一种面向公路网络的位置隐私保护方法

薛 姣[1),2)] 刘向宇[1),2)] 杨晓春[1),2)] 王 斌[1),2)]

[1)](东北大学医学影像计算教育部重点实验室 沈阳 110819)

[2)](东北大学信息科学与工程学院 沈阳 110819)

摘 要 移动用户经常会发出基于目前位置的最近邻查询.通常移动终端(用户)向不可信的位置服务器发送查询请求,请求中包含移动终端的位置信息,因而导致位置隐私的泄露.给移动用户提供位置服务的同时,保护移动用户的位置隐私也至关重要,而这种需求在公路网络应用中尤为明显.根据公路网络的结构特点,提出了隐匿环和隐匿树这两种子图结构.利用隐匿环和隐匿树模糊移动用户在公路网络中的位置信息,可以有效地保护位置隐私.文中提出了一种新的位置隐私保护方法——隐匿环与森林(CCF),即利用宽度优先搜索在图中寻找满足一定要求的环和森林.对于包含单行线的公路网络,CCF 依然能够保护移动用户的位置隐私.在基于真实与模拟数据集的实验测试中,CCF 方法显示了其在保护位置隐私方面的有效性以及在提供服务质量方面的高效性.

关键词 位置隐私;基于位置服务;公路网络;子图隐匿;单行线

中图法分类号 TP309 **DOI 号**: 10.3724/SP.J.1016.2011.00865

A Location Privacy Preserving Approach on Road Network

XUE Jiao[1),2)] LIU Xiang-Yu[1),2)] YANG Xiao-Chun[1),2)] WANG Bin[1),2)]

[1)](*Key Laboratory of Medical Image Computing (Northeastern University) of Ministry of Education, Shenyang* 110819)

[2)](*College of Information Science and Engineering, Northeastern University, Shenyang* 110819)

Abstract Mobile users often issue the nearest neighbor query based on their current locations. Generally, a mobile terminal (user) sends the query including his location information to an untrusted location-based server and leads to his location privacy is leaked. So, it is important to protect mobile user's location privacy while providing location-based service, especially in a road network application. According to the structural characteristics of the road network, two cloaking subgraph structures, cloaking cycle and cloaking tree, are proposed. They blur the user's location information and protect the user's location privacy effectively. Furthermore, a novel location privacy preserving approach, called cloaking cycle and forest (CCF for short), is proposed. CCF finds cycles and forest that satisfy certain conditions using breadth-first search in graph. When the road network contains one-way streets, CCF can still protect the mobile user's location privacy. In an experimental test based on real and simulated datasets, the effectiveness on protecting location privacy and efficiency on providing service quality of CCF are given.

Keywords location privacy; location-based services; road network; subgraph cloaking; one-way street

收稿日期:2010-06-11;最终修改稿收到日期:2010-07-05.本课题得到国家自然科学基金(60973018,60973020,60828004)、中央高校基本科研业务费专项资金(N090504004)资助.**薛 姣**,女,1986 年生,硕士研究生,主要研究方向为位置隐私.**刘向宇**,男,1981 年生,博士研究生,主要研究方向为数据安全、数据世系管理.**杨晓春**,女,1973 年生,教授,博士生导师,主要研究领域为数据库理论与技术、数据质量.E-mail: yangxc@mail.neu.edu.cn.**王 斌**,男,1972 年生,讲师,主要研究方向为分布式数据管理、系统结构.

1 引 言

近年来,基于位置服务(LBS)逐渐走入人们的生活,以智能化的互动方式给人们的生活带来了极大便利,使人们对其需求出现了大量增长.例如,基于位置服务可以给移动用户提供感兴趣地点的查询("离我最近的公交站牌在哪?")、位置导航("去火车站怎么走?")、实时路况信息查询("车辆目前在三好街是否能畅行?")等.

为了获得基于位置服务,移动用户需要向位置服务提供商发送包含他们精确位置信息的查询请求.通常情况下,位置服务提供商的服务器都是不可信的,用户的位置信息很容易被攻击者窃取.在窃取到移动用户的位置信息之后,攻击者可能会通过位置追踪或者链接其它的一些公开信息(例如地理编码数据库、电话本等)重新确认用户的身份,进而得知用户更多的隐私信息,例如,移动用户的生活方式、健康状况、政治背景等[1].

如何保护移动用户的位置隐私?这个问题已经引起了专家、学者们的高度重视,很多解决方案也相继提出.目前绝大多数的研究工作都假设移动用户在欧式空间(自由空间)中移动,即他们的移动方向不受任何约束,采用的主要保护方法是将用户的精确位置用一个至少包含其它 $k-1$ 个人的空间区域来代替,从而使攻击者无法将某一个人的位置信息通过推理攻击的方式与其身份相匹配.同时,这种方法利用空间区域面积的大小以及所包含人数的多少来衡量为用户提供的隐私保护力度.

但是在现实生活中,人们无论是步行还是乘坐某种交通工具,总是要遵循固定的公路网络.例如,图 1 显示的就是一个简单的公路网络模型,模型中有移动用户以及他们感兴趣的地点.显然,在公路网络环境下,不能再利用空间区域面积的大小来衡量隐私保护的强弱,因为对于公路网络上两个面积相等的空间区域,如果一个包含一条路段①,而另一个包含三条路段,那么后者的保护强度要大于前者.此外,只包含一条路段的空间区域也很容易导致用户被攻击者跟踪[2].所以,空间区域隐匿的方法不再适用于公路网络环境下.

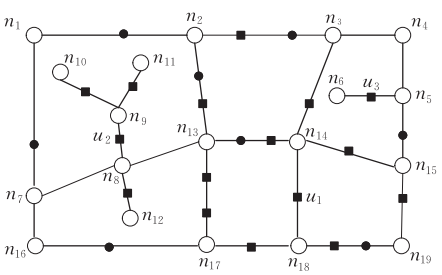


图 1 一个简单的公路网络模型

此外,为了缓解交通压力,目前已有很多城市都在实行单行化交通,即规定车辆在某些路段上只能朝一个方向行驶.国内外的很多实践也均证明了单行化交通对解决城市交通问题发挥着重要作用.那么在包含单行线的公路网络(复杂公路网络)中如何保护移动用户的位置隐私?这是一个新挑战.

挑战. 现实生活中的公路网络有其自身结构特点;位置隐私保护的效能与基于位置服务的质量之间也存在着一种权衡.如何挖掘公路网络中潜在的结构特点以保护移动用户的位置隐私安全并为他们提供高质量的位置服务,这将是一个很大的挑战.

贡献. 通过观察简单和复杂公路网络的结构特点,定义了两种隐匿子图结构,即隐匿环和隐匿树(移动用户分布和位置隐私都满足一定要求的环和树).基于这两种隐匿子图结构,提出了一种新的位置隐私保护方法——隐匿环与森林(Cloaking Cycle and Forest,CCF).CCF 不仅能够有效地保护移动用户的位置隐私安全,而且还能够给移动用户提供高质量的位置服务.此外,首次考虑了包含单行线的公路网络,并使用 CCF 成功解决了这种复杂公路网络环境下的位置隐私保护问题.

本文第 2 节回顾位置隐私保护技术的相关工作;第 3 节介绍一些背景知识并给出要解决问题的正式定义;第 4 节和第 5 节分别介绍简单公路网络模型中的隐匿环构造和隐匿树构造;第 6 节介绍复杂公路网络中的位置隐私保护,并给出完整的位置隐私保护方法 CCF;系统的实验结果与分析将在第 7 节中介绍;第 8 节总结全文.

2 相关工作

近年来,关于如何保护移动用户的位置隐私已经提出了很多方法.这些方法可以大体分为空间区域隐匿[1,3-9]和假位置[10-11]两种技术.大多数采用空间区域隐匿技术的文章都使用了位置 k-匿名模型.该模型是在文献[1]中被首次提出的,位置 k-匿名是指当一个人的位置信息至少和其它 $k-1$ 个人的

① 路段是指一个前后相继的边序列,除了两个端点的度等于 1 或者大于 2,其余顶点的度都等于 2.

位置信息不可区分时，那么这个人的位置就满足了位置 k-匿名. 文献[1]中还提出了一个基于四叉树的 Interval-Cloak 算法. 由于该文献中的 k 值是系统设定的，不符合位置隐私个性化需求的特点，所以文献[3]提出了用户可以定制的 k-匿名模型，允许用户自己指定匿名程度的高低，同时文献提出了 Clique-Cloak 算法. 文献[4]针对文献[3]中匿名成功率低的缺点，提出了一种改进的基于有向图的隐匿算法. 文献[5]利用完全金字塔数据结构以及不完全金字塔数据结构来维护移动用户的位置信息，并基于这两种数据结构分别提出了基本的和自适应的隐匿算法. 文献[6]则提出了动态自底向上和动态自顶向下的格隐匿算法. 文献[7]分别提出了 Nearest-Neighbor-Cloak 算法和 Hilbert-Cloak 算法. 以上文献中采用的系统结构都是中心服务器结构. 而文献[8-9]则采用的是分布式点对点结构. 文献[8]中，移动用户在向位置服务器发送查询之前，先通过向其它对等节点发送成组请求来组成一个空间区域，然后再将此区域和查询一同发送给服务器. 由于该方法在很多情况下都会匿名失败，所以文献[9]提出了一个基于 Hilbert 填充曲线的高级 k-匿名空间范围构建机制，提升了系统的匿名成功率.

文献[10-11]采用了假位置技术. 文献[10]中，移动用户自行生成一些假位置，然后将自己的真实位置和生成的假位置一同发送给服务器，由于攻击者辨别不出用户的真实位置，所以用户的位置隐私得到了保护. 在文献[11]中，移动用户只发送一个他指定的假位置，然后服务器根据这个假位置进行增量最近邻查询，并将查询结果返回给用户，用户再根据返回的结果检索出自己想要的答案. 因为攻击者得到的并不是用户的真实位置，所以用户的位置信息也得到了保护.

以上提到的工作中都假设用户在一个自由空间中移动. 但是在现实生活中，人们经常是沿着公路网络行进. 文献[2]首次注意到了这个问题并提出了 XStar 的位置隐私保护模型. 但是文献[2]中只考虑了简单公路网络(所有道路都是双行线)环境下的位置隐私保护，而本文则是在同时考虑简单和复杂公路网络(包含单行线的情况)的基础之上，提出了一种新的基于隐匿子图的位置隐私保护方法.

3　背景知识与问题定义

本节将介绍与本文工作密切相关的公路网络模型、位置匿名系统结构、移动用户位置隐私以及问题定义.

3.1　简单公路网络模型

使用无向图 $G=(V,E)$ 来表示一个简单的公路网络. 例如，图 1 显示的是一个简单的公路网络模型. 模型中的每条边可以看作是一条双行线公路. 度为 1 的顶点可以看作是公路的尽头；度为 2 的顶点可以看作是路弯；度大于等于 3 的顶点可以看作是公路的交叉口. 此外模型中还用小正方形来表示移动用户，用小圆点来表示移动用户感兴趣的地点，例如商店、加油站、旅馆等等.

3.2　位置匿名系统结构

采用中心服务器结构[12]，即在移动客户端和位置服务器端之间增加一个可信的服务器，通常称作位置匿名器. 如图 2 所示，位置匿名器将对用户的精确位置信息进行匿名处理，同时对位置服务器端返回的候选结果进行求精处理. 本文的主要工作就是为位置匿名器设计一个有效的位置匿名算法.

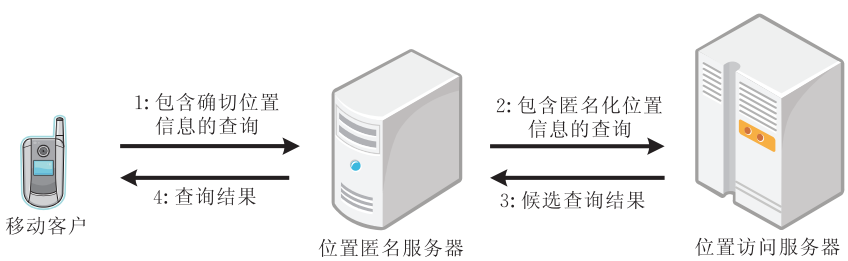


图 2　位置匿名体系结构

3.3　移动用户位置隐私

目前，已经有两种位置隐私保护模型被提出，一种是位置 k-匿名，一种是路段 l-多样性，本文结合了这两种模型. 路段 l-多样性由 Wang 在文献[2]中提出的. 它是指如果一个用户的位置信息满足位置 k-匿名，并且在匿名位置中至少包含了 l 条不同的路段，那么这个匿名位置就满足路段 l-多样性. 试想如果一个匿名位置只包含一条路段，那么攻击者将

会很容易跟踪沿这条路段行进的人.相反,如果有三条或者更多条路段出现在此匿名位置中,那么攻击者的跟踪难度就会增大.所以,路段 l-多样性是公路网络中保护移动用户位置隐私的一个必不可少的条件.

k 和 l 都是衡量位置隐私保护强弱的参数,本文还给出另外一个参数,即 $l_{\max}$.$l_{\max}$ 表示一个匿名位置中所包含路段数的上限.虽然较多的路段数能够给用户提供较强的位置隐私保护,但是这会导致位置服务器端较高的查询处理代价,并且给用户提供的服务质量下降.为了达到隐私保护、查询处理代价以及服务质量三者间的平衡,提出 $l_{\max}$ 路段数量约束参数.

定义 1(位置隐私). 以 $(k,l,l_{\max})$ 来表示一个移动用户的位置隐私.其中 k 表示匿名位置中至少包含的移动用户数量,$l(l_{\max})$ 表示匿名位置中至少(至多)包含的路段数量.用户 u 的位置隐私 $(k,l,l_{\max})$ 可以简单地表示成 $pp(u)=(k,l,l_{\max})$.

一个用户的位置隐私由他自己决定,并且该位置隐私会随着用户的精确位置信息以及查询一同发送给位置匿名器.

3.4 问题定义

首先对攻击者的背景知识做一些假设:假设攻击者预先知道位置隐私保护算法,并且能够获知公路网络上每条路段上移动用户的数量.基于此,给出本文的问题定义.

问题. 假设有一个请求 LBS 的移动用户 u,他的位置隐私为 $(k,l,l_{\max})$.那么位置匿名器应如何为 u 找到一个隐匿路段集 S(S 中包含了 l' 条路段,k' 个移动对象),使得 S 不仅覆盖 u 的精确位置,而且要满足 $S.k'\geq k$,$l\leq S.l'\leq l_{\max}$,此外,攻击者不能够以很高的概率(>0.5)从 S 中推断出 u 的具体位置.

那么攻击者在窃取到一个隐匿路段集 S 后,他是如何对用户 u 的具体位置做出推断呢?通常情况下,攻击者会对 S 中的每条路段 $s_i(i=1,2,\cdots,l')$ 都依次假设 u 位于该路段上,接着他会对该路段执行位置隐私保护算法,并得到一个包含该路段的隐匿路段集 S_{ai},然后他将 S_{ai} 与 S 进行比较,得出这两个集合中相同路段的个数与 S 中所有路段个数的比值 r_i,最后攻击者推断 u 在路段 s_i 上的概率就是 r_i 与 $r_1+r_2+\cdots+r_{l'}$ 之间的比值[2].

4 简单公路网络下的隐匿环构造

本节首先分析简单公路网络中环结构所具有的保护性,然后介绍为用户寻找最优隐匿环的 3 个步骤.

4.1 隐匿环

通过观察简单公路网络模型的结构特点发现,环因其对称性而能够保护公路网络上移动用户的位置隐私.因为对于一个环,攻击者无论对该环上的哪条边执行找环算法,他都会得到和该环相同的一个环,进而他推断出用户在该环上每条边上的概率都相等,所以,攻击者无法确定出用户所在的边(具体位置),那么用户的位置隐私也将会得到很好的保护.

根据对攻击者背景知识的假设,攻击者是能够得知每条路段上移动用户的数量,所以攻击者很有可能还会利用这点去排除用户不可能存在的路段.所以,经上述分析,定义出了一种能够保护移动用户位置隐私的隐匿环.

定义 2(隐匿环). 无向图中的一个环,它所包含的移动用户数量和路段数量都满足用户的位置隐私,并且该环中至少有两条路段上有移动用户存在.

在无向图中,覆盖一个用户位置的环可能有多个.例如,在图 1 中,覆盖 u_1 的环除了〈$n_{13},n_{14},n_{18},n_{17},n_{13}$〉之外,还有环〈$n_{14},n_{15},n_{19},n_{18},n_{14}$〉、环〈$n_{14},n_3,n_4,n_5,n_{15},n_{19},n_{18},n_{14}$〉等.所以在为用户找到了覆盖他的环之后,还要接着判断这些环是否是隐匿环.如果有多个环都符合隐匿环的条件,那么选择移动用户数量和路段数量均最接近用户位置隐私的环(称这种环为最优隐匿环)来作为用户的隐匿位置.这是因为在所有的隐匿环中,移动用户数量和路段数量相对较多的环会给位置服务器端造成较高的查询处理代价,同时也会使给移动客户端提供服务的质量下降.所以选择最优隐匿环是很有必要的一步.

为了能够快速找到最优隐匿环,先找到覆盖用户位置的最小环,然后判断这些环是否满足隐匿环的条件,如果只有一个环满足条件,那么此环就是所要寻找的最优隐匿环;如果有多个环同时满足条件,那么就从中选出一个最优隐匿环;如果所有环都不满足条件,那么就对那些移动用户数量和(或)路段数量小于用户位置隐私的环先进行扩展,然后再判断扩展后的环是否满足隐匿环条件,最后从符合条件的扩展环中选择出最优的隐匿环.

4.2 发现最小环

利用无向图中宽度优先搜索的方法为用户发现最小环.为了使最小环能够覆盖用户所在位置,以用

户所在边的两个端点分别作为搜索最小环的起始点和目标点. 所以在搜索开始之前，先根据用户所在的边指定出搜索的起始点和目标点(因为在无向图中，无论指定哪个端点为起始点，最后为用户发现的最小环都是一样的，所以起始点可以是两端点中的任意一个. 如果起始点选定了，那么另外一个端点就是目标点)，然后再从起始点开始进行宽度优先搜索，当搜索到的顶点是目标点时，覆盖用户位置的最小环就被发现了，接着再利用递归的方法从访问过的边中找出构成该最小环的各条边. 发现最小环的具体细节见算法 1.

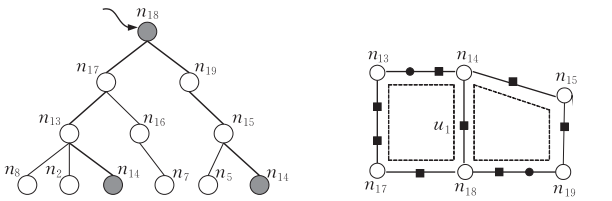
算法 1.　发现最小环.

输入：无向图 G；用户 id；访问顶点队列 $\boldsymbol{Q}$；访问边向量 $\boldsymbol{V}$

输出：构成最小环的各条边

1. 根据用户 id 定位用户所在的边；
2. 取得边上的相关信息 ue_{info}；
3. 指定搜索的起始点 n_{start} 和目标点 n_{stop}；
4. 置 G 中所有顶点和边为未被访问；
5. 访问顶点 n_{start}、n_{stop} 以及边 $n_{\text{start}}n_{\text{stop}}$；
6. n_{start} 入队列 Q；
7. while 队列 Q 不空
8.　　对头元素出队列，并赋值给变量 u；
9.　　for each u 的邻接顶点 w
10.　　if w 未被访问
11.　　访问 w，并把 w 插入队列 Q 中；
12.　　访问边 uw，得到边 uw 上的信息 e_{info}，并保存 (u,w,e_{info}) 至向量 $\boldsymbol{V}$ 中；
13.　　else
14.　　if 边 uw 未被访问
15.　　访问边 uw，得到边 uw 上的信息 e_{info}，并保存 (u,w,e_{info}) 至向量 $\boldsymbol{V}$ 中；
16.　　end if
17.　　end if
18.　　if $w=n_{\text{stop}}$ 并且 $u\neq n_{\text{start}}$
19.　　将 $(n_{\text{stop}},n_{\text{start}},ue_{\text{info}})$ 插入向量 $\boldsymbol{V}$ 中；
20.　　利用递归的方法找到构成该最小环的各条边，并输出；
21.　　end if
22.　end for
23. end while

下面以为用户 u_1 发现最小环为例来说明发现最小环的过程. 首先，因为 u_1 所在边是 $n_{14}n_{18}$，所以指定 n_{18} 为起始点，n_{14} 为目标点(在图 3(a)中，用灰色实心圆圈来表示搜索的起始点和目标点). 然后将无向图图 G 上所有顶点和边都置为未被访问，并将顶点 n_{14}、n_{18} 以及边 $n_{14}n_{18}$ 都置为已被访问. 接着搜索从 n_{18} 开始，首先访问 n_{18} 未被访问的邻接顶点 n_{17}、n_{19} 以及邻接边 $n_{18}n_{17}$、$n_{18}n_{19}$，同时保存边 $n_{18}n_{17}$、$n_{18}n_{19}$ 的信息，同理，n_{17}、n_{19} 的未被访问的邻接顶点 n_{13}、n_{16}、n_{15} 和邻接边 $n_{17}n_{13}$、$n_{17}n_{16}$、$n_{19}n_{15}$ 也会被依次访问，当访问到的邻接顶点是 n_{14} 时，最小环就被发现了. 最后，本例中为 u_1 发现了两个最小环，即 $\langle n_{18},n_{17},n_{13},n_{14},n_{18}\rangle$ 和 $\langle n_{18},n_{19},n_{15},n_{14},n_{18}\rangle$. 图 3(a)中用粗实线表示从起始点 n_{18} 到目标点 n_{14} 所访问过的路径. 同时，图 3(b)中用虚线显示了在简单公路网络模型中为 u_1 发现的两个最小环.



(a) 为 u_1 发现最小环　　(b) 简单公路网络中的最小环

图 3　利用宽度优先搜索为 u_1 发现最小环示意

4.3　选择最优隐匿环

在对每一个最小环上的移动用户数量和路段数量进行统计之后，对于那些符合隐匿环条件的最小环，利用式(1)来衡量每一个隐匿环与用户位置隐私的接近程度，也就是为每一个隐匿环进行打分，得分越高就说明越接近用户的位置隐私.

$$score_c = \beta_k.k_p/k_c + \beta_l.l_p/l_c \tag{1}$$

在式(1)中，β_k 和 β_l 是两个权重系数，它们的取值都在 0 到 1 之间，并且 $\beta_k+\beta_l=1$. k_p 和 l_p 代表了用户位置隐私中要求的移动用户数量和路段数量，而 k_c 和 l_c 代表了隐匿环上包含的移动用户量和路段数量. 当 $k_p=k_c$，$l_p=l_c$ 时，$score_c$ 达到最大值 1. 理论上分析，路段数量的多少对查询处理代价有较大的影响，因此更看重 l_c 与 l_p 的接近程度. 所以实验中设置 $\beta_l=0.6$，$\beta_k=0.4$.

4.4　扩展最小环

扩展最小环是针对那些环上移动用户数量、路段数量都小于(其中一个小于)用户位置隐私以及(或者)环上有移动用户的路段数小于 2 的最小环. 采用类似于在最小环上起跑的方法. 首先保存最小环上顶点度数大于 2 的顶点；然后从这些顶点中选出搜索的起始点和目标点(可能是多对)；接着再基于每一对起始点和目标点搜索最小环的扩展部分.

例如，对于 u_1 的最小环 $\langle n_{18},n_{17},n_{13},n_{14},n_{18}\rangle$，如果选择 (n_{14},n_{13}) 为起始点和目标点，那么可以得到环 $\langle n_{14},n_3,n_2,n_{13},n_{17},n_{18},n_{14}\rangle$，如果选择 (n_{13},n_{17})，则可以得到环 $\langle n_{14},n_{13},n_8,n_7,n_{16},n_{17},n_{18},n_{14}\rangle$. 为

了保证扩展后的环依然能覆盖用户的位置，不能再次选择(n_{14}，n_{18})为搜索的起始点和目标点.

5 简单公路网络下的隐匿树构造

本节首先定义出与隐匿环互补的子图隐匿树，然后介绍如何为用户寻找隐匿树的两个步骤.

5.1 隐匿树

隐匿环虽然能够保护移动用户的位置隐私，但是并不是为所有的用户都能够发现覆盖他们位置的最小环.

例如，图 4 显示的是为用户 u_2 发现最小环的过程. 因为 u_2 所在边是 n_8n_9，所以指定搜索的起始点为 n_9，目标点为 n_8. 算法从 n_9 开始搜索，并期望搜索到顶点 n_8，但是很遗憾，直到搜索结束也没有发现 n_8，也就是说在公路网络中没有覆盖 u_2 位置的环.

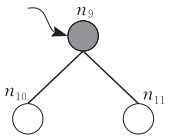


图 4 为用户 u_2 发现最小环示意

通过观察没有发现最小环的子图(例如，图 4 中的子图)，发现它们其实都是自由树，所以联想到可否用自由树来作为用户的隐匿位置. 下面就逐渐定义出能够保护移动用户位置隐私的自由树.

定义 3(树边). 无向图中不被任何环覆盖的边.

定义 4(边界树). 无向图中一个仅由树边构成的自由树.

例如，在图 4 中显示的就是一个边界树. 因为边 n_9n_{10}、n_9n_{11} 都是树边.

定义 5(相对最大边界树——RMBT). 无向图中的一个边界树，如果再向它添加一条边，那么它就不是边界树了.

例如，在图 1 中，由边 n_9n_{10}、n_9n_{11}、n_9n_8、n_8n_{12} 所构成的边界树就是一个相对最大边界树.

因为对于每一条构成一个相对最大边界树的边来说，它们的相对最大边界树都是唯一且相同的，攻击者无论对哪条边执行寻找相对最大边界树算法，他都会得到同样的一个相对最大边界树，所以，在攻击者还不能够知道公路网络上每条路段上移动用户数量的情况下，他会得出用户在每条边上的可能性都有.

那么在攻击者能够得知每条路段上移动用户数量的情况下，如何保护移动用户的位置隐私呢？下面给出隐匿树的定义.

定义 6(隐匿树). 一个相对最大边界树，它所包含的移动对象数量和路段数量都满足用户的位置隐私，并且该树中至少有两条路段上有移动用户存在.

根据隐匿树的定义，要为用户寻找覆盖他位置的隐匿树，首先应该找到覆盖他位置的相对最大边界树，然后对此相对最大边界树进行判断，看它是否满足隐匿树的条件，如果满足，那么就以此相对最大边界树作为用户的隐匿树，如果不满足，则将多个相对最大边界树进行组合，即通过构造隐匿森林，来达到满足用户位置隐私和移动用户在路段上分布的要求.

5.2 寻找相对最大边界树

寻找相对最大边界树的过程其实就是在宽度优先搜索的过程中逐渐搜索树边的过程. 因为用户所在的边本身就是一条树边，所以首先保存用户所在的边的信息，然后访问该边的两个顶点，并将它们先后插入到队列中. 对于出队列的每一个顶点，先判断该顶点与它未被访问的邻接顶点之间的边是否是树边，如果是树边，那么就访问此邻接顶点，并将此邻接顶点插入队列(表明此路径还要继续搜索下去)，然后保存该树边的信息；如果不是树边，那么就不对该顶点的邻接顶点做任何操作(表明这条路径的搜索就到此为止). 重复以上步骤，直到队列为空停止. 算法 2 给出了为用户寻找相对最大边界树的具体细节.

算法 2. 寻找相对最大边界树.

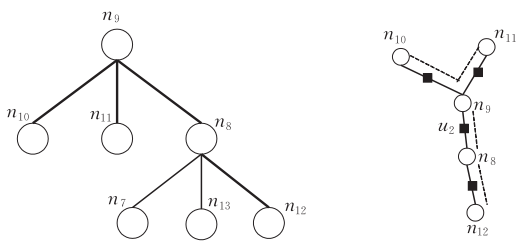
输入：无向图 G；用户 id；访问顶点队列 $\boldsymbol{Q}$；保存树边向量 $\boldsymbol{V}$

输出：构成相对最大边界树的各条边

1. 根据用户 id 得到用户所在的边 n_in_j；
2. 取得边 n_in_j 的相关信息 ue_{info}；
3. 将(n_i，n_j，ue_{info})存入向量 $\boldsymbol{V}$ 中；
4. 把无向图 G 上的所有顶点都置为未被访问；
5. 访问顶点 n_i、n_j；
6. n_i、n_j 入队列 $\boldsymbol{Q}$；
7. while 队列 $\boldsymbol{Q}$ 不空
8. 对头元素出队列，并赋值给 u；
9. for each u 的邻接顶点 w
10. if w 没有被访问
11. 访问 w；
12. 判断边(u，w)是否是树边；
13. if 边(u，w)是树边
14. w 入队列；
15. 取得到边(u，w)上的相关信息 e_{info}；

16.　　将(u,w,e_{info})存入向量中；
17.　　end if
18.　　end if
19.　end for
20. end while

下面以为用户 u_2 寻找相对最大边界树为例来说明上述过程. 因为 u_2 所在的边是 n_9n_8，所以先保存边 n_9n_8 的信息，然后访问顶点 n_9、n_8，并将它们先后插入队列中. 随后 n_9 出队列，宽度优先搜索就从 n_9 开始. 如图 5(a)所示，n_9 未被访问的邻接顶点有 n_{10}、n_{11}，由于边 n_9n_{10}、n_9n_{11} 都是树边，所以将 n_{10}、n_{11} 置为已被访问之后，就将它们也先后插入到队列中去，并保存边 n_9n_{10}、n_9n_{11} 的信息，图中用粗实线来表示保存的边. 接着 n_8 出队列，由于边 n_8n_7、n_8n_{13} 不是树边，所以不对顶点 n_7、n_{13} 再做任何操作. 而边 n_8n_{12} 是树边，所以访问 n_{12} 之后，就将其插入队列，并保存边 n_8n_{12} 的信息. 因为顶点 n_{10}、n_{11} 以及 n_{12} 已再没有未被访问的邻接顶点了，所以算法到它们都出队列之后就终止了. 最后，为用户 u_2 找到的相对最大边界树就是由边 n_9n_{10}、n_9n_{11}、n_9n_8、n_8n_{12} 所构成的子图. 图 5(b)中用虚线显示了在简单公路网络模型中为 u_2 找到的相对最大边界树.



(a) 为u_2寻找RMBT　　(b) 简单公路网络模型中的RMBT

图 5　为 u_2 寻找相对最大边界树

5.3　构造隐匿森林

为了减少位置服务器端的查询处理代价，要求构造的隐匿森林中所包含的移动用户数量和路段数量要尽可能最接近用户的位置隐私. 那么如何达到这个要求？本文首先保存路段数分别为 1、3、5 的具有低查询处理代价的相对最大边界树信息，并将每种路段数的相对最大边界树信息按照树中所包含移动用户数量的多少从小到大进行排序. 如果是由于移动用户数量少的原因而去构造隐匿森林，那么先计算出移动用户数量的差值 dif_k，然后从路段数为 1 的相对最大边界树信息中选择一个移动用户数量最接近 dif_k 的相对最大边界树；如果是由于路段数量少的原因而去构造隐匿森林，那么先计算出路段数量的差值 dif_l，然后根据 dif_l 计算出所需各路段数的相对最大边界树的个数，再根据这些个数顺序选取每种相对最大边界树；如果是由于移动用户数量和路段数量都小于位置隐私而去构造隐匿森林，那么除了根据 dif_l 计算出所需各路段数的相对最大边界树的个数之外，还要选出移动用户数量最接近于 dif_k 的相对最大边界树的组合.

6　复杂公路网络下的位置隐私保护

前面两节主要介绍了如何通过在无向图中构造隐匿环和隐匿树的方法来保护简单公路网络模型中移动用户的位置隐私安全，那么本节将要介绍在复杂公路网络模型中的位置隐私保护问题. 完整的位置隐私保护算法在本节最后给出.

在现实生活中的公路网络中，单行路也是一种很常见的道路. 它在缓解城市交通拥挤、减少交叉口的冲突点、提高车辆运行速度等方面起着非常大的作用. 目前，全世界已有很多城市在实行单行化交通，例如纽约、伦敦、新加坡等等.

对于包含单行线的公路网络，可以将其抽象为一个有向图. 例如，图 6 显示的就是一个复杂的公路网络模型. 在该模型中，有 4 条单行线，即弧 $n_{17}n_{13}$、$n_{13}n_2$、n_3n_{14} 以及 $n_{14}n_{18}$. 那么在复杂的公路网络模型中，如何保护移动用户的位置隐私安全，经研究发现，隐匿环与隐匿树的子图结构依然能够给用户提供位置隐私保护.

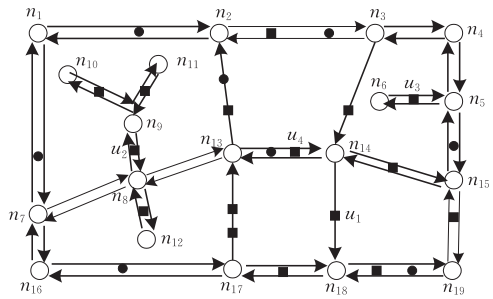


图 6　一个复杂的公路网络模型

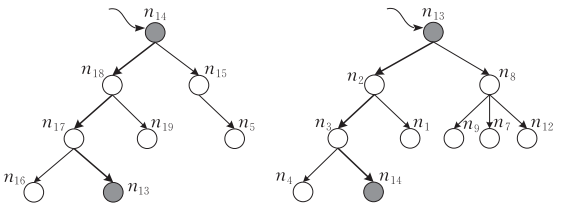
6.1　有向隐匿环构造

在有向图中，顶点与顶点之间由弧相连接，所以有相图中的环都是有方向的(顺时针方向或者逆时针方向)，但是这并不妨碍它们所具有的保护性. 因为对于一个有向环，攻击者无论对该环中的哪条弧执行找环算法，他都会得到和此环相同的一个有向环，所以，如果一个有向环满足用户的位置隐私，并且它至少有两条弧上有移动用户存在，那么这个有

向环就能够保护用户的位置隐私，将这种环称为有向隐匿环.

为用户寻找一个最优的有向隐匿环同样也分为发现最小有向环、选择最优的有向隐匿环和扩展最小有向环三步.其中选择最优的有向隐匿环与无向图中选择最优隐匿环的方法一样，而其它两步与无向图中的稍有差别，因此下面就主要介绍发现最小有向环和扩展最小有向环.

对于发现最小有向环，如果用户位于双行线上，那么要为用户指定两次起始点，并进行两次搜索.因为在有向图中，基于不同的起始点得到的有向环可能会不一样.例如，图 7(a)、(b)中显示的就是在基于不同起始点的情况下为用户 u_4 发现的两个最小有向环，即环〈n_{14}，n_{18}，n_{17}，n_{13}，n_{14}〉和〈n_{13}，n_2，n_3，n_{14}，n_{13}〉.



(a)〈n_{14}，n_{13}〉 (b)〈n_{13}，n_{14}〉

图 7 基于不同的起始点为用户 u_4 发现最小有向环示意

对于扩展最小有向环，因为在有向图中，一个顶点的度有出度和入度之分，所以在选择搜索扩展环的起始点和目标点时和无向图中的会有所不同.首先将最小有向环上各顶点的出度和入度分别减 1，然后保存剩余出度大于 0 的顶点和剩余入度大于 0 的顶点，再从这两类顶点中挑选出搜索的起始点和目标点.

例如，在图 6 中，假设覆盖 u_1 的最小有向环是〈n_{18}，n_{17}，n_{13}，n_{14}，n_{18}〉，将环上各顶点的出度和入度分别减 1 之后，得到顶点 n_{18}、n_{17}、n_{13}、n_{14}的出度和入度都大于 0.如果选择〈n_{13}，n_{14}〉为搜索的起始点和目标点，那么将得到扩展有向环〈n_{18}，n_{17}，n_{13}，n_2，n_3，n_{14}，n_{18}〉；如果选〈n_{17}，n_{13}〉，那么将得到扩展有向环〈n_{18}，n_{17}，n_{16}，n_7，n_8，n_{13}，n_{14}，n_{18}〉.

6.2 有向图中的隐匿树

在有向图中，也会碰到为用户找不到最小环的情况，但是发生这种情况的用户一般都位于双行线上.因为公路网络中单行线的布局要符合单行互补理论，即在一条单行线的旁边必须要设计一个与之方向相反的单行线，而且它们要尽可能相近、相似、相平行.单行互补的理论保证了用户从某条单行线出发之后，他还能够通过其它线路而又重新回到此单行线上.因此，如果用户位于单行线上，肯定能够发现覆盖他位置的有向环.既然用户只有在双行线上时才会发生找不到最小有向环的情况，那么就可以将这类双行线看作是树边.要保护位于树边上用户的位置隐私，就可以为用户构造隐匿树或者隐匿森林.

例如，对于图 6 中的用户 u_2，由于发现以〈n_8，n_9〉为起始点和目标点时没有发现最小环，而且以〈n_9，n_8〉为起始点和目标点时也没有发现最小环，所以确定 n_8n_9 为树边.那么如何为 u_2 寻找相对最大边界树呢？如图 8 所示，先访问顶点 n_9、n_8，并将它们插入队列中，然后宽度优先搜索从顶点 n_9 开始.首先得到 n_9 未被访问的邻接顶点 n_{10}，由于弧 n_9n_{10} 不被任何有向环覆盖，因此接着判断弧 $n_{10}n_9$ 是否也不被任何环覆盖，如果是，那么 n_9n_{10} 就是树边，然后访问顶点 n_{10}，将其插入队列，并保存树边 n_9n_{10} 的信息.以同样的方法，判断出 n_9n_{11}、n_8n_{12} 也都是树边，所以为用户 u_2 找到的相对最大边界树就是由树边 n_9n_{10}、n_9n_{11}、n_9n_8、n_8n_{12} 所构成的子图.

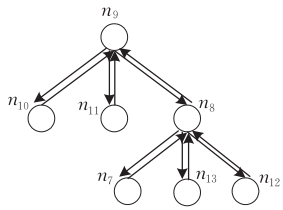


图 8 复杂公路网络下为 u_2 寻找 RMBT 示意

如果为用户找到的相对最大边界树所包含的移动用户数量、路段数量小于用户位置隐私并且(或者)不满足移动用户在路段上的分布要求，那么同样也是为用户去构造隐匿森林，构造的方法与无向图中构造隐匿森林的方法基本相同，这里就不再介绍了.

6.3 隐匿环与森林保护方法

当位置匿名器收到一个用户的匿名请求消息时，首先试图为用户构造隐匿环，如果发现了最小环，就接着执行选择最优隐匿环、扩展最小环的算法；如果没有发现最小环，那么位置匿名器将为用户构造隐匿树或隐匿森林.算法 3 描述了具体细节.

算法 3. 隐匿环与森林——CCF.

输入：无向图 G；用户 id；$pp(u)$

输出：隐匿路段集

1. 为用户 id 寻找最小环(使用算法 1)；
2. if 找到最小环
3. 选择最优隐匿环；
4. if 选到了最优隐匿环

```
5.      返回此最优环作为用户的隐匿位置；
6.    else
7.      if 存在可以被扩大的环
8.        扩展最小环；
9.        选择最优隐匿环；
10.     end if
11.   end if
12. else
13.   为用户 id 寻找相对最大边界树(使用算法 2)；
14.   判断相对最大边界树是否是隐匿树；
15.   if 是隐匿树
16.     返回此隐匿树作为用户的隐匿位置；
17.   else
18.     基于 pp(u)为用户构造隐匿森林；
19.   end if
20. end if
```

7　实验及结果分析

本文算法使用 C++编程语言实现，编程环境为 Microsoft Visual C++6.0. 实验硬件环境为 1.86GHz 的 Intel 双核 CPU，1GB 内存. 操作系统平台是 Microsoft Windows XP Professional.

7.1　实验数据集和参数设置

实验数据集采用美国加州圣华金郡的公路网络数据. 该公路网络数据包括 18496 个顶点，24017 条边. 为了模拟包含单行线的公路网络，实验特别设置了 6569 条边为单行线，并且这些单行线都符合单行互补理论. 利用 Brinkhoff[13] 的基于网络的移动对象生成器生成了 10000 个移动对象. 此外还生成了 14849 个在公路网络中均匀分布的感兴趣点，包括商店和加油站. 同时，实验设置了 1000 个移动用户的最近邻查询请求消息.

表 1　请求消息中各参数的默认值

参数	平均值	方差
k	5	1
l	5	1
$l_{\max}$	20	1
k_{nnp}	5	1

用户的查询请求消息中包括了 4 个参数，即 $(k,l,l_{\max},k_{nnp})$. 其中 $k,l,l_{\max}$ 代表了用户的位置隐私，k_{nnp} 代表用户要查询的最近邻感兴趣点的个数. 在所有的查询请求中，假设这 4 个参数都服从正态分布. 表 1 给出了这 4 个参数默认的平均值和方差. 实验中，令其中一个参数取不同的平均值，而另外的 3 个参数分别取各自的默认值并且保持不变.

7.2　算法衡量标准

实验分别从平均信息熵、匿名成功率、平均匿名执行时间、相对匿名度、平均查询执行时间以及平均候选结果大小 6 个方面对提出的算法进行衡量.

(1) 平均信息熵. 攻击者通过计算会得出用户在隐匿位置中每条路段上的概率 $p_i(i=1,2,\cdots,l')$，那么给用户提供的保护强度就可以用此概率分布的信息熵来衡量[2]，其计算公式如式(2)所示：

$$H=-p_1\log_{10}p_1-p_2\log_{10}p_2-\cdots-p_{l'}\log_{10}p_{l'} \quad (2)$$

信息熵越大，攻击者猜出用户具体位置的平均不确定性就越大，进而用户所得到的保护强度也就越大.

(2) 匿名成功率，指算法成功匿名的消息数在所有移动用户发送的消息数中所占的百分比[12]. 它能够反应出位置隐私保护算法对用户查询请求的响应能力，其值越高，表明算法越好.

对于不能成功匿名的消息，本文将采用假位置[10]的方法，即在公路网络中随机生成其它 $k-1$ 个假位置，然后将用户的真实位置和这些假位置一同发送给位置服务器，由于攻击者辨别不出哪个位置是用户的真实位置，因此用户的位置信息得到了保护.

(3) 平均匿名执行时间，指位置匿名器平均对一个用户的精确位置进行匿名处理所花费的时间. 它是用来衡量一个位置隐私保护算法执行效率，其值越小，算法效率就越高.

(4) 相对匿名度，指匿名位置中包含的移动对象数量 k'(路段数量 l')与用户位置隐私中的移动对象数量 k(路段数量 l)的比值. 它可以表示成式(3).

$$RAL_k=k'/k,\ RAL_l=l'/l \quad (3)$$

当 l'、$l_{\max}$不变的情况下，RAL_k 越大越好，因为固定路段上移动对象数量越多，用户所得到的保护强度就会越大. 当 k'、$l_{\max}$不变的情况下，RAL_l 越接近于 1 越好，因为太多的路段会导致较高的查询处理代价，用户所享受服务的质量也会下降.

(5) 平均查询执行时间，指位置服务器平均对一个匿名位置进行查询时所花费的时间. 它是用来衡量位置服务器端的查询执行代价，查询花费的时间越少，服务器端的查询执行代价就越低，表明算法的性能越好. 本文实验中模拟位置服务器，对各隐匿子图进行了 kNN 查询，以测试提出算法的性能.

(6) 平均候选结果大小，指位置服务器对一个匿名位置查询之后，返回给位置服务器端的平均候选结果大小. 它是用来衡量位置服务器端与位置匿名器之间的通信代价. 平均候选结果数量越少，通信

代价就越低,表明算法的性能越好.

7.3 实验结果分析

(1)平均信息熵. 图 9(a)、(b)显示的当参数 k 和 l 取不同的平均值时,本文算法对移动用户提供的隐私保护强度.

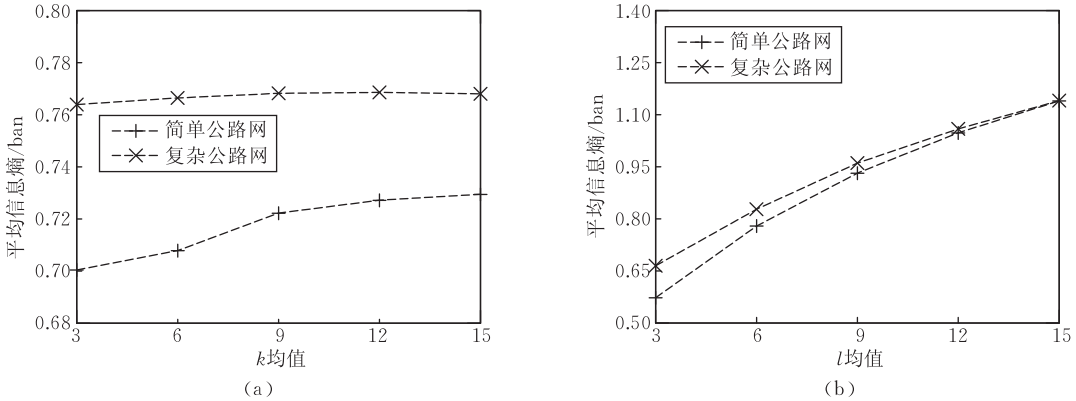


图 9 平均信息熵

从图中可以看到无论是简单公路网络还是复杂公路网络,平均信息熵都大于 0.5,而且复杂公路网络中的平均信息熵要比简单公路网络中的高,说明本算法在复杂公路网络中给用户提供的隐私保护强度要大于简单公路网络. 此外,从图 9(a)、(b)的对比中还可以看出,隐匿位置中路段数量越多,用户所得到的保护强度就越大.

(2)匿名成功率. 图 10 显示了匿名成功率在各参数不同设置情况下的变化. 从这 4 个图中可以看到简单公路网络中的匿名成功率比复杂公路网络中的要高. 这是因为加上单行线之后,某些道路在交通行进方向上有了约束,导致了公路网络中环的数量减少. 尤其是原本能够扩展的环,在加上单行线之后就不能被扩展了,从而导致匿名失败的情况比较多.

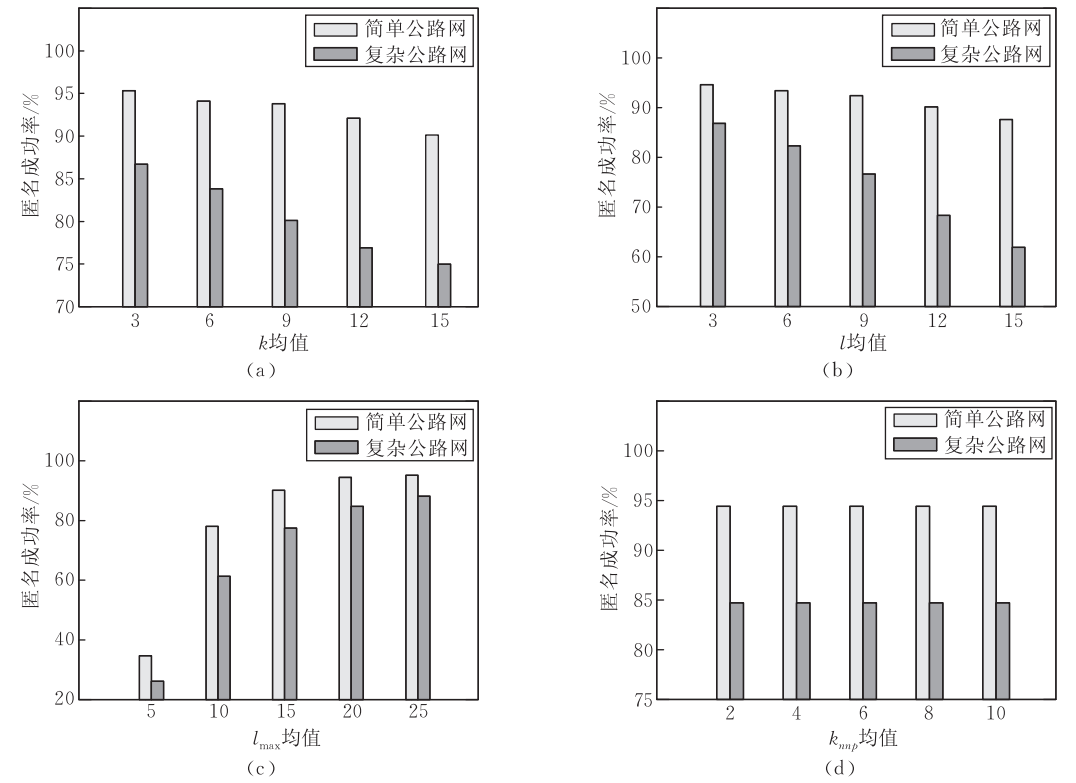


图 10 匿名成功率

在图 10(a)、(b)中,随着 k 和 l 的增大,简单公路网络中匿名成功率都略微有所下降,而复杂公路网络中匿名成功率的下降幅度较大,尤其是当 l 的平均值等于 15 时,匿名成功率不到 65%. 这说明本

算法在复杂公路网络中对过高的位置隐私(k 和 l 的平均值都大于等于 12)的保护效果不好. 图 10(c)中，随着 $l_{\max}$ 的增大，匿名成功率都呈上升趋势. 当 $l_{\max}$ 的平均值等于 5 时，匿名成功率异常的低，这可能是因为在为用户找到的隐匿位置中，路段数量大于 5 的情况很多，所以导致很多隐匿位置都因 $l'>l_{\max}$ 而导致匿名失败，可见对参数 $l_{\max}$ 的默认值设置不能太小. 从图 10(d)中可以看到当各隐私参数分别取默认值时，简单公路网络中的匿名成功率接近于 95%，而复杂公路网络中的匿名成功率接近于 85%.

(3)平均匿名执行时间. 由于发现最小环和扩展最小环都利用了宽度优先搜索的方法，所以它们的时间复杂度均为 $O(|V|+|E|)$. 对于寻找相对最大边界树的算法，因为在每一次搜索中都要判断一条边是否是树边，所以时间复杂度要比寻找隐匿环的时间复杂度高. 为了降低服务器端的查询处理代价，本文在选择最优隐匿环和隐匿森林的过程中花费了相对较多的时间.

从图 11 中可以看到随着各参数值的增大，位置匿名器花费的匿名执行时间都有所增多，尤其是在简单公路网络中. 这主要是因为当 k 和 l 的平均值都大于等于 9 时，简单公路网络中的匿名成功率比复杂公路网络中的匿名成功率高很多，所以平均匿名执行时间会花费得相对多一些.

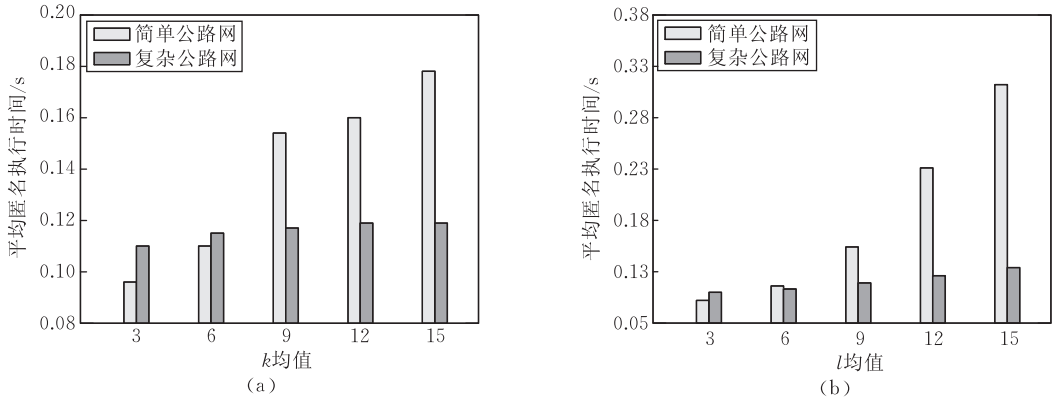


图 11　平均匿名执行时间

(4)相对匿名度. 图 12、图 13 显示的是 k 的相对匿名度和 l 的相对匿名度分别在参数 k 和 l 不同设置情况下的变化.

在图 12 中，随着 k 的增大，RAL_k 在迅速下降，而 RAL_l 略微有所升高. 这说明移动用户数量的增多，并没有引起太多路段数量的增加，因而也不会导致较高的查询处理代价.

从图 13 中可以看到，随着 l 的增大，RAL_k 在逐渐升高，而 RAL_l 在急剧下降. 说明隐匿位置中包含路段数量越多，移动用户的数量也就越多，用户所得到的保护强度也就越大.

(5)平均查询执行时间. 由于 kNN 查询是位置服务中最常见的一种查询，所以本文对各隐匿子图进行了 kNN 查询处理. 从图 14 中可以看到，只有当 l 和 k_{nnp} 增大时，平均查询执行时间是逐渐增大的，说明隐匿位置中路段数量越多，查询花费的时间就越长，而且随着用户要求查询最近邻感兴趣点数量的增多，查询所需要的时间也要增多.

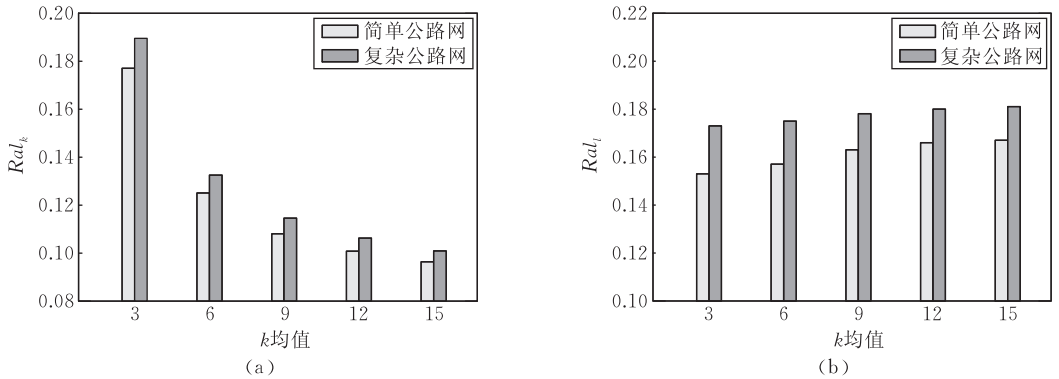


图 12　RAL_k、RAL_l 相对于参数 k 的不同设置

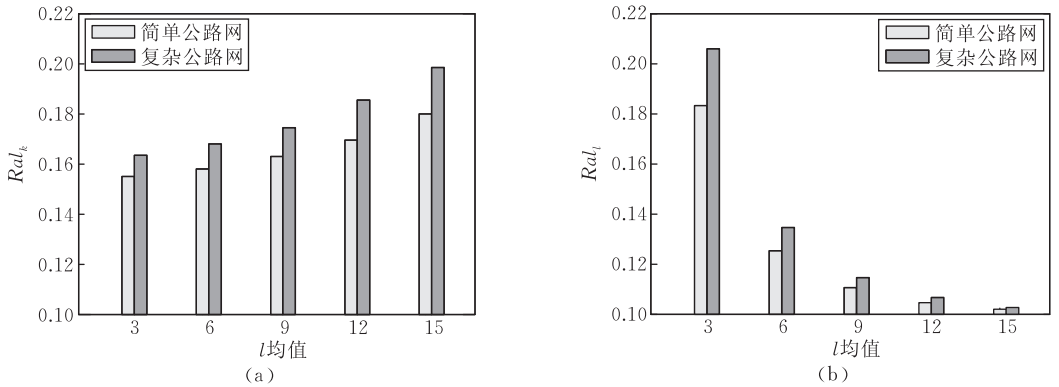


图 13 RAL_k、RAL_l相对于参数 l 的不同设置

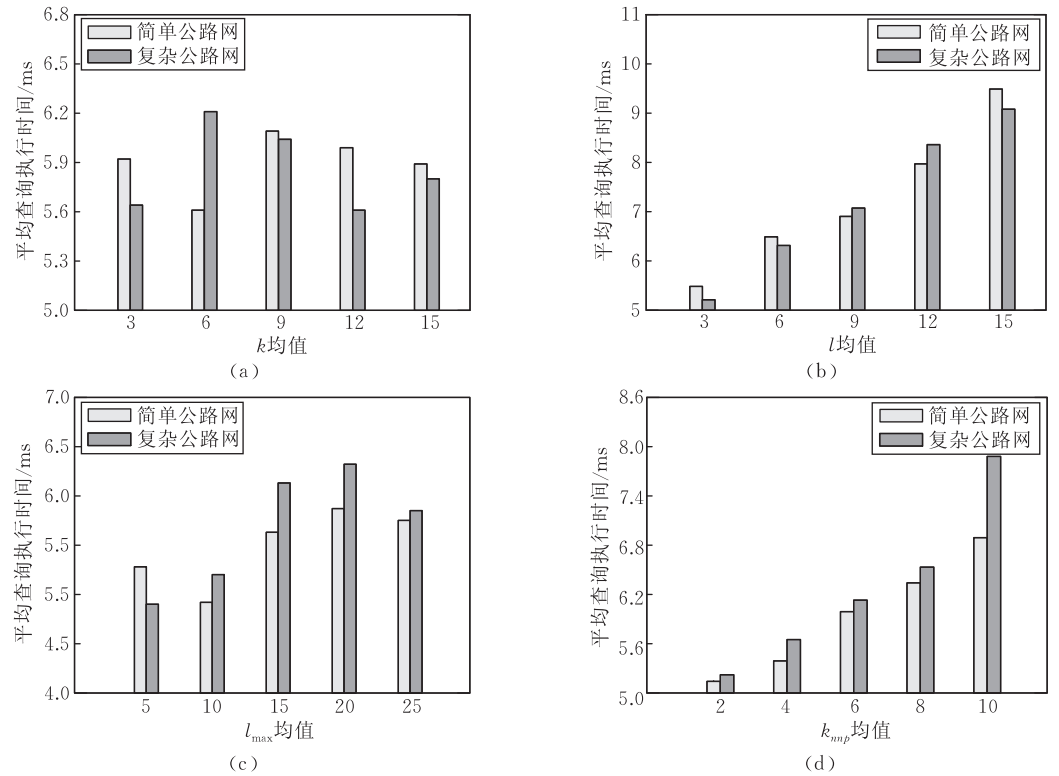


图 14 平均查询执行时间

(6)平均候选结果大小. 从图 15、图 16 中可以看到,随着各参数的增大,查询返回的候选结果数量基本都在增大. 在图 15(a)中,复杂公路网络中得到的候选结果数量随着 k 的增大没有改变,但是在图 15(b)中,候选结果数量随着 l 的增大也逐渐增大,说明在复杂公路网络中,路段数量的多少对通信代价的影响要比移动对象数量的多少对通信代价的影响大.

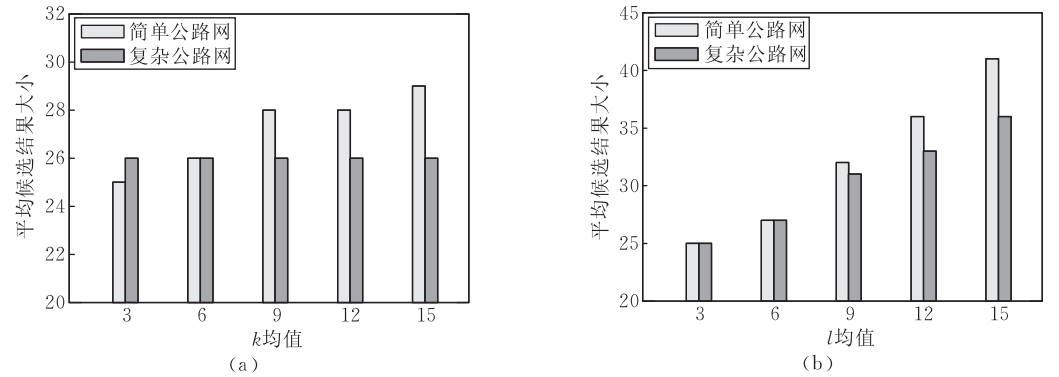


图 15 平均候选结果 1

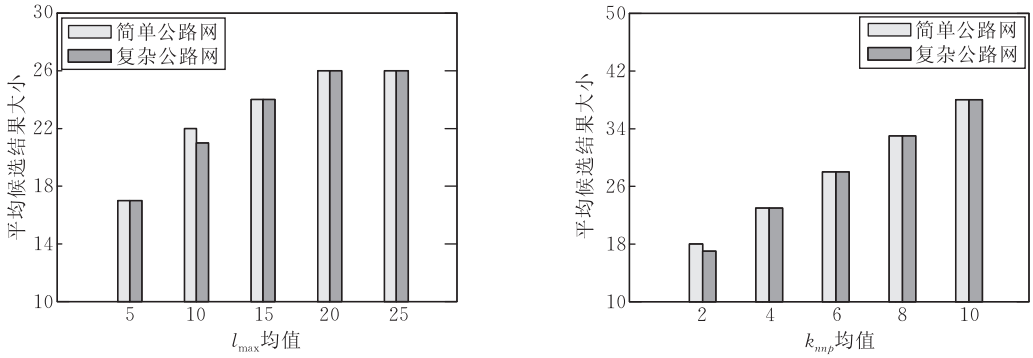


图 16　平均候选结果 2

8　结　论

以往关于位置隐私保护的研究工作很少关注公路网络环境下的位置隐私保护，尤其是包含单行线的公路网络. 通过同时观察简单公路网络和复杂公路网络的结构特点，发现了两种可以用于保护用户位置隐私的隐匿子图，即隐匿环和隐匿树. 基于这两种隐匿子图结构，提出了一个新的保护位置隐私方法—隐匿环与森林(CCF). CCF 不仅可以用于简单的公路网络中，而且它能够解决包含单行线的公路网络中的位置隐私保护问题. 在基于真实与模拟数据集的实验测试结果中，CCF 方法显示出了其在位置隐私保护方面的有效性以及在提供服务质量方面的高效性.

参　考　文　献

[1] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking//Proceedings of the first International Conference on Mobile Systems, Applications, and Services. San Francisco, CA, USA, 2003: 163-168

[2] Wang T, Liu L. Privacy-aware mobile services over road networks//Proceedings of the 35th International Conference on Very Large Data Bases. Lyon, France, 2009: 1042-1053

[3] Gedik B, Liu L. A customizable k-anonymity model for protecting location privacy//Proceedings of the International Conference on Distributed Computing Systems. Columbus, OH, USA, 2005: 620-629

[4] Xiao Z, Meng X, Xu J. Quality aware privacy protection for location-based services. Advances in Database: Concepts, Systems and Applications, 2007, 10(33): 434-446

[5] Mokbel M, Chow C, Aref W. The new Casper: Query processing for location services without compromising privacy//Proceedings of the 32nd International Conference on Very Large Data Bases. Seoul, Korea, 2006: 763-774

[6] Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with privacy grid//Proceedings of the 17th International World Wide Web Conference. Beijing, China, 2008: 237-246

[7] Kalnis P, Ghinita G, Mouratidis K. Preventing location-based identity inference in anonymous spatial queries. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(12): 1719-1733

[8] Chow C, Mokbel M, Liu X. A peer-to-peer spatial cloaking algorithm for anonymous location-based services//Proceedings of the 14th International Symposium on Advances in Geographic Information Systems. Arlington, VA, USA, 2006: 171-178

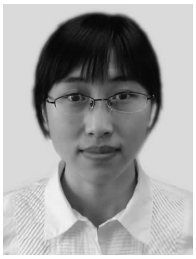
[9] Ghinita G, Kalnis P, Skiadopuslos S. PRIVE: Anonymous location-based queries in distributed mobile systems//Proceedings of the 16th International World Wide Web Conference. Banff, Alberta, Canada, 2007: 1-10

[10] Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services//Proceedings of the IEEE International Conference on Pervasive Services. Santorini, Greece, 2005: 88-97

[11] Yiu M, Jensen C, Huang X, Lu H. Spacetwist: managing the trade-offs among location privacy, query performance, and query accuracy in mobile services//Proceedings of the 24th International Conference on Data Engineering. Cancun, Mexico, 2008: 366-375

[12] Pan Xiao, Xiao Zhen, Meng Xiao-Feng. Survey of location privacy-preserving. Journal of Computer Science and Frontiers, 2007, 1(3): 268-281(in Chinese)
(潘晓，肖珍，孟小峰. 位置隐私研究综述. 计算机科学与探索，2007，1(3)：268-281)

[13] Brinkhoff T. A framework for generating network-based moving objects. GeoInfomatica, 2002, 6(2): 153-180



XUE Jiao, born in 1986, M. S.. Her research interest is location privacy.

LIU Xiang-Yu, born in 1981, Ph. D.. His main research interests include data security and data provenance.

YANG Xiao-Chun, born in 1973, professor, Ph. D. supervisor. Her main research interests include database theory and technique, and data quality.

WANG Bin, born in 1972, lecturer. His main research interests include distributed database management and system structure.

Background

Location privacy preserving problem is an important branch of data privacy and security. By linking other public information or tracking the location of a mobile user, an adversary can reidentify the mobile user and learn the user's private information. Aiming the problem, M. Gruteser and D. Grunwald first presented the location k-anonymity model to prevent linking-attack and tracking-attack. Then many works have been done to design efficient, scalable and flexible system architectures and algorithms, which balance the trade-off between privacy and service as well as possible. The system architectures of these works adopted include non-cooperative, centralized trusted and peer-to-peer cooperative. Landmark objects and false dummies are two main methods using non-cooperative architecture, where mobile users send generated false positions or the locations of their closest landmarks to the server. There are also many works adopted centralized trusted architecture, which blur the mobile user's exact location using a trusted third party (anonymizer) and send the blurred location to the server. The classic data structures that the anonymizer utilized to divide the space include quadtree and hierarchical gird index data structure. The algorithms these works proposed include Interval-Cloak, Clique-Cloak, Bottom-Up and Top-Down Dynamic Grid Cloak, Hilbert-Cloak and so on. In peer-to-peer cooperative architecture, each mobile user is a peer and they collaborate with each other to protect their own location privacy. Most works supposed that the mobile users move in the Euclidean space. But in real life, people often travel along a fixed road network. XStar model was first proposed to solve the problem of location privacy on simple road network, which only includes two-way streets. But through carefully observing, we find that one-way street is also a most common road type, it plays a significant role in alleviating traffic pressure, decreasing intersection conflict and improving the speed of vehicles. And at present, many domestic and overseas cities have implemented one-way traffic. So, aiming to this complex road network, the authors propose a novel method, called cloaking cycle and forest, which can not only provide exact location services for mobile users, but also protect their location privacy effectively.

The paper focuses on the field of location privacy preserving on road network. The research group has done much work in designing effective and efficient protecting algorithm and other techniques in data privacy, such as privacy preserving data publishing, privacy preserving data outsourcing and so on.

The work is supported by the National Natural Science Foundation of China (Nos. 60973018, 60973020, 60828004), the Fundamental Research Funds for the Central Universities (No. N090504004).