一种新型的防范历史攻击的k-匿名算法

李 响 孙华志
(天津师范大学计算机与信息工程学院 天津 300387)

摘 要 针对位置信息服务(LBS)中出现的连续查询的隐私问题,提出了一种新型的防范历史攻击的k-匿名算法。该算法根据周围用户的位置、移动速度和移动方向,预测这些用户将来的位置,利用这些位置计算出未来不同时间点上将某用户加入匿名集使匿名区域增大的面积,利用贪心算法优先选择增大面积之和最小的用户加入匿名集。在OPNET 14.5平台下进行了仿真实验,实验结果证明了该算法所形成的匿名区域大小适当,在历史攻击的情况下,既能保护用户的隐私,又能保证一定的服务质量。

关键词 位置信息服务,k-匿名,连续查询,攻击算法
中图法分类号 TP393.0 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.8.041

New k-anonymization Algorithm for Preventing Historical Attacks

LI Xiang SUN Hua-zhi
(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China)

Abstract Aiming at the problem of continuous query privacy in location-based services, a new historical attacks prevention k-anonymization was proposed. The algorithm uses the position, speed and direction of users around to predict these users' future position, then uses these positions to calculate the increased area of the anonymous region caused by joining users into the set at different future time points. The smaller the sum of these increased areas is, the higher the user's priority to add into the anonymous set is. This paper simulated the k-anonymity algorithm on OPENT 14.5 platform. The simulation results show that the size of anonymity region formed by the proposed algorithm is appropriate, which can both protect the privacy of users and guarantee a certain quality of service.

Keywords Location-based service(LBS), k-anonymization, Continuous query, Attacking algorithm

1 引言

随着GPS和无线网络技术的快速发展,人们可以更加方便地测定自己所处的地理位置,一种新的应用服务——基于位置的服务(Location Based Service,LBS)应运而生。基于位置的服务在给人们带来便利的同时,又对人们的个人隐私构成了威胁。人们在使用基于位置的服务时,需要向服务提供商提供自己的位置和查询内容,如果这些信息被不法攻击者获取,那么攻击者就可以了解用户在某时刻的位置、用户习惯的行进路线以及用户的生活隐私。

为了保护用户的隐私,需要阻止攻击者获得用户身份和位置信息之间的一一对应关系。k-匿名是一种常用的方法,它将一个精确位置扩大到一个区域,在这个区域内至少存在k个用户,形成用户身份和位置信息的多对一关系,使攻击者无法从k个用户中区分哪一个才是真正的查询发起者,从而有效保护了用户的隐私。

2 LBS中的私密保护

匿名空间区域对用户的精确位置进行泛化,从一个坐标点扩展为一片包含该点的区域,它的形状是任意的,可以是矩形、圆形或者多边形,其中矩形区域最为常用[2]。

因为查询位置信息的不精确性,LBS提供商在从地图数据库中查找结果时,也从查找距离一个点最近的目标转变成查找所有可能距离一个区域内的某个未知点最近的目标集合[3]。如图1所示,圆点表示发起请求的用户,H1到H5表示用户周围分布的某种目标建筑物。如果收到的是用户的匿名区域(图中矩形区域),那么只能查出可能离用户最近的建筑物集合{H2,H3,H4},最后由用户自己根据自己的坐标筛选出正确的结果H3。

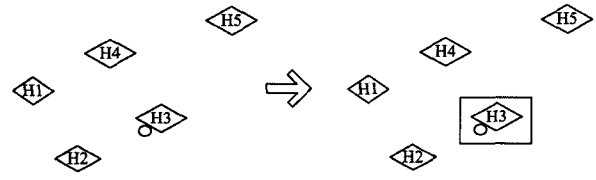

图1 找出离用户最近的某种目标建筑物

匿名区域越大,它可能包含的用户数量就越多,所提供的隐私保护力度就越强,但是LBS服务器的负担就越重,LBS的服务质量就越低。所以,需要权衡好私密性保护力度和LBS服务质量(Quality of Service,QoS)之间的矛盾,既要保

到稿日期:2014-09-03 返修日期:2014-11-21 本文受国家自然科学基金(61103074)资助。
李 响(1985-),男,硕士,实验师,主要研究方向为数据安全与计算机安全、计算机网络,E-mail:lxlindaxyz@hotmail.com;孙华志(1961-),男,博士,教授,主要研究方向为操作系统、网格计算。

证用户的隐私得到有效的保护，又要保证 LBS 服务质量不能太低[4-6]。

3 防范历史攻击的 k-匿名算法的具体实现

3.1 邻居表

每个用户定期广播一个邻居探测数据包，格式如图 2 所示，其中包类型为邻居探测包，源用户 ID 为用户自身 ID，目的用户 ID 为广播。收到别人发来的邻居探测数据包的用户，查看该用户是否在自己的邻居表中。如果在邻居表中，则保持不变；如果不在邻居列表中，那么将其加入邻居表。这样每个用户将始终拥有在某一范围内存在通信用户的邻居表。

1B	2B	2B
包类型	源用户ID	目的用户ID

图 2 邻居探测数据包格式

3.2 周围用户位置信息表

请求发起用户向自己邻居表中广播一个单跳信息请求数据包，收到单跳信息请求数据包的用户向源用户 ID 返回一个信息响应数据包；请求发起用户收到邻居发来的信息响应数据包后，将邻居用户的信息保存在如表 1 所列的周围用户信息表中[7,8]。

表 1 周围用户信息表

用户	邻居数	位置信息				移动速度			
		x_{min}	x_{max}	y_{min}	y_{max}	v_{minx}	v_{maxx}	v_{miny}	v_{maxy}
u_1	5	1.5	2.6	3.0	5.0	1	3	−5	−2
u_2	4	4.0	6.0	4.5	6.0	4	6	3	8
u_3	10	4.5	5.5	0.6	0.8	−5	−1	5	10
…	…	…	…	…	…	…	…	…	…

3.3 形成匿名集

根据周围用户信息表中的数据，首先计算每个用户在当前时刻加入匿名集后使匿名区域增加的面积，式(1)给出了计算方法，其中 $AREA_{i0}$ 表示在当前时刻将用户 u_i 加入匿名集后匿名区域增加的面积，$AREA(D(u_i))$ 表示包含原匿名集中用户和新用户 u_i 在内的匿名区域的面积，$AREA(D)$ 表示未加入 u_i 用户前匿名区域的面积。

$$AREA_{i0}=AREA_0(D(u_i))-AREA_0(D),i=1,2,\cdots \quad (1)$$

计算未来 Δt 时间后，用户 u_i 的位置坐标：

$$(x_i+v_{iminx}\times\Delta t,\ x_i+v_{imaxx}\times\Delta t,\ y_i+v_{iminy}\times\Delta t,y_i+v_{imaxy}\times\Delta t),i=0,1,2,\cdots$$

根据新得到的坐标，按照式(2)计算出用户在 Δt 时间后加入匿名集使匿名区域增加的面积。

$$AREA_{i1}=AREA_1(D(u_i))-AREA_1(D),i=1,2,\cdots \quad (2)$$

若干时间后，将匿名区域增加的面积叠加，就得到了将某用户加入匿名集使匿名区域在一段时间内增加的面积，如式(3)所示，其中 $AREA_i$ 表示将用户 u_i 加入匿名集后匿名集增加的总面积，$AREA_{ij}$ 表示用户 u_i 在第 j 个 Δt 后使匿名集增加的面积。

$$AREA_i=\sum_{j=0}^{k}AREA_{ij},i=1,2,\cdots \quad (3)$$

获得了周围用户信息表中每个用户的匿名代价 $AREA_i$ 后，利用贪心算法，先将匿名代价最小的用户加入匿名集，得到一个新的匿名区域。如果匿名集中的用户数目小于 $k-1$，则说明匿名工作还没完成，这个新的匿名区域只是一个中间临时的匿名区域。

在这个新的匿名区域的基础上，按照式(1)－式(3)计算周围用户信息表中剩余用户的匿名代价，从中选出代价最小的用户加入匿名集，产生另一个新的匿名区域。在最新匿名区域的基础上，反复计算剩余用户的匿名代价，直到匿名集中用户数目达到 $k-1$ 个。

3.4 未来一段时间内的匿名

请求发起用户从匿名集中随机选择一个用户，向这个用户发送一个匿名请求数据包，格式如图 3 所示，其中包类型为匿名请求数据包，匿名集用户数目为去除随机选出的用户之后匿名集剩余用户的数目，匿名集用户 1 到 n 为匿名集剩余用户的 ID。

1B	2B	2B	2B	2B	2B	2B
包类型	源用户ID	目的用户ID	x方向最小坐标	x方向最大坐标	y方向最小坐标	y方向最大坐标

2B	2B		2B
匿名集用户数目	匿名集用户1	…	匿名集用户n

图 3 匿名请求数据包格式

收到匿名请求数据包的用户将源用户 ID 保存起来，用于以后返回匿名响应数据包。先根据自己的位置和请求包中携带的匿名区域的坐标，计算出一个包含自己在内的新的匿名区域。然后随机地从请求包中携带的剩余匿名集用户中选择一个，向选择的用户发送一个新的匿名请求数据包，其中原用户 ID 为用户自身 ID，目的用户 ID 为随机选出的用户，匿名区域更新为新产生的匿名区域。匿名集变为除去新选出的用户后剩下的集合。

以此类推，直到某用户收到的匿名请求数据包中的匿名集为空，说明该用户已是匿名集中最后一个用户。根据自己的位置对匿名区域进行更新。然后向匿名请求数据包中的源用户发送一个匿名响应数据包，格式如图 4 所示，其中源用户 ID 为用户自身 ID，目的用户 ID 为收到的请求包中的源用户，匿名区域坐标为最后形成的匿名区域。收到匿名响应包的用户更新包中的源用户和目的用户后，向收到请求包时保存下来的源用户发送匿名响应数据包，直到被请求发起用户接收到，请求发起用户从响应包中得到最后形成的匿名区域。

1B	2B	2B	2B	2B	2B	2B
包类型	源用户ID	目的用户ID	x方向最小坐标	x方向最大坐标	y方向最小坐标	y方向最大坐标

图 4 匿名响应数据包格式

4 模拟实验和性能评价

本节在 OPNET 14.5 平台上对本文提出的防范历史攻击的 k-匿名算法进行模拟仿真，其中网络协议的物理层和 MAC 层的通信机制采用 IEEE 的 802.11b 标准，并在上层搭建以防范历史攻击的 k-匿名算法为核心的应用层[5]。分别在不同匿名等级 k 的情况下，对历史攻击 k-匿名算法、带阈值的历史攻击匿名算法和快照攻击匿名算法所产生的匿名区域大小进行比较和分析。

4.1 场景设置

本文实验在 500000m² 的区域内随机生成 50 个移动用

户，如图 5 所示，在图中黑色边框表示整个区域边框，50 个用户随机地在这个区域内沿着随机的方向移动。随机选取一个用户，该用户每隔 100s 对自己的位置进行匿名，模拟时间为 1 小时(3600s)。

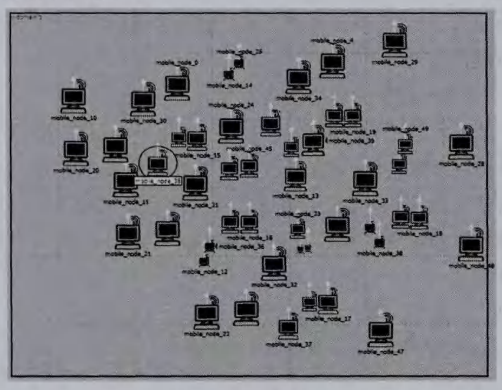

图 5　移动用户分布

4.2　防范快照攻击和匿名集不变时防范历史攻击算法的匿名区域的大小

比较防范快照攻击匿名算法和匿名集不变时防范历史攻击的匿名算法形成的匿名区域的面积。图 6 为匿名等级 k 等于 3、10 和 20 时，历史攻击 k-匿名算法在保持匿名集不变情况下的匿名区域面积和快照攻击 k-匿名算法的匿名区域面积随时间变化的曲线。

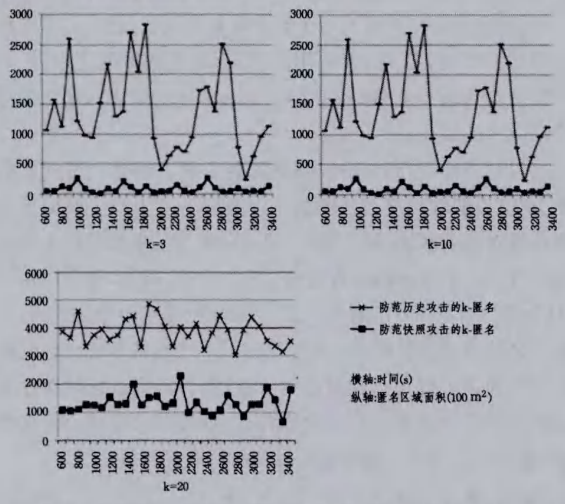

图 6　k 为 3、10、20 时两种 k-匿名算法的匿名区域面积变化曲线

可以看出，历史攻击的 k-匿名算法形成的匿名区域的面积要比快照攻击的大，这是因为历史攻击匿名算法除了要找到形成最小匿名区域的 $k-1$ 个用户外，更重要的是要保证这 $k-1$ 个用户在之后一段时间内所形成的匿名区域不会过大，所以防范历史攻击的 k-匿名算法在选择匿名集用户时，要在快照攻击的基础上附加一个条件，从而导致历史攻击匿名算法的匿名区域略大。

4.3　预测之后 n 个时间段的用户位置

本文算法在选择用户组成匿名集时，对每个用户在之后几个时间段内的位置进行了预测，将不同阶段的匿名代价叠加，使用这个叠加值来衡量用户加入匿名集的优先级。为了测试预测的有效性，对不同情况下，k 等于 3、10 和 20 时的匿名区域面积进行比较，如图 7 所示。图中横坐标表示时间(单位：s)，纵坐标表示匿名区域面积(单位：100m²)，3 条曲线分别为不预测用户将来位置，只考虑当前匿名代价最小化的匿名集所形成的匿名区域；预测今后 5 个时间段匿名代价总和最小化的匿名集所形成的匿名区域；预测今后 5 个时间段匿名代价总和最小化的匿名集并且加入面积最大阈值所形成的匿名区域。

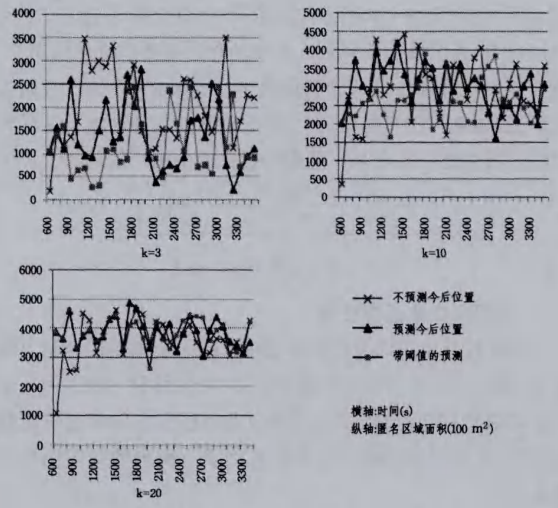

图 7　k 分别为 3、10、20 时不预测、预测和带阈值预测的匿名区域面积变化曲线

可以看出，经过预测今后几个时间段匿名代价总和的匿名集所形成的匿名区域面积在很多时刻都小于只考虑当前匿名代价的匿名区域，说明该预测能对今后一段时间内匿名区域的大小有所反映，通过预测可以选择出在一段时间内形成的匿名区域都不会过大的匿名集。而加入匿名区域面积最大阈值后，所形成的匿名区域面积在很多时刻都小于前面两种，说明加入阈值后，匿名区域面积得到了进一步的控制，从而保证了一定的服务质量。

结束语　在 OPNET 14.5 平台下对本文提出的防范历史攻击的 k-匿名算法进行了模拟实验，分别在匿名等级 k 等于 3、10 和 20 时，统计在不同参数下本文 k-匿名算法和传统防范快照攻击的 k-匿名算法所产生的匿名区域面积数据变化，从而证明了本文提出的防范历史攻击的 k-匿名算法能够通过预测用户未来的位置，找到某一段时间内都可使用的固定匿名集，并且这个匿名集所形成的匿名区域大小会得到一定的控制，当区域过大时算法会更新匿名集，既能帮助用户防范历史攻击，又能保证良好的服务质量。

参 考 文 献

[1]　钟世明，张胜，辜志力，等. 基于移动 Agent 的 LBS 应用平台设计与实现[J]. 计算机应用，2005，25(10)：2306-2309
Zhong Shi-ming, Zhang Sheng, Gu Zhi-li, et al. Design and implementation of mobile agent-based LBS application platform [J]. Computer Applications, 2005, 25(10): 2306-2309

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

[3] Gedik B, Liu Ling. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms [J]. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18

[4] Gruteser M, Grunwald D. Anonymous usage of locationbased services through spatial and temporal cloaking [C] // ACM/USENIX MobiSys. 2003: 1-8

[5] Lu Zhao, Lin Xin. A Data Privacy-Oriented Multi-Parities Location Collect Scheme in Location Based Services [C] // 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology. 2009: 964-969

[6] Ghinita G, Kalnis P, Skiadopoulos S. PRIVE: anonymous location based queries in distributed mobile systems[C] // Proceedings of International Conference on World Wide Web (WWW'07). Banff, Alberta, Canada, 2007: 1-10

[7] Tang Ming, Wu Qian-hong, Zhang Guo-ping, et al. A New Scheme of LBS Privacy Protection [C] // 5th International Conference on WiCom'09. 2009: 1-6

[8] Gallery E, Mitchell C J. Trusted Mobile Platform[M] // Foundations of Security Analysis and Design IV. Springer, 2007: 282-323

(上接第 169 页)

与其他策略静态语义处理的实验对比表明，本文算法具有很高的准确率且消耗时间少，即时间复杂度小，因此实用性强。

结束语 安全策略语义的一致性检测效果直接影响到安全策略的正确执行，并体现安全策略对网络与信息安全的保障能力是否有效。因此保证安全策略语义一致性至关重要。本文在研究静态安全策略语义冲突的基础上提出一种语义相似度计算模型，利用本体提取特征因子，在计算其语义相似度后，对安全策略进行处理，对是否具有冲突的安全策略进行不同的标记，来作为管理者后期改进的目标，从而保证安全策略规则库的一致性。该模型和算法计算简单且有效，检测结果具有很高的准确率，使得安全策略对网络与信息具有很好的安全保障能力。本文仅研究了一般网络环境下的静态安全策略语义一致性检测，进一步的工作是优化算法的过程以提高其时间效率，以及策略执行环节中的动态检测研究。

参 考 文 献

[1] David B, Vincent J, Felix K, et al. Enforceable security policies revisited [J]. ACM Transactions on Information and System Security, 2013, 16(1): 31-56

[2] Mohan A, Blough D M, Kurc T, et al. Detection of conflicts and inconsistencies in taxonomy-based authorization policies [C] // Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Atlanta, GA, 2011: 590-594

[3] Li Zang, Chu Chao-hsien, Yao Wen. A semantic authorization model for pervasive healthcare [J]. Journal of Network and Computer Applications, 2014, 38: 76-87

[4] 李瑞轩，鲁剑锋，李添翼，等. 一种访问控制策略非一致性冲突消解方法[J]. 计算机学报，2013，36(6)：1210-1223
Li Rui-xuan, Lu Jian-feng, Li Tian-yi, et al. An approach for resolving inconsistency conflicts in access control policies [J]. Chinese Journal of Computers, 2013, 36(6): 1210-1223

[5] Bao Yi-bao, Yin Li-hua, Fang Bin-xing, et al. A novel logic-based automatic approach to constructing compliant security policies [J]. Science China: Information Sciences, 2012, 55(1): 149-164

[6] 包义保，殷利华，方滨兴，等. 基于良基语义的安全策略表达与验证方法[J]. 软件学报，2012，23(4)：912-927
Bao Yi-bao, Yin Li-hua, Fang Bin-xing, et al. Approach of security policy expression and verification based on well-founded semantic [J]. Journal of Software, 2012, 23(4): 912-927

[7] Basile C, Cappadonia A, Lioy A. Network-level access control policy analysis and transformation [J]. IEEE/ACM Transactions on Networking, 2012, 20(4): 985-998

[8] 沈国华，张伟，黄志球，等. 基于描述逻辑的特征语义建模及验证[J]. 计算机研究与发展，2013，50(7)：1501-1512
Shen Guo-hua, Zhang Wei, Huang Zhi-qiu, et al. Description-logic-based feature modeling and verification [J]. Journal of Computer Research and Development, 2013, 50(7): 1501-1512

[9] 王腾，朱青，王珊. 基于语义相似度的 Web 信息可信分析[J]. 计算机学报，2013，36(8)：1668-1681
Wang Teng, Zhu Qing, Wang Shan. Fact statements verification based on semantic similarity [J]. Chinese Journal of Computers, 2013, 36(8): 1668-1681

[10] 程勇，黄河，邱莉榕，等. 一个基于相似度计算的动态多维概念映射算法[J]. 小型微型计算机系统，2006，27(6)：975-979
Cheng Yong, Huang He, Qiu Li-rong, et al. Similarity-based dynamic multi-dimension concept mapping algorithm [J]. Mini-Micro Systems, 2006, 27(6): 975-979

[11] 郑晓洁，张琳. 本体映射中相似度计算的改进[J]. 计算机科学，2013，40(12)：108-112
Zheng Xiao-jie, Zhang Lin. Modification of similarity computation in ontology mapping [J]. Computer Science, 2013, 40(12): 108-112

[12] Kobra E, Amin R D, Mahmoud N. Overlapped ontology partitioning based on semantic similarity measures [C] // Proceedings of the 5th International Symposium on Telecommunications. Tehran, Iran, 2010: 1013-1018

[13] Pirro G. A semantic similarity metric combining features and intrinsic information content [J]. Data & Knowledge Engineering, 2009, 68(11): 1289-1308

[14] Kunal V, Rama A, Richard G. Semantic matching of web service policies[C] // Proceedings of the 2nd International Workshop on Semantic and Dynamic Web Processes. Orlando, USA, 2005: 1-12

[15] Gruber T R. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 1993, 5(2): 199-220

[16] 倪俊，陈晓苏，刘辉宇，等. 网络安全策略求精一致性检测和冲突消解机制的研究[J]. 计算机科学，2011，38(2)：32-37
Ni Jun, Chen Xiao-su, Liu Hui-yu, et al. Research on network security policy refinement consistency of detection and conflict resolution mechanisms [J]. Computer Science, 2011, 38(2): 32-37