高技术通讯 2013 年 第23 卷 第11 期:1148 ~1153

doi:10.3772/j.issn.1002-0470.2013.11.007

路网条件下基于用户协作的 LBS 隐私保护①

毛典辉② 蔡 强 李海生 曹 健 黄今慧 程叶椋

(北京工商大学 计算机与信息工程学院 北京 100048)

摘 要 针对移动终端在缺乏固定通信基础设施的路网环境下消费 LBS 服务时的用户位置隐私保护问题,提出了一种基于用户协作的位置隐私保护方法:用户间通过 Chord 协议自组织成 P2P 网络,当用户消费服务时,由其选择代理节点进行转发增量近邻查询。为了提高该方法的服务质量,提出了最大连通稳定度理论以实现代理用户优选,保证 P2P 通讯的稳定性,同时引入通信成本估算方法实现查询点的自适应选择,以减少代理用户的查询通讯开销。对此算法进行了理论分析,并通过实验验证了其性能,实验结果表明该方法在路网条件下具有较高的隐私保护度,而且实现了隐私保护度与服务质量的平衡。

关键词 P2P 网络,基于位置的服务(LBS),位置隐私,网络安全,隐私保护

0 引 言

在基于位置的服务(location based service,LBS)应用中,如何妥善保护用户的位置隐私成为亟待解决的重要课题[1]。从目前已取得的研究成果来看,大都采用可信第三方匿名器(trusted third party anonymization,TTP)[2]应用 k -匿名模型[3]将 k 个用户组成空间区域,从而使得攻击者无法推测单个用户身份与位置,但是事实上并没有一个权威机构来评估 TTP 的可信度,因此 TTP 并不完全可信[4],而且中心匿名器容易成为系统性能瓶颈和集中攻击点,因此越来越多的研究趋向无 TTP 模式。无 TTP 模式的隐私保护方法根据用户合作与否分为两类:非用户合作与用户合作。前者一般由终端直接发送假位置点至服务器进行增量近邻查询,如 SpaceTwist 方法[6]、CAP 方法[7]等,但该类方法在实际应用中要求终端已存储完整的路网信息,且无法隐藏用户身份与查询内容,因此对终端性能以及应用场景均有较高要求;后者一般通过用户协作构建 P2P(Peer-to-Peer)通信网络并形成匿名区域[8],如 MobiHide 方法[9]、匿名链方法[10]、CoPrivacy 方法[11]等;该类方法能实现用户身份以及位置隐私的双重保护,但是已有的 P2P 解决方案通常要求终端维护复杂的数据结构,而且依赖固定基础通信设施。而在动态 P2P 网络中和路网环境下,终端通常缺乏固定通信基础设施,移动用户只能通过互相协作、多跳路由进行通信.在这样的环境中用户位置隐私保护的问题有一些独特性,需考虑终端带宽、运动受限等现实因素(如车辆在某些路段上只能朝一个方向行驶)。针对上述动态 P2P 网络及路网环境中的用户位置隐私保护问题,本文提出了一种新的位置隐私保护方法:移动用户在自组织的 P2P 通信网络中消费服务时,通过寻找具有最大连通稳定度的近邻用户进行代理,然后由代理用户通过通信成本估算进行自适应的增量近邻查询。该方法不仅切断了服务消费者与服务提供商之间的直接关联,而且减少了 P2P 网络节点间的通信开销,使得系统在保护位置隐私的同时具有较高的服务质量。

1 隐私保护的技术条件和思路

该隐私保护方法实现系统由两部分组成:移动终端和 LBS 服务器。假设每个移动终端具备 2 个无线接口卡:一个通过移动基站与 LBS 服务器通信,支持 2G/3G 网络;另一个用于在缺乏足够固定

① 国家自然科学基金(61170113),北京市自然科学基金(4112016),北京市属高等学校人才强教计划(PHR201108075),北京市属高等学校科学技术与研究生教育创新工程建设项目(PXM2012_014213_000079)和北京市属高等学校教师队伍建设——青年英才计划(YETP1452)资助项目。

② 男,1979 年生,博士;研究方向:空间信息服务;联系人,E-mail:maodianhui@gmail.com

(收稿日期:2012-12-22)

通信基础设施时进行 P2P 通信,具有 802.11b/g 通信能力。移动用户根据在服务过程中扮演的角色不同分为消费用户与代理用户,消费用户与代理用户间通过 P2P 网络实现通信,而代理用户通过 2G 或 3G 移动基站与 LBS 服务器进行增量近邻查询。

消费用户在服务过程中,通过终端定位设备取得真实位置点信息,然后根据路网状况自主设置模糊区域、查询内容以及最大允许时延等参数,形成消费请求,同时在 P2P 网络中选择协作意愿高、连通稳定性强的近邻节点作为代理用户;代理用户根据通信成本估算方法自适应情境取得查询点,然后与服务提供商进行增量近邻查询,并返回查询结果至消费用户;消费用户如果在最大允许时延内收到返回结果,则根据真实位置点计算出最终结果,否则重新选择代理节点重复上述过程,直至查询结束为止。

2 隐私保护方法的实现

移动终端在无固定通信基础设施条件下消费服务时,需搜索近邻节点组建或加入 P2P 网络,Chord 协议作为一种可靠的分布式路由算法与资源协议,具有高扩展性以及节点快速定位等优点[12],因此,本文选择 Chord 协议实现节点通信并自组织成 P2P 网络。考虑到隐私保护方法在极端应用情况下,移动终端的加入与退出都较为频繁,因此,为了保证 P2P 网络节点通信的可靠性,设计 Chord 协议组建的环结构中每个节点均为一簇用户,而非单个用户,每个簇中最多包含 α 个用户(根据网络状况设定),且必须含有一个头节点,该头节点称为簇头节点。为了保护 Chord 环结构中关键字的顺序,簇头节点为簇成员中 *key* 值最大的节点,簇成员组成信息由簇头节点维护[12]。P2P 网络的稳定性需考虑节点的加入、重定位或节点退出等情况,具体状态维护机制可参看文献[12]。

2.1 代理用户的优化选择

移动用户在服务消费过程时,不断改变位置区域或者退出网络,因此,为了保证通信的稳定性,消费用户需选择连通稳定性强的近邻节点作为代理用户,而在实际路网条件下,移动节点间的通信稳定性受两点间的距离、运动方向与运动速度的影响,为了描述某一时刻用户间的通信能力,在此定义用户连通性[10]。

定义 1(连通性 C_t) 对于任一时刻 t,移动用户间的连通性定义为两者间的最大通信距离与实际距离之比,记为 $C_t = R/d_t$。其中 R 为用户间的最大通信距离,d_t 表示在 t 时刻用户间的实际距离。

如图 1 所示,在动态 P2P 网络中,假设两个移动用户 U_1、U_2 在 t 时刻的速度分别为 $\vec{v}_1$、$\vec{v}_2$,运动方向夹角为 α,用户间距离为 $\vec{d}_t$,U_2 与 $\vec{d}_t$ 间夹角为 γ,具有的连通性为 C_t;经过时间 $\tau = t' - t$ 后,移动用户间的距离为 $\vec{d}_t'$,速度变化夹角为 β,则在 t' 时刻,用户间的距离 $\vec{d}_t' = \vec{d}_t + (\vec{v}_2 - \vec{v}_1)\tau = \vec{d}_t + \Delta\vec{v}\tau$,其值满足公式

$$\begin{cases}\Delta d = \Delta v \times \tau \\ d_{t'} = \sqrt{d_t^2 + \Delta d^2 - 2d_t\Delta d\cos(\gamma - \beta)}\end{cases} \tag{1}$$

则移动用户间具有的连通性概率为

$$P(C_{t'} \geqslant 1) = P(d_{t'} \leqslant R)$$

$$= P(|\Delta v| \leqslant \frac{d_t\cos(\gamma - \beta) + \sqrt{R^2 + (d_t\sin(\gamma - \beta))^2}}{\tau}) \tag{2}$$

假设移动用户间的速度变化值 Δv 服从均值为 $\sqrt{v_2^2 + v_1^2 - 2v_1v_2cos\alpha}$、方差为 σ 的正态分布,则

$$P(C_{t'} \geqslant 1) = \frac{1}{\sqrt{2\pi}\sigma}\int_0^{l/\tau} \exp(-\frac{(\Delta v - u)^2}{2\sigma^2})\,\mathrm{d}\Delta v \tag{3}$$

其中 $l = d_t\cos(\gamma - \beta) + \sqrt{R^2 + (d_t\sin(\gamma - \beta))^2}$,$u = \sqrt{v_2^2 + v_1^2 - 2v_1v_2\cos\alpha}$。

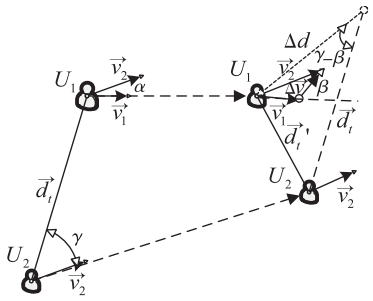


图 1 移动用户间连通性分析

为了衡量连续时间内移动用户间的连通稳定性,本文定义式(3)为移动用户在时间 τ 内的连通稳定度,因此,对于消费用户而言,为了保证通信质量,需在近邻用户列表中评估邻居节点的连接稳定度,选择一个在容忍时延内具有最大连通稳定度的用户进行服务代理,其代理节点选择算法实现过程如算法 1 所示。

算法 1 代理节点优选算法

1. $R \leftarrow$ 设置移动用户间最大通信距离
2. $\mu \leftarrow$ 设置最大容忍查询时延
3. $\gamma \leftarrow$ 设置移动对象间距离与其运动方向间夹角为固定值
4. $Max = 0 \leftarrow$ 设置最大连通稳定度初值为 0
5. $U_p = \phi \leftarrow$ 设置初始代理用户对象为空
6. $UList \leftarrow$ 取得近邻用户列表
7. *foreachpinUList*
8. $d = O_p - O_c \leftarrow$ 求取两个移动对象间的距离
9. $\alpha \leftarrow$ 求取两个移动对象间运动方向夹角
10. $u \leftarrow$ 求取两个移动对象间速度变化值
11. $\beta \leftarrow$ 求取两个移动对象间速度变化夹角
12. $C_p \leftarrow$ 求取两个移动对象间连通稳定度
13. $ifC_p > Maxthen \leftarrow$ 求取最大连通稳定度对象
14. $Max = C_p \leftarrow$ 保存当前最大连通稳定度值
15. $U_p = p \leftarrow$ 保存当前对应的移动对象
16. *endif*
17. *endfor*
18. $returnU_p$ //返回最优的代理用户

2.2 代理用户自适应查询实现

代理用户与服务提供商进行增量近邻查询[6]时,其通信开销由服务器返回的路网拓扑信息以及查询兴趣点(points of interest,POIs)组成,而通信开销受查询点位置选取影响,传统的查询点随机选取策略容易导致通信开销不可控,因此,本文在保证服务质量基础上,在代理用户端引入通信成本估算方法,并自适应情境确定查询点,从而降低整体服务过程中计算开销与通信代价。

自适应情境的增量近邻查询算法实现流程如下:

(1)代理用户仅发送查询主题(如医院、餐厅等)至 LBS 服务器,服务器根据空间分布统计方法将查询空间分成若干个子域,统计每个子域内 POIs 数量形成空间直方图 H 并返回(空间直方图文件极小,可直接存储于服务器端缓存中,避免多次重复统计)。

(2)代理用户依据 H, 判断模糊区域 Q 内 POIs 值是否满足供应条件(如图 2 中深色区域),否则扩大 Q 范围形成 Q', 使得 Q' 内 POIs 值满足供应条件,此时 Q' 即为最小通信成本估计的需求空间(如图 2 中灰色区域),覆盖 Q' 的最小外接圆区域即为最小通信成本估计的供应空间,其圆心为所求查询点位置,该过程具体实现详见算法 2。

(3)代理用户端发送查询点、Q' 至服务器进行增量近邻查询,该查询过程可参看文献[6],并将查询结果转发至消费用户,至此服务消费过程结束。

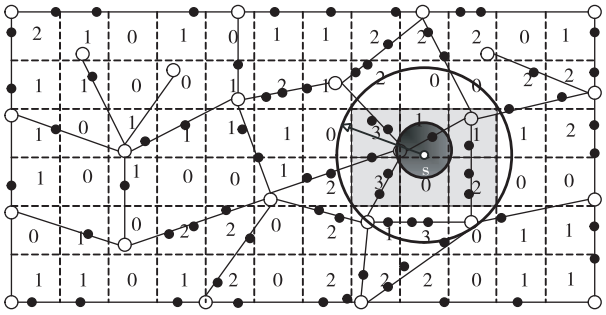


图 2 通信成本估算示意图

算法 2 自适应情境的查询点选择算法

1. $o \leftarrow$ 取得模糊区域 Q 中心点坐标
2. $r \leftarrow$ 取得模糊区域 Q 半径
3. $Serv \leftarrow$ 取得服务提供商地址
4. $con \leftarrow$ 取得消费用户查询内容
5. $SendMsg(Serv,con)$ //发送查询内容 con 至 $Serv$
6. $H \leftarrow$ 取得服务提供商返回查询结果估计空间直方图
7. $h \leftarrow getBin(H,o)$ //取得 Q 圆心所在的子空间
8. $\hat{m} \leftarrow 0$ //需求结果估计值
9. $m* \leftarrow$ 取得请求用户设定值
10. *do*
11. $d = dist(h,o)$ //取得 o 与 h 边界的 8 个方向距离
12. $ifd < 3r$ //判断距离条件是否满足
13. $h = Enlarge(H,h,d)$ //沿不满足距离条件方向扩展 h
14. *endif*
15. $\hat{m} \leftarrow Count(h)$ //取得需求空间估计值
16. $until(d \geqslant 3r)and(\hat{m} > m*)$
17. $bins \leftarrow check(h,0)$ //剔除 h 中统计值为 0 的子空间
18. $o \leftarrow MBC(bins)$ //求取需求空间最小外接圆圆心 o
19. $returno$ //返回查询点

3 隐私保护方法分析

终端在服务消费过程中,其位置信息存在两种途径被攻击者截取:一是攻击者伪装成代理用户窃取消费请求;二是攻击者在服务器端截取查询请求以及返回结果。下面分析这两种途径下用户的隐私保护情况。

(1)攻击者伪装为代理用户服务

假设攻击者伪装成代理用户截取消费用户的消费请求,攻击者据此可解析出消费用户设置的模糊区域 Q、查询内容等信息,由于用户真实位置已被模糊化,如图 3 所示,攻击者无法准确推测用户真实位置,故其位置隐私得到保护,但就隐私保护效果而言, Q 中包含交叉路口可提高隐私保护强度。

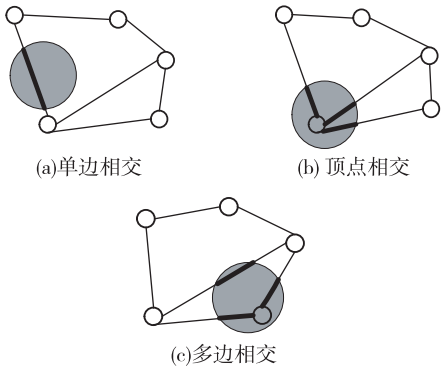


图 3　代理端攻击模式下用户隐私区域

(2)攻击者在服务器端截取信息

假设攻击者在服务器端截取到查询请求以及返回结果,则可从查询请求数据包中解析出查询内容以及查询点等信息。对于代理用户而言,该信息与其隐私无关,故代理用户不存在隐私泄露。对于消费用户而言,攻击者无法与其直接关联,故其身份信息以及查询内容得到保护。而对于位置隐私,假设攻击者已知查询点 s 以及 POIs 有序结果集 $\{p_1,p_2,p_3,\cdots,p_m\}$,根据增量查询结束条件[6]可知,查询结果集满足下式:

$$\begin{cases} dist(Q,s)+\min\limits_{1\leqslant i\leqslant(m-1)}^{k} dist(Q,p_i) > dist(s,p_{m-1}) \\ dist(Q,s)+\min\limits_{1\leqslant i\leqslant m}^{k} dist(Q,p_i) \leqslant dist(s,p_m) \end{cases} \tag{4}$$

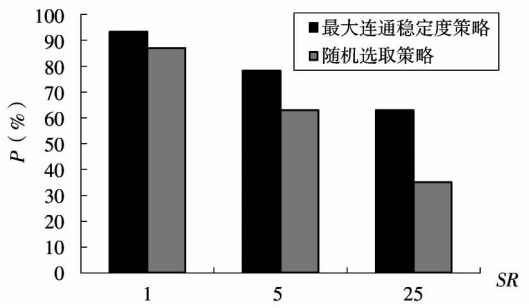
上述表达式中 k 为消费用户期望的 POI 值在查询集合中顺序,攻击者无法确定具体 k 值,故该方程存在多个解,每个解均表示消费用户的一个隐私区域,因此,消费用户位置隐私也得到保护。

4　实验及结果分析

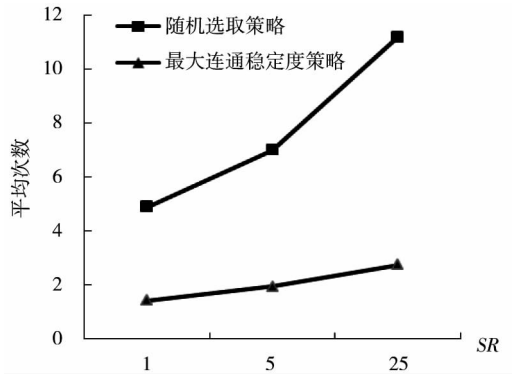
实验选用业界认可的 Thomas Brinkhoff 路网数据生成器[13]模拟不同交通状况条件下的用户服务查询环境,交通路网为 Oldenburg 城市道路地图(区域面积为 23.57km × 26.92km),5000 个 POIs 由模拟平台随机生成;实验机器配置为 Intel 双核1.73 GHz CPU、2GB 内存,本文算法均采用 Java 语言实现。

4.1　代理用户优化选择实验分析

实验中设置初始移动用户数为 1000,簇规模为 10,容忍查询时延为 5 个模拟时间,移动用户的运动路径为随机选取,用户间的速度比率 SR 为 1、5、25[13]($SR = V_{消费用户}/V_{代理用户}$)。假设消费用户在移动过程中以某个固定的概率发起查询,为了比较代理用户选取策略对通信开销的影响,在同一数据集上分别开展最大连通稳定度选取策略与随机选取策略实验比较,统计两种策略下消费用户与代理用户间的通信稳定性以及服务完成的平均连接次数,比较结果如图 4 所示。



(a)通信稳定性比较



(b)服务完成平均连接次数

图 4　最大连通稳定度与随机选取策略比较图

从图 4 的分析可知:在通信稳定性上,消费用户基于最大连通稳定度策略选择代理用户优于随机选择策略,且二者间的连通稳定性与速度比负相关,因此,前者保证了消费用户在不同交通路况下均能选择到可靠性高的代理节点,保证了查询过程中信息传输的稳定性。另外,从通信开销来看,由于随机选取策略导致二者间通信稳定性不高,因此消费用户需多次寻找代理用户才能完成一个完整的服务,故在总体通信开销上,最大连通稳定度选择策略远远低于随机选取策略,因此前者既保证了通信的稳定性,也减少了系统的整体通信开销。

4.2　代理用户自适应情境的查询实验分析

代理用户在增量近邻查询中,为了比较查询点位置选取对算法性能以及通信开销的影响,实验中

分别采用自适应情境的查询点确定算法与随即选择策略进行对比。前者实验参数为:用户模糊区域半径 r 为 100、500、1000(单位:m)、路网空间划分子域尺度为 1:500;查询点随机选取策略中实验参数为查询锚点与模糊区域距离 l,值为 0、100、250、500、1000(单位:m)。每组参数设置均运行 100 次,统计两种方法下算法的供应空间 *Sct* 平均值(单位:个)、需求空间 *Dct* 平均值(单位:个)以及服务器端查询时间 *Qt*(单位:ms)、代理用户处理时间 *Pt*(单位:ms),运行结果如表 1 和表 2 所示。

表 1 查询点随机选取策略的查询统计结果表

r(m)	l(m)	*Sct*(个)	*Dct*(个)	*Qt*(ms)	*Pt*(ms)
100	0	12	6	23.20	0.23
	50	17	6	22.68	0.42
	100	25	6	22.72	0.39
	250	32	6	24.43	0.59
	500	53	6	23.12	1.21
	1000	95	6	25.42	2.24
500	0	175	68	27.05	4.98
	50	192	64	26.84	4.54
	100	212	72	28.35	5.65
	250	246	74	28.50	6.87
	500	274	77	27.57	5.70
	1000	358	71	32.14	9.09
1000	0	683	248	40.06	18.22
	50	678.	252	39.32	17.40
	100	698	263	40.68	18.32
	250	719	234	40.30	20.00
	500	765	239	44.88	21.04
	1000	921	241	49.91	25.57

表 2 查询点自适应算法的查询统计结果表

r(m)	*Sct*(个)	*Dct*(个)	*Qt*(ms)	*Pt*(ms)
100	33	6	25.23	25.79
500	178	70	27.65	30.08
1000	392	242	39.30	39.89

从表 1、表 2 的数据分析可知:在供应空间数量上,查询点自适应算法较随机选取策略优势明显,而二者在需求空间数量上几乎一致。这是因为增量近邻查询算法的需求空间只与模糊区域大小相关,而与查询点距离远近无关,而供应空间除了与模糊区域大小正相关外,还与查询点距离成正比,因此,查询点自适应算法在保证用户隐私保护效果的同时,有效降低了系统通信开销。

从算法的计算效率来看,随机选取查询点策略算法比自适应情境的查询点选取算法更优,由于后者在查询之前需根据空间直方图确定查询点位置,因此产生了额外的计算开销,但从服务的整体过程来看,后者减少了供应空间数量,缩短了代理用户与服务器间的通讯开销,因此在整个服务过程中,后者的计算时间开销并不比前者差。

5 结 论

本文针对移动终端在缺乏固定通信基础设施的实际路网环境中消费 LBS 服务时位置隐私保护问题,提出了一种用户协作的位置隐私保护方法,同时为了提高系统服务质量,通过最大连通稳定度理论实现了代理用户的优选,保证了用户间通信的稳定性;引入通信成本估算方法实现了查询点的自适应选择,控制了增量近邻查询中的通信开销,实现了隐私保护度与服务质量间的平衡。但由于消费用户的隐私保护意识容易因身份、环境、应用不同而改变,具有较强主观性、个性化、随意性,因此如何根据查询情境,提高隐私保护方法在消费用户端的自适应性将是下一步研究重点。

参考文献

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

[2] 周傲英,杨彬,金澈清等. 基于位置的服务:架构与进展. 计算机学报,2011,34(7):1155-1171

[3] GedikB, Liu L. Location privacy in mobile systems: A personalized anonymization model. In: Proceedings of the 25th International Conference on Distributed Computing Systems, Columbus, USA, 2005. 620-629

[4] 薛姣,刘向宇,杨晓春等. 一种面向公路网络的位置隐私保护方法. 计算机学报,2011,34(5):865-878

[5] 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. 计算机研究与发展,2010,47(1):121-129

[6] Man L Y, Christian S J, Jesper M, et al. Design and analysis of a ranking approach to private location-based services. *ACM Transactions on Database Systems*, 2011, 36(2):1-46

[7] Aniket P, Yu W, Zhang N. A context-aware scheme for privacy-preserving location based services. *Computer Networks*, 2012, 56(11): 2551-2568

[8] Chow C, Mokel M, Liu X. A peer to peer spatial cloaking algorithm for anonymous location based service. In: Proceedings of 14th ACM International Symposium on Geographic Information Systems, Arlington, USA, 2006. 171-178

[9] Ghinita G, Kalnis P, Skiadopouslos S. MobiHide: A mobile peer to peer system for anonymous location based queries. In: Proceedings of the 10th International Symposium Advances in Spatial and Temporal Databases, Boston, USA, 2007. 221-238

[10] 徐建,黄孝喜,郭鸣等. 动态 P2P 网络中基于匿名链的位置隐私保护. 浙江大学学报(工学版),2012,46(4):712-718

[11] 黄毅,霍峥,孟小峰. CoPrivacy:一种用户协作无匿名区域的位置隐私保护方法. 计算机学报,2011,34(10):1975-1985

[12] 毛典辉,蔡强,李海生等. 协作代理增量查询的 LBS 隐私保护方法. 华中科技大学学报(自然科学版),2013,41(3):73-77

[13] Brinkhoff T. A framework for generating network based moving objects. *GeoInformatica*, 2000, 6(2): 153-180

A collaborative LBS privacy protective method in road-network

Mao Dianhui, Cai Qiang, Li Haisheng, Cao Jian, Huang Jinhui, Cheng Yeliang

(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048)

Abstract

A new user collaboration-based location privacy-preserving method for location based service (LBS) users is presented to realize the location privacy protection for mobile users when mobile devices are in the road networks without enough fixed communication infrastructures. According to the method, mobile users are self-organized as a P2P network according to the Chord protocol, and a user consuming LBS services blurs his real position and selects a proxy user from the P2P network for forwarding incremental nearest-neighbor query. Meanwhile, in order to improve the QoS(quality-of-service) of the method, the maximum connect-stability theory is proposed to select an optimized proxy user for stable communication of the network, and a communication-cost estimation method is introduced to determine the query-anchor for eliminating communication overhead. The theoretical analysis and experimental results validate the technique's effectiveness in LBS accuracy, privacy protection and communication QoS.

Key words: P2P network, location based service(LBS), location privacy, network security, privacy protection