doi:10.11959/j.issn.1000-436x.2016212

面向移动社交网络内容分享的位置隐私保护方法

李超[1]，殷丽华[1]，耿魁[1]，方滨兴[2]
（1. 中国科学院信息工程研究所信息安全国家重点实验室，北京 100093；
2. 东莞电子科技大学电子信息工程研究院，广东 东莞 523808）

摘 要：针对移动社交网络内容分享中内容参与者的位置信息泄露问题，提出一种面向内容分享的位置隐私访问控制模型，细粒度控制分享内容中所涉及用户的敏感位置信息的访问，设计一种针对位置隐私设置的 *k*-匿名算法，保证设置的敏感位置信息在内容分享服务提供商的服务器端不被推测，给出一种基于位置敏感度的位置偏移算法，以平衡隐私与服务质量。最后通过仿真实验验证该方法的有效性。

关键词：位置隐私；隐私计算；隐私保护；社交网络隐私
中图分类号：TN929 **文献标识码**：A

Location privacy preservation approach towards to content sharing on mobile online social network

LI Chao[1], YIN Li-hua[1], GENG Kui[1], FANG Bin-xing[2]
(1. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
2. Institute of Electronic and Information Engineering, University of Electronic Science and Technology of China in Dongguan, Dongguan 523808, China)

Abstract: A privacy access control model for content sharing was presented to fine-grained control users' location information associated with sharing content in mobile social network. A *k*-anonymity privacy algorithm for privacy settings was given to protect against inference attack on a content sharing service provider server. To balance the privacy and quality of service, a location shifting method was presented. Finally experimental results demonstrate the validity and practicality of the proposed approach.

Key words: location privacy, privacy computing, privacy preservation, mobile social network

1 引言

随着 3G/4G 网络的迅猛发展和移动设备的广泛普及，移动社交网络应用已成为人们生活中的重要部分。移动社交网络是传统的基于位置的服务与在线社交网络服务的结合，使用户通过移动设备进行社交和内容分享。它目前已有许多不同的应用，如 Loopt、Buzz、Facebook 等交友应用，Foursquare 和 Gowalla 等签到应用，Twitter 和微博等结合位置信息的内容分享应用。这些应用服务可以用于查找周围的朋友、在用户间分享带位置信息的照片、视频和文字，使用户了解身边发生的事件。移动社交网络为互联网用户提供了便捷的人机交互接口，使人们可以和朋友间分享多媒体信息（如照片、视频）变得简单。移动社交网络是传统社交网络服务在移动环境下的扩展和延续，被认为是未来社交网络服

收稿日期：2016-08-12；**修回日期**：2016-09-22
通信作者：耿魁，gengkui@iie.ac.cn
基金项目：广东省产学研合作基金资助项目（No.2016B090921001）；国家自然科学基金—广东联合基金资助项目（No.U1401251）；国家高技术研究发展计划（“863”计划）基金资助项目（No.2015AA016007）；国家“核高基”基金资助项目（No.2015ZX01029101）

Foundation Items: The Industry-University-Research Cooperation Project of Guangdong Province (No.2016B090921001), The National Natural Science Foundation of China-Guangdong Union Foundation (No.U1401251), The National High Technology R&D Program of China (863 Program) (No.2015AA016007), Core Electronic Devices, High-end General Purpose Chips and Basic Software Products (No.2015ZX01029101)

务发展的重要趋势。以 Facebook 为例，2015 年 Facebook 每月活跃在线用户已超过 14.9 亿，平均每天在线用户为 9 680 万[1]。

但是，当用户在分享内容、共享位置信息的同时，用户隐私泄露也不可避免。大量文献指出通过暴露的位置信息，可以确定用户身份[2]。隐私信息可能会被攻击者或恶意用户利用，造成无法估量的损失[3,4]，这使用户开始关注自身的隐私保护。2010 年，文献[5]指出 55%的 LBS 用户表现出对位置隐私泄露的忧虑，文献[6]指出，在社交网络上有个人档案的美国公民中，有 50%的人对隐私非常关注。

事实上，自从 LBS 的概念被提出，位置隐私问题就是一个热点学术问题，研究者提出了大量位置隐私保护方法[7~11]。但是，移动社交网作为 LBS 的一种特例，它的位置信息隐私保护需求与 LBS 不同。传统 LBS 中，用户的目的是获得位置相关的服务，在主观上并不主动暴露位置，而在移动社交网中，用户自愿地分享位置信息，且分享的位置信息可以通过社交网络进行传播。这一区别给用户隐私保护带来了新的挑战。

已有的内容分享隐私保护方法主要针对信息发布者的隐私信息。而实际上，当用户分享的内容中包含其他用户的相关信息，且这些信息其他用户并不想公开，则这将造成其他用户在不知情的情况下的隐私泄露。如 Alice 和 Bob 是好友，Alice 在移动社交网络应用中分享了一张和 Bob 在酒吧的合影，而 Bob 并不希望自己出现在“酒吧”等场所，Alice 的行为无意中对 Bob 的位置隐私带来了损害。本文所要解决的问题就是对发布内容中非发布者的位置信息进行隐私保护。

本文提出一种针对内容分享中用户的隐私保护方法，避免用户在不知情的情况下，被他人分享的内容中暴露其位置隐私；提出一种面向内容分享的位置隐私控制模型，支持用户对敏感位置的访问控制；提出一种基于位置敏感度的位置生成方法，根据访问者的不同权限，给出分享内容的不同位置信息，保护内容中用户的位置隐私；提出一种针对内容分享服务商（CSSP, content sharing service provider）的用户敏感位置配置信息的隐私保护方法，防止用户配置的敏感位置信息被不可信的 CSSP 获得。

2 相关工作

面向移动社交网络内容分享的位置隐私保护研究刚刚起步，将传统的 LBS 位置隐私保护方法引入到移动社交网络中是一种主要思路。本文的方法结合了访问控制和 LBS 位置隐私保护。

2.1 LBS 位置隐私保护

学术界在面向 LBS 的位置隐私保护方面已开展了大量工作，文献[12, 13]对这些研究工作进行了梳理。经典的位置隐私保护方法主要有 2 类：基于匿名的方法和基于模糊化的方法。

Gruteser 和 Grunwald 最早提出了位置 *k*-匿名的概念[8]，并设计了一种算法调整位置信息的解析，使所得的位置区域满足 *k*-匿名的需求，从而使攻击者无法对一个区域内的 *k* 个用户进行区分。Bettini 等[14]提出了一个评估框架来评估敏感位置信息被暴露的风险，Bettini 等认为，用户提交的请求中所包含的位置信息的历史记录，可以看作是一个准鉴别器，因为一组属性集可以和外部信息关联，以减少用户身份的不确定性。Mokbel 等[15]提出了一个框架，在该框架中每个用户通过 *k* 匿名参数及位置信息的最小可接受解析区域 $A_{\min}$，定义其隐私偏好。框架包含位置匿名器及隐私感知的查询处理器。前者用于扰乱用户的位置信息实现其隐私，后者用于匿名查询串的管理。

基于模糊化的技术是通过降低位置信息的准确性来提供隐私保护。Hong 等利用地标信息来代替用户基于坐标的地理位置信息，使用户暴露更少的准确位置信息。但是，若用户位置信息的准确性太低，对服务质量会产生严重影响，因此，Duckham 和 Kulik 提出了一个位置模糊化协商框架[16]，通过协商算法来平衡服务质量和位置隐私间的矛盾。由于不同用户对同一位置的隐私敏感程度不同，因此，针对不同用户的隐私偏好应生成不同准确程度的位置信息，Damiani 利用语义感知的模糊化技术[17,18]，通过对用户不同位置敏感程度的设定，来生成满足用户隐私需求的模糊位置信息。

2.2 隐私信息访问控制

为在照片分享服务中，对照片浏览者访问的照片进行限制，Klemperer 等[19]通过对用户分享的照片进行语义标注，并基于这些标注设置访问控制规则。但该方法仅适用于内容分享者对自己空间中的内容进行隐私设置，而无法保证其他内容中涉及用户的隐私。

针对非意愿隐私泄露的可能，Xu 等[20]提出了一种基于多用户协商决定图片分享访问权限的隐

私保护机制，通过人脸识别技术，对上传的图像进行分析识别，找出图像中的用户，通知并提醒这些用户与上传用户共同协商决定图像的发送、评论的访问权限，但该方法需要人为参与决策过程。Ilia 等[1]将 Facebook 中的人脸标记功能与人脸识别结合起来，通过识别出的人脸自动关联到 Facebook 中用户的访问控制策略，实现了基于人脸的细粒度访问控制机制，并将其与现有的应用系统很好地结合起来。但该方法未考虑同框用户的隐私策略冲突问题。Wishart 等[21]提出了一种方法来协商定义隐私策略。该方法基于隐私策略的协同定义，参与到协同定义中的各方可以定义强隐私和弱隐私偏好，并定义了隐私语言通过强弱条件来描述用户偏好，基于此进行隐私冲突检测。但是，这一方法无法通过自动方法来解决冲突，仅为用户提供一些建议。

针对移动社交网络的特点，文献[22~24]提出了基于关系的访问控制方法，这些新的访问控制方法以用户的关系作为核心概念，目的是更好地解决社交网中的信息共享。大量研究[25,26]证实用户关系是驱动用户暴露个人信息的主要因素，在定义社交媒体访问控制机制时，这些关系应当扮演重要角色。在基于关系的访问控制模型中，隐私策略不是基于用户个体或角色，而是基于用户与其他用户的关系紧密程度。

3　问题描述与方法框架

在内容分享应用中，用户可以通过设置隐私策略来限定特定内容的访问，如微信中，用户可以通过指定分组或特定用户，来限定朋友圈的分享内容是否对这些分组或用户可见。但这类方法仅能保护内容发布者自己分享内容的隐私，而无法保护分享内容中涉及的用户的隐私。即对某个用户 u 而言，若其他用户在 u 不知情或不愿意的情况下分享了内容 m，且 m 中涉及了 u 的隐私，以上方法将无法保护 u 的隐私不被泄露。如在照片分享应用中，用户 u_1、u_2 是好友关系，u_1 在自己的空间中分享了一张和 u_2 的合影。用户 u_2 是 u_1 的好友，u_2 可以浏览这张照片。假设 u_2 不希望其照片或照片的位置信息被陌生人看到，则用户 u_1 分享的内容，侵犯了用户 u_2 的隐私。现有上传者指定分享内容访问权限的方法，无法保护 u_2 的隐私。本文的方法旨在解决这一问题。

3.1　内容与用户

在移动社交网络中，用户通过移动设备在朋友圈、微博等应用中与他人分享照片、文字、签到数据等信息，本文将以上信息统一抽象称为分享内容，或在不引起歧义的情况下简称为内容，用 $M=\{m_1,m_2,\cdots,m_n\}$ 表示内容 $m_i(1\leqslant i\leqslant n)$ 的集合。

$U=\{u_1,u_2,\cdots,u_n\}$ 表示移动社交网络中的用户 $u_i(1\leqslant i\leqslant n)$ 的集合，用户根据在内容 m 分享中的角色可分为发布者、参与者。当用户 $u\in U$ 将内容 m 分享到移动社交网络中时，称 u 是内容 m 的发布者，记为 u_p，u_p 分为内容的上传者、转发者。

为表示内容 m 中涉及的用户集合，引入一个从内容到用户的映射函数 $Extract{:}M\rightarrow 2^U$。$Extract(m)=\{u_1,u_2,u_3\}$ 表示内容 m 中含有用户 u_1、u_2 和 u_3。本文中 *Extract* 是一个抽象函数，它根据实际内容的类型，实例化为具体的用户识别方法。如当内容 m 的类型为照片时，*Extract* 实例化为人脸识别方法[1]，通过人脸识别方法，提取照片中包含的用户。

定义 1　（参与者）。给定内容 $m\in M$，称 $U\subseteq Extract(m)$ 为内容 m 的参与者。

在用户 u_i 分享的内容 m 中，可能包含除 u_i 外的其他用户 $U'\subseteq U$，当 m 被发布时，U' 中的用户可能在不知情的情况下被提及，若内容 m 中包含了 U' 中的隐私信息，那么，u_i 在 U' 中用户不知情的情况下，非恶意地造成 U' 中用户的隐私泄露。本文称内容 m 中除 u_i 外的其他用户 $U'\subseteq U$ 为共现用户。

定义 2　（共现用户）。给定 u_p 是移动社交网络中内容 m 的发布者，称 $U_C=Extract(m)\setminus u_p$ 为内容 m 的共现用户集，$u_c\in U_C$ 为内容 m 的一个共现用户。

U_C 是一个用户 U 的子集，在不引起歧义的情况下，用 $U_C(m)$ 表示内容 m 的共现用户集。由于对所有的 $u_i\in U_C(m)$，u_i 的上下文特征都来自内容 m，因此，所有的 $u_i\in U_C(m)$，具有相同的时空信息。

3.2　时空表示

移动社交网络中，用户通过智能手机等带有定位功能的移动设备，分享带有位置信息的内容。位置信息包括用户的物理位置信息（如 GPS 坐标），以及标注的位置语义信息（如餐厅、酒吧等 POI 信息）。用户的物理位置用 l 表示，l 可以是一个二维的点用以描述 GPS 坐标，或者是一个包含该点的区域。$L=\{l_1,l_2,\cdots,l_n\}$ 表示移动社交网络的位置空间。对每个位置，不同用户可能有不同的语义标注类

型，如给定位置 l、用户 u_i 和 u_j，l 可能是用户 u_i 的家庭住址，而对 u_j 而言，l 可能是其工作地点。不难发现，位置标注对用户隐私而言是敏感的。用户位置语义标注的集合记为 $A=\{a_1,a_2,\cdots,a_n\}$。物理位置和位置语义标注存在多对多的映射关系。

定义 3 （位置包含关系）。若位置 l_i 被位置 l_j 所包含，记为 $l_i \leqslant_l l_j$。

为了描述用户在给定时间内的隐私偏好，利用日历（calendar）的概念[27]来描述时间。一个日历是一个由连续间隔构成的可数集。令 CT 为一个日历时间集。通常，2 个日历时间是可比较的，对于 CT 中任意的 ct_x 和 ct_y，有 $ct_x \leqslant ct_y$ 或 $ct_y \leqslant ct_x$。

定义 4 （时间包含关系）。若时间 ct_i 在时间 ct_j 内，则 ct_i 和 ct_j 满足时间包含关系，记为 $ct_i \leqslant_t ct_j$。

3.3 敏感度与隐私度量

在发布的内容中，位置的敏感度与用户特征、时空特征相关，主要体现在以下几点。

1) 位置信息对不同用户而言，敏感程度并不相同。如用户 u_i 和 u_j 出现在分享内容 m 中，假设 $a(m)$=医院，u_i 的职业是医生，u_j 是一名商业领袖或政治人物，显然 u_j 对医院的隐私敏感程度要远高于 u_i。

2) 同一位置信息对同一用户而言，敏感程度也可能因时间的不同而不同。如用户 u_i 不希望在非周末时间出现在 $a(m)$=酒吧的内容 m 中，显然对用户 u_i 而言，在周末和非周末的午夜，酒吧具有不同的隐私敏感程度。

3) 同一用户的同一位置信息，对不同的内容访问者而言，敏感程度可能不同。如用户 u_i 是一名学生，他的朋友分享了 $a(m)$=网吧的内容 m，u_i 出现在 m 中。u_i 不希望他的老师知道他的位置，显然 u_i 对朋友和老师对网吧的敏感程度不同。

因此，用四元组$\langle u,l,ct,senL\rangle$表示用户 u 在位置 l 的敏感度为 $senL$。用函数 $sensitive$: $U\times L\times CT\rightarrow senL$ 表示用户对位置 $l\in L$ 的敏感度。如 $sensitive(Alice, l_1, ct_1)$=0.8 表示用户 Alice 在 l_1 和 ct_1 下的敏感度为 0.8。

3.4 攻击者假设

假设攻击者的目标是获得特定用户的敏感位置信息。当前移动互联网应用服务商除百度、阿里、腾讯等大型服务商外，还存在大量安全保障参差不齐的服务商。这些 CSSP 若被攻击，则会导致包括用户隐私策略信息在内的用户数据被泄露。因此本文假设，攻击者可以攻击 CSSP 服务器并获得服务器上的所有信息。基于此假设，若用户设置隐私策略不加保护，用户的隐私会遭受严重威胁。如用户 u 若在 CSSP 上设置对位置 l 进行了访问权限设置，则攻击者可直接根据该知识推测出 l 为用户 u 的敏感位置。

3.5 方法框架

方法框架如图 1 所示，当用户 u 上传分享内容 m 时，首先识别出内容 m 中的用户集 $U_C(m)$，查询每个用户 $u_i\in U_C(m)$的隐私设置，通过第 4 节中提出的隐私访问控制模型，根据内容 m 的上下文环境，对每个共现用户 $u_i \in U_C(m)$的隐私配置进行策略评估，合并每个策略的决策结果，判断用户 u 是否对内容的位置可见。若可见，返回用户对位置的敏感度，并根据敏感度，通过第 6 节中提出的位置生成方法，生成内容 m 的偏移位置，返回给浏览者。同时，为保护用户设置的敏感位置不被 CSSP 非法获取，对用户设置的敏感位置进行隐私保护。

4 隐私信息控制模型

为保护分享内容 m 中涉及的用户 $u\in Extract(m)$

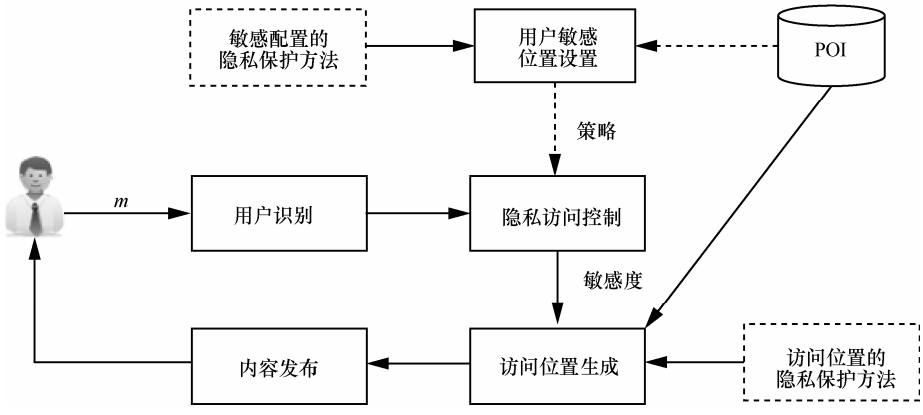


图 1 内容分享位置隐私保护方法框架

的敏感位置不被非授权的浏览者获得，提出一种面向内容分享的位置隐私保护模型 CS-LPPM，该模型在基于关系访问控制的基础上，引入内容上下文约束。CS-LPPM 根据访问者 u_a 请求访问的内容 m 中所提及参与者 $u\in Extract(m)$的隐私配置，判断 u_a 是否能获得 m 的位置信息。

4.1　CS-LPPM 模型描述

CS-LPPM 基于访问者 u_a 与内容参与者 $u\in Extract(m)$的关系，以及内容 m 的时空、共现上下文约束，判断 u_a 对 m 的位置服务质量。CS-LPPM 模型如图 2 所示。

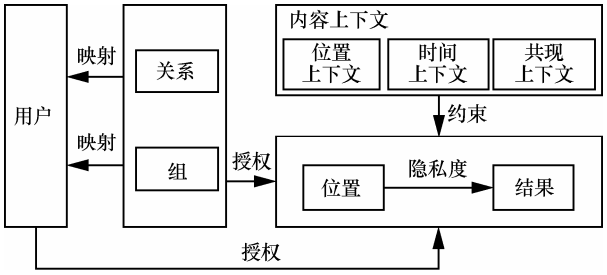


图 2　CS-LPPM 模型

其基本元素定义如下。

$U=\{u_1,u_2,\cdots,u_n\}$是用户集合。

$RT=\{r_1,r_2,\cdots,r_n\}$ 是用户间关系类型的集合，如 $RT=\{\text{friend, family}\}$。

$int:U\times U\to[0,1]$表示 2 个用户间亲密度，如 $int(u_1,u_2)=w$ 表示用户 u_1 和 u_2 间的亲密度为 w。

r: $U\times U\to R$ 表示 2 个用户间关系的映射函数，如 $r(u_1,u_2)$表示用户 u_1 和 u_2 存在关系 R。

$G=\{g_1,g_2,\cdots,g_n\}$表示用户建立的用户组集合。

$g:U\to G$ 是一个从用户到用户组的函数映射，表示用户属于的用户组，如 $g(u)$表示 u 属于的用户组为 G。

$LC\subseteq\langle 2^L,2^A\rangle$：位置上下文约束，表示内容 m 中的位置信息需满足的物理位置或语义位置的约束条件。

$TC\subseteq 2^{CT}$：时间上下文约束，表示内容 m 中的时间信息需满足的时间约束条件。

$UC\subseteq 2^U$：用户上下文约束，表示内容 m 中设计的用户需满足的约束条件。

$Effect$:{show, off}。

$Permission:L\times Effect$ 权限函数，是位置到效果的映射关系，如(l_1,show) 表示 l_1 被授权为 show。

$PA:U\to Permission$ 授权访问者访问位置的权限。

在社交网络数据共享中，关系类型和用户分组是 2 个重要特征。因此，模型需要支持用户描述关系类型和用户分组。Rong 等[23]利用关系类型对内容的可见权限进行授权，通过判断用户是否属于指定的关系类型，来决定用户是否有权浏览内容。但通常内容的所有者对同一关系类型和用户组里的用户有不同程度的分享意愿。大量证据显示，在社交网络中，一个用户向另一个用户分享特定内容的意愿，取决于这 2 个用户间的关系亲密程度[26,28,29]。亲密程度是指直接相连的用户的关系强度。因此，本节利用文献[30]中的亲密度阈值思想，支持用户基于不同的分享意愿设置访问控制策略。

例 1　给定社交网络中的一个用户集 $U=\{u_1,u_2,u_3,u_4,u_5,u_6\}$，$r_1=$ friend，$r_2=$ family 是 U 中用户间的关系类型，l 是内容中包含的位置信息，用户 u_1、u_2 与其他用户的关系类型及亲密度如表 1 所示。假设用户 u_1 设置 l 的隐私策略，规定对朋友中亲密度大于 0.8 的用户或家人中亲密度大于 0.2 的用户，位置 l 的敏感度为 s_1=0.3；对朋友中亲密度小于 0.5 的用户或家人中亲密度小于 0.2 的用户，位置 l 的敏感度为 s_1=0.8。若 u_3 请求访问内容 m，其中，$Extract(m)=\{u_1\}$，$loc(m)=l$，则 u_3 访问 m 时 l 的敏感度为 0.8。

表 1　　用户关系亲密度示例

l	$r_1=$ friend				$r_2=$ family	
	u_1	u_2	u_3	u_4	u_5	u_6
u_1	1	0.8	0.4	0	0.4	0.7
u_2	0.5	1	0.7	0	0.5	0.3

在 CS-LPPM 中，关注 3 种类型的内容上下文：位置、时间及共现用户，表述内容 m 产生的时空信息和内容中涉及的参与者。定义该上下文约束 $Constraint$ 为三元组$\langle LC,TC,UC\rangle$。

定义 5　(FPMAC 策略)。隐私策略为一个三元组 $P=\langle U,Constraint,Effect\rangle$，其中，$U$ 表示访问内容的用户集，$Constraint$ 是内容的上下文约束，$Effect$ 是请求的结果。

每一个用户 u 包含一组隐私策略 $P=\{p_1,p_2,\cdots,p_n\}$，每一条策略 $p_i\in P$ 规定了与 u 相关的内容 m（$u\in Extract(m)$）的位置信息 $loc(m)$，在哪些约束下以何种精确度可见。

4.2 隐私策略匹配算法

当用户 u 访问某个分享内容 m 时，u 向 CSSP 发起访问请求，CSSP 接收到 u 的请求后，通过算法 1 评估 u 是否满足请求内容 m 中相关用户的隐私配置，返回内容 m 相对于 u 的敏感度 $senL$。

算法 1 隐私策略匹配算法

输入 访问者 u, 请求浏览的内容 m

输出 m 的位置敏感度$\langle loc, senL\rangle$

1) $U_C \leftarrow Extract(m) \setminus u_p$

2) foreach $u_i \in U_C$ do

3) $PU_i \leftarrow get\_policy(u_i)$

4) $DV_i \leftarrow evaluate(re,p)$

5) end foreach

6) $DV_m \leftarrow \sum_{i=1}^{n} w_i DV_i$

7) if *Decision* = allow then

8) $\langle loc, senL\rangle \leftarrow \langle LC(m), p(senL)\rangle$

9) end if

10) output $\langle loc, senL\rangle$

定义 6（内容访问请求）。$Request = (u,m, \langle location,time,co\text{-}occurance\rangle) \in U \times M \times Contraint$，其中，$U$ 表示用户，M 表示访问的内容，$Contraint$ 表示内容的上下文约束。

定义 7（策略判定函数）。策略判定函数，$evaluate: Request \times P \rightarrow \{allow, deny\}$通过访问请求和用户策略的评估，判断是否允许访问。评估方法可描述为

$$evaluate(re,p)=\begin{cases} allow, & r(u,u_p)\in p.R \wedge \\ & int(u,u_p)\geqslant w \wedge \\ & LC(m)\leqslant_l pl \wedge \\ & TC(m)\leqslant_t pl \wedge \\ & UC(m)\cap pu_C \neq \varnothing \\ deny, & 否则 \end{cases} \tag{1}$$

其中，$re \in Request, p \in P$。

判定过程如下。当 CSSP 收到访问请求，它通过 $Extract(m)$提取参与者用户 U，对用户 u 的每个策略 p，检查请求者是否满足 p 中的用户关系，再根据 m 产生的位置、时间元数据信息和 m 中共现用户，判断是否满足 p 中的时空上下文和共现用户上下文约束。

由于多个用户出现在内容中时，每个用户都有自己独立的访问控制策略。这些策略可能会发生冲突。如 2 个用户 u_1、u_2 在 m 中共现，u_1 的策略规定其位置 l_1 不能被陌生人浏览，u_2 的策略规定 l_1 可以被任何人浏览。假设一个访问者 u 访问 m 时，u_1 和 u_2 的策略发生冲突。

本节通过一种基于亲密度投票方法来实现多方决策冲突的消解。

定义 8 （投票函数）。给定一个策略 p，$evaluate(re, p)$表示策略 p 对访问请求 re 的判定结果，投票函数 DV 通过策略判定定义如下

$$DV=\begin{cases} 1, evaluate(re,p)=allow \\ 0, evaluate(re,p)=deny \end{cases} \tag{2}$$

通常内容 m 的参与者与访问者间具有不同的亲密度，使参与者的策略对访问者的重要程度不同，决策过程中通过亲密度作为决策时的权重，来计算最终的决策值。

定义 9 （基于亲密度投票的冲突消解）。用 w_i 表示访问者 u 与共现用户 u_i 间的亲密度，带权重的投票函数为

$$DV_m=\sum_{i=1}^{n} w_i DV_i \tag{3}$$

假设敏感度阈值为 st，定义冲突消解如下

$$Decision=\begin{cases} allow, DV_m \geqslant st \\ deny, 否则 \end{cases} \tag{4}$$

5 面向位置隐私偏好设置的 k-匿名算法

在隐私模型中，CSSP 需根据用户 u 提供的位置信息生成相应的位置隐私策略，用于判断用户浏览的内容 m 是否满足用户的隐私设置 p。由于假设 CSSP 不完全可信，当用户直接将所设置的准确位置 l 发送给 CSSP 时，直观上 CSSP 可认为 l 可能是用户不想暴露的敏感位置。因此，若不对 l 进行隐私保护，则 CSSP 可直接获得用户所有的敏感位置。

本文采用 k-匿名的思想，在移动端生成多个候选位置，并将基于候选位置生成的候选区域发送给 CSSP，使 CSSP 无法准确获得用户设置的敏感位置信息。

与经典的 k-匿名方法[31]通过获得用户 u 附近的 k–1 个用户的区域不同，本文的目的是在整个地图范围内，生成 k–1 个与位置 l 不可区分的位置，使 CSSP 无法推测用户的敏感位置。

通常在整个地图范围内，不同位置或区域内的历史内容数量存在差异，如 k=2 时，随机生成位置 l_1 和 l_2。如图 3(a)所示，在位置 l_1 共包含 40 条分享内容，而在位置 l_2 无分享内容（如 l_2 为不可达位置）。由于 CSSP 具有分享内容数量的知识，因此，CSSP 可判断用户提供的位置信息 l_2 是一个生成的假位置，从而推测出用户设置的敏感位置为 l_1。而图 3(b)中，l_1 和 l_2 处分享内容数量相等，从而无法准确推测出用户设置的位置是 l_1 还是 l_2。

10	20	32	l_2 0
20	40	40	20
20	l_1 40	30	30
30	20	10	10

(a) 随机k-匿名

0	0	10	10
10	l_1 10	10	10
40	20	40	30
0	0	l_2 10	10

(b) 最大熵k-匿名

图 3　随机 k-匿名与最大熵 k-匿名（$k=2$）

本文利用熵来度量匿名程度，它可看作是从所有候选位置中选出用户当前位置的不确定度。由于每个位置都存在一个内容分享的概率 p_i，且$\sum_{i\in\{1\cdots k\}} p_i = 1$。则从候选位置获得准确位置的熵 H 为

$$H = -\sum_{i=1}^{k} p_i \mathrm{lb} p_i \tag{5}$$

为获得最大的熵，使设置的位置 l 具有最大的不确定性，所有的 k 个位置具有相同的概率 $\frac{1}{k}$。

假设地图被划分为 $n\times n$ 的单元格，将位置 l 处进行内容分享的概率，记为 l 所处单元格 c_i 内的内容分享概率。

$$p_i = \frac{l\text{所在}c_i\text{的内容分享数}}{\text{总的内容分享数}} \tag{6}$$

其中，

$$\sum_{i=1}^{n^2} p_i = 1 \tag{7}$$

本文的架构中，CSSP 需要请求 POI 数据中心获取 POI 语义信息。因此，将 POI 数据中心在某点被 CSSP 请求查询的次数可视为该点共享内容的次数。

为达到最大熵值，所选位置单元 c_i 中分享内容的数量应当相等。算法 2 中给出最大熵 k-匿名单元的选择方法。

算法 2 的主要思想是生成一个含有 k 个位置区域的集合 DL_k。为使集合 DL_k 中的 k 个位置满足最大熵以保证最大的不确定性，算法将隐私设置中位置 l 处历史分享内容的数量 n 作为不可区分的基准，获取所有单元格中内容数量为 n 的单元格，并从中随机选取 k 个单元格，使选取的单元格在内容数量上与 l 的相同。

由于隐私配置的位置隐私保护目的是保证 CSSP 无法获得用户在服务器上设置的敏感位置，所选单元格间的距离应尽量分散。

算法 2　最大熵位置 k-匿名算法

输入　隐私配置位置 l,匿名参数 k

输出　假位置集合 DC_k

1) $m \leftarrow m\_count(l)$

//获取位置 l 所在单元格历史分享内容数量

2) $candidates \leftarrow getcells(n)$//获得所有数量为 n 的单元格

3) $k' \leftarrow 0$

4) $d_{\text{base}} \leftarrow dist(l, rand(candidates))$

5) while $k' < k$ do

6)　　for $c_i \in candidates$ do

7)　　$c_i \leftarrow rand(candidates)$

8)　　if　$dist(l, c_i) \geqslant d_{\text{base}}$ do

9)　　$k' \leftarrow k'+1$

10)　　$DC_k \leftarrow add(c_i)$

11)　　end if

12)　　end for

13)　　if　$k' < k$ do

14)　　$d_{\text{base}} \leftarrow dist(l, rand(candidates))$

15)　　end if

16) end while

17) output DC_k

6　基于位置敏感度的位置生成算法

当访问者 u 请求内容 m 时，算法 1 根据内容共现者的隐私策略，判断 m 的位置信息是否能够

被访问，若能够被访问则返回该位置的敏感度 *senL*。本节利用该返回的敏感度 *senL* 生成对应的偏移位置，随内容 *m* 返回给访问者，从而保护 *m* 的位置隐私。敏感度越低，返回的位置准确性越高，反之亦然。

由于位置 *l* 对应的 POI，可能跨越多个邻接单元格，如大型社区、公园等。若生成的偏移位置与原位置仍对应同一 POI，则用户的敏感位置无法得到有效保护。如用户 *u* 设置其家庭位置 *l* 为敏感位置，其对应的 POI 为某小区。若内容 *m* 为一张照片，照片上涉及用户 *u*，若 *u* 的策略规定访问者在浏览该内容时，位置 *l* 需进行偏移处理，以保护其位置隐私。若偏移的位置 *l′*对应的 POI 仍为该小区，则偏移的隐私保护目的并未达到。因此，本节的方法将考虑 POI 的多样性。

同时，假设浏览者具有相关背景知识，知道 *l* 周围的 POI 信息。返回的 POI 若与内容的背景上下文信息（如照片中的场景、文字中反映位置的关键字）存在较大差异，则浏览者推测返回的 POI 发生了错误，并能根据内容背景推测周围可能的 POI 信息。如偏移位置 *l′*对应的 POI 为某学校，而内容中的背景上下文为医院，浏览者发现 POI 发生偏差，并根据其先验知识，知道 *l′*附近只有一家医院，则可推测该内容的真实位置为 *l*。因此，本节的方法将考虑 POI 的相似性。

定义 10 (POI *n*-多样性)。POI *n*-多样性表示用户所在区域内的有兴趣点类型数量至少为 *n*。

定义 11 （POI σ-相似性）。给定 2 个 POI_i 和 POI_j，若 $sim(POI_i, POI_j) \leqslant \sigma$，则 POI_i 和 POI_j 是σ-相似性的，其中，$sim: POI \times POI \to \sigma$为相似度计算函数。

相似度函数 *sim* 可通过 POI 类别相似性、文本相似性等方法进行计算。

本节提出基于位置敏感度的位置生成算法，如算法 3 所示，以 *l* 为中心，按螺旋线方式不断生成区域 *region*(*l*)，判断 *region*(*l*)是否满足 *n*-多样性和σ-相似性。

算法 3 基于位置敏感度的位置生成算法

输入 位置 *l*, 敏感度 *senL*,多样性 *n*,相似性σ

输出 偏移位置 *l′*

1) $c_i \leftarrow current\_cell(l)$

2) $cloacked\_region \leftarrow \{c_i\}$

3) while $poi\_count\ (cloacked\_region) \leqslant 1 + senL \cdot n \wedge sim(poi(\text{foreach } c_i \in cloacked\_region, cell(l)) \leqslant \sigma$ do

4) $c_i \leftarrow c_i.neighbor$

5) $cloaked\_region \leftarrow cloaked\_region \cup c_i$

6) end while

7) $l' \leftarrow loc(c_i)$

8) output l'

7 安全性分析

本文的方法从两方面保护用户的位置隐私。一方面保护用户的隐私设置不被 CSSP 推测，另一方面保护用户设置的隐私位置不被非授权的用户访问和推测。

对 CSSP 而言，本文提出的最大熵 *k*-匿名方法将用户设置的隐私位置 *l* 生成了 *k*–1 个在概率上不可区分的假位置。针对每一个位置生成一个包含 *n* 个单元格的区域作为隐私配置里的位置。对每一隐私设置里的位置 *l* 生成 *k* 个区域后发送给 CSSP。由于本文提出的方法满足最大熵，即 CSSP 攻击者猜测出 *l* 所在区域的概率为 $\frac{1}{k}$，又由于在每一个假位置上生成了 *n* 个单元格的区域，因此，攻击者猜测出 *l* 的概率为 $\frac{1}{nk}$。

对访问者而言，对不满足访问控制策略的用户，将完全无法看到内容 *m* 的位置。对于满足访问控制策略，且将位置 *l* 的敏感度设置非 0 的配置，将返回隐私保护后的位置，这部分主要考虑了用户分享需求的服务质量与隐私保护间的平衡。返回的位置根据隐私配置中敏感度的不同，而产生不同偏差的位置。当敏感度最高时，产生位置与原位置间组成的区域内包含了 *n* 个不同的 POI，同时，产生位置的 POI 与原位置 POI 语义相似，攻击者无法通过关于 POI 的先验背景知识和内容背景推测用户的准确位置。

8 实验

为测试本文提出的位置隐私保护方法，本次实验所处的硬件环境为 Inter(R) i3、500 GB 硬盘、8 GB 内存、Win7 64 位系统。实验数据集为 Gowalla 的签到数据集[32]，该数据集包含 319 063 个用户在 2011 年 6 月前的 36 001 959 个签到位置/时间和好友关系，统计结果如表 2 所示。实验将 check-ins

数量作为分享内容的数量。由于数据集中不含 POI 的类型信息，为满足 POI 多样性和相似性的实验，本文另外下载了一个美国芝加哥的 POI 数据集[33]，共包含 211 141 条 POI 记录。实验选取 Gowalla 数据集芝加哥部分的数据进行测试。

表 2　　Gowalla 数据集统计

类型	Berlin	Chicago	London	S.F.
#users	5 510	13 852	17 112	21 591
#locations	15 528	38 505	63 466	66 142
#check-ins	238 972	490 998	942 877	1 545 407
avg.#check-ins/loc	15.39	12.75	14.86	23.36

首先通过实验验证最大熵 *k*-匿名的效果。位置利用 Geohash 进行索引，编码长度取 7 位，每一条签到内容对应单元格 c_i，并计算每一单元格内的内容数。1) 通过随机 *k*-匿名的方式选取假名位置[31]，并计算对应的熵；2) 随机产生的一个单元格 c_i，在与其内容数量相等的单元格中随机产生 *k*–1 个单元格，并计算对应的熵。进行 1 000 次实验，取熵的平均值作为实验结果，如图 4 所示。

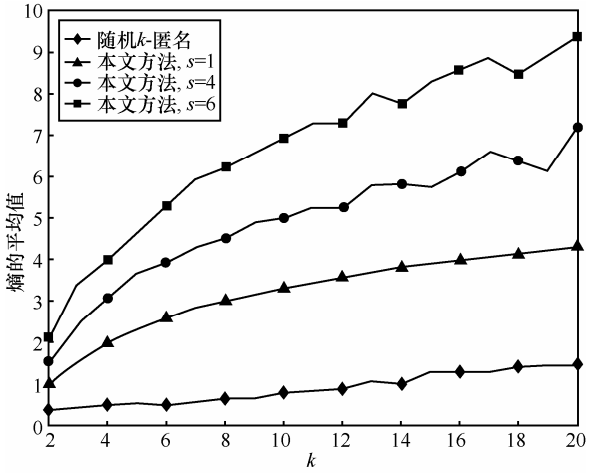


图 4　假名位置随机选取与本文方法的对比

实验结果表明本文方法选择的单元格，具有最大的熵，随着 *k* 的增加，沿最大熵的曲线可以看出，匿名效果越明显。实验中，随机 *k*-匿名的效果最差，这是由于其忽略了内容数量差异给攻击者推理带来的信息增益，攻击者通过过滤掉内容数量较少的假名单元格，可大大降低候选单元格的不确定性。

图 5 所示为位置偏移生成算法的性能实验结

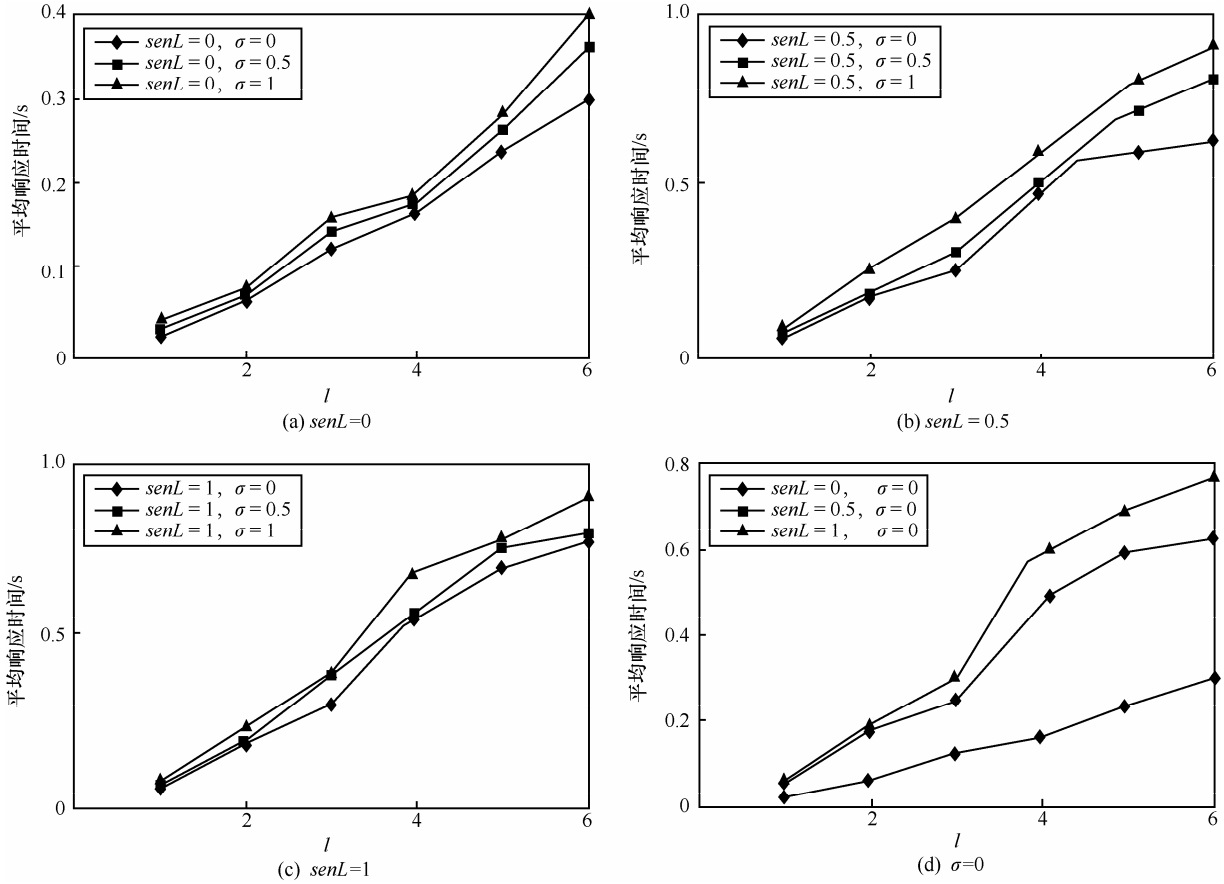


图 5　偏移位置生成性能

果，实验随机选取 100 个单元格，进行 100 次实验。每一次实验，对选取其中一个单元格进行位置偏移生成实验，每一个单元格的位置偏移实验，选取敏感度 $senL = 0$、$senL = 0.5$、$senL = 1$ 及$\sigma = 0$、$\sigma = 0.5$、$\sigma = 1$ 分别进行实验。

实验表明，当 POI 多样性相等时，位置的敏感度要求越高，所消耗的时间越多，这是由于随着敏感度的增加，算法需更多次迭代生成更大的偏移区域，并从中选择偏移点，以保护敏感位置的隐私，因此耗时会有所增加。随着 POI 多样性的增加，算法的时间增加，这是由于算法为保证候选区域内的 POI 多样性数量，需不断尝试扩大单元格的搜索范围，使耗时随之增加。从图 5(a)~图 5(c)中可看出 3 条曲线比较紧密，表明相似度对性能的影响比多样性参数对性能的影响小，这说明利用本方法进行隐私保护时，在多样性参数固定的情况下，位置偏移生成算法在根据隐私信息访问控制系统返回的敏感度进行计算时，敏感度的不同对算法的影响不大，这一特征保证了方法的可用性。

9　结束语

本文针对移动社交网络内容分享中的位置隐私保护问题进行了研究。提出了一种面向内容分享的位置隐私访问控制模型，能够根据用户与访问者的关系、内容的时空/共现上下文，细粒度控制用户分享内容和他人提及自己的分享内容的位置信息。本文设计了一种满足该位置隐私访问控制模型的系统方法，提出了面向位置隐私设置的 *k*-匿名算法，使用户在位置隐私设置中提交的位置信息满足最大熵，保护用户敏感位置信息不被 CSSP 推测。并提出了一种基于敏感度的 POI 保护方法，使访问者在访问用户分享内容时，根据用户不同的敏感度显示当前内容所在的不同 POI 信息。最后通过实现分析了本方法的效率和可用性。

参考文献：

[1] ILIA P, POLAKIS I, ATHANASOPOULOS E. Face/Off: preventing privacy leakage from photos in social networks[C]//The 22nd ACM SIGSAC Conference on Computer and Communications Security. 2015: 781-792.

[2] NAINI F M, UNNIKRISHNAN J, THIRAN P. Where you are is who you are: user identification by matching statistics[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(2): 358-372.

[3] DATELINE NBC. Tracing a stalker[EB/OL]. http://www.nbcnews.com/id/19253352/ns/dateline_nbc-crime_reports/t/tracing-stalker/,2013.

[4] VOELCKER J. Stalked by satellite: an alarming rise in gps-enabled harassment[J]. IEEE Spectrum, 2006, 47(7): 15-16.

[5] WEBROOT SOFTWARE. Webroot survey finds geolocation apps prevalent amongst mobile device users, but 55\% concerned about loss of privacy[EB/OL]. http://pr.webroot.com/threat-research/cons/social-networks-mobile-security-071310.html,2010.

[6] Half of social networkers online concerned about privacy[EB/OL]. http://maristpoll.marist.edu/714-half-of-social-networkers-online-concerned-about-privacy/,2010.

[7] CHOW C Y, MOKBEL M F, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service[C]//The 14th Annual ACM International Symposium on Advances in Geographic Information Systems. 2006: 171-178.

[8] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//The 1st International Conference on Mobile Systems, Applications and Services. New York, USA: ACM, 2003: 31-42.

[9] KALNIS P, GHINITA G, MOURATIDIS K. Preventing location-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, IEEE, 2007, 19(12): 1719-1733.

[10] GEDIK B, LIU L. Protecting location privacy with personalized *k*-anonymity: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, IEEE, 2008, 7(1): 1-18.

[11] HU H, XU J, ON S T. Privacy-aware location data publishing[J]. ACM Transactions on Database Systems (TODS), ACM, 2010, 35(3): 1-40.

[12] 李凤华, 李晖, 贾焰. 隐私计算研究范畴及发展趋势[J]. 通信学报, 2016, 37(4): 1-11.
LI F H, LI H, JIA Y. Privacy computing: concept, connotation and its research trend[J]. Journal on Communications, 2016, 37(4): 1-11.

[13] KRUMM J. A survey of computational location privacy[J]. Personal and Ubiquitous Computing, London, UK, UK: Springer-Verlag, 2009, 13(6): 391-399.

[14] BETTINI C, WANG X, JAJODIA S. Protecting privacy against location-based personal identification[J] Secure Data Management, 2005, 3674: 185-199.

[15] CHOW C Y, MOKBEL M F. Trajectory privacy in location-based services and data publication[C]//SIGKDD 2011, 19-29.

[16] DUCKHAM M, KULIK L. A formal model of obfuscation and negotiation for location privacy[J]. Pervasive Computing, 2005, 3468: 152-170.

[17] DAMIANI M L. Privacy enhancing techniques for the protection of mobility patterns in LBS: research issues and trends[C]//European Data Protection: Coming of Age. Springer, 2013: 223-239.

[18] DAMIANI M, BERTINO E, SILVESTRI C. Protecting location privacy through semantics-aware obfuscation techniques[C]//IFIPTM, 2008: 231-245.

[19] KLEMPERER P, LIANG Y, MAZUREK M. Tag, you can see it!: using tags for access control in photo sharing[C]//The SIGCHI Conference on Human Factors in Computing Systems. 2012: 377-386.

[20] XU K, GUO Y, GUO L. My privacy my decision: control of photo sharing on online social networks[J]. IEEE Transactions on Dependable and Secure Computing, 2015, (99): 1.

[21] WISHART R, CORAPI D, MARINOVIC S. Collaborative privacy policy authoring in a social networking context[C]//Policies for

Distributed Systems and Networks. 2010 IEEE International Symposium. 2010: 1-8.

[22] WANG Y, ZHAI E, LUA E K. ISAC: Intimacy based access control for social network sites[C]//Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC). 2012: 517-524.

[23] FONG P W L. Relationship-based access control: protection model and policy language[C]//The First ACM Conference on Data and Application Security and Privacy. 2011: 191-202.

[24] CARMINATI B, FERRARI E, PEREGO A. Enforcing access control in web-based social networks[J]. ACM Transactions on Information and System Security (TISSEC), 2009, 13(1): 6.

[25] GATES C. Access control requirements for web 2.0 security and privacy[J]. IEEE Web, 2007, 2.

[26] HOUGHTON D J, JOINSON A N. Privacy, social network sites, and social relations[J]. Journal of Technology in Human Services, Taylor & Francis, 2010, 28(1–2): 74-94.

[27] JOSHI J B, BERTINO E, LATIF U. A generalized temporal role-based access control model[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(1): 4-23.

[28] GREENE K, DERLEGA V J, MATHEWS A. Self-disclosure in personal relationships[M]. The Cambridge Handbook of Personal Relationships, 2006: 409-427.

[29] WIESE J, KELLEY P G, CRANOR L F. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share[C]//The 13th international conference on Ubiquitous computing. 2011: 197-206.

[30] SUCH J M, ROVATSOS M. Privacy policy negotiation in social media[J]. ACM Trans Auton Adapt Syst, New York, NY, USA: ACM, 2016, 11(1): 4:1-4:29.

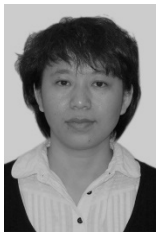
[31] SWEENEY L. *k*-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(7): 557-570.

[32] LIU Y, WEI W, SUN A. Exploiting geographical neighborhood characteristics for location recommendation[C]//The 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14. 2014: 739-748.

[33] Real datasets for spatial databaes: road networks and points of Internet[EB/OL]. http://www.cs.fsu.edu/~lifeifei/SpatialDataset.html.

作者简介：

李超（1981-），男，重庆人，博士，中国科学院信息工程研究所博士后，主要研究方向为隐私保护、访问控制。



殷丽华（1973-），女，辽宁朝阳人，博士，中国科学院信息工程研究所副研究员、硕士生导师，主要研究方向为信息安全、安全评估。

耿魁（1989-），男，湖北红安人，博士，中国科学院信息工程研究所助理研究员，主要研究方向为网络安全。

方滨兴（1960-），男，江西万年人，中国工程院院士，东莞电子科技大学教授、博士生导师，主要研究方向为计算机体系结构、计算机网络与信息安全。

doi:10.11959/j.issn.1000-436x.2016218

高稳定被动群集车联网连通性研究

邱恭安，包志华，章国安，张士兵
（南通大学电子信息学院，江苏 南通 226019）

摘 要：提出了基于车辆相对运动速度的车联网被动群集车间通信模型，通过选择平均相对速度和邻居节点数作为输入参数，应用模糊逻辑推举群首，建立高稳定性车节点群集，推导了群集模型的存活时间函数。交通安全消息在车间通过群内广播和群间中继进行传播，针对群内链路连通性，推导了群内节点连通概率与车辆分布密度间的关系。针对群间路径连通性，推导了群间连通概率与群间间距、车辆分布密度间的关系。最后，在满足车辆分布密度前提下计算仿真结果，验证了理论分析结论的合理性，分析结论为高速交通安全消息的车间传播机制应用提供理论参考。

关键词：被动群集；模糊逻辑；连通概率；相对速度；车载无线网络
中图分类号：TP393 **文献标识码**：A

Connectivity analysis of passive cluster with high stability in vehicular wireless network

QIU Gong-an, BAO Zhi-hua, ZHANG Guo-an, ZHANG Shi-bing
（School of Electronics and Information, Nantong University, Nantong 226019, China）

Abstract: A passive cluster model with the maximum lifetime was proposed for vehicle to vehicle communication based on the relative velocity. The cluster head was elected based on the average relative velocity and the neighbor list. The cluster lifetime was deduced as the function of the average relative velocity. The traffic safety messages were disseminated to all cluster members by inter-cluster message broadcasting and intra-cluster message relaying in interconnected vehicular network. The link connectivity probability between the cluster head and members were deduced as the function of the vehicle density for inter-cluster broadcasting. The path connectivity probability between the cluster head and the neighbor cluster head was deduced as the function of the vehicle density and intra-cluster distance for on intra-cluster dissemination. Simulation results show that the connected probability is suitable for vehicular network under the traffic density constraints.

Key words: passive cluster, fuzzy logic, connectivity probability, relative velocity, vehicular wireless network

1 引言

智能交通系统（ITS, intelligent transportation system）能够提高交通安全和交通效率，减小能源消耗和环境污染。车联网是 ITS 交通状态消息实时获取与传播的载体，它依靠车载专用短距离通信（DSRC, dedicated short-range communications）设备实现车与车（V2V, vehicle to vehicle）、车与路边设施（V2I, vehicle to infrastructure）间的通信，具有高效计算、持续能量、既定拓扑和车辆定位等优势，也因高速移动特征带来动态网络拓扑和多径衰落，存在控制信道带宽和通信距离受限的问题，导致 V2V 通信链路随机中断，难以维持稳定、可靠的车间通信[1]。将行驶模式相

收稿日期：2015-09-08；**修回日期**：2016-10-12
基金项目：国家自然科学基金资助项目(No.61371111, No.61371112, No.61371113); 交通运输部应用基础研究基金资助项目(No.2013319825110)
Foundation Items: The National Natural Science Foundation of China (No.61371111, No. 61371112, No. 61371113), Application Foundation of the Ministry of Transport of China (No.2013319825110)