QR-TCM:具有质量保证的位置服务隐私保护模型

胡文领 王永利
(南京理工大学计算机科学与技术学院 南京 210094)

摘 要 针对传统的基于位置服务的隐私模型匿名时间较长的情况,建立了QR-TCM模型。该模型提出了隐私保护算法CRCA。通过分析影响匿名时间的因素,提出了解决用户服务延迟的方法以及位置服务质量评价模型。实验采用了标准数据集上的数据,通过响应时间、隐私性等多个维度去衡量QR-TCM模型。实验结果证明,该方法适用于连续查询位置隐私保护,可有效保护用户的位置隐私和提供及时的服务。

关键词 基于位置服务,隐私保护,k-anonymity,k-means,质量保证
中图法分类号 TP391 **文献标识码** A

QR-TCM:A Privacy Protection Model with Quality Guarantee for Location-based Services

HU Wen-ling WANG Yong-li
(Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract To solve the problem that the traditional location services anonymity model takes a lot of time to produce anonymous region, the quasi real-time cloak model (QR-TCM) was established. The model proposes a privacy protection method called clock rotation cloak algorithm (CRCA). After comprehensive analysis of the reasons that influence the anonymity, a model that can solve the users' service delayed and the method of measuring the quality of service were proposed. The experiment uses the standard data sets, and measures the QR-TCM model with multiple dimensions such as the response time and the degree of privacy. The experiment results confirm that the method is suitable for continuous query location privacy protection, and can effectively protect the user's privacy and offer fast service.

Keywords Location-based services, Privacy protecting, K-anonymity, K-means, Quality guarantee

1 引言

基于位置服务[1](Location Based Services, LBS)的广泛应用给人们生活增添了很多便利[2]。但是,用户在享受这些便捷服务时,也面临隐私泄漏的威胁[3]。恶意的位置服务提供商或其它针对位置服务器的攻击者根据用户位置和查询内容[4],运用数据挖掘和机器学习等方法,就可以轻松地获得用户的隐私信息,从而给用户的隐私带来严重挑战[5]。

本文对目前基于位置服务的隐私保护模型[6]进行分析,对匿名失败的情况[7,8]进行深入研究,采用了近实时的隐私模型[9](Quasi Real-Time Cloak Model, QR-TCM),对匿名失败的区域进行重新分组,并提出了时钟轮转的匿名算法(Clock Rotation Cloak Algorithm, CRCA),解决了匿名失败情况下基于位置服务不能实时提供服务的问题。通过实验证明:(1)解决了匿名失败后位置服务延迟的问题;(2)通过综合考虑匿名度和用户满意度,建立了有效的位置服务质量评价模型。

2 预备知识

假设用户拥有定位功能的移动终端,向服务器发出位置服务请求,给出如下定义:

定义1 k-匿名(k-anonymity)

k-匿名即发布的数据中存在至少 k 个数量在准标识符上不可区分的记录,使攻击者不能判别出隐私信息所属的具体个体,从而保护了个人隐私,k-匿名通过参数 k 指定用户可承受的最大信息泄露风险。

定义2 基于位置服务查询(Location-based Services Query)

用户通过可定位的移动终端,向服务器提交一个查询 Q,即一个三元组,记为 $Q(L,T,C)$:

L(location)表示查询 Q 当前所在位置经纬度;T(timestamp)表示发出查询 Q 的当前时刻(时间戳);C(content)表示查询内容。

定义3 最大值 k 约束

匿名区域中,每个用户对应的匿名度 $K=\{k_1,k_2,\cdots,$

到稿日期:2013-04-25 返修日期:2013-07-18 本文受国家自然科学基金(61170035, 61272420),江苏省自然科学基金重大专项(BK2011022),江苏省自然科学基金(BK2011702),江苏省青蓝工程创新团队,江苏省高校2010年"青蓝工程"优秀青年骨干教师项目,南京市科技计划重点项目(020142010)以及南京理工大学2009年度紫金之星项目资助。

胡文领(1988—),男,硕士生,工程师,主要研究方向为隐私保护、位置服务、数据库技术等,E-mail:hwltony@126.com;**王永利**(1974—),男,博士,副教授,CCF会员,主要研究方向为数据库技术、情境感知、物联网数据处理、模式识别等。

$k_n\}$，真实的用户匿名度为 k_i；满足：

$\max\{k_1,k_2,\cdots,k_n\}>k_i$

定义 4 组团匿名约束

对应用户 n 个用户信息集合 $M=\{m_1,m_2,\cdots,m_n\}$，每个用户对应的匿名度 $K=\{k_1,k_2,\cdots,k_n\}$，n 个用户能够生成匿名区域，当且仅当满足下面的约束：

$\max\{k_1,k_2,\cdots,k_n\}\leqslant n$

定义 5 用户的熵

假设在一个匿名区域 R 中的 n 个用户的位置分别是 L_1，L_2，$L_3\cdots L_n$，则对 R 中的任意用户 u 来说，其熵 $h(u)$ 定义为：

$$h(u)=-\sum_{i=1}^{n}p_i\log_2 p_i \tag{1}$$

其中，p_i 是用户 u 在位置 L_i 上出现的概率。如果用户在匿名区域中均匀分布，假设有 k 个用户，则 $p_i=1/k$，当 $k=5$ 时，$p_i=0.2$。

定义 6 熵匿名度

定义用户 u 的熵匿名度为：

$$PA(u)=2^{h(u)-1} \tag{2}$$

定义 7 服务质量 QoS

根据 SERVQUAL 理论[10]，基于位置服务的质量模型的评价公式见式(3)：

$$QoS(u)=\sum_{i=1}^{5}w_i(P_i-E_i) \tag{3}$$

其中，$\sum_{i=1}^{5}w_i=1$，$0\leqslant P_i\leqslant 1(1\leqslant i\leqslant 5)$。

变量 P_1 表示用户服务质量的隐私性 $PA(u)$，由系统计算得出。

变量 P_2 表示用户服务质量的可靠性，由用户进行评价。

变量 P_3 表示用户服务质量的响应性，由系统计算得出的时间的倒数表示。

变量 P_4 表示用户服务质量的保证性，即系统提供服务的能力，由用户进行评价。

变量 P_5 表示用户服务质量的移情性，即系统为每个用户提供个性化位置服务的能力，由用户进行评价。

变量 E_i 表示每个特性在用户中的期望，默认是 1。

参数 w_i 表示每个位置服务测量值的权重，系统默认为 0.2。

每个参数权重和取值范围，如表 1 所列。

表 1 位置服务质量评价参数

参数名称	变量名称	权重	取值范围
隐私性	P_1	w_1	[0,1]
可靠性	P_2	w_2	[0,1]
响应性	P_3	w_3	[0,1]
保证性	P_4	w_4	[0,1]
移情性	P_5	w_5	[0,1]

定义 8 匿名集 CR

匿名集由用户、查询内容组成，即 $CR=\{(m_1,m_2,\cdots,m_n),(c_1,c_2,\cdots,c_n)\}$，$m_i$ 表示用户信息，c_i 表示查询内容。这是用户的一次位置查询经过匿名服务器后，发送给服务器的查询内容。$|CR|$ 表示匿名集大小。

3 QR-TCM 模型

3.1 QR-TCM 模型介绍

基于位置服务的隐私保护，最主要的就是隐藏用户信息，对用户的位置隐私保护是典型的隐私策略。针对不同的隐私要求，模型可以对用户的匿名度进行个性化设置。在实际运用中，该模型弥补了服务质量下降的缺点。

3.2 QR-TCM 模型定义

QR-TCM 模型的处理步骤如下：

(1)用户使用带有位置定位功能的移动终端向匿名服务器发送位置服务连续查询请求 Q；

(2)根据用户个性化匿名度 k，匿名服务器根据组团匿名约束和最大值 k 约束生成匿名区域，当成功得到匿名区域后，就可以将该区域发送给位置服务提供商，等待返回结果；如果匿名区域生成失败，那么先对区域中的用户根据匿名度 k 进行 k-means 分组，将真实用户所在的组 R 找出来；

(3)对 R 中的用户应该采用时钟旋转匿名算法，对匿名失败区域中的其他用户，按照时钟旋转方向，依次生成匿名区域 Ri，如果匿名区域生成成功，那么将匿名区域 Ri 中的用户合并到 R 中得到新的 R'，此时判断新的 R' 是否满足组团匿名约束和最大值 k 约束，如果满足，那么发送 R' 给位置服务器；如果不满足，那么继续找到下一个用户，尝试生成匿名区域，如果成功，则合并到 R' 中，一直执行这个操作，直到 R' 满足组团匿名约束和最大值 k 约束，然后将 R' 发送到位置服务器。虽然用户的连续查询是精确的，但是攻击者无法通过匿名区域中的用户来判断查询是哪个用户提出的，从而达到了保护用户隐私的目的。

(4)匿名服务器收到服务提供商返回的服务信息后，将查询结果返回给用户。

本文采用典型的中心服务器型的架构，主要组件有移动终端、可信的匿名服务器和位置服务提供商，如图 1 所示。

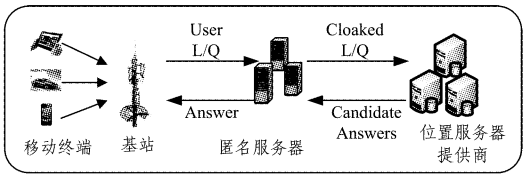

图 1 基于位置服务中心架构

4 CRCA 位置隐私保护算法

4.1 算法描述

Step1 当用户信息集匿名失败后，为了实时地得到位置服务信息，模型 k-means 对用户的匿名度 k 进行聚类，将用户分为 k' 个组。对真实用户所在的组，按照时钟旋转对其他用户生成匿名区域，若能成功，则将两个分组合并，若合并满足组团匿名约束和最大值 k 约束，则将新的匿名区域发送到位置服务提供商。若不能生成匿名区域，则继续找到下一个用户，尝试生成匿名区域，如果成功，则合并，一直执行这个操作，直到 R' 满足组团匿名约束和最大值 k 约束，然后将 R' 发送到位置服务器。

Step2 根据 GetQos 方法，模型将系统中的 5 个测量位置服务中的用户挑选出来，然后根据位置服务计算模型，将位置服务质量评价返回给位置服务提供商。

Algorithm 1 CRCA

Input 用户的查询 Q，个性化匿名度 k，k-means 聚类中的中心个数 k′；

Output k′个用户聚类；

初始化 $M=\{m_1,m_2,\cdots,m_n\}$，对应的匿名度集合是 $K=\{k_1,k_2,\cdots,k_n\}$，初始化 k' 组连续位置服务查询用户 $\{M_1,M_2,\cdots,M_{k'}\}$；

1. Begin
2. for i=1,…,k′
3. $M_i=\{k_i\}$;
4. $R_i=k_i$;// R_i represents the center of each group
5. Do{
6. j=1;
7. for i=1…n:
8. minDist=Dist (k_i,R_1);
//minDist represents the minimal distance
9. minIdx=1;
//minIdx represents the index of k′ groups
10. for j=1,…,k′
11. if:minDist<=Dist(k_i,R_j):
12. min Idx=j
13. $M_{minIdx}=M_{minIdx}\cup k_i$;
14. for j=1,…,k′:
15. $R_j=\frac{k_{j1}+\cdots+k_{jq}}{|M_j|}$,(1≤q≤n)
// k_{jq} represents the k in M_j
16. }while k′ groups of M_j are changed:
//下面进行时钟旋转匿名处理
//假设真实用户所在集合 $M_a=\{m_{a1},m_{a2},m_{a3},\cdots,m_{an}\}$
//假设 m_{a1}就是真实用户，从 m_{a1}的 x-轴正方向顺时钟旋转，假设得到 m_{a2}，对 m_{a2}进行区域匿名
17. $M_a'=M_a$//初始化新的真实用户所在的 M_a'
18. flag=false;
19. for i=2,…,n
20. If m_{ai}匿名成功:
//得到集合 $M_b=\{m_{b1},m_{b2},m_{b3},\cdots,m_{bn}\}$
21. $M_a'=M_a\cup M_b$
//将 M_a 和 M_b 合并，得到新的 M'_a
22. If(M_a'能生成匿名区域) {flag=true;break;}
23. Endif
24. Endif
25. Endfor
26. Return M_a'
27. End

Algorithm 2 GetQos

Input $M_i=\{m_1,m_2,\cdots,m_p\}(p=|M_i|)$，对应的匿名是 $K=\{k_1,k_2,\cdots,k_p\}$，真实的用户在集合 M_i 中，$k_{max}=\max\{k_1,k_2,\cdots,k_p\}$，

Output Qos 值

1. Begin
2. h_total=0;//初始化变量
3. For i=1…p{
4. h_total+=−(1/k_i)$\log_2$(1/k_i);
5. //循环获取得到熵值
6. }
7. $PA(u)=2^{h\_total-1}$;//得到熵匿名度
8. $P_1\sim P_5=\{PA(u),0.5,-,0.5,0.8\}$;
9. //初始化系统实际概率，其中 P1 和 P3 待定
10. $E_1\sim E_5=\{1,1,1,1,1\}$;
11. //初始化期望概率，默认全为 1
12. $w_1\sim w_5=\{0.2,0.2,0.2,0.2,0.2\}$;
13. //初始化权重，默认均分，即 0.2
14. qos_total=0;//计算服务质量
15. For i=1…5{
16. qos_total+=w_i*(P_i-E_i)
17. }
18. Return qos_total //将服务质量返回
19. End

4.2 算法分析

(1)CRCA 算法分析

通过 k-means 和 CRCA 方法，模型按照匿名度 k 将用户信息划分为 k' 组，然后采用时钟旋转合并匿名区域处理方法，直到成功生成匿名区域。算法的时间复杂度是 $O(N*k'*t)$，其中 N 是用户个数，k' 是划分组个数，t 是迭代次数。

(2)GetQos 方法将位置服务中的 5 个参数提取出来，根据位置服务计算方法，计算出了位置服务质量 QoS。算法的时间复杂度是 $O(n)$，其中 n 是区域中用户的个数。

5 实验

(1)实验材料：实验使用著名的 Network-Based Generator of Moving Objects 模拟器来模拟用户运动轨迹，采用 Oldenburg 市区的交通网络作为模拟程序输入。实验使用 Java 技术、Eclipse 开发工具，在官方开源代码 CompleteSource21. zip 基础上进行修改，达到满足本实验的场景，如图 2 所示。

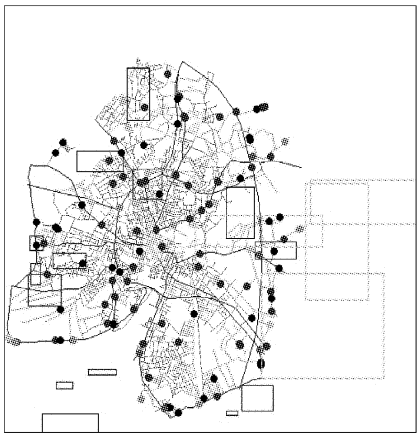

图 2 物体运动轨迹实验截图

(2)实验配置，如表 2 所列。

表 2 实验数据参数

参数名称	参数值
θ	0.5
网络图	Oldenburg 市区图
最大移动速度(米/秒)	50
初始化移动终端(个)	8
额外移动终端(个)	3
时间戳(秒)	1
持续时间(秒)	200
每个时间戳产生终端(个)	6

(3)实验主要内容：(a)匿名成功率检测，参数可以衡量本文工作的重要性；(b)匿名失败后，比较 CliqueCloak 模型和 QR-TCM 模型的匿名处理时间；(c)检测 QR-TCM 模型下不同权重时用户的位置服务质量 QoS。

(4)实验分析：

实验 1 匿名成功率

如图 3 所示，横坐标是匿名度 k，纵坐标是匿名成功率。

从图中可以看出，当 k 逐渐变大时，匿名成功率越来越低，因为随着用户数量增多，组团匿名约束条件和最大值 k 约束使失败率升高。

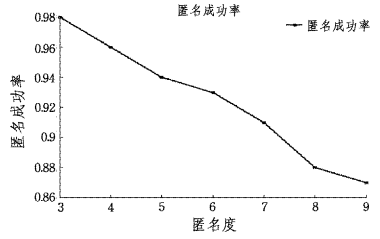

图 3　匿名度和匿名成功率关系

实验 2　两个模型的匿名处理时间

如图 4 所示，横坐标是匿名度 k，纵坐标是匿名时间。当 $k=3$ 和 4 时，两个模型匿名处理时间大致相同，因为此时都是第一次生成匿名区域；当 $5<k<9$ 时，两个模型的匿名处理时间上升，但是 CliqueCloak 模型用户等待时间总体较长，上升比较快。QR-TCM 模型采用了新的 CRCA 算法，虽然需要多次匿名，但每次匿名处理时间较短，因此缩短了用户等待时间，对于连续查询节省了大量时间。

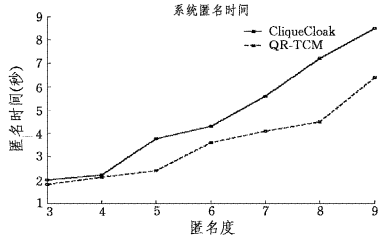

图 4　CliqueCloak 和 QR-TCM 匿名时间

实验 3　不同权重对 QR-TCM 模型的影响

表 3　A 组实验数据参数设置

参数名称	参数值	参数名称	参数值
P_1	—	w_1	0,5
P_2	0.7	w_1	0,1
P_3	—	w_1	0,2
P_4	0.6	w_1	0,1
P_5	0.5	w_1	0,1

表 4　B 组实验数据参数设置

参数名称	参数值	参数名称	参数值
P_1	—	w_1	0,1
P_2	0.5	w_2	0,2
P_3	—	w_3	0,5
P_4	0.5	w_4	0,1
P_5	0.8	w_5	0,1

如图 5 所示，横坐标是匿名度 k，纵坐标是两组不同参数下的位置服务质量。在表 3 和表 4 中，设置了不同的参数。从图中可以看出，随着匿名度增加，服务质量不断下降，A 组实验侧重隐私性，而 B 组实验侧重匿名处理时间，不同权重对 QR-TCM 模型影响比较大。

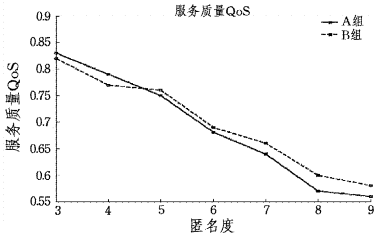

图 5　权重不同时，服务质量 QoS 不同

结束语　通过研究连续查询下的隐私保护问题，提出了 QR-TCM 用户隐私保护模型。该模型分析现有的位置隐私保护方法，发现位置服务延迟的缺点，提出了 CRCA 算法。该算法使得用户在使用位置服务信息时，能够在不降低匿名度的情况下，实时地得到精确位置服务。该模型平衡了匿名度和服务质量之间的矛盾，极大地提高了基于位置服务的质量。移动终端和服务器交互的过程中能量消耗比较多[11]，这些不足是我们下一阶段改进的重点。

参 考 文 献

[1] 潘晓，郝兴，孟小峰. 基于位置服务中的连续查询隐私保护研究[J]. 计算机研究与发展，2010，47(1)：121-129

[2] 王彩梅，郭亚军，郭艳华. 位置服务中用户轨迹的隐私度量[J]. 软件学报，2012，23(2)：352-360

[3] 王智慧，许俭，汪卫，等. 一种基于聚类的数据匿名方法[J]. 软件学报，2010，21(4)：680-693

[4] 魏志强，康密军，贾东，等. 普适计算隐私保护策略研究[J]. 计算机学报，2010，33(1)：128-138

[5] 周水庚，李丰，陶宇飞，等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报，2009，32(5)：843-861

[6] 林欣，李善平，杨朝晖. LBS 中连续查询攻击算法及匿名性度量[J]. 软件学报，2009，20(4)：1058-1068

[7] 彭志宇，李善平. 移动环境下 LBS 位置隐私保护[J]. 电子与信息学报，2011，33(5)：1211-1216

[8] Pingley A，Wei Yu，Zhang Nan，et al. A context-aware scheme for privacy-preserving location-based services [J]. Computer Networks，2012，56(11)：2551-2568

[9] Lin Yu-bao，Chen Xiu-wei，Li Zhan，et al. An efficient method for privacy preserving location queries[J]. Front Computer Science，2012，6(4)：409-420

[10] 艾小淞，孙红，孙西国. SERVQUAL 和 SERVPERF 方法在 GPS 服务质量中的应用研究[J]. 北京航空航天大学学报：社会科学版，2010，23(4)：76-78

[11] Vergara-Laurens I J，Labrador M A. Preserving privacy while reducing power consumption and information loss in lbs and participatory sensing applications[C] // GLOBECOM Workshops (GC Wkshps)，2011 IEEE. 2011：1247-1252

(上接第 89 页)

[7] Zhao H，Ansari N. Wavelet Transform-based Network Traffic Prediction：A Fast On-line Approach[J]. Journal of Computing and Information Technology，2012，20(1)：15-25

[8] Maurya C K，Minz S. Fuzzy inference system for Internet traffic load forecasting [C] // Computing and Communication Systems (NCCCS)，2012 National Conference on. IEEE，2012：1-4

[9] 姜明，吴春明，张旻，等. 网络流量预测中的时间序列模型比较研究[J]. 电子学报，2009，37(11)：2353-2358

[10] http://ita. ee. lbl. gov/html/contrib

[11] http://datamarket. com/data/list/? q=time+series

[12] Mallat S G. A theory for multiresolution signal decomposition：the wavelet representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，1989，11(7)：674-693