基于位置服务隐私自关联的隐私保护方案

李维皓，曹进，李晖
（西安电子科技大学网络与信息安全学院，陕西 西安 710071）

摘 要：随着移动智能终端的普遍运用，基于位置服务（LBS）成为了人们生活中必不可少的部分，在提供便捷生活服务的同时，也引发了用户隐私信息泄露的隐患。在考虑背景信息存在的同时，进一步地考量了用户自身和服务提供商短期缓存的查询记录，避免了攻击者利用查询信息的可能性对用户的隐私信息进行猜测并实现推断攻击。基于用户隐私信息自关联的前提下，提出了 2 种隐私保护方案——简易隐私自关联的隐私保护算法（Ba-2PS）和扩展隐私自关联的隐私保护算法（En-2PS），其中 En-2PS 从时间和查询范围 2 个维度扩展了简易隐私自关联的隐私保护算法，提高了从匿名位置单元和匿名查询内容中推测用户真实信息的不确定性。最后，通过隐私性证明和实验结果证明了方案的有效性和安全性。

关键词：位置服务；隐私保护；位置隐私；查询隐私；*k* 匿名
中图分类号：TN929.5
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2019110

Privacy self-correlation privacy-preserving scheme in LBS

LI Weihao, CAO Jin, LI Hui
School of Cyber Engineering, Xidian University, Xi’an 710071, China

Abstract: The prevalence of mobile intelligent terminals gives the location-based service (LBS) more opportunities to enrich mobile users’ lives. However, mobile users enjoy the convenience with the cost of personal privacy. The side information and mobile user’s recent requirement records were considered, which were obtained or stored by the service provider. Based on the existence of recent requirement records, adversary can employ the inference attack to analysis mobile user’s personal information. Therefore, two schemes were proposed, including of basic privacy self-correlation privacy-preserving scheme (Ba-2PS) and enhanced privacy self-correlation privacy-preserving scheme(En-2PS). In En-2PS, the privacy-preserving scheme was designed from two dimensions of aspects of time factor and query region, which increased the uncertainty inferring out the real information. Finally, the privacy analysis was illustrated to proof En-2PS’s privacy degree, then the performance and privacy evaluation results indicate that En-2PS is effective and efficient.

Key words: location service, privacy preservation, location privacy, query privacy, *k*-anonymity

1 引言

随着移动智能终端的普遍应用，社交网络已经成为了人们生活中必不可少的一部分，其中基于位置服务（LBS, location-based service）的使用最为频繁。如今，智能终端可以下载多样化的应用程序，增添多样化的服务和个性化的操作。然而，这种便利也带来了安全隐患。基于位置服务中，用户的位

收稿日期：2018−11−06；**修回日期**：2019−04−06
通信作者：李维皓，vivianlee90xd@gmail.com
基金项目：国家重点研发计划基金资助项目（No.2017YFB0802203）；国家自然科学基金资助项目（No.61732022, No.61672411, No.61772404, No.U1401251）；陕西省自然科学基础研究计划重大基础研究基金资助项目（No.2016ZDJC-04）
Foundation Items: The National Key Research and Development Program of China (No.2017YFB0802203),The National Natural Science Foundation of China (No.61732022, No.61672411, No.61772404, No.U1401251),The Major Basic Research Program of Shaanxi Province Natural Science Foundation Research Project (No.2016ZDJC-04)

置信息被用户视为一项“无关紧要”的信息无偿地发送给了服务提供商来换取相关的服务。然而这些信息一旦被恶意用户或是具有强大计算能力的服务提供商获取，那么用户的个人信息安全就会受到威胁。位置信息能够泄露用户的诸多敏感信息，因此保护用户的位置信息隐私也成为了必不可少的研究课题。

在基于位置服务中，用户的位置信息分为 2 种，即单点的位置信息和连续的位置信息。其中，单点的位置信息为在某个时间点发起的基于位置的请求，例如，寻找附近的商家、查询某个城市的天气等。连续的位置信息则为在某个时间段内多次的位置信息的记录，例如，路线导航、查询交通拥堵情况等。无论是单点的位置信息还是连续的位置信息，都存在很多隐私保护方案[1-4]，在保证用户的位置隐私不被泄露的同时，也使用户能够享受相应的基于位置的服务。背景信息的存在提高了攻击者实现攻击的精准度，因此，在设计隐私保护方案时，背景信息是必须考虑的因素。现有的隐私保护方案中，背景信息是指地图中位置点的历史查询概率[5-7]。根据历史查询概率的大小，攻击者可以将低概率的位置过滤掉。随着移动智能终端的不断升级，存储成本已经大幅降低，用户在享用基于位置服务的同时，移动终端也存储了短时间内的查询记录，同样，服务提供商也有相关的信息存储，可以通过分析用户的查询记录获取用户位置信息，因此用户自身信息也成为了隐私保护方案设计中不可忽略的一项。

本文方案的创新点包括以下 3 个方面。

1) 现有的隐私保护方案大多基于历史的背景信息，通过普遍存在性来实现隐私保护方案，然而针对个体差异性的考虑并不充分。本文方案从历史信息的关联性和用户自身信息的关联性 2 个方面设计来保护用户的隐私，提出了一个简易隐私自关联的隐私保护算法（Ba-2PS, basic privacy self- correlation privacy- preserving scheme）实现对用户的隐私保护。

2) 在考虑历史信息的关联性和用户自身信息的关联性的同时，从时间的维度和查询范围的维度扩展了简易隐私自关联的隐私保护算法，提出了扩展隐私自关联的隐私保护算法（En-2PS, enhanced privacy self-correlation privacy-preserving scheme）。本文提出的隐私保护方案基于匿名技术，由于发送给服务提供商的请求信息中包含了用户的真实请求，并将真实信息隐藏在匿名信息中，因此不影响用户所享有的服务质量。

3) 不同于现有隐私保护方案的单一保护用户的位置隐私或者查询隐私，本文方案在考虑到背景信息存在的前提下，同时个性化地保护用户的位置隐私和查询隐私。

2 相关工作

基于位置服务中，用户的隐私信息包含位置隐私[8-10]和查询隐私[11-13]，现有的隐私保护方案主要分为基于密码学的隐私保护方案[14-15]、基于模糊的隐私保护方案[16-17]、基于差分的隐私保护方案[3,10]、基于匿名的隐私保护方案[18-19]。其中，基于匿名的隐私保护方案分为基于匿名中心的隐私保护方案[14-15]和基于移动终端的隐私保护方案[7-8,16]。在基于匿名中心的隐私保护方案中，匿名中心位于用户终端和服务提供商之间，承担了性能和安全的责任，对用户的信息进行处理，经过匿名化操作的请求信息由匿名中心发送至服务提供商，降低了用户的计算、存储开销。然而，匿名中心成为了性能和安全的瓶颈，大量的隐私请求由匿名中心来处理，执行效率会大幅降低，同时一旦匿名中心被攻破，那么所有的隐私信息将会泄露。目前，移动网络不断发展，智能终端不断智能化，存储能力不断增加，计算处理能力不断加强，为了避免匿名中心的单点泄露问题，基于移动终端的隐私保护方案也不断涌现。Niu 等[5]设计实现了基于匿名的隐私保护方案，考虑到背景信息中每一个位置的历史查询概率，同时为了避免所选的匿名位置位于相同的建筑之内降低隐私保护效果，通过计算匿名位置之间的距离，确保所选取的匿名位置单元两两之间的距离尽可能远。基于混合区域（Mixzone）的隐私保护方案中，Palanisamy 等[20]通过更换用户离开 Mixzone 时的假名，使攻击者不能够通过分析进入 Mixzone 和离开该区域的关联性来获取用户的真实信息，以保护用户的位置隐私。CacheCloak 方案[21]利用缓存用户请求的服务信息来避免和服务提供商的多次交互，通过减少交互次数保护用户的隐私信息。Xu 等[18]提出了一种基于感知的隐私保护模型，利用信息熵来衡量位置区域的受欢迎程度，并且利用四叉树来分离请求信息和特定用户，针对不同的分类提供个性化的隐私保护。

在保护用户位置隐私的同时，用户请求中的查

询内容同样包含了用户的隐私，根据基于匿名的隐私保护算法演变出了 *l*–多样性方案[23-25]和 *t*–接近性方案[26-28]。其中，*l*–多样性方案在基于 *k* 匿名方法的基础上，保证了每个等价类中敏感属性达到阈值，避免了敏感属性的泄露。*k* 匿名技术是一种利用混淆的方法使观测者无法辨别出真实信息的技术，例如，将用户的真实信息和 *k*–1 个匿名信息一起发送给服务提供商，服务提供商不能够分辨出接收到的 *k* 个请求中哪一个是真实的请求信息。He 等[25]提出了空间多样性，结合用户的路径和随机模型，分析得到用户的空间差异度，提出了一个优化停顿的隐私访问方案，实现了较高的隐私保护效果。Li 等[28]分析了 *l*–多样性的隐私保护算法的局限性，设定在等价类中敏感属性的分布距离不大于预设的阈值，实现对敏感属性的保护。

现有的隐私保护方案为保护用户的位置隐私和查询隐私奠定了深厚的基础。本文根据现有的隐私保护方案，考虑了用户自身信息之间的关联，在保证位置信息不被泄露的同时，从时间的维度和查询范围的维度保护了用户的查询隐私。

3　准备工作

3.1　基本概念

1) POI

POI（point of interest）是指在某一个位置点区别于坐标信息来辨别该位置的性质。例如，Alice 在某商场请求周围公交车站，那么此商场就是该位置的 POI，而该位置的坐标则是位置信息，公交车站是查询内容。

在某种程度上，POI 和查询内容在语义上具有重合性。例如，在上述的例子中，商场是该位置的 POI，公交车站是查询内容，然而从公交车站的位置坐标来说，公交车站是该位置的 POI。综上所述，POI 是位置坐标上基础建设（或者其他能够区别辨识的设施）的语义内容，能够作为用户进行查询的关键字成为查询内容。

2) 背景信息

随着科技的发展和信息的海量增加，任何一个具有计算、处理和存储功能的个体都能够获取地图中用户所处区域的历史查询数据，其中包含了某个地点曾经发生和当前发生的请求记录、该地的查询概率以及该地所处的商圈和 POI。现有的隐私保护方案单一地考虑了背景信息中历史数据的集合，而未关注某个特定时间点的查询概率以及相邻时间点的查询概率。本文所指背景信息不仅包含了查询概率，同时包含了区别位置的兴趣点 POI。本文中的背景信息是通过 Google 地图的 API（application programming interface）来获取区域内每个位置的查询概率，并将背景信息存储于智能终端中。

3.2　LBS 基本框架结构

基于位置服务的基本架构主要由 4 个部分组成：GPS 卫星、移动终端、通信基站和服务提供商，如图 1 所示。

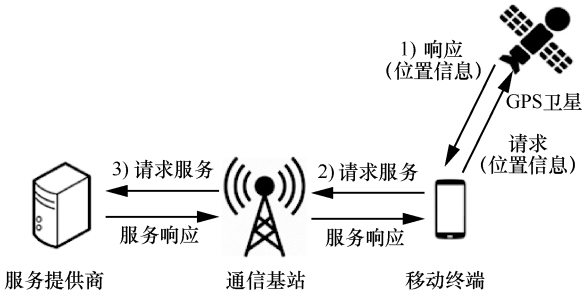


图 1　LBS 基本框架结构

用户通过移动终端获取其所处的位置信息，然后通过移动蜂窝网络或者 Wi-Fi 将服务请求发送至服务提供商的服务器。服务提供商在接收到由通信基站转发来的请求后，将用户所需的服务信息作为响应发送给用户的移动终端，从而完成一个完整的请求服务和响应服务的过程。

1) GPS 卫星为移动终端提供当前的地理位置信息。基于位置服务的隐私保护算法对此过程中的信息不进行考虑，默认在位置获取的过程是安全的。

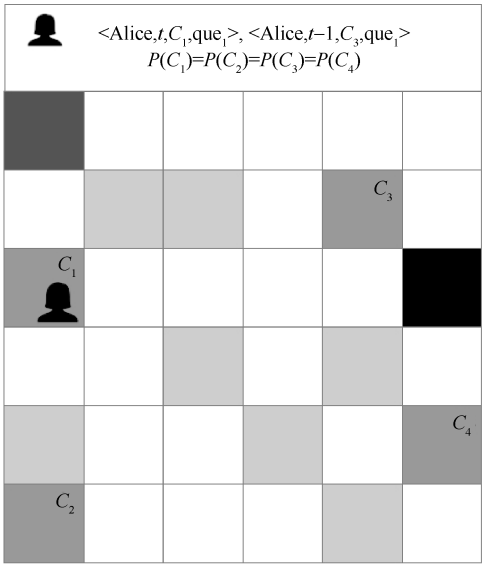
2) 用户通过移动终端对请求信息进行隐私保护，将保护后的信息发送至通信基站，在该过程中，用户的请求信息已经过隐私保护处理。

3) 通信基站将接收到的信息发送至相应的 LBS 提供商的服务器，通信基站只对用户的请求和收到的响应进行转发，不对请求信息进行修改。同样，通信基站在收到服务器的响应后将其转发至移动终端，不对响应信息进行修改。

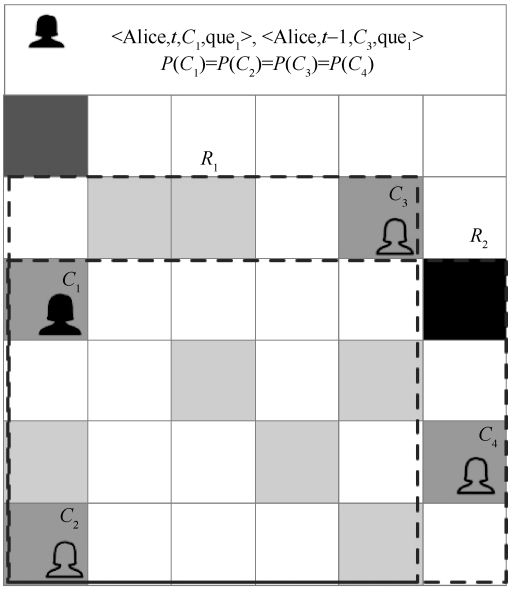
3.3　研究动机

在设计基于位置服务的隐私保护方案时，背景信息的存在为攻击者提供了线索来分析推测出用户的真实信息。本节通过一个例子来解释本文的研究动机。如图 2 所示，将地图等分为 6×6 的位置单元，根据背景信息中的历史查询概率，每一个位置单元用不

同的灰度来表示该位置单元历史查询概率的不同，灰度越高表示在该位置单元的历史查询概率越高，灰度越低表示该位置单元的历史查询概率越低。



(a) 地图划分



(b) 位置区域选择

图 2 研究动机

现有的隐私保护方案为了达到最佳的隐私保护效果，所选取匿名位置的历史查询概率和用户（Alice）真实位置的历史查询概率相同，从而使隐私度量值最高。在 t 时刻，Alice 位于位置区域 C_1 中，该地图中和 Alice 所处位置区域 C_1 具有相同的历史查询概率的位置区域为 C_2、C_3、C_4，即

$$P(C_1)=P(C_2)=P(C_3)=P(C_4)$$

其中，$P(C_i)$ 表示位置区域 C_i 的历史查询概率，如图 2(a)所示。假设在该例子中，选取的 k 匿名算法中的 k=3，则需要选择 $k-1=2$ 个匿名位置区域，则从 3 个具有相同历史查询概率的位置区域中选出 2 个，即 C_2、C_3，如图 2(b)所示的虚线区域 R_1。

然而，有很多 APP 对用户的历史查询信息进行存储。例如，在 $t-1$ 时刻，Alice 曾在位置区域 C_3 发起过对于 que_1 的查询请求，由于 2 个时间点之间的时间差值足够小，即 $\Delta t=\varepsilon$，$\varepsilon\in \mathrm{N}^+$，那么攻击者则会认为 Alice 在短期内搜索过该位置的服务，且服务内容相同，则该位置区域 C_3 为匿名位置，从而用户的泄露概率由 $\frac{1}{3}$ 增加为 $\frac{1}{2}$，增加了隐私泄露的风险。

因此，单一地考虑地图中位置的历史信息无法抵御攻击者从用户自身信息的关联性来推断出用户真实信息的攻击。本文从历史信息和用户自身信息 2 个方面同时考虑，提出了 2 个隐私保护算法：简易隐私自关联的隐私保护算法（Ba-2PS）和扩展隐私自关联的隐私保护算法（En-2PS）。其中，Ba-2PS 考虑到用户自身关联的隐私信息，在选取匿名位置区域时，将区域 C_3 过滤，从而选取区域位置 C_4 为匿名区域，如图 2(b)中的虚线区域 R_2 所示。

3.4 隐私度量

在基于 k 匿名的隐私保护方案中，用户发送给服务提供商的请求包含了用户的身份、发送请求的时间戳、用户的当前位置、查询内容和查询范围。本文为了衡量隐私保护的程度，采用 k 匿名概率和信息熵来衡量隐私信息被推测的不确定性[28]。其中，k 匿名概率表示用户隐私信息的泄露概率，泄露概率越大则表示隐私保护的程度越低。在基于信息熵的隐私度量中，信息熵值越大，说明计算的信息的不确定性越高，则隐私保护的程度就越高。信息熵的具体定义如下。

定义 1 基于 k 匿名概率的隐私度量。已知所选取的匿名位置构成区域为 R，$R=<\text{loc}_1,\text{loc}_2,\cdots,\text{loc}_r>$，即 $|R|=r$，则基于 k 匿名概率中 k 的取值为该区域中位置单元的个数，即 $k=|R|$，则泄露概率为

$$Q=\frac{1}{k}=\frac{1}{|R|}$$

定义 2 基于信息熵的隐私度量。已知用户真实的位置区域为 j，经过 k 匿名算法得到 k–1 个匿名位置区域，则信息熵为

$$H=-\sum_{i}^{k}\Pr_i \mathrm{lb}\Pr_i \tag{1}$$

其中，$\Pr_i$ 表示根据用户的真实位置区域 j 选取的匿名区域 i 的概率。信息熵具有极值性，即事件的发生概率相同，信息熵值达到最大值，概率事件的不确定性最高，在隐私保护算法中则说明用户的隐私保护程度最高。

4　简易隐私自关联的隐私保护算法

Ba-2PS 算法由地图预处理算法和双筛选算法组成。其中，地图预处理算法实现了对地图的初始化分割，根据历史背景信息初始化位置单元的历史查询概率，进而通过双筛选算法选取匿名位置发送给服务提供商来换取相应的服务。

4.1　地图预处理算法

在地图预处理算法中，获取地图，将地图等分为 m 个位置单元，每个位置单元作为后续进行位置选取的最小位置单元。

$$\mathrm{MAP}=\{\mathrm{loc}_1,\mathrm{loc}_2,\cdots,\mathrm{loc}_m\}$$

其中，m 的数值决定了地图划分的粒度，loc_i 表示第 i 个位置单元，且 $i,m\in \mathrm{N}^+$。

在 t 时刻，用户 Alice 位于位置单元 loc_u 中，请求附近的 POI 的信息，其中请求范围记作 rad_u，所查询的内容记作 que_u。用户 Alice 的真实请求向量为 $\mathbf{Req}=<\mathrm{ID}_u,t,\mathrm{loc}_u,\mathrm{que}_u,\mathrm{rad}_u>$。根据背景信息获取位置单元 loc_u 的查询概率 $\Pr(\mathrm{loc}_u)$，遍历地图中的位置单元，选择和 Alice 当前位置单元具有相近查询概率的位置单元，即 $|\Pr(\mathrm{loc}_u)-\Pr(\mathrm{loc}_i)|<\gamma$，其中 $\gamma\geqslant 0$，将满足条件的位置单元存储在集合 η 中，并将集合 η 作为双筛选算法的一个输入，则 $\mathrm{loc}_i\in\eta$。具体如算法 1 所示。

算法 1　地图预处理算法

输入　地图，用户当前位置信息

输出　集合 η

1) 等分地图为 m 个位置单元；

2) 选取满足 $|\Pr(\mathrm{loc}_u)-\Pr(\mathrm{loc}_i)|<\gamma$ 的位置单元；

3) 存储于集合 η 中；

4) 输入集合 η。

4.2　双筛选算法

根据地图预处理算法得到一个位置单元的集合 η，在该集合中每个位置单元的历史查询概率和用户当前所处位置单元具有相同的历史查询概率。在双筛选算法中，假设用户的移动终端中缓存了用户在 T 时间段内的查询记录，该时间段的查询记录存储于集合 ζ 中，其中用户 u 在 T 时间内的查询记录为 $\mathrm{local}_i<\mathrm{ID}_u,t_i,\mathrm{loc}_i,\mathrm{que}_i,\mathrm{rad}_i>$，其中 $\mathrm{local}_i\in\zeta$，$|t_i-t_u|<T$, $T>0$。

根据已知位置单元集合 η 和查询记录集合 ζ，对位置单元集合 η 和查询记录集合 ζ 中的向量求交集，如果集合 η 中的位置单元存在于集合 ζ 中，则将该位置单元从集合 η 中删除。经过第一次筛选，集合 η 中的位置单元个数小于或等于最初从地图预处理算法中得到的位置单元个数。此次筛选得到的位置单元避免了因用户自身信息关联而降低隐私强度的问题。

然而，此时集合 η 中的位置单元不能满足距离用户的位置尽可能远，而一旦选取的匿名位置单元和用户的真实位置十分接近时，无法避免两者位于同一座建筑内，从而失去进行匿名保护的意义。第二次筛选操作的目的则是尽可能地选取距离较远的位置单元，保证匿名位置单元尽可能分散。

为了方便第二次的筛选操作，本文采用四叉树地图进行存储，根据背景信息迭代划分地图，每次四等分，如图 3 所示。查询概率高的区域位置单元的密度较高，查询概率低的区域位置单元的密度较低，因此，在 $\mathrm{MAP}=<\mathrm{loc}_1,\mathrm{loc}_2,\cdots,\mathrm{loc}_m>$ 中，设置 $m=4n,n\in\mathrm{N}^+$。根据迭代划分，对应由四叉树来进行存储，为了保证所选的匿名位置尽可能分散，则所选的匿名位置单元中不存在任意 2 个位置单元属于同一个父节点，且深度检索该位置单元所属分支的深度满足 $\mathrm{Dep}(\mathrm{loc}_i)\geqslant\xi$ 的节点不存在相同的父节点。

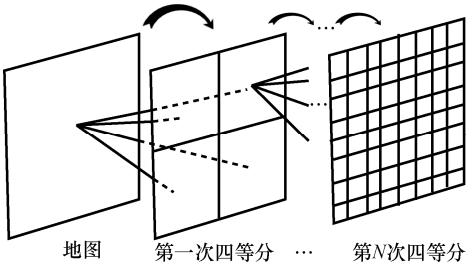


图 3　迭代四等分划分

最后将得到的 $k-1$ 个位置单元作为匿名位置单元构成匿名查询请求 $\mathbf{Req}_i=<\mathrm{ID}_u,t,\mathrm{loc}_i,\mathrm{que}_u,\mathrm{rad}_u>,\mathrm{loc}_i\in\eta$ 和用户的真实位置 $\mathbf{Req}=<\mathrm{ID}_u,t,$

$\langle \mathrm{loc}_u, \mathrm{que}_u, \mathrm{rad}_u \rangle$ 一起发送给服务提供。具体如算法 2 所示。

算法 2 双筛选算法

输入 集合 η、ζ，k 值

输出 集合 η'

1) if ($\eta \cap \zeta \neq \varnothing$)
{
2) if($\mathrm{loc}_i \in \eta \,\&\&\, \mathrm{loc}_i \in \zeta$)
{
3) 从集合 η 删除 loc_i；
i++;
} end if ;
4) else
{
5) i++;
}
} end if;
6) 用户当前位置的父节点 node_u；
7) 选取与节点 node_u 同深度的节点 node_a，node_b，node_c；
8) while (k−1)
{
9) if ($\mathrm{loc}_i \in \eta$)
{
10) if (node_i 是 node_a 节点下任一节点 ‖ node_i 是 node_b 节点下任一节点 ‖ node_i 是 node_c 节点下任一节点)
{
11) loc_i 为匿名位置单元；
12) 存储 loc_i 于集合 η'；
} end if;
} end if;
} end while;
13) 输出集合 η'；

5 扩展隐私自关联的隐私保护算法

En-2PS 算法是 Ba-2PS 的扩展方案，在考虑到历史背景信息和用户自身信息的同时，分析位置单元、查询内容、查询半径之间的关联关系，设计匿名查询内容生成算法和请求生成算法。

5.1 匿名查询内容生成算法

在匿名查询内容生成算法中，从时间的维度分析位置单元和查询内容的关系，从而确定匿名查询内容。时间维度的不同，所查询的 POI 的种类也不相同。用户请求服务的时间为 t，根据 Ba-2PS 选取的 k−1 个匿名位置单元满足以下 3 个要求。

1) 根据背景信息所选取的位置单元和用户所处的位置单元有相近的查询概率。

2) 不存在用户在 T 时段内的查询记录中。

3) 所选取的位置单元分散在地图中且互不相邻。

为了保护用户的查询隐私，在发送给服务提供商的请求信息中的查询内容不能和用户真实的查询内容相同，同时避免在特定时间点发送不可能的请求信息（例如，在清晨时间查询酒吧）。已知用户 Alice 在位置单元请求的查询内容 que_u，集合 η' 中的匿名位置单元 loc_i，根据背景信息，在 t 时刻，匿名位置单元曾查询的 POI 中选取查询概率较高的 POI 作为查询请求中匿名位置单元 loc_i 的查询内容 que_i，则该查询内容的查询概率满足 $\mathrm{Pr}_{\mathrm{loc}_i}(\mathrm{que}_i) > \theta$，其中 $0<\theta<1$。具体如算法 3 所示。

算法 3 匿名查询内容生成算法

输入 集合 η'，背景信息

输出 集合 δ

1) while (k−1)
{
2) if ($\mathrm{Pr}_{\mathrm{loc}_i}(\mathrm{que}_i) > \theta$)
{
3) 存储 que_i 于集合 δ；
4) k−−;
} end if;
else
{
5) i++;
6) 跳转至 2);
}
} end while;
7) 输出集合 δ；

5.2 请求生成算法

在请求生成算法中，从查询范围的维度重新选择每一个匿名请求对应的查询半径，从而避免攻击者利用背景信息推测匿名位置单元的查询半径过大或者过小，进而降低匿名位置单元的个数，提高用户隐私的泄露概率。例如，位置单元的查询半径过大，包含了不可能的地方（例如河、湖），从而

攻击者推测此位置为匿名位置单元。

已知 Ba-2PS 中对地图利用四叉树进行存储，假设 k=3，选取的 2 个匿名位置单元分别为 loc_1 和 loc_2，如图 4 所示。根据深度优先搜索四叉树，得到位置单元的深度分别为 $\mathrm{Dep}(\mathrm{loc}_1)$、$\mathrm{Dep}(\mathrm{loc}_2)$ 和 $\mathrm{Dep}(\mathrm{loc}_u)$。如果所选的位置单元的深度满足 $\mathrm{Dep}(\mathrm{loc}_i)>\theta$，则说明该位置单元所处区域密度较高，进而相应的查询半径为 $\mathrm{rad}_i=\mathrm{rad}_i+\beta$，其中 $\beta<0$，如果 $\mathrm{Dep}(\mathrm{loc}_1)\leqslant\theta$，则查询半径为 $\mathrm{rad}_i=\mathrm{rad}_i+\beta$，其中 $\beta>0$。

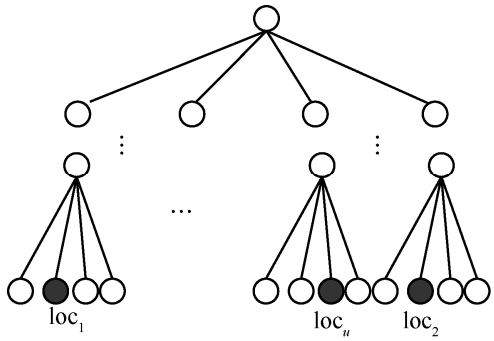


图 4　四叉树

最后，得到查询半径和相应的位置单元、查询内容，重新构成匿名请求和用户的真实请求一起发送给服务提供商。具体如算法 4 所示。

算法 4　请求生成算法

输入　集合 η'，集合 δ

输出　真实请求和匿名请求

1) while (k−1)
{
2)　if ($\mathrm{Dep}(\mathrm{loc}_1)>\theta$)
{
3)　　$\mathrm{rad}_i=\mathrm{rad}_i+\beta$, 其中 $\beta<0$;
4)　　$\mathrm{Req}_i=<\mathrm{ID}_u,t,\mathrm{loc}_i,\mathrm{que}_i,\mathrm{rad}_i>$;
5)　　k−−;
} end if;
else
{
6)　　$\mathrm{rad}_i=\mathrm{rad}_i+\beta$，其中 $\beta>0$;
7)　$\mathbf{Req}_i=<\mathrm{ID}_u,t,\mathrm{loc}_i,\mathrm{que}_i,\mathrm{rad}_i>$;
8)　　k−−;
}
} end while;
9) 输出 $\mathbf{Req}_u$ 和 $\mathbf{Req}_i$;

6　隐私性分析和实验验证

本节分别从隐私性分析、性能分析以及隐私性验证对 Ba-2PS 和 En-2PS 方案的性能和隐私性进行测试和验证。通过分析证明 En-2PS 方案能够抵御的攻击类型，同时通过隐私度量对保护隐私的保护程度证明了方案的隐私性。

本文实验采用 Matlab 软件在 PC 机（2.9 GHz Intel i7 CPU, 16 GB 内存）上进行模拟仿真，并对方案的性能进行验证。

选取了某城市内 8 km×8 km 的地图，划分为 160×160 个位置单元，每个位置单元为 50 m×50 m 的矩形。实验中参数 k 是指 k–匿名中发送给服务提供商的请求数量，选取范围为[5,50]。

6.1　隐私性分析

本节分别从共谋攻击和推断攻击 2 个方面分析了 En-2PS 方案的安全性。在基于位置服务中，用户利用无线信道和基于位置的服务提供商进行交互。无线信道是开放的，基于密码学的隐私保护算法可以避免恶意用户窃取无线信道中的隐私信息，然而基于密码学的隐私保护算法（例如，AES、3DES、RSA 等）计算和处理开销较大。

定义 3　已知 $\forall i\in C$，满足 $\Pr(\widetilde{\mathrm{loc}}\,|\,\mathrm{loc}_i)=\phi$，其中集合 C 为共谋集合，$\widetilde{\mathrm{loc}}$ 是攻击者通过接收到的信息推测出用户的位置信息，loc_i 为攻击者接收到的位置信息，且 $\mathrm{loc}_i\in O$，集合 O 是攻击者所观测到的位置信息集合，$|O|=k$，$\Pr(\widetilde{\mathrm{loc}}\,|\,\mathrm{loc}_i)$ 表示攻击者通过收集到的位置信息推测出位置信息的概率。

引理 1　En-2PS 可抵御共谋攻击。

证明　攻击者为了获取用户的真实位置信息，可能联合其他用户甚至服务提供商实现共谋攻击。在本文方案中，无论多少用户或者服务提供商联合，通过公布的位置信息推断出真实位置信息的概率为一常数。En-2PS 利用了 k 匿名技术来实现匿名位置单元的选取，则从 k 个公布出的位置单元中得出用户真实位置单元的概率为 $\frac{1}{k}$。

En-2PS 是根据真实的背景信息来选取匿名位置单元的，其中背景信息可以被任意用户或服务提供商获取，因此，无论共谋用户有多少，用户的真实信息始终隐藏在 k–1 个匿名位置信息中。综上所述，En-2PS 能够抵御共谋攻击。证毕。

定义 4　已知 $\mathrm{loc}_i\in O$，集合 O 是攻击者所观测

到的位置信息集合，且$|O|=k$，攻击者推断出的位置信息为$\widetilde{\text{loc}_a}$、$\widetilde{\text{loc}_b}$，如果方案满足$\Pr(\widetilde{\text{loc}_a}|\text{loc}_i)=\Pr(\widetilde{\text{loc}_b}|\text{loc}_i)$，则表明攻击者观测到的位置单元两两相同，因概率相同从而不可分区；如果满足$\Pr(\widetilde{\text{loc}_a}|\text{loc}_i)\neq\Pr(\widetilde{\text{loc}_b}|\text{loc}_i)$，则表明攻击者观测到的位置单元两两不同，使攻击者不能分辨。因此，没有孤立点和通过背景关系可分析出的特殊点，说明该方案能够抵御推断攻击。

引理 2 En-2PS 可抵御推断攻击。

证明 在 LBS 中，服务提供商具有强大的计算、处理和存储能力，在推断攻击中视服务提供商为主动攻击者，可以基于背景信息来分析推断用户的隐私。在 En-2PS 中，根据背景信息选取的匿名位置单元满足$|\Pr(\text{loc}_u)-\Pr(\text{loc}_i)|<\gamma$，其中$\gamma>0$，攻击者不能根据位置单元的查询概率对观测到的信息进行筛选来分析得出用户的真实位置单元。从而，攻击者推断用户的真实信息满足$\Pr(\widetilde{\text{loc}_a}|\text{loc}_i)\neq\Pr\widetilde{\text{loc}_b}|\text{loc}_i)$。此外，En-2PS 中对用户的查询内容和查询范围进行个性化选择，成功地避免了攻击者通过背景信息和事件发生的可能性来推断用户真实信息的情况，因此，En-2PS 能够抵御推断攻击。证毕。

基于隐私度量的隐私性分析，根据第 3.4 节中提出的隐私度量，利用信息熵计算隐私保护方案所实现的隐私保护程度。参照 3.4 节中用户真实的位置区域为 j，所对应的熵值如式(1)所示。

在地图预处理算法中，所选取位置单元loc_u的查询概率为$\Pr(\text{loc}_u)$，遍历地图中的位置单元，选择和当前位置单元具有相近查询概率的位置单元，即$|\Pr(\text{loc}_u)-\Pr(\text{loc}_i)|<\gamma$，其中$\gamma\geqslant 0$。根据信息熵的定义，当选择的位置单元具有相同的查询概率时，信息熵值达到最大。假设loc_x和loc_y满足$\text{loc}_x+\text{loc}_y=p$，其中$0<p<1$，当$\text{loc}_x$越接近$\frac{p}{2}$时，$z=-(\text{loc}_x\text{lb}\,\text{loc}_x+(p-\text{loc}_x)\text{lb}(p-\text{loc}_x))$的值越大，求导可得

$$z'=-(\text{lb}\,\text{loc}_x+\text{lb}(p-\text{loc}_x))-(p-\text{loc}_x)\cdot\frac{1}{p-\text{loc}_x})=-\text{lb}\frac{\text{loc}_x}{p-\text{loc}_x}$$

则有$\text{loc}_x=\frac{p}{2}$时，$z'=0$。且$\text{loc}_x<\frac{p}{2}$时，$z'>0$，$\text{loc}_x>\frac{p}{2}$时，$z'<0$。故而$\text{loc}_x=\frac{p}{2}$时z为最大值。

假设$p=\frac{1}{n}$，则可推出

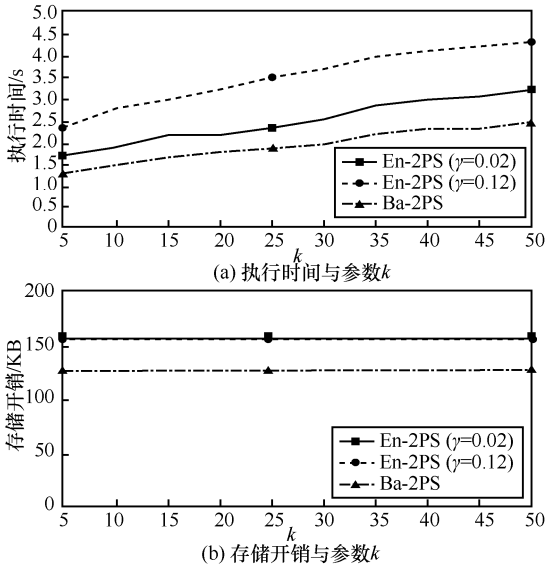
$$-(p\text{lb}\,p+(\max(p)+\min(p)-p)\text{lb}(\max(p)+\min(\text{p})-p))>-(\max(p)\text{lb}\max(p))+\min(p)\text{lb}\min(p)$$

在地图预处理算法中参数$\gamma\geqslant 0$，则有当γ=0时，上述式子得到$p=\frac{1}{n}$，即当所选取的位置单元的查询概率相同时，$H=-\sum_i^k \Pr_i\,\text{lb}\,\Pr_i$取最大值。

6.2 性能验证

1) 执行时间和存储开销与参数 k 关系

本节评估了 Ba-2PS 和 En-2PS 的执行时间和存储开销随 k 匿名算法中参数 k 值变化的情况，如图 5 所示。Ba-2PS 中没有参数γ的参与，故只考虑了 En-2PS 中γ=0.02 和γ=0.12 的情况。参数γ决定了选取匿名位置单元的查询概率的偏差度，参数γ越大，则集合η中的元素就越多，从而影响了算法的执行时间。Ba-2PS 和 En-2PS 的计算和存储都在用户的智能移动终端中进行，终端中存储的数据大小并不会因为选取的 k 值大小而改变，故在图 5(b)中算法的存储开销保持不变。En-2PS 中，参数γ不影响通信开销，因此γ=0.2 与γ=0.12 这 2 条曲线重合。



(a) 执行时间与参数k

(b) 存储开销与参数k

图 5 执行时间和存储开销与参数 k 关系

2) 执行时间和通信开销与参数γ关系

本节评估了执行时间和通信开销和参数γ的关系，如图 6 所示。Ba-2PS 没有参数γ的参与，故只选取了$k=20$和$k=50$对 En-2PS 进行性能评估。用户利用移动终端将请求信息发送给服务提供商来换

取服务，其中请求信息的大小决定了通信开销的大小，而请求信息的大小取决于匿名算法中参数 k 值的大小，而和参数 γ 无关。然而，参数 γ 的大小影响了集合 η 的大小，参数 γ 的值越大，则满足 $|\Pr(\mathrm{loc}_u)-\Pr(\mathrm{loc}_i)|<\gamma$ 条件的位置单元就越多，故集合 η 中元素越多，在进行迭代遍历时消耗的时间更多。

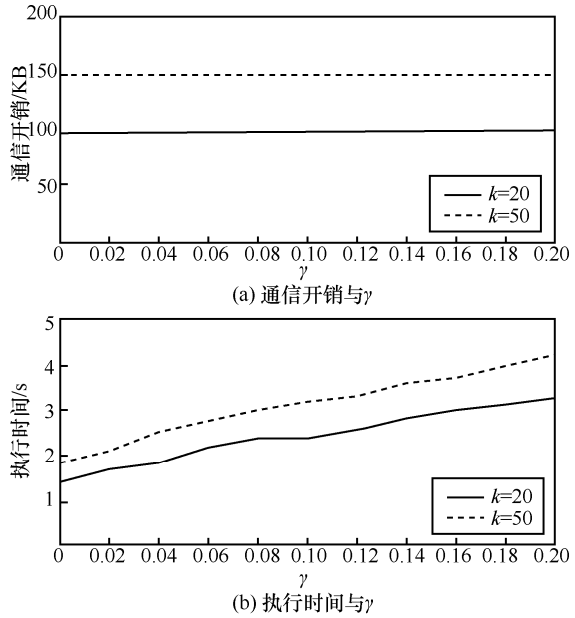


图 6　执行时间和通信开销与参数 γ 关系

6.3　隐私验证

本节分别从泄露概率、位置隐私和查询隐私 3 个方面对 Ba-2PS 和 En-2PS 的隐私性进行评估，同时与 Spati-PPM 算法[6]进行比较，验证本文方案的安全性。

实验比较了理论上的最佳值（optimal）、随机值（baseline）和 Ba-2PS、En-2PS、Spati-PPM 方案随参数 k 变化的泄露概率，如图 7 所示，k 值越大，发送给服务提供商的请求信息就越多，则攻击者能推测出真实信息的概率就越低。

本文采用信息熵作为用户位置隐私和查询隐私的度量标准，具体的计算方式如定义 2 所示。信息熵 $H=-\sum_i^k \Pr_i \mathrm{lb}\Pr_i$，其中熵值越大，对应的位置隐私度量和查询隐私度量越高。因为 En-2PS 在构造请求时对用户的请求范围进行了个性化选取，减小了攻击者利用查询范围和位置单元的可能性，进而实现推断攻击，降低了分析出用户敏感信息的可能性，提高了位置隐私和查询隐私的不确定性，如图 8 和图 9 所示。其中，Spati-PPM 算法考虑了时间和空间之间的关联性，为攻击者推断用户的真实信息增加了难度，因此图 8 中该方案的位置隐私效果优于 Ba-2PS，但因其未涉及对查询隐私的保护，图 9 中该方案的效果最差。En-2PS 在考虑请求范围的同时，也在选取匿名位置单元时将时间因素考虑在内，避免了利用时空关联因素来推断出用户的真实位置信息和查询内容的可能性。

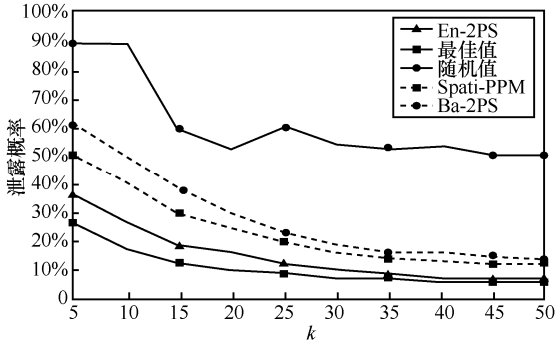


图 7　参数 k 与泄露概率关系

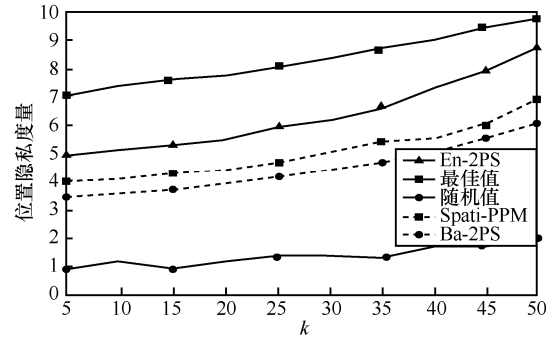


图 8　参数 k 与位置隐私度量关系

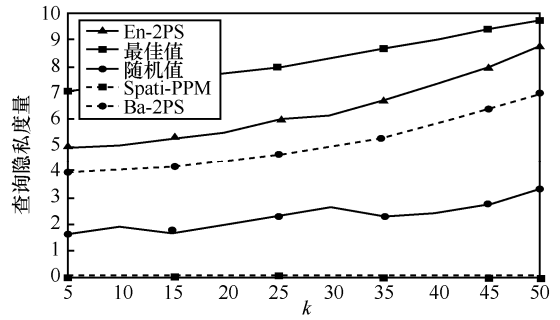


图 9　参数 k 与查询隐私度量关系

7　结束语

在基于位置服务中，用户和服务提供商都会存储一定时间内的用户请求信息，因此，攻击者可以通过对历史查询信息的分析降低匿名效果。本文提出了一种隐私信息自关联的隐私保护方案，在选取匿名位置单元时考虑到用户自身的查询历史的自

关联关系，避免了攻击者利用该信息来实现推断攻击。本文首先提出了一个基本的隐私保护方案 Ba-2PS 考虑了用户信息的自关联，进而从查询范围出发，提出了 En-2PS，实现了对用户位置隐私和查询隐私的保护。最后，本文通过详细的性能和安全性实验验证了方案的有效性和安全性。

参考文献：

[1] LAI C, ZHOU H, CHENG N, et al. Secure group communications in vehicular networks: a software-defined network-enabled architecture and solution[J]. IEEE Vehicular Technology Magazine, 2017, 12(4): 40- 49.

[2] HE X, JIN R, DAI H. Leveraging spatial diversity for privacy-aware location-based services in mobile networks[J]. IEEE Transactions on Information Forensics and Security, 2018,13(6):1524-1534.

[3] ZHANG J, CHOW C Y. Enabling probabilistic differential privacy protection for location recommendations[J]. IEEE Transactions on Services Computing, 2018: 1.

[4] 李凤华,李晖,贾焰,等.隐私计算研究范畴及发展趋势[J].通信学报, 2016,37(4): 1-11.
LI F H, LI H, JIA Y, et al.Privacy computing:concept,connotation and its research trend[J].Journal on Communications,2016,37(4): 1-11.

[5] NIU B, LI Q, ZHU X, et al. Achieving *k*-anonymity in privacy-aware location-based services[C]//International Conference on Computer Communications. IEEE, 2014: 754-762.

[6] LIU H, LI X, LI H, et al. Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services[C]// International Conference on Computer Communications. IEEE,2017.

[7] ZHANG P, HU C, CHEN D, et al. Shiftroute: achieving location privacy for map services on smartphones[J]. IEEE Transactions on Vehicular Technology, 2018,67(5):4527-4538.

[8] ZHENG X, CAI Z, LI J, et al. Location-privacy-aware review publication mechanism for local business service systems[C]//International Conference on Computer Communications. IEEE, 2017: 1-9.

[9] JIANG H, ZHAO P, WANG C. Roblop: towards robust privacy preserving against location dependent attacks in continuous lbs queries[J]. IEEE/ACM Transactions on Networking,2018,26(2):1018-1032.

[10] XIAO Y, XIONG L. Protecting locations with differential privacy under temporal correlations[C]//The 22nd ACM Conference on Computer and Communications Security.2015: 1-7.

[11] SHOKRI R, THEODORAKOPOULOS G, BOUDEC J Y L, et al. Quantifying location privacy[C]//IEEE Security and Privacy. 2011: 247-262.

[12] WANG J, LI Y, YANG D, et al. Achieving effective *k*-anonymity for query privacy in location-based services[J]. IEEE Access, 2017, 5: 24580-24592.

[13] LI W, NIU B, LI H, et al. Privacy-preserving strategies in service quality aware location-based services[C]//IEEE International Conference on Communications. 2015: 7328-7334.

[14] YI X, PAULET R, BERTINO E, et al. Practical approximate K nearest neighbor queries with location and query privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(6): 1546-1559.

[15] PAULET R, KAOSAR M, YI X, et al. Privacy-preserving and content-protecting location based queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2014,26(5): 1200-1210.

[16] ANDRES M, BORDENABE N, CHATZIKOKOLAKIS K, et al. Geo-indistinguishability: differential privacy for location-based systems[C]//The ACM Conference on Computer and Communications Security. 2013: 901-914.

[17] PERAZZO P, DINI G. A uniformity-based approach to location privacy[J]. Computer Communications, 2015, 64(1): 21-32.

[18] XU T, CAI Y. Feeling-based location privacy protection for location-based services[C]//ACM Conference on Computer and Communications Security. 2009: 348-357.

[19] XU T, CAI Y. Exploring historical location data for anonymity preservation in location-based services[C]//International Conference on Computer Communications. IEEE, 2008: 1220-1228.

[20] 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.

[21] MEYEROWITZ J, CHOUDHURY R. Hiding stars with fireworks: location privacy through camouflage[C]//The Annual International Conference on Mobile Computing and Networking. 2009: 345-356.

[22] MACHANAVAJJHALAA, GEHRKEJ, KIFERD, et al. *L*-diversity: privacy beyond *k*-anonymity[C]//International Conference on Data Engineering. IEEE, 2006: 24-26.

[23] TRIPATHYB K, MAITY A, RANAJIT B, et al. A fast psensitive l-diversity anonymisation algorithm[C]// Recent Advances in Intelligent Computational Systems. IEEE, 2011: 741-744.

[24] HE X, JIN R, DAH I. Leveraging spatial diversity for privacy- aware location-based services in mobile networks[J]. IEEE Transactions on Information Forensics and Security, 2018,13(6):1524-1534.

[25] HAO G, BINX Y. Research on privacy preserving method based on t-closeness model[C]//The International Conference on Communications in China. IEEE, 2017: 1455-1459.

[26] MONEDERO D R, FORNE J, FERRER J D. From t-closeness-like privacy to post randomization via information theory[J]. IEEE Transactions on Knowledge and Data Engineering, 2010,22(11):1623-1636.

[27] LI N H, LIT C, VENKATASUBRAMANIAN S. T-closeness: privacy beyond k-anonymity and l-diversity[C]//International Conference on Data Engineering. IEEE, 2007: 106-115.

[28] CHEN Z, HU X, JU X, et al. Lisa: location information scrambler for privacy protection on smartphones[C]// Conference on Communications and Network Security. IEEE, 2013: 296-304.

[作者简介]

李维皓（1990− ），女，辽宁沈阳人，西安电子科技大学博士生，主要研究方向为社交网络中的隐私保护。

曹进（1985− ），男，陕西西安人，博士，西安电子科技大学副教授、博士生导师，主要研究方向为无线网络安全和应用密码学。

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。