户的路径，达到保护用户轨迹隐私的目的。在基于虚假轨迹技术的研究方面，Wu 等[17]提出了一种基于移动终端生成虚假轨迹以实现轨迹 *k*-匿名的方法。该方法通过调整虚假轨迹和真实轨迹之间的角度来产生满足约束条件的虚假轨迹集合，避免虚假轨迹与真实轨迹重合，有效实现轨迹 *k*-匿名。然而，该方法忽略了对用户轨迹隐私的综合考虑，故而对于拥有背景信息的敌手而言，其匿名效果难以得到保障。Niu 等[18]提出了基于背景信息的虚假位置生成算法。该算法充分考虑用户在各个位置发送请求的概率信息，通过信息熵形式化描述背景信息，确保生成的（*k*–1）个虚假位置能够有效迷惑敌手。然而，该算法只适用于 snapshot 场景中的位置隐私问题，无法解决连续查询场景和用户移动轨迹的隐私保护问题。针对如何隐藏连续查询服务属性的问题，Pingley 等[19]提出了 DUMMY-Q 机制，该机制通过构造虚假查询内容和用户运动模型实现对查询内容的 *k*-匿名，但是该方案未考虑攻击者已知的背景信息。Hara 等[20]针对移动车载网络中的轨迹隐私保护问题进行了进一步研究，设计了一种基于车辆移动轨迹的虚假轨迹生成算法，由于该算法综合考虑了车辆移动的轨迹特征，因而在一定程度上降低了虚假轨迹被猜测出的概率。然而，该算法仍未考虑相关背景信息。

综上所述，由于现有轨迹隐私保护方案存在上述若干问题，难以满足移动用户越来越复杂的隐私保护需求。本文以此为目标，旨在设计一种基于虚假轨迹的高效轨迹隐私保护方案。

2.2 轨迹隐私的衡量标准

在移动社交网络尤其是基于位置服务的隐私保护问题中，一种普遍使用的衡量标准是采用 *k*-匿名技术。文献[9]方案中以 *k*-匿名为基础，基于连接 2 个特定用户的假名设计隐私保护方案。作为 *k*-匿名标准的扩展，文献[21]方案通过计算信息熵来衡量用户 *k*-匿名的隐私保护效果。基于失真的衡量标准同样也被广泛用于位置隐私的衡量中，文献[22]方案便采用了这种衡量标准。为了更好地量化用户的轨迹隐私，最重要的步骤之一是需要找出敌手是如何准确地推断出用户的真实信息。轨迹 *k*-匿名[5,13]作为轨迹隐私中最为广泛使用的衡量标准，其目的在于提高轨迹的不确定性，通过将用户的真实轨迹藏匿于其他(k−1)条虚假轨迹或者历史轨迹中以实现轨迹 *k*-匿名。然而此类方案大都忽略了背景信息有可能被敌手所掌握这一现状，例如道路的密度，用户从某一特定位置或特定区域发送服务请求的概率等。上述潜在的背景信息一旦被敌手利用，将会对用户的轨迹隐私产生很严重的威胁。这就要求在设计轨迹隐私的衡量标准时需要充分考虑背景信息的存在。

3 预备知识

3.1 基本概念

3.1.1 背景信息(BI, background information)

本文的背景信息是指用户在某一具体位置或者特定区域发送 LBS 服务请求的概率[18]。

例如，在一个划分为 $n\times n$ 个网格的地图中，每个网格的服务请求概率可以表示为

$$q_i=\frac{\text{网络}i\text{中曾经发送过的请求数}}{\text{所有网络中发送过的请求数}}$$

其中，$i=1,2,\cdots,n^2$，且满足 $\sum_{i=0}^{n^2}q_i=1$。

3.1.2 轨迹相似性 (trajectory similarity)

考虑到用户的行为模式和移动轨迹的规律性，本文拟采用轨迹相似性函数度量虚假轨迹和用户真实轨迹的相似程度，从而衡量所生成虚假轨迹的区分困难度。

如图 1 所示，假设用户的真实轨迹为

$$U=\{(x_0,y_0),(x_1,y_1),\cdots,(x_m,y_m)\}$$

其中，点 (x_i,y_i) 表示用户在第 i 个时间点所处位置，m 表示轨迹 T 上时间点的个数。记用户在第 i 个点 $(1\leqslant i\leqslant m)$ 相对于初始点的行动方向变化为 θ_i，则 $\tan\theta_i=\frac{y_i-y_0}{x_i-x_0}$，所以 $\theta_i=\arctan\frac{y_i-y_0}{x_i-x_0}$，轨迹 T 可表示为 $U=\left\{(x_0,y_0),\langle\theta_1,\theta_2,\cdots,\theta_m\rangle\right\}$。

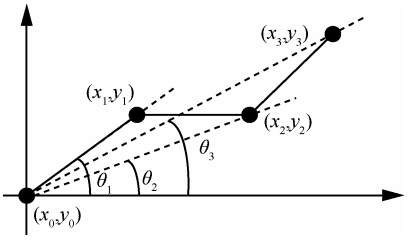

图 1　轨迹相似性示意

同理，终端产生的(k−1)条潜在虚假轨迹

$$T_j=\left\{(x_0^j,y_0^j),(x_1^j,y_1^j),\cdots,(x_m^j,y_m^j)\right\}$$

可表示为

$$T_j=\left\{(x_0^j,y_0^j),\left\langle\theta_1^j,\theta_2^j,\cdots,\theta_m^j\right\rangle\right\}$$

其中，$\theta_i^j=\arctan\dfrac{y_i^j-y_0^j}{x_i^j-x_0^j}(1\leqslant i\leqslant m,1\leqslant j\leqslant k-1)$。

那么，轨迹相似性为

$$\sigma^2=\frac{\sum_{j=1}^{k}E_m\left[\left(\dfrac{\theta_i^j-\theta_i}{2\pi}\right)\right]}{k}=\frac{\sum_{j=1}^{k}\sum_{i=1}^{m}\left(\dfrac{\theta_i^j-\theta_i}{2\pi}\right)^2}{km}$$

显然，$\sigma^2\in[0,1]$ 内，σ^2 越小，对应的虚假轨迹的轮廓与用户的真实轨迹轮廓越接近，从而越难被区分。

3.1.3　轨迹泄漏概率(trajectory exposing probability)

轨迹泄漏概率是通过计算用户在每一个静态快照(snapshot)上发送请求的位置与对应虚假轨迹上位置的交叉程度来衡量。

假设用户的真实轨迹 U 被分为 m 个 snapshot，记第 i 个 snapshot $(1\leqslant i\leqslant m)$ 中的假位置 (x_i^j,y_i^j) $(1\leqslant j\leqslant k-1)$ 和真实位置 (x_i,y_i) 所属网格的服务请求概率分别为 q_i^j 和 q_i。为了达到轨迹 k-匿名，在每个 snapshot 中，最优的情况是 k 个位置相互独立且无相交，即对于任意$1\leqslant j\leqslant k-1$, 有$(x_i^j,y_i^j)\neq(x_i,y_i)$，与这些位置点所处的服务请求概率与真实用户所处位置的服务请求概率非常相似，否则将会被敌手依据背景信息而排除，即有效假位置需满足 $\left|q_i^j-q_i\right|<\Delta$，$\Delta$ 为极接近 0 的正数，可由用户自定义，例如 $\Delta=0.05$。

由以上分析可知，对某个 snapshot 而言，有效的虚假位置数目越多，该 snapshot 的真实位置隐藏效果越好。极端地，当某个 snapshot 中 k 个位置完全相交于一点时，真实位置将被泄漏。通常而言，这样的情况在 m 个 snapshot 中出现的次数越多，用户的轨迹隐私保护程度越差。基于此，用户在每个 snapshot 中泄漏概率可以表示为

$$p_i=\frac{1}{snapshot_i\text{中所观测到的有效位置数}}$$

因此，轨迹泄漏概率可以表示为

$$Pr_i=\frac{\sum_{i=1}^{m}p_i}{m}$$

一般来说，轨迹泄漏概率越大，用户的隐私保护程度越差。

3.2　攻击模型

本文依据敌手的不同攻击能力，将其划分为被动攻击者和主动攻击者。

被动攻击者。此类攻击者可以通过监听无线信道，从而获取用户与 LBS 服务器之间(或者移动用户之间)的交互信息，发动窃听攻击。另外，被动攻击者可以与其他用户一起发动合谋攻击，威胁合法用户的隐私安全。

主动攻击者。此类攻击者的能力相对较强，可以俘获其他用户，甚至可以获取 LBS 服务器上保存的相关移动用户的所有信息，以此为基础发动推理攻击。本文中，考虑到 LBS 服务器拥有用户的所有信息，包括任意用户的当前和历史请求，故而直接假定其为主动攻击者。值得注意的是另一个常见假设为主动攻击者可以获取用户所采用的轨迹隐私保护策略。

3.3　研究动机

在基于轨迹 k-匿名的用户轨迹隐私保护方案中，通过生成虚假轨迹的策略不失为一种有效的方法。然而，现有基于虚假轨迹的解决方案大都通过随机生成的虚假轨迹达到轨迹 k-匿名，而忽略了上文提及的背景信息以及用户的行为模式等重要因素，从而导致用户隐私的泄漏。以图 2 为例，通过对一个现有方案[13]的解决策略进行详细的分析观测，以下列出了 3 条重要的观测结果。

1) 现有轨迹隐私保护方案大都基于可信第三方，从而使此类可信第三方成为整个系统的安全、隐私和性能瓶颈，影响用户隐私安全及服务质量。

2) 地图的背景信息在现有隐私保护方案中已经得到重视[18]，并以此为基础设计了相应的解决方案。而在轨迹隐私保护中，由于未能充分考虑此类信息的泄漏对于用户隐私的影响，现有方案大都采用随机生成的虚假轨迹，而未能考虑虚假轨迹中各个时间点的位置的有效性，最终导致用户的轨迹隐私难以得到保护。

3) 为了降低过于分散的虚假轨迹带来的系统开销，图 2 提及的方案期望参与轨迹 *k*-匿名的轨迹之间尽可能多的相交，产生尽可能多的交点以起到迷惑敌手的作用。如图 2 所示，虚假轨迹 T_1 由 5 个三角形组成，箭头方面代表用户移动方面，同理，虚假轨迹 T_2 由 5 个正方形组成，用户转迹 U_1 由 5 个实心圆形组成，3 条轨迹在坐标(2, 3)和(3, 3)处有 2 个交点，然而，轨迹之间的交点会引起虚假轨迹的泄漏概率增加，从而降低轨迹隐私的保护效果。一种极端的情况是 *k* 条轨迹完全重合，则用户的真实轨迹则 100%的泄漏于敌手面前。

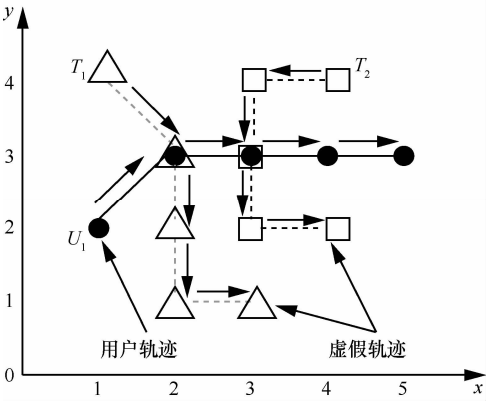

图 2　一条用户轨迹 U_1 和 2 条虚假轨迹 T_1 和 T_2

基于以上观测，敌手通过对用户进行短期或者长期观测，很容易采用数据挖掘[1]和机器学习等技术分析出用户的行动轨迹，从而以超出用户期望值的概率猜测出用户的真实轨迹。

在全面考虑上述 3 条基本观测的基础上，本文旨在设计能够同时兼顾背景信息，用户运动模式及虚假位置有效性的轨迹隐私保护方案。

3.4　基本思路

本方案主要包括 2 部分，真实轨迹旋转（如图 3 所示）和虚假轨迹调整（如图 4 所示）。以下以实现轨迹 *k*-匿名为例描述本方案的基本思路，其中，*k*=3。

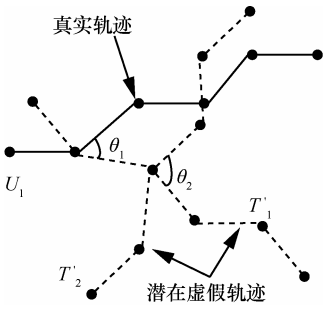

图 3　真实轨迹旋转

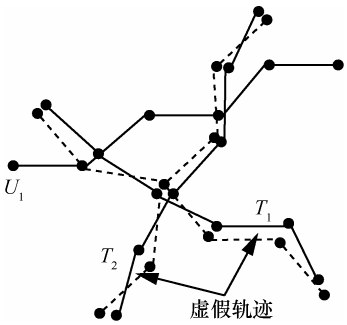

图 4　虚假轨迹调整

如图 3 所示，考虑到用户行为模式，本方案首先在用户的真实轨迹 U_1 上随机选择一个参考点，通过将用户的真实轨迹旋转角度 θ_1 和 θ_2，从而依次生成其他 2 条潜在虚假轨迹 T_1' 及 T_2'。由于生成的 2 条潜在虚假轨迹是用户真实轨迹旋转的产物，从而有效地保证了轨迹之间的相似性。

更进一步，考虑到背景信息对用户轨迹隐私保护的影响，图 4 给出了虚假轨迹的调整策略。在轨迹旋转的过程中，通过将选定旋转点进行基于背景信息的偏移，从而尽可能地降低用户在各条轨迹之间交点上重叠的可能性。最后，通过将图 3 中潜在虚假轨迹上各个位置点进行基于背景信息相似程度的偏移，从而获取参与匿名的虚假轨迹(如图 4 中的 T_1 及 T_2)，并使最终生成的虚假轨迹的泄漏概率最小，提高用户的轨迹隐私。虚假轨迹生成算法的细节将在下一节给出。

4　基于虚假轨迹的隐私保护方案

4.1　系统架构

本文提出一种基于虚假轨迹的隐私保护方案，其系统架构如图 5 所示。以用户通过移动网络访问 LBS 服务器并获取相关导航服务为例对系统中各个实体进行解释。

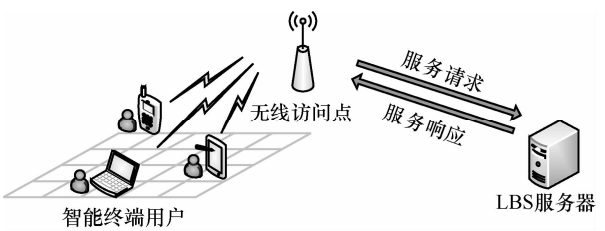

图 5　系统架构示意

智能终端用户。本系统中，智能终端用户可以为任何带有 GPS 模块并具有网络访问功能的智能设备，例如智能手机、平板电脑或相关车载设备等。

LBS 服务器。该实体为用户提供相关服务数据，主要包括基于用户位置的相关服务信息，比如导航数据等。常见的 LBS 服务器多种多样，广为人知的包括 GPS、中国的北斗导航系统，百度地图以及大众点评网等社交服务及应用。

移动网络。该网络为任意可以基于移动设备访问的互联网，例如 3G/4G 网络、无线局域网等。

4.2　虚假轨迹生成算法

本节以图 6 为例，详细描述虚假轨迹的生成过程。图中灰色格子代表地图上某些可能发送 LBS 请求的区域，比如道路等。空白区域代表一些理论上不能发送出 LBS 请求的区域，例如湖泊、河流或移动设备到达可能性比较小的地方。在本节的算法描述过程中，首先假设背景信息已经为各个用户所知，在 4.4 节将给出背景信息的获取方法。作为算法的初始化，用户首先根据当前位置和目的地确定出其真实的行动轨迹，然后将该轨迹根据用户的移动速度等信息划分为若干个时间点，并确定出用户在每个时间点的位置。时间点的划分问题亦可以通过设置系统的时间间隔来确定，例如，GPS 系统每隔 1 s 或者 2 s 发送一次请求并获取相关服务信息。基于以上信息，本方案执行以下几步以生成虚假轨迹，实现轨迹 *k*-匿名。

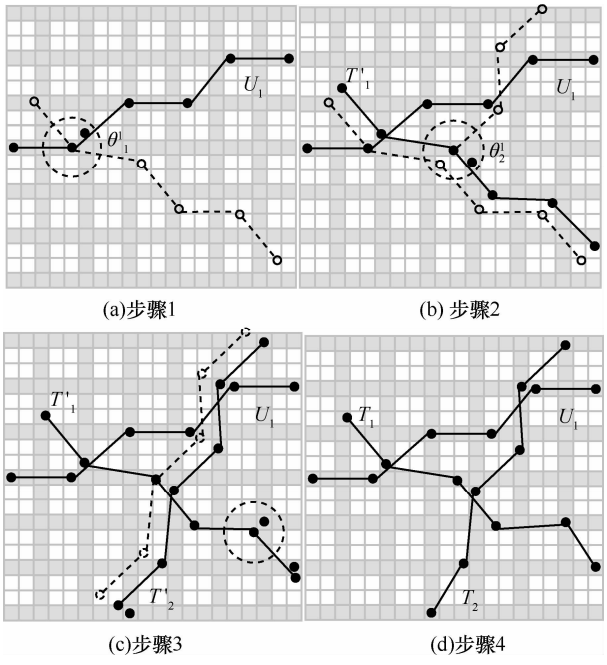

图 6　虚假轨迹的生成和调整算法

步骤 1　首先在用户的真实轨迹上随机选取一个时间点所对应的位置，然后将该真实轨迹进行随机选择角度(理论上讲，该角度的取值范围为 $0 \leqslant \theta_1^1 \leqslant \pi$，然而旋转角度过小会导致生成的轨迹相距很近，引起隐私问题；旋转角度过大会导致生成的轨迹占据地图面积过大，带来额外的系统开销)的旋转以获取一条旋转后的虚拟轨迹，如图 6(a)中虚线曲线所示。同时，以所选择的旋转点为基准，将其在一个虚拟圆范围内进行随机偏移，目的在于降低虚假轨迹在不同时间点上处于同一位置的概率，减少交叉点的数目。其中，虚拟圆的半径由用户的平均移动速度和真实轨迹上 2 个位置点间的时间间隔决定。在本文的实验部分，该半径为用户平均速度与两位置点间的时间间隔的乘积的一半。进而将该旋转后的虚拟轨迹平移至图 6(b)中的带有灰度的实线处以获取一条潜在的虚拟轨迹 T_1'。

步骤 2　如图 6(b)所示，从已选定的轨迹(包括真实轨迹 U_1 和已选定的潜在虚拟轨迹 T_1')上的各个时间点所对应的位置集合中随机选取一个位置，重复执行步骤 1，将所选定轨迹进行随机选定角度 (θ_2^1) 的旋转，并通过虚拟圆对参考点进行偏移和对旋转后的虚拟轨迹平移，以获取第二条潜在虚拟轨迹 T_2'。

步骤 3　如图 6(c)所示，重复执行步骤 1 和步骤 2，直至选够剩余(k−1)条虚拟轨迹。

步骤 4　在图 6(d)中，结合地图的背景信息以及用户真实轨迹上各个时间点所对应的位置的服务请求概率，对上述所生成的(k−1)条潜在虚拟轨迹对应的各个时间点上的值进行遍历。如果其服务请求概率与真实轨迹上对应点的概率相差较大(两概率值相减大于某一个由用户自定义的门限值，例如 0.05)，则将潜在虚拟轨迹上的该位置进行基于背景信息的偏移，其目的在于将该位置偏移至附近的一个最接近真实位置上的服务请求概率的位置，进而保证新的位置能够更好地保护用户隐私，抵御各种基于背景信息的攻击。与此相反，如果该点对应的服务请求概率与真实轨迹上的概率值相似，则算法继续运行直至结束。

4.3　算法安全性讨论

由于常见的密码学技术，比如公钥密码基础(PKI, public key infrastructure)体系，这些技术应用到本方案通信通道之上，通过对各个实体间所传输的内容进行加密操作，便可以很容易地保证通信信道的安全，避免窃听等被动攻击的发生。因此本方案忽略上述通信信道安全问题，旨在应对一些来自

被动攻击者和主动攻击者所发动的合谋攻击和推理攻击等。

4.3.1 抵御合谋攻击

合谋攻击往往发生在一组合法用户之间或者合法用户与后台服务器(LBS 服务器)之间。它们通过相互之间的信息共享，从而以更大概率计算或者猜测出其他合法用户的隐私信息。通过轨迹 k-匿名，用户的真实轨迹 U 被隐匿于另外(k−1)条虚拟轨迹 $T_j(1\leqslant j\leqslant k-1)$ 之中，从而将敌手能成功猜测出真实轨迹的概率降低到 $\frac{1}{k}$。从用户所拥有的信息角度而言，由于本方案无需通过用户之间的信息交互来实现轨迹 k-匿名，故而通过用户之间的合谋来实现对合法用户隐私信息猜测的概率不会随着参与合谋用户的增多而提升。

从敌手的角度而言，最好的情况是能够获取所有用户和 LBS 服务器的信息，从一个被动攻击者将升级为一个主动攻击者，可以发动推理攻击。

4.3.2 抵御推理攻击

假设 LBS 服务器为主动攻击者，它可以通过监视用户所发送过的所有 LBS 服务请求，包括当前的请求以及历史请求等，并根据这些信息推理论断合法用户的其他隐私信息。

定理 1 设 $negl(k)$ 表示关于 k 的可忽略函数，若 3.1.3 节中 $\Delta\leqslant negl(k)$，本方案能抵御推理攻击。

证明 记参与轨迹 k-匿名的轨迹集合为 $C=\{T_1,T_2,\cdots,T_k\}$，其中，$1\leqslant j\leqslant k$，$T_j=\{(x_0^j,y_0^j),(x_1^j,y_1^j),\cdots,(x_m^j,y_m^j)\}$。

由步骤 1)和步骤 4)可知，对于任意$1\leqslant i\leqslant m$ 和 $1\leqslant j\leqslant k$，参与轨迹 k-匿名的轨迹在第 i 个 snapshot 上有 $\left|q_i^j-q_i\right|<\Delta$，而 $\Delta\leqslant negl(k)$，所以对于任意 $1\leqslant j\neq s\leqslant k$，有

$$\begin{aligned}&\left|Pr\left((x_i^j,y_i^j)\in U\right)-Pr\left((x_i^s,y_i^s)\in U\right)\right|=\\&\left|q_i^j,q_i^s\right|=\left|(q_i^j-q_i)-(q_i^s-q_i)\right|\\&<2\Delta\leqslant negl(k)\end{aligned}$$

即对于任意$1\leqslant i\leqslant m$，第 i 个 snapshot 上的位置点是不可区分的，又因为本方案中对于整个轨迹集合 C 而言，每个 snapshot 相互独立，所以参与轨迹 k-匿名的 k 条轨迹是不可区分的，即本方案能抵御推理攻击。

4.4 算法实现

虚拟轨迹能否有效生成需要以用户获取背景信息为基础，本节给出 2 种有效的方案以实现背景信息的获取。一种最简单有效的方案是由一些广为人知的 LBS 服务器来负责发布地图上的背景信息，这样移动终端用户可以通过访问这些知名站点来获取背景信息，例如 Google Latitude、百度地图等。另一种方法是采用先前提出的基于分布式无线访问点(WiFi access points)的策略[18]，通过用户与公共基础设施的交互来共同维护一个为广大用户所共享的背景信息库。最后，用户从无线访问点处获取背景信息。

4.5 时间复杂度分析

该方案包括虚假轨迹生成和调整算法。虚假轨迹生成算法包括一个递归过程，每次递归包含位置点的随机选取、轨迹旋转和轨迹平移 3 次运算。而虚假轨迹生成算法的递归过程包含(k−1)次递归运算，因此递归过程的时间复杂度为 $O(3k-3)$。虚假轨迹调整算法中(k−1)条轨迹遍历算法的时间复杂度为 $O(k-1)$，因此算法的时间复杂度为 $O(4k-4)$。

5 实验及仿真

5.1 基本设置

为了验证本方案的有效性和高效性，本文以 Borlange data set (该数据集由瑞士联邦理工学院研究小组收集，是 Borlange 城市交通拥堵研究的重要组成部分，并已成功验证了多项相关工作[23~25]。该数据集在 Borlange 城市范围内，收集了 200 辆配有 GPS 设备车辆的行驶路线，用来帮助司机合理选择出行线路，避免交通拥堵等相关问题。详情请参阅瑞士联邦理工学院 Frejinger 博士的毕业论文[26]。)为基础，随机选择位于该城市中心区域 8 000 m×8 000 m 范围内的 5 000 条行驶路线进行实验。由于 Borlange data set 对于用户轨迹信息的记录比较全面，适用于本方案的验证工作，故本文选用此数据集进行实验。实验环境为 Lenovo Thinkpad T430，i7 CPU，8 GB 内存，操作系统为 Windows 8 64 bit 专业版。

下述实验涉及到的参数包括：1) k 为匿名等级，与轨迹 k-匿名息息相关，通常设置为 3~30；2) m 为轨迹上时间点的个数，默认为 m=10；3) 产生的虚假轨迹点的背景信息与真实轨迹背景信息的差值门限值，比如 0.05。实验中轨迹旋转角度的取值范

围为 $0 \leqslant \theta \leqslant \frac{\pi}{4}$。与文中参与比较的方案包括：基于随机选择的方案(Random)[3]、基于轨迹旋转的方案(Rotation)[13]、基于用户停—走模型的解决方案(Pauses)[27]、以及理论上的最优方案(Optimal)。为了保证实验结果的准确性，以下所有的实验结果均为 1 000 次运行的平均值。

5.2　实验结果

5.2.1　k 对轨迹相似度的影响

轨迹相似度在一定程度上体现了轨迹 k-匿名的隐私保护效果。图 7 验证并给出了 k 的变化对轨迹相似性的影响。不难看出，文献[3]方案的效果最差，原因在于该方案不考虑用户的行动模式，只是随机的生成虚假轨迹。文献[27]方案的效果次之，原因在于该方案在生成虚假轨迹的时候考虑了用户在每一次停顿后继续行进的方向问题，使得所生成虚假轨迹的相似性相比用户真实的轨迹相差不是很大。由于文献[13]的方案未考虑地图的背景信息，仅仅通过对用户真实轨迹的旋转来产生虚假轨迹，故而其轨迹相似性和理论上的最优值保持一致，然而，忽略背景信息所带来对用户隐私的泄漏问题将在下一实验中指出。在本文提出的方案(Ours)中，由于在生成潜在虚假轨迹的时候充分考虑了用户行动模式，即使在最后对潜在虚假轨迹上的相关位置进行了基于背景信息的调整，但由于合理选择了偏移的最大半径，从而一定程度上保证了轨迹相似度接近理论的最优值。

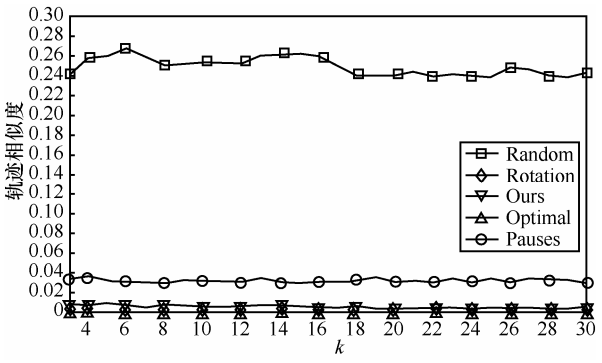

图 7　k 对轨迹相似度的影响

5.2.2　k 对轨迹泄漏概率的影响

轨迹 k-匿名隐私保护效果的另一个优点体现在轨迹泄漏概率上。本节给出了不同方案中，k 的变化对轨迹泄漏概率的影响。如图 8 所示，基于随机选择的文献[3]方案的轨迹泄漏概率最高，原因在于，虽然该方案中每个 snapshot 中的 k 个位置点很难发生重合，但由于其生成过程中忽略了背景信息，从而导致相当一部分位置点会被拥有背景信息的敌手所排除，进而导致轨迹 k-匿名的效果大幅下降，真实用户的轨迹泄漏概率大增。在文献[13]方案中，为了保证轨迹的相似性，虚假轨迹的生成仅仅是通过将真实轨迹进行一定角度的旋转而得到，每个位置上的背景信息难以同时兼顾，从而导致所生产的轨迹难以应对来自拥有背景信息的主动攻击者的推理攻击，从而难以提供较高的轨迹隐私保护效果。文献[27]方案的轨迹隐私泄漏概率相对较低，原因在于其选择每一跳位置的时候考虑到了可达性要求，但由于对服务请求概率因素考虑的不充分，其轨迹隐私保护程度仍然难以达到较高水平。在本方案中，步骤 1 的旋转点偏移过程(如图 6(a)所示)保证了轨迹之间尽可能少地在各个 snapshot 中发生位置重叠，另一方面，步骤 4 旨在将潜在虚假轨迹上的位置点进行基于服务请求概率的偏移，从而进一步的降低了用户轨迹泄漏概率。

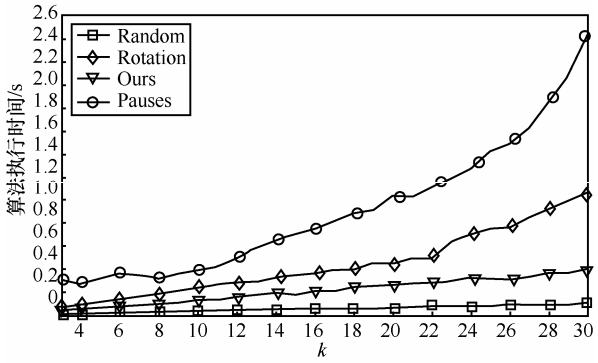

图 8　k 对轨迹泄漏概率的影响

5.2.3　k 对算法执行时间的影响

本节验证了 k 的变化对算法执行时间的影响，实验结果如图 9 所示。其中，文献[3]方案由于未考虑其他任何信息，故而执行效率最高，且近乎于线性，但较快的执行效率是以牺牲用户的轨迹隐私为代价。文献[27]方案由于考虑了较多的因素，包括用户轨迹中存在的停顿过程，目标位置偏移角度和可达性等因素，故而在对虚假轨迹每一跳的位置选择时需要遍历更多种的可能性，从而导致算法执行时间相对较长。文献[13]方案中在生成虚假轨迹的环节期望产生更多的轨迹交点以达到迷惑敌手的目的，故而对虚假轨迹生成算法就有更多的要求，也使其算法的执行时间比较长。在本方案中，宏观

而言，虚假轨迹的生成仅需要通过轨迹旋转和位置偏移 2 个步骤，无需过多地考虑其他因素，故而使得其算法执行效率较高，执行时间相对较短，且亦趋近于线性。

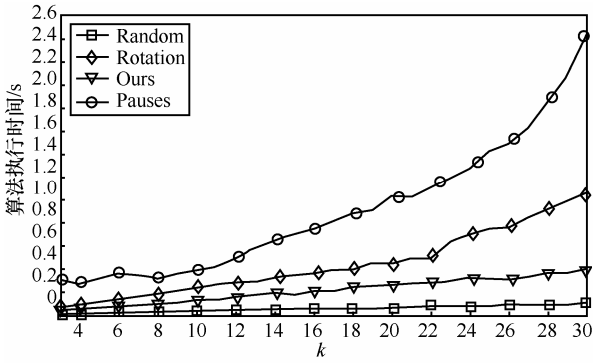

图 9　k 对算法执行时间的影响

图 10 给出了本文所提出的方案在真实智能手机环境下的实验结果。所采用设备的硬件信息为：小米智能手机，内存 1.7 GHz，Cortex-A8 处理器，1 GB RAM，操作系统为 Android 4.1.2。具体的实验结果如图 10 所示。

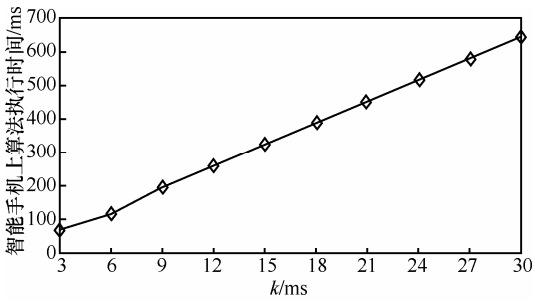

图 10　k 对智能手机上的算法执行时间的影响

5.2.4　旋转角度对占据地图面积的影响

图 11 给出了本方案中轨迹旋转角度与方案执行完时所有轨迹占据地图面积的情况，用以说明方案收敛性。不难看出，一方面，旋转角度越大，会造成最终参与匿名的轨迹集合占据地图面积过大；

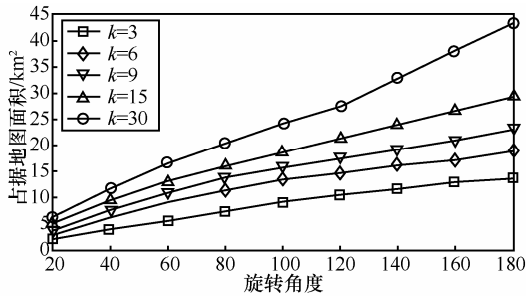

图 11　旋转角度对虚假轨迹占据地图面积(平方千米)的影响

另一方面，k 的取值越大，同样会致使该面积增大，导致整个方案收敛较慢。因此，在选择旋转角度时，建议用户选择较小的轨迹旋转角度，例如 20°或 40°。

6　结束语

本文针对基于位置服务中智能终端用户的轨迹隐私保护问题，提出了一种基于虚假轨迹的高效轨迹隐私保护方案。本方案不依赖于任何可信第三方，在综合考虑用户所处区域的背景信息、用户行动模式和轨迹相似性等特征的基础上，通过轨迹旋转及虚假轨迹调整等方法构建 k−1 条难以被拥有背景信息的敌手所区分的虚假轨迹，实现对用户轨迹 k-匿名级别的隐私保护。安全性分析和实验结果进一步验证了所提出方案的有效性和高效性。

参考文献：

[1] ZHANG H, XU L, HUANG H, et al. Mining spatial association rules from LBS anonymity data set for improving utilization [A]. Proceedings of the 21th IEEE International Conference on Geoinformatics [C]. Kaifeng, China, 2013. 1-6.

[2] WERNKE M, SKVORTSOV P, DURR F, et al. A classification of location privacy attacks and approaches [J], Personal & Ubiquitous Computing, 2014, 18(1):163-175.

[3] HIDETOSHI K, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services [A]. Proceedings of International Conference on Pervasive Services [C]. Santorini, Greece, 2005. 88-97.

[4] JIA J, ZHANG F. Nonexposure accurate location *k*-anonymity algorithm in LBS [J], The Scientific World Journal, 2014, 2014(1): 619357-619357.

[5] PAN X, MENG X F, XU J L. Distortion-based anonymity for continuous queries in location-based mobile services [A]. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems [C]. New York, USA, 2009. 256-265.

[6] GAO S, MA J F, SHI W, et al. TRPF: a trajectory privacy-preserving framework for participatory sensing [J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887.

[7] XU T, CAI Y. Exploring historical location data for anonymity preservation in location-based services [A]. Proceedings of the 27th IEEE International Conference on Computer Communications 2008 [C]. Phoenix, USA, 2008. 1220-1228.

[8] HWANG R H, HSUEH Y L. A novel time-obfuscated algorithm for trajectory privacy protection [J]. IEEE Transactions on Services Computing, 2014, 7(22): 126-139.

[9] PALANISAMY B, LIU L. Attack-resilient mix-zones over road networks: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2015, 14(3): 495-508.

[10] YANG N, CAO Y, LIU Q, et al. A novel personalized TTP-free loca-

tion privacy preserving method [J]. International Journal of Security & Its Applications, 2014, 8(2): 387.

[11] CHOW C Y, MOKBEL M F, LIU X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments [J]. Geoinformatica, 2011, 15(2): 351-380.

[12] NIU B, ZHU X Y, LI W H, et al. A personalized two-tier cloaking scheme for privacy-aware location-based services [A]. Proceedings of International Conference on Computing, Networking and Communications [C]. Garden Grove, USA, 2015. 94-98.

[13] YOU T H, PENG W C, LEE W C. Protecting moving trajectories with dummies [A]. Proceedings of IEEE 14th International Conference on Mobile Data Management[C]. Mannheim, Germany, 2007. 278-282.

[14] DONG K, GU T, TAO X, et al. Jointcache: collaborative path confusion through lightweight P2P communication[A]. Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops[C]. San Diego, USA, 2013. 352-355.

[15] PALANISAMY B, LIU L. Road network mix-zones for anonymous location based services[A]. Proceedings of the 29th International Conference on Data Engineering[C]. Brisbane, Australia, 2013. 1300-1303.

[16] HOH B, GRUSTESER M. Protecting location privacy through path confusion [A]. Security and Privacy for Emerging Areas in Communications Networks [C]. Athens, Greece, 2005. 345-356.

[17] WU X, SUN G. A novel dummy-based mechanism to protect privacy on trajectories [A]. Proceedings of IEEE International Conference on Data Mining series [C]. Shenzhen, China, 2014. 1120-1125.

[18] NIU B, LI Q H, ZHU X Y, et al. Achieving *k*-anonymity in privacy-aware location-based service [A]. Proceedings of the 33rd IEEE International Conference on Computer Communications [C]. Toronto, Canada, 2014. 754-762.

[19] PINGLEY A. Protection of query privacy for continuous location based services [A]. Proceedings of the 30th IEEE International Conference on Computer Communications 2011 [C]. Shanghai, China, 2011. 1710-1718.

[20] HARA T, YAMAMOTO A, ARASEL Y, et al. Location anonymization using real car trace data for location based services [A]. Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication [C]. Siem Reap, Cambodia, 2014.1-8.

[21] SHARMA V, SHEN C. Evaluation of an entropy-based *k*-anonymity model for location based services [A]. Proceedings of International Conference on Computing, Networking and Communications [C]. Garden Grove, USA, 2015. 374-378.

[22] SHOKRI R, THEODORAKOPOULOS G, BOUDEC J Y L, et al. Quantifying location privacy [A]. Proceedings of the 32nd IEEE Symposium on Security & Privacy [C]. Claremont Resort, USA, 2011. 247-262.

[23] FREUDIGER J. When Whereabouts is no Longer Thereabouts: Location Privacy in Wireless Networks [D]. PhD thesis EPFL, 2011.

[24] FREUDIGER J, SHOKRI R, HUBAUX J P. Evaluating the Privacy risk of Location-based Services [M]. Financial Cryptography and Data Security. Springer Berlin Heidelberg, 2012. 31-46.

[25] NIU B, LI Q H, ZHU X Y, et al. Enhancing privacy through caching in location-based services [A]. Proceedings of the 34rd IEEE International Conference on Computer Communications [C]. Hong Kong, China, 2015. 1017-1025.

[26] FREJINGER E. Route Choice Analysis: Data, Models, Algorithms and Applications [D]. PhD thesis EPFL, 2008.

[27] KATO R, IWATA M, HARA T, et al. A dummy-based anonymization method based on user trajectory with pauses[A]. Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems[C]. Redondo Beach, USA, 2012. 249-258.

作者简介：

李凤华（1966-），男，湖北浠水人，博士，中国科学院信息工程研究所副总工程师、研究员、博士生导师，主要研究方向为网络与系统安全、隐私计算、可信计算。

张翠（1985-），女，江西抚州人，中国科学院信息工程研究所博士生，主要研究方向为网络安全、隐私保护。

牛犇[通信作者]（1984-），男，陕西西安人，博士，中国科学院信息工程研究所助理研究员，主要研究方向为网络安全、隐私计算。E-mail: niuben@iie.ac.cn。

李晖（1968-），男，河南灵宝人，博士，西安电子科技大学教授、博士生导师，主要研究方向为密码学、无线网络安全、云计算安全、信息论与编码理论。

华佳烽（1989-），男，湖北黄冈人，西安电子科技大学博士生，主要研究方向为医疗隐私。

史国振（1974-），男，河南济源人，博士，北京电子科技学院副教授，主要研究方向为信息安全。