

OrientPrivacy: 移动环境下的隐私保护服务器

黄毅 潘晓 孟小峰

(中国人民大学信息学院 北京 100872)

(westyi@ruc.edu.cn)

OrientPrivacy: An Anonymizer for Privacy Preserving in Mobile Services

Huang Yi, Pan Xiao, and Meng Xiaofeng

(School of Information, Renmin University of China, Beijing 100872)

Abstract This demo presents OrientPrivacy, an anonymizer for privacy preserving in mobile services. It consists of three main components: 1) Simulated moving objects generator. It generates locations of mobile users and also accepts input of points of interested (POI) and the roadmap in which the users are. 2) Location cloaking module, in which a user's exact location is extended into a region according to the users' privacy requirements. 3) Anonymized results output module. It shows the cloaked regions on the roadmap, as well as the performance parameters for cloaking algorithms.

Key words privacy protection; location based service; mobile service

摘要 系统展示了移动计算服务中的隐私保护服务器——OrientPrivacy. 它主要由3部分组成: 1) 移动数据生成模块. 可以模拟生成移动用户的查询和位置信息, 导入用户兴趣点(POI)和用户所在城市的地图; 2) 隐私处理模块. 根据用户的隐私保护需求, 采用隐私处理算法, 将用户的精确位置转换成匿名区域, 同时将用户的敏感查询进行隐匿; 3) 匿名结果展示模块. 展示隐私处理的结果, 并展示移动用户的匿名区域和隐私服务质量参数.

关键词 隐私保护; 移动计算; 位置服务

中图法分类号 TP311.13

近年来,随着无线通信技术和移动定位技术的发展,基于位置的服务(location-based services, LBS)日益普遍. 学术界很多研究者关注如何在保证服务质量的前提下保护移动用户隐私. 总的来说,在位置服务中有2种隐私类型: 用户位置隐私^[1](保护特定用户的位置^[2])和查询内容隐私^[3](保护特定用户的查询内容^[4]). 例如: 假设张三通过手机向服务提供商(例如百度地图)发起了一个查询“离我最近的皮肤病医院在哪儿?”. 从保护位置隐私角度看,张三想隐藏他的精确位置(例如他在一个餐厅或者酒吧);

从查询内容隐私角度看,张三想隐藏他具体想查询的内容——最近的皮肤病医院.

同很多已有的工作^[5-7]一样,本系统采用中心服务器结构. 它由移动用户、可信的隐私处理服务器和不可信的位置服务商(service provider, SP)组成. 移动用户向隐私处理服务器发起格式为 (id, l, q, p) 的查询,其中 id 为用户的标识, l 为用户的位置, q 代表查询内容, p 是用户的最低隐私需求. 在收到用户查询后,隐私处理服务器将用户的标识 id 换成伪造的标识 id' ,同时根据用户的隐私需求,调用隐私

收稿日期: 2010-06-25

基金项目: 国家自然科学基金项目(60833005, 60573091); 国家“八六三”高技术研究发展计划基金项目(2007AA01Z155, 2009AA011904); 教育部博士学科点专项科研基金项目(200800020002)

保护算法为用户生成一个匿名区域 R , 然后将匿名后的查询 (id', R, q) 发给位置服务提供商. 位置服务提供商经过查询后, 将查询结果返回给隐私处理服务器. 最后, 隐私处理服务器将该结果求精后返回给用户. 在本演示系统中, 我们模拟构建了隐私处理服务器 OrientPrivacy, 模拟实现了根据用户隐私需求和场景保护移动用户隐私的目的.

1 OrientPrivacy 系统结构

OrientPrivacy 系统主要由 3 个部分组成(见图 1): 1) 移动数据生成模块. 输入移动用户的位置, POI 和用户所在城市的地图数据; 2) 隐私处理模块. 运用隐私算法将用户的精确位置 l 处理为匿名区域 R , 并且将用户的敏感查询进行隐匿; 3) 匿名结果展示模块. 它将用户的匿名区域在地图上展示出来, 同时显示匿名处理的性能参数. 本节主要关注隐私处理模块.

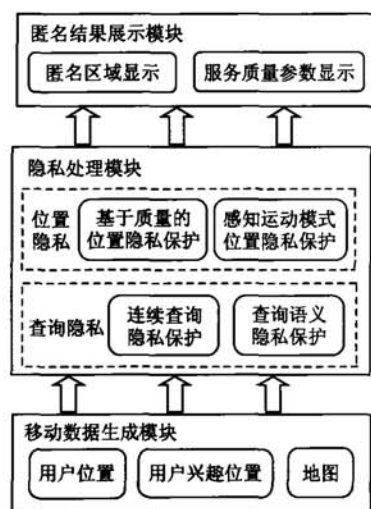


图 1 OrientPrivacy 系统结构

OrientPrivacy 将最近的工作中^[1-4]集成在一起, 包括位置隐私保护和查询隐私保护. 其中基于质量的位置隐私保护算法和感知运动模式隐私保护算法针对位置隐私, 连续查询隐私保护和查询语义隐私保护算法针对查询隐私保护. 本节剩余部分将分别介绍隐私处理模块中的关键技术.

1.1 基于质量的位置隐私保护算法

为了平衡位置隐私和服务质量二者之间的关系, 文献[2]提出了基于质量的匿名模型. 在这个算法中, 用户可以指定个性化的隐私需求(即最小匿名级别)和服务质量需求(即最大匿名区域大小). 算法

维护一个有向图, 根据这个有向图计算出用户的匿名集. 假设新到请求为 r , 有向图 r 邻居的匿名集合的出节点集和入节点集的大小分别是 k_o 和 k_i , r 的最小匿名级别为 r . k . 如果 $k_o \geq r$, $k-1$ 和 $k_i \geq r$, $k-1$ 同时成立, 那么 r 的出节点集的最小边界矩形(minimum boundary rectangle, MBR)就是 r 的匿名区域. 详细的算法请参考文献[2].

1.2 感知运动模式的位置隐私保护算法

基于质量的位置隐私保护算法可以保护移动用户的位置隐私, 但是, 它没有考虑用户的连续移动对位置隐私泄露的影响. 如果一个攻击者(例如 SP)可以收集用户一段时间的历史匿名区域并获知运动模式(例如速度), 那么他就有可能攻破用户的位置隐私. 这种攻击被称为位置依赖攻击. 现有的大多数位置 K 匿名算法只考虑了快照式的隐私处理, 并不能有效地防止位置依赖攻击. 系统中使用 ICliqueCloak 算法, 利用增量维护极大团集寻找匿名集的方法, 从极大团中直接寻找匿名集. 该算法考虑了移动对象的运动模式和连续位置更新. 详细的算法请参考文献[1].

1.3 连续查询隐私保护算法

基于质量的位置隐私保护算法和感知运动模式的位置隐私保护算法都只适用于快照式查询, 在不同的时间, 匿名集里面的用户是不一样的. 在连续查询的情形下直接应用上述算法是不足以保护查询隐私或者导致很差的服务质量. 文献[3]采用 δp 隐私模型和 δp 差异模型来衡量用户隐私需求和服务质量. 位置信息扭曲度(distortion)用匿名区域周长表示, 进而映射为两个连续查询相似度(包括初始位置、运动速度和有效期):

$$SimDis(Q_1, Q_2) = \int_{T_s}^{T_{exp}} Distortion_{R_{12}, t}(CS_{12}, RL_{12}, t) dt, \quad (1)$$

其中, CS_{12} 是由连续查询 Q_1 和 Q_2 组成的匿名集, R 是 Q_1 和 Q_2 在 t 时刻的匿名区域, T_s 是连续查询成功处理的开始时刻, T_{exp} 是连续查询失效的时间. 系统将查询根据上述公式计算出来的相似距离进行聚类, 使得一个聚集中查询位置信息扭曲度最小. 当有新查询到来或离开时, 系统增量维护这些查询组成的聚集, 并根据从这些聚集中直接寻找匿名集. 详细的算法请参考文献[3].

1.4 查询语义隐私保护算法

连续查询隐私保护算法只关注连续查询隐私,

并没有考虑查询语义,因此可能会遭受查询同质性攻击(query homogeneity attack).极端情形下,如果一个匿名集中的所有查询的内容都是相同的,那么尽管连续查询保护了位置隐私,查询内容依然会泄露.为了防止查询同质性攻击,文献[4]提出了一种新的保护模型,称为 p -敏感度模型.它在保护用户位置隐私的同时考虑了查询敏感度和查询语义信息.通过在匿名集中加入一些非敏感的查询,达到迷惑攻击者的目的. p -敏感度模型可以用下面的公式来表示:

$$P(u^* \rightarrow Q_i) = \frac{|\{u^*.S_r | u^*.S_r \in Q_i\}|}{|u^*.S_r|} < p, \quad (2)$$

其中, $|u^*.S_r|$ ($|\{u^*.S_r | u^*.S_r \in Q_i\}|$) 表示匿名集中(敏感)查询的个数.攻击者获知用户提出敏感查询内容的概率不会超过用户指定的概率 p .在算法的实现中,使用划分枚举树来找到符合 p -敏感度模型要求的匿名集.详细的匿名算法请参考文献[4].

2 演示环境与场景

2.1 演示环境

演示系统是一种在线访问的 Web 应用,利用 Java 语言开发,使用在线的 Google Maps for Flex 接口来展示地图.它的运行环境为:2.0 GHz CPU, 2 GB 内存, Windows 7 操作系统.

2.2 演示场景

图 2 是系统的界面.它主要由 4 部分组成:A 部分显示地图、移动用户的位置以及经过隐私保护处理返回的匿名区域;B 部分是移动用户的控制界面,可以选择不同的隐私保护算法和指定不同的隐私保

护参数;C 部分以表格的形式显示模拟用户的位置和服务质量参数;D 部分展示了模拟移动用户的位置更新频率.整幅图显示了使用基于质量的位置隐私保护算法处理 200 个移动用户中 30 个查询请求的匿名结果.当 OrientPrivacy 启动的时候,将会自动加载北京的地图和 POI,并且生成模拟的用户位置,用户可以设定系统中的隐私级别 K ,最长匿名时间和最大匿名边界大小,然后, OrientPrivacy 运用用户指定的隐私保护算法计算出用户的匿名区域并在地图上显示.

2.3 演示步骤

首先展示移动数据的生成.通过对服务器参数的修改,可以生成不同数量、不同运动模式和不同运动速度的移动物体,同时可以指定生成具有不同隐私需求的查询请求.

其次展示不同的隐私保护算法.点击地图上的移动物体,设定好隐私需求参数,系统会模拟向隐私处理服务器发起查询请求,然后服务器会根据选择的算法对查询进行隐私处理,并在界面展示隐私保护结果,即模拟移动用户的匿名区域和服务质量参数.基于用户的请求,可以几种不同的隐私保护之间切换,以查看不同隐私保护算法的差异.

最后展示隐私保护的结果.设定不同的移动用户位置和查询请求更新速度,界面的更新速度也会随之改变.通过点击用户的信息,可以查看用户的匿名结果和服务器的隐私处理参数.

3 未来工作

目前, OrientPrivacy 系统可以对用户发起的基于位置的查询进行位置隐私和查询隐私进行保护.其中用户的位置数据和查询都是模拟生成的,并非实际应用中的真实数据.以后考虑使用真实的移动数据和查询信息.此时为了能够处理大量真实用户的查询请求,服务器的性能会是一个瓶颈,可以考虑将它部署在云计算平台上.同时,应该智能地为用户配置隐私需求.目前系统中需要用户自行指定隐私需求,而现实中大多数用户并不了解隐私级别、最大匿名区域这些专业术语,隐私处理器应该根据用户所处的情景信息自动地配置用户的隐私需求,并选择适用的隐私保护算法进行处理.

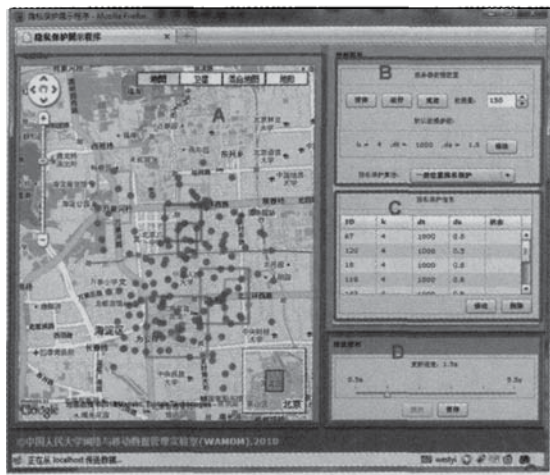


图 2 系统界面

参 考 文 献

- [1] Pao Xiao, Xu Jianliang, Meng Xiaofeng. Protecting location privacy against location-dependent attack in mobile services // Proc of the 17th ACM Conf on Information and Knowledge Management. New York: ACM, 2008: 1475-1476
- [2] Xiao Zhen, Meng Xiaofeng, Xu Jianliang. Quality aware privacy protection for location-based services //Proc of the 12th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2007: 434-446
- [3] Pan Xiao, Meng Xiaofeng, Xu Jianliang. Distortion-based anonymity for continuous query in location-based mobile services //Proc of the 17th ACM SIGSPATIAL Int Conf on Advances in Geographic Information Systems. New York: ACM, 2009: 256-265
- [4] Xiao Zhen, Xu Jianliang, Meng Xiaofeng. p -Sensitivity: A semantic privacy-protection model for location-based services //Proc of the 9th Int Conf on Mobile Data Management Workshops. Piscataway, NJ: IEEE, 2008: 47-54
- [5] Chow C, Mokbel M F, Tian H. TinyCasper: A privacy-preserving aggregate location monitoring system in wireless sensor networks //Proc of the 28th ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2008: 1307-1310
- [6] Du Jing, Xu Jianliang, Tang Xueyan, et al. iPDA: Supporting privacy-preserving location-based mobile services //Proc of the 8th Int Conf on Mobile Data Management. Piscataway, NJ: IEEE, 2007: 212-214
- [7] Mokbel M F, Chow C, et al. The new casper: A privacy-aware location-based database server //Proc of the 23rd Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2007: 1499-1500

黄 毅 男,1987 年生,硕士研究生,主要研究方向为数据隐私保护、移动数据管理等。

潘 晓 女,1981 年生,博士,主要研究方向为移动数据管理等。

孟小峰 男,1964 年生,教授、博士生导师,主要研究方向为 Web 数据管理、移动数据管理、XML 数据管理等。