基于距离跃变的“探探”恶意用户定位方法

郭家山[1,2], 杜少勇[2,3], 时文旗[2], 刘瑞婷[2], 罗向阳[2]

[1]郑州大学网络空间安全学院 郑州 中国 450001
[2]河南省网络空间态势感知重点实验室 郑州 中国 450001
[3]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 中国 100093

摘要 近年来, “探探”作为国内知名即时通信平台之一, 常常被恶意用户用来实施诈骗、策反等各类不法活动。为有效发现和打击探探平台上的此类恶意用户, 亟需针对探探恶意用户的定位技术。然而, 当前国内外尚未见针对探探用户定位的相关报道, 现有针对其他即时通信平台的定位方法由于没有考虑探探平台的特点(即, 通告距离的特性), 难以高效发现用户和实施高精度定位。为此, 本文提出了一种针对探探的基于通告距离跃变分析的恶意用户定位方法。本文结合探探平台通告距离的数据特点, 分析了通告距离与实际距离关系, 建立了探探平台通告距离模型; 为了有效逼近恶意用户, 结合分段式通告距离模型特点, 本文设计了恶意用户潜在区域发现算法; 为了提升定位精度, 本文结合探探的通告距离跃变特点, 设计了一种基于距离跃变的定位算法, 通过探针移动策略以更好地获得跃变距离, 并结合三边测量方法得到目标的理论位置。在实际网络环境下开展了一系列实验, 结果表明: 本文所提方法平均定位误差为 20.92 m, 且 95%的定位误差小于 50 m, 相比于现有基于空间分割、加权最小二乘、启发式数论等针对微信、陌陌等的典型定位方法, 定位误差降低 34.99%～60.60%。

关键词 即时通信; 探探; 恶意用户定位; 通告距离模型; 距离跃变
中图法分类号 TP309.2 DOI 号 10.19363/J.cnki.cn10-1380/tn.2024.11.09

A Malicious User Geolocation Method on Tantan App with Distance Transition

GUO Jiashan[1,2], DU Shaoyong[2,3], SHI Wenqi[2], LIU Ruiting[2], LUO Xiangyang[2]

[1] School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China
[2] Key Laboratory of Cyberspace Situation Awareness of Henan Province, Zhengzhou 450001, China
[3] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract In recent years, as one of the well-known instant messaging platforms in China, Tantan is often used by malicious users to carry out various illegal activities such as fraud and countermeasures. To effectively discover and combat such malicious users on the Tantan platform, it is urgent to target the positioning technology of Tantan malicious users. However, at present, there are no relevant reports at home and abroad on Tantan user positioning. The existing positioning methods for other instant messaging platforms do not consider the characteristics of Tantan platform (i.e., the characteristics of report distance), which makes it difficult to efficiently find users and implement high-precision positioning. Therefore, this paper proposes a malicious user geolocation method on Tantan app with distance transition. Based on the data characteristics of report distance of the Tantan, the relationship between the report distance and actual distance is analyzed, and the report distance model of the Tantan is established. Aiming at the characteristics of segmented report distance model and approaching malicious users, this paper designs a potential area discovery algorithm for malicious users. To improve the positioning accuracy, this paper designs a geolocation algorithm based on distance transition by combining the characteristics of probe to announce distance transition, and the probe movement strategy is used to better obtain the transition distance, and the theoretical position of the target is obtained by combining the trilateration method. A series of experiments are carried out in the actual network environment, and the results show that the average positioning error of the proposed method is 20.92 m, and 95% of the positioning error is less than 50 m, which is reduced by 34.99%～60.60% compared with the existing typical positioning methods for WeChat and Momo based on space partition, weighted least squares, and heuristic number theory.

Key words instant messaging; Tantan; malicious user geolocation; model of report distance; distance transition

通讯作者: 罗向阳, 教授, Email: luoxy_ieu@sina.com。
本课题得到国家自然科学基金(No. 62002386, No. U1804263, No. 62172435)、中原科技创新领军人才项目(No. 214200510019)、河南省重点研发专项(No. 221111321200)资助。
收稿日期: 2022-12-04; 修改日期: 2023-02-24; 定稿日期: 2024-09-20

1 前言

智能手机承载的基于位置的社交网络(Location-Based Social Networks, LBSN)提供了众多基于位置的服务(Location-Based Service, LBS)方便用户使用[1], 如基于位置的社交用户发现(Location-Based Social Discovery, LBSD)、导航、兴趣点(Point of Interest, POI)等[2-6]。然而这些服务也为恶意用户从事如赌博、诽谤, 甚至是窃密、渗透、策反等危害国家安全的非法活动提供了新的隐蔽渠道。其中探探作为拥有 2.7 千万月活跃用户的 LBSN 平台[7]也面临此类问题。据国家最高人民检察院“检察机关依法惩治危害国家安全犯罪典型案例”披露的案例表明, 探探已经成为了这些不法活动的重灾区。近年来已有数起境外敌对势力通过探探等 LBSN 进行渗透策反事件发生[8], 同时探探上出现的非法集资、网赌网诈等事件层出不穷[9]。为了有效打击这些不法活动, 亟需针对探探平台上的恶意用户开展定位技术研究。

现有 LBSN 用户定位的研究可分为基于文本和社交关系推断的用户定位与基于 LBSD 服务的用户定位。前者利用用户在 LBSN 中发布的文帖、签到数据与社交关系等推断用户的位置[10-13]; 后者则研究如何利用 LBSD 服务中展示的用户间距离信息等位置相关信息定位目标用户上传的最新位置, 具有更高的定位精度[14-22]。本文基于探探平台提供的 LBSD 服务, 重点研究针对探探恶意用户的定位方法。LBSD 用户定位技术主要基于 LBSD 服务公开展示的附近用户通告距离、相对次序等特征, 通过部署位置已知的探针在不同位置获取目标用户被混淆后的通告距离等信息, 进而推断目标用户的位置。根据定位方法原理的差异, 基于 LBSD 服务的用户定位方法可分为两大类: 基于位置迭代的定位方法[14-20]和基于数论的定位方法[16, 21-22]。

基于位置迭代的定位方法首先根据目标用户的大致信息, 确定用户所在的潜在区域; 然后在区域内不断调整探针位置, 对探针与目标之间通告距离迭代, 逐步计算得到目标用户的理论位置。Ding 等[14]通过经典三边测量方法对用户进行地理定位, 该方法在目前 LBSD 服务对通告距离混淆不断加强的情况下误差往往较大; 为了突破最小通告距离对定位精度的限制, Li 等[15]提出了基于空间划分的定位(Space Partition Based Geolocation, SPBG)方法, 该方法通过对最小通告距离对应地理区间的二分查找, 实现对微信用户更高精度的定位, 后续工作以更大的开销进一步实现了定位精度的提升[16-17]。利用用户在有序距离列表中的失序现象, Shi 等[18]提出了基于失序现象的用户定位方法, 最大程度减少通告距离混淆机制对定位精度的影响。利用二分搜索, Wang 等[19]通过在潜在区域建立笛卡尔直角坐标系, 通过多次调整探针逼近目标用户。Shi 等[20]提出了基于加权最小二乘的定位(Weighted Least Squares Based Geolocation, WLBG)方法, 利用基本的通告距离统计特性, 针对微信平台提高了定位方法的效率与精度。

基于数论的定位方法利用了同心圆中相对距离的特点, 建立了通告距离与实际距离的数学关系模型。一维数论定位方法[16,21] 基于通告距离和实际距离之间数学模型的假设, 在一维直线上研究了对目标进行定位的方法; 二维数论定位方法[16,21]由一维数论定位方法扩展至二维得来, 在模拟环境下实现较高的地理定位精度, 但缺乏实际验证, 同时没有考虑到系统误差的影响。Peng 等[22]提出了更为贴近实际环境的基于启发式数论的定位(Heuristic Number Theory Based Geolocation, HNBG)方法。HNBG 方法在 1 km×1 km 的潜在区域创建笛卡尔直角坐标系, 并沿两坐标轴部署大量探针。将每个坐标轴上通告距离最小的探针位置作为目标用户初始理论位置, 同时采用一维数论方法对初始理论位置修正, 得出用户最终理论位置。与二维数论方法相比, HNBG 方法在实际环境中有更高的定位精度。

然而, 上述方法在定位过程中存在一定的局限性: (1) 缺乏针对现实平台的深度挖掘(如探探平台的特点); (2) 缺少对通告距离分段特性的充分利用; (3) 最后精度仍会受到最小通告距离限制。因此, 上述方法无法直接应用于探探恶意用户的定位。

针对上述现有研究存在的问题, 本文提出一种基于距离跃变的探探恶意用户定位方法。本文通过引入探探的通告距离模型来匹配最接近目标用户实际距离的通告距离并结合文献[18]改进的三边测量模型确定目标用户的实际位置。本文主要工作如下:

1) 建立了探探平台通告距离与实际距离的关系模型。本文通过 14600 次不同位置的实地测试, 获得大量<通告距离, 实际距离>数据, 通过分析建模, 构建出探探通告距离与实际距离的关系模型。

2) 设计了一种探探恶意用户潜在区域的发现方法。本文充分利用通告距离模型的分段特点, 结合三边测量方法与渐进式探针策略, 有效逼近探探恶意用户所在潜在区域。

3) 提出了适用于“探探”平台的基于跃变现象的定位方法。该方法利用通告距离模型, 结合本文设计的探针移动策略, 利用高效的三边测量模型对目

标进行定位。实际测试结果表明，本文方法平均定位误差为 20.92 m，显著优于现有定位方法。

本文后续章节安排如下：第 2 节介绍探探平台与其 LBSD 服务；第 3 节解释本文涉及的概念与符号定义；第 4 节介绍本文定位方法原理与具体流程；第 5 节讨论探针移动策略；第 6 节给出实验结果并对结果进行分析；第 7 节总结论文。

2 探探 LBSD 服务原理及特点分析

本节将介绍探探在用户发现上的特点，并分析探探 LBSD 服务展示其他用户位置时采用的策略。

2.1 探探 LBSD 服务

LBSD 服务可以让用户根据智能手机的位置以不同形式发现附近的用户。如图 1 所示，LBSD 服务器存储维护众多<用户，位置>的数据。查询者通过操作移动应用界面选择具体 LBSD 服务类型，并发送相关 LBSD 请求；请求会将查询者当前的位置(使用手机内置的位置相关 API)与服务类型上传至 LBSD 服务器。LBSD 服务器收到请求后，会记录该查询者的位置，通过对服务器中的数据进行筛选，选出符合要求的其他用户。然后，服务器根据 LBSD 服务类型，返回含有不同信息的用户列表。最后，该查询者根据返回的列表，进行后续社交活动。

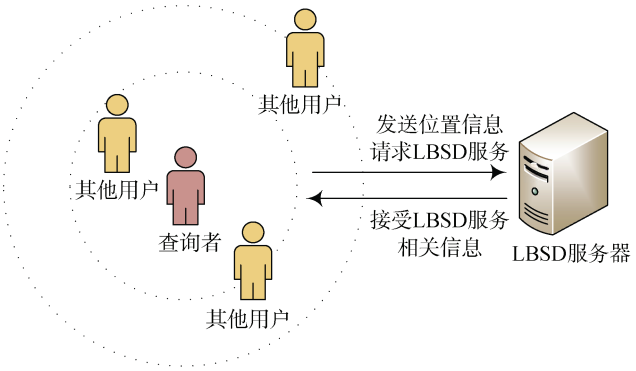

图 1 LBSD 服务模型
Figure 1 Service model of LBSD

“探探”主要提供“推荐”、“娱乐”和“发现”等 LBSD 服务；查询者利用这些服务关注某个用户后，可以方便地获取被关注用户与查询者间的通告距离。

“推荐”功能主要向查询者推荐附近的用户，附近用户列表以整页卡片形式逐个呈现，如图 2 左所示。该列表存储于服务器，仅当查询者对当前附近用户卡片操作(如“喜欢”和“不喜欢”)后才会返回后续附近用户卡片。根据真实测试发现，相同位置相同标签的用户进行操作时，初始附近用户列表基本一致，但对卡片不同的操作会改变列表。卡片会展示当前该用户的照片、昵称以及与当前查询者间距离等。点击“详情”后如图 2 右所示，卡片会展示当前用户详细信息以及“关注”等其他操作。该服务下，查询者可以通过添加贴合特定用户特征的相关信息等社会工程学[23-24]方法，提高特定类型用户发现概率。

图 2 “推荐”功能详情
Figure 2 Details of function “Recommend”

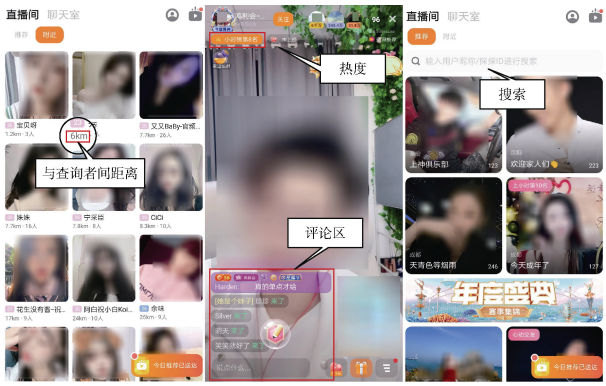

图 3 “娱乐”功能详情
Figure 3 Details of function “Entertain”

“娱乐”功能主要向查询者推荐主播(主播对于探探而言也属于用户)，主播列表以卡片列表形式展示，该列表内容及顺序与主播直播间热度和距离同时相关。如图 3 左所示，每个卡片中都包含主播的照片、昵称和主播与查询者间距离等信息。点击卡片可以进入主播的直播间，如图 3 中间所示，直播间内有当前直播间热度，评论等。直播间内可以关注该主播以及留言用户。同时该服务提供主播搜索功能，如图 3 右所示，也可以直接搜索主播名或 UID(用户唯一标识)直接关注某特定主播。

“发现”功能主要向查询者推荐附近用户的帖子, 以普通列表形式展示, 该列表的顺序同时被帖子热度和与查询者距离两方面影响。如图 4 左所示, 每个列表项包含用户的帖子概览、发帖时帖主与查询者间距离等信息。点击列表项, 如图 4 右所示, 可以查看帖子详情与评论详情。帖子内可以关注帖主与帖子评论区用户。

图 4 “发现”功能详情

Figure 4 Details of function “Discovery”

2.2 探探 LBSD 服务分析

随着对位置保护意识的提高, 探探对附近用户的通告距离进行了混淆。经过真实测试发现, 探探采用的混淆策略如下:

1) 只提供用户间距离: 探探只提供用户之间瞬时距离, 隐藏准确的经纬度坐标。

2) 限制最小通告距离: 探探提供用户之间的通告距离以 100 m(1 km 以内)为单位。使得通告距离小于 1 km 时为 100 m 的倍数, 否则为 1 km 的倍数。

3) 弱化用户列表的距离相关性: 探探在提供各种用户列表时, 优先使用其他信息决定列表顺序。

4) 限制登录设备类型: 探探在现有保护策略的基础上, 进一步限制模拟器设备用户最小通告距离为 2 km。

3 概念与定义解释

在介绍本文方法前, 本节首先介绍一些定义与术语。为表述方便, 若无特殊说明, 本文各种距离默认统一使用米(meter, m)为单位。本文涉及的符号定义如表 1 所示。提及的概念与解释如下:

探针: 用来观测与目标用户之间距离的设备, 该设备需要拥有自由修改位置的能力, 是定位过程中数据采集的关键设备。

通告距离: LBSD 服务提供的用户间距离, 是定位过程中的关键数据。

实际距离: 在实际地理空间中用户间距离, 该距离可藉由球面距离公式等计算得到。

通告距离跃变: 指在实际距离连续变化, 通告距离从一个值跃变为另一个值的现象。

跃变点: 发生跃变现象时探针所在位置。

步长: 探针触发跃变距离 D_t 移动的单位距离称为步长。

潜在区域: 指用户当前所在的大致区域; 一般该区域是边长为 1～2 km 的正方形区域。

表 1 符号描述

Table 1 Explanation of notation

符号定义	描述
p_t	跃变点(<经度, 纬度>)
p_o	探针初始位置(<经度, 纬度>)
p_a	潜在区域代表点(<经度, 纬度>)
U	最小通告距离单位, 探探为 100 m/1 km
D_r	通告距离
D_a	两个点之间的实际距离
D_t	发生跃变现象时的通告距离
t	步长
Th	潜在区域间距阈值

4 基于距离跃变的探探定位方法

本文第 2 节详细介绍了探探 LBSD 服务。通过探探的 LBSD 服务关注特定用户后, 探探会返回该用户与查询者之间的通告距离。通过分析跃变现象发生时通告距离与实际距离的关系, 可以确定查询者与目标探探用户的实际距离区间。本文提出方法如图 5 所示。方法分为通告距离模型构建与基于通告距离跃变的定位过程两部分。

通告距离模型构建由数据采集与跃变现象分析两部分组成。其中数据采集通过大量统计探探的<通告距离, 实际距离>数据对, 分析构建通告距离模型。跃变现象分析部分利用通告距离模型分析跃变现象发生条件, 即跃变点 p_t 出现的区间。

基于通告距离跃变的定位过程由潜在区域发现、跃变距离获取与精确用户定位组成。

步骤 1: 潜在区域发现部分关注目标用户, 并利用通告距离模型结合经典三边测量逼近用户潜在区域;

步骤 2: 跃变距离获取部分通过本文设计的探针移动策略, 在不同位置触发跃变现象并记录多组 D_t;

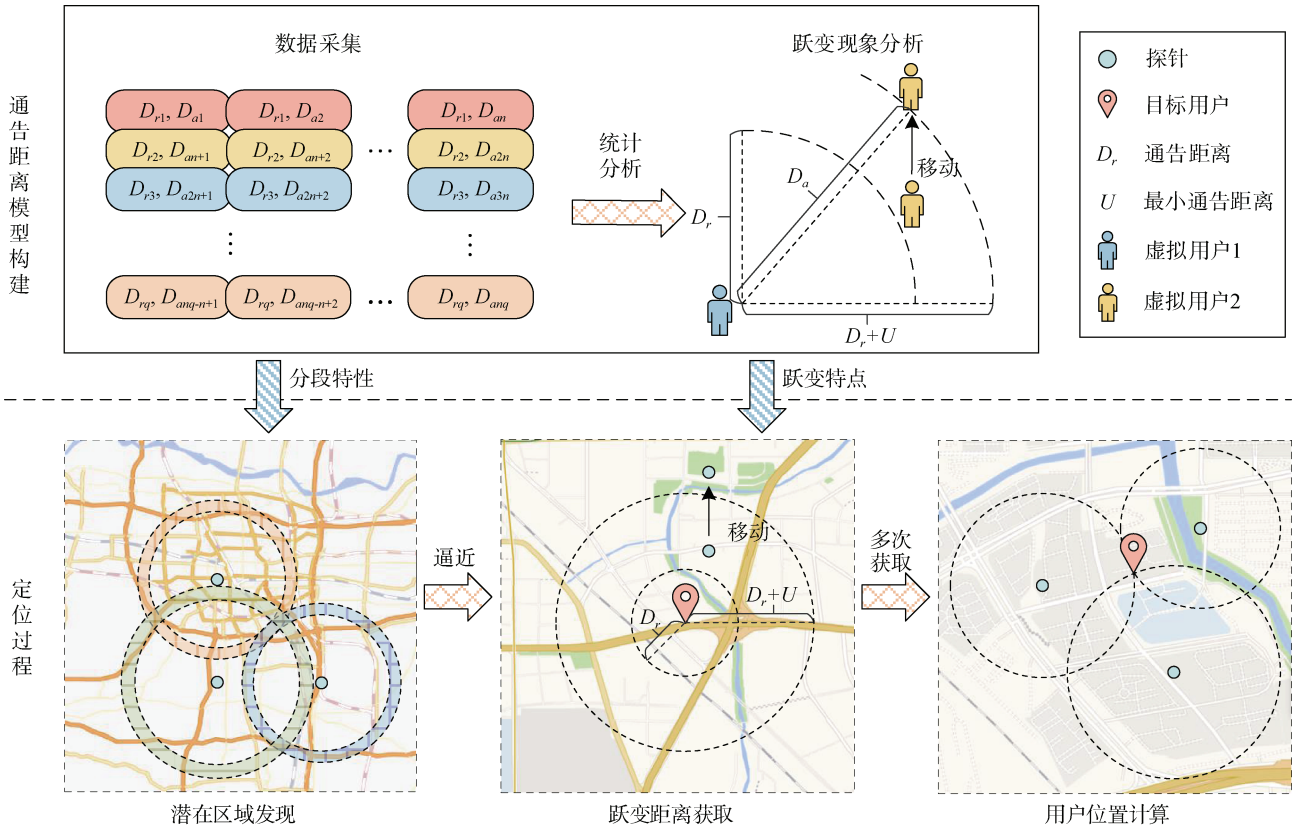

图 5　基于距离跃变的探探恶意用户定位方法原理示意图

Figure 5　Schematic of malicious user geolocation method on Tantan with distance transition

步骤 3: 精准用户定位部分使用 D_t 并结合三边测量模型对目标用户进行地理定位。

下面将针对通告距离模型构建、潜在用户发现、跃变距离获取和用户位置计算等关键步骤进行具体阐述。

4.1　通告距离模型构建

该部分利用多个探探账号确定通告距离模型，为了便于表述，以用户 A 和用户 B 来表示该过程。首先固定用户 A 的位置作为目标，用户 B 作为探针并随机选一个方向 θ 固定。A 与 B 初始位置相同，之后沿 θ 以 10 m 为单位移动 B，并记录此时 B 显示对 A 的通告距离；当通告距离超过 1 km 时，继续沿 θ 以 100 m 为单位移动 B，直至通告距离显示为 20 km。重复上述过程 50 次，根据数据特性构建通告距离模型。基于上述设置，共计得到 14600 条<通告距离，实际距离>关系。由数据可得当实际距离 D_a ≤1000 m 时，查询得到的通告距离为 $\left(\lfloor D_a/100 \rfloor+1\right)\times 100$ m。当 D_a > 900 m 时，探探的通告距离单位从米变更为千米(即显示为 1 km)；当 D_a > 1000 m 时，查询得到的通告距离为 $\left(\lfloor D_a/1000 \rfloor+1\right)\times 1000$ m。以此构建通告距离与实际距离之间的距离关系模型如下:

$$D_r=\left(\left\lfloor \frac{D_a}{U} \right\rfloor+1\right)\times U, \tag{1}$$

其中有:

$$U=\begin{cases}100, & 0\leqslant D_a<1000,\\ 1000, & D_a\geqslant 1000.\end{cases} \tag{2}$$

定位过程中通常仅利用通告距离模型中 U = 100 m 的部分，但可以利用 U = 1000 m 的部分进行目标的潜在区域发现，这是现有研究往往忽视的一部分。

4.2　潜在区域发现

由本文 4.1 节可知，通告距离模型并没有被充分利用。在 U = 1000 时的通告距离模型下，定位时的跃变现象难以触发，同时直接定位误差较大；但可以结合三边测量逼近目标用户，找到目标所在潜在区域。这样在充分利用通告距离特性的同时，提高了方法的适用范围。

如图 5“潜在区域发现”部分所示，该步骤在经典三边测量模型的基础上设计了渐进式探针部署: 首先，部署探针至初始点 p_o，记录此时探针与目标的通告距离 D_{r0}；然后，将 p_o 的正北方向与正东方向距离 D_{r0} 处设置为探针下一步移动的两个位置 p_1 与 p_2，移动探针至 p_1 与 p_2 并记录此时探针与目标的通告距

离; 后续的探针位置通过上述过程的定位结果决定, 每次定位时选择最新的三个探针位置观测目标用户通告距离, 直到下一轮迭代开始时三个探针对目标的通告距离均小于 Th, 选取当前该组探针为中心, 对应通告距离为半径的圆环相交区域的中心作为用户所在潜在区域的代表点 p_a。

在实验中现有方法设置潜在区域为边长为 1000 m 的正方形区域, 为了同现有方法保持一致, 这里 Th 设置为 1000。

4.3 跃变距离获取

由本文 4.1 节可知, 目标用户与探针间实际距离变化的过程中, 产生跃变现象的点为跃变点 p_t; 但实际过程中, 由于 p_t 不可被直接观测, 探针需要通过一定的移动(如沿着某方向移动一定距离)到达 p_t 附近观测 D_t。令移动的步长为 t, 根据 $U = 100$ 时通告距离 D_r 模型的阶梯特征可知 $D_t = 100 \times n$ ($n \in 1, 2,\ldots,9$)。为了保证定位精度, 充分触发每一段 D_t, 步长 t 不应大于 100, 即 $t \in [1, 100)$。令跃变前通告距离为 D_{r0}, 跃变后为 D_{r1}, 此时有 $D_t = \min(D_{r0}, D_{r1})$。结合前文提及的探探通告距离模型可知, 此时有实际距离 $D_a \approx D_t$。最终定位时, 使用 D_t 作为探针与目标间的实际距离。

对于移动过程而言, 步长 t 与探针的移动方式是触发跃变现象时的两个关键部分, 影响最终定位过程中的精度与效率。这里记:

$$e = \left| D_a - D_t \right| \tag{3}$$

e 表示将 D_t 作为实际距离时与实际距离 D_a 的误差。直观的有当 e 越小 D_t 与 D_a 越接近, 最终定位时精度越高。

如图 6 所示, 探针初始点位于目标正北方向 x m 处(三角处)出发沿正北方向移动, 且有 $t_1 < t_2$; 虚弧线表示通告距离模型下, 触发跃变现象时的最远实际距离, 在该弧线上有 $e = 0$。当探针以 t_1 为步长移动时, 如图 6a 所示, 探针触发跃变现象时 e 较小, 但需要移动 3 次。当探针以 t_2 为步长移动时, 如图 6b 所示, 探针触发跃变现象时 e 相比图 6a 较大。但只需要移动 2 次。

在图 6 的基础上, 设此时探向北方向移动一次就会触发跃变现象, 移动步长为 t, 则触发跃变现象时 $D_a = x + t$, 代入公式 3 有:

$$e = x + t - D_t \tag{4}$$

推广情况至探针位于目标东北方向, 与正北方向夹角为 $\theta(\theta \in [0, 90])$, 此时对 e 有:

$$e = \sqrt{(x \cdot \sin\theta)^2 + \left(x \cdot \cos\theta + t\right)^2} - D_t \tag{5}$$

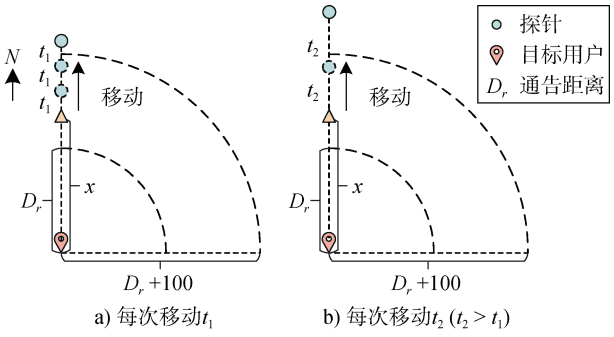

图 6 步长对定位效率与误差的影响

Figure 6 Effect of step on positioning efficiency and error

对 t 求导得:

$$e'(t) = \frac{t + x\cos\theta}{\sqrt{(t + x\cos\theta)^2 + (x\sin\theta)^2}} \tag{6}$$

由公式 6 可知当 t 增加时, e 同时增加; 推广至其他情况有类似结论。

当加上移动次数 n 时, 此时触发跃变现象时有 $D_a = \sqrt{(x \cdot \sin\theta)^2 + \left(x \cdot \cos\theta + nt\right)^2}$。根据公式可知 D_a 与目标位置和探针初始点位置有关, 初始点对于目标的相关位置与通告距离区间之间并非线性关系, 难以分析 D_a 与 D_t 之间关系, 无法对 e 的分布进行直观的分析; 但通过定性分析可知, 当初始点在同一通告距离段内均匀分布时, e 的期望与 t 成正相关。考虑到 t 过小时获取位置的次数过高, 影响定位效率, 这里取 $t \in [10, 50]$。为了权衡实际定位过程的效率与最终的精度, t 的具体取值将通过实验来论证。

移动方向同样决定触发跃变现象的效率与定位精度。移动探针时, 通常沿着单一确定方向移动, 但实际上单一方向移动时会遇到难以触发跃变现象的问题。这部分问题本文将在第 5 节详细讨论。

4.4 用户位置计算

利用本文 4.3 节获取的多对跃变距离 D_t 与 p_t, 结合经典三边测量模型进行精准定位。三边测量模型是一种高效定位模型, 该模型仅利用探针在三个不同位置得到与目标的通告距离区间, 即可得到目标的理论位置[14, 25]。为了提高最终定位效率, 本文采用三边测量模型计算目标位置。三边测量模型的基本流程如下: 首先, 利用三个探针位置并记录对应位置下探针与目标间的通告距离。然后, 分别以三个探针位置为圆心对应通告距离为半径做圆, 计算相交区域, 并将相交区域中心视作定位结果。

本文为充分利用定位过程数据, 提高定位精度, 在经典三边测量的基础上, 结合了本文第 5 节的探

针移动策略, 并在对目标用户定位时结合文献[18]引入的最小二乘法, 优化目标用户的理论定位结果。

5 探针移动策略

本文 4.2 节部分讨论了探针移动时步长 t 对定位的影响, 本节主要针对探针移动策略深入讨论。实际操作过程中, 探针的移动对跃变距离获取有较大的影响。

如图 7 所示, 虚弧线表示通告距离模型下, 触发跃变现象时的最远实际距离; 两条虚弧线所围成的区间称为通告距离区间。当探针处在通告距离区间, 如图 7a 所示, 探针沿箭头方向移动时可以更好的逼近目标用户, 触发跃变现象。但如图 7b 的探针视角, 对探针而言目标用户的位置在阴影区间内, 探针难以观测目标用户具体的方位。因此探针移动时通常沿某固定方向移动。

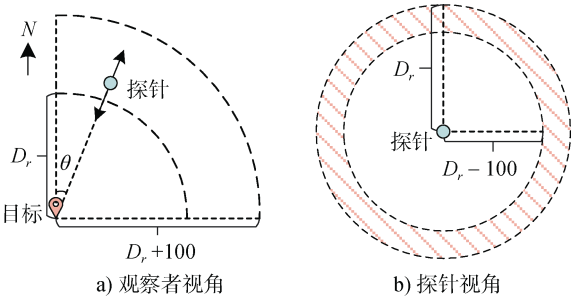

图 7 真实空间下探针移动问题

Figure 7 Problem of probe movement in real space

为了更好的触发跃变现象, 同时充分利用定位过程中产生的通告距离数据, 本文结合定位中间结果, 设计了一种方向自修正的探针移动策略。根据本文 4.2 节, 我们得到潜在区域代表点 p_a。首先, 将探针移动至 p_a 观察此时 D_r 的大小。然后, 移动探针时仍旧先沿着固定正北方向移动, 观察触发跃变现象时 D_r 的变化情况。令跃变现象触发前的通告距离为 D_{r0}, 触发得到的跃变距离为 D_{t0}, 探针此时跃变点 p_{t0}。这里为了便于讨论引入 ΔD, 其中有:

$$\Delta D = D_{t0} - D_{r0} \tag{7}$$

由式(7)可知, 当 $\Delta D > 0$ 时, 探针离目标越远; 当 $\Delta D < 0$ 时反之, 由此我们可以构建目标可能的位置。

如图 8 所示, 第 1 步移动表示当 $\Delta D > 0$ 时, 触发跃变距离时探针和目标用户可能的位置。由于 ΔD 仅能确定目标与探针之间南北方向上的关系, 因此目标可能在关于移动路径轴对称东西两侧; 当 $\Delta D < 0$ 时同理。接下来本文以 $\Delta D > 0$, 目标在西侧讲解方法思路。

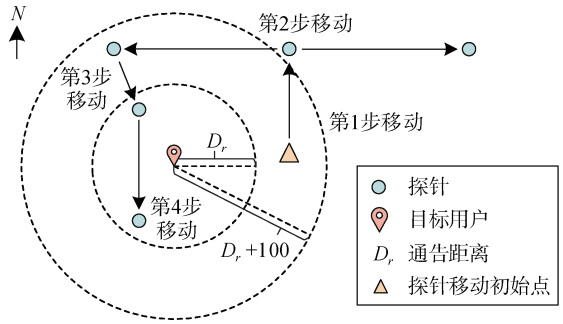

图 8 $\Delta D > 0$ 时部分探针移动策略

Figure 8 Part of probe movement strategy when $\Delta D > 0$

在得到 D_{t0} 后, 接着以 p_{t0} 为初始点, 沿东西方向距离 D_{t0} 处分别观察通告距离, 如图 8 第二步移动所示, 取两者较小者作为获取第二个跃变距离 D_{t1} 的基本点 p_1。此时可以排除关于第 1 步移动路径轴对称东西两侧中其中一侧的目标的可能位置。并在此基础上向着推断位置移动, 如图 8 第三步移动所示, 并获取此时的 D_{t1} 与 p_{t1}。

获取 D_{t1} 后, 以 p_{t1} 为初始点沿南方向距离 D_{t1} 处, 将离推测点较近的位置作为获取第三个跃变距离 D_{t2} 的基本点 p_2, 如图 8 第四步所示。此时需要对通告距离的情况进行讨论。

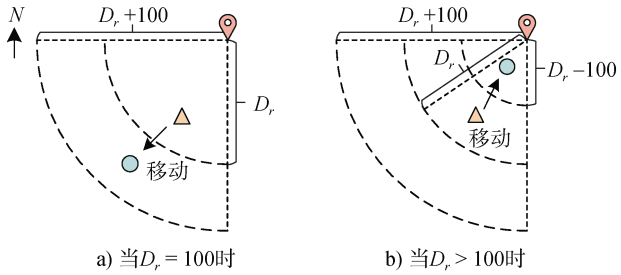

图 9 第五步探针移动策略

Figure 9 Step 5 of probe movement strategy

第五步移动时, 若 $D_r = 100$, 如图 9a 所示, 则移动探针时沿推断位置的反方向移动; 若此时 $D_r > 100$, 如图 9b 所示, 则移动探针时沿推断位置方向移动, 获取此时的 D_{t2} 与 p_{t2}。

最后根据生成的三组跃变距离与跃变点, 结合本文 4.4 的用户位置计算完成对目标探探用户的定位部分。

6 实验结果与分析

为了验证本文定位方法的有效性, 本文对探探恶意用户进行了实际网络环境下的定位实验。实验结果将与现有代表算法 HNBG[22]、WLBG[20]和 SPBG[15]

进行对比, 同时 HNBG、WLBG 和 SPBG 采用文献[22]、文献[20]和文献[15]的实验设置基础上, 均结合了本文的潜在区域发现方法, 在评估效率时潜在区域发现部分带来的开销不计入比较。

6.1 实验设置

实验部分包含步长确定与探探目标用户定位两部分。实验设置如表 2 所示。步长实验部分选用郑州市市区 1 km× 1 km 的区域作为测试区域, 将探针与目标随机的部署在该区域内。通过将跃变步长 t 设置为不同的值带入本文方法中执行 300 次以直观地观察步长 t 对定位方法的影响。实验中记录定位过程中对 LBSD 服务的查询次数、理论定位结果与定位误差, 最后选出最优步长 t 作为最终定位使用步长。

表 2 实验环境

Table 2 Experience environment

条目	配置
自动化工具	Appium 1.22.3
账号数量	10
测试用手机	3 部小米, 2 部华为, 1 部欧珀
安卓系统版本	Android 10
实验城市	北京、郑州

目标用户定位实验部分选取北京与郑州两座城市, 初始探针位置设置在郑州。为了实验, 本文开发了虚拟位置应用“越”, 可以修改手机上如基站、经纬度坐标、Wi-Fi 等和位置相关的信息。利用安装了“越”的安卓手机将多个探探账号的位置随机设置在实验区域中, 一部分作为定位的目标用户, 另一部分作为探针。探针利用本文的潜在区域逼近方法先对目标进行逼近, 当所有探针显示的通告距离 $D_r \leqslant 1000$ m 时, 对目标用户进行定位, 得到目标用户的理论位置。成功定位目标用户后, 按上述方法重置目标用户位置, 并再次定位。

6.2 步长确定

本文从定位误差与查询次数两个角度评价步长 t。定位误差是定位得到的目标理论位置与目标实际位置之间的偏移值。查询次数指定位时请求 LBSD 服务返回探针与目标间通告距离的次数。使用查询次数评估定位方法的效率可以客观地比较方法之间的用时, 避免因设备差异、代码实现等导致定位时长难以统一衡量的问题。定位误差越小, 定位时精度越高; 查询次数越少, 定位时效率越高。不同 t 下的定位误差越低, 定位精度越高; 定位精度越高、查询次数越少, 定位算法越优。

在上述实验设置与数据的基础上, 该部分进行了 6000 次实验。首先设置步长 t 从 10 m 开始, 每隔 5 m 进行一次测量, 并记录平均定位精度与平均查询次数, 实验结果如图 10 所示。

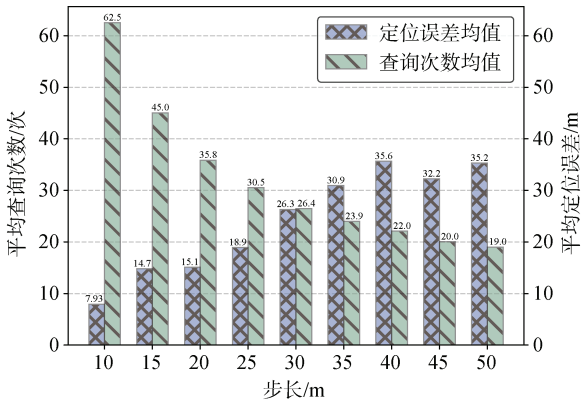

图 10 t 统计特性

Figure 10 Statistic feature of t

由图 10 可知, 当 $t \in [25,35]$, 有较好的效率与较低的定位误差表现。接下来以 $t \in [25,35]$为范围, 每隔 1 m 进行一次测量, 对 t 进行更为详细的测试。

根据图 11 得出的 t 统计特性, 不难看出当 t 取 32、33 时, 定位误差与查询次数均有较好的表现; 同时根据图 12 的中位数表现, 当 $t = 32$ 时定位误差较小, 表现更稳定。因此接下来的对比实验中本文所提方法将令 t 取 32 与其他方法进行对比。

6.3 定位结果对比与分析

该部分同样从定位精度与定位效率两个方面将 HNBG[22]、WLBG[20]和 SPBG[15]与本文方法进行对比与分析, 通过各方法所得定位结果与已知位置目标之间的偏移值作为定位误差评价定位精度; 以定位目标过程中探针查询通告距离的次数评价定位效率。

6.3.1 定位精度

基于对 300 个探探目标恶意用户的定位结果, 分别计算所提方法和对比方法的定位误差, 结果如图 13 所示。

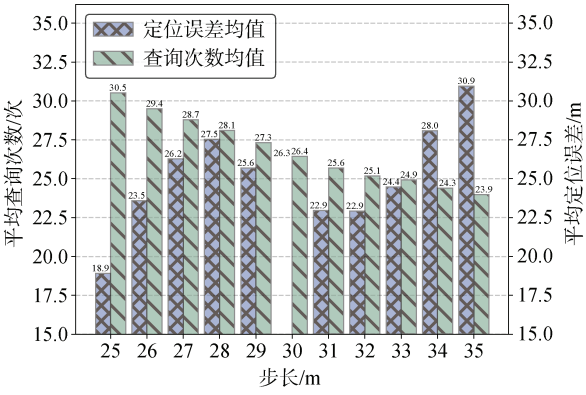

图 11 t 均值比较

Figure 11 Comparison of average value in different t

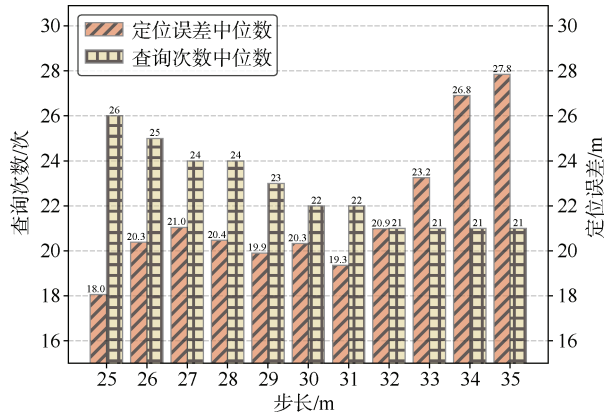

图 12　*t* 中位数比较

Figure 12　Comparison of middle value in different *t*

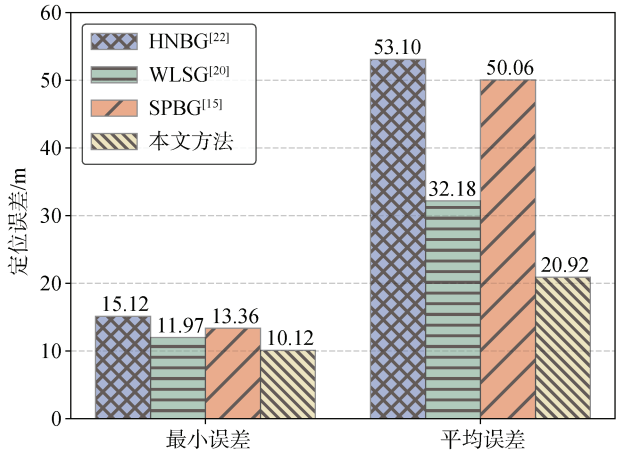

图 13　定位误差对比

Figure 13　Comparison of geolocation error

图 13 展示了本文方法与对比方法的最小误差和平均定位误差，纵坐标表示定位误差大小。由图 13 可以看出，本文方法最小定位误差为 10.12 m，平均定位误差为 20.92 m；与现有定位方法相比，平均定位误差分别降低了 60.60%、34.99%、58.21%。为了更好的评估定位方法的效果，本文针对定位误差分布做了进一步分析。

图 14 展示了本文方法与对比方法的定位误差分布，纵坐标表示各方法在对应定位误差区间对应的占比。实验结果表明，本文所提方法在定位精度分布方面表现更优，可以实现 51%的目标用户定位误差小于 20 m, 95%(51%+44%)的目标用户定位误差小于 50 m，均高于所对比方法。

6.3.2　定位效率

本文所提方法和对比方法在定位过程中，通告距离的平均查询次数如图 15 所示。本文在计算查询次数时，由于 HNBG 将目标限制在 1 km×1 km 的区域。因此在对比方法时首先采用本文提出的潜在区域发现方法，确定目标锁定 1 km×1 km 的区域；再分别利用各方法对目标进行定位，并统计定位过程中的查询次数。

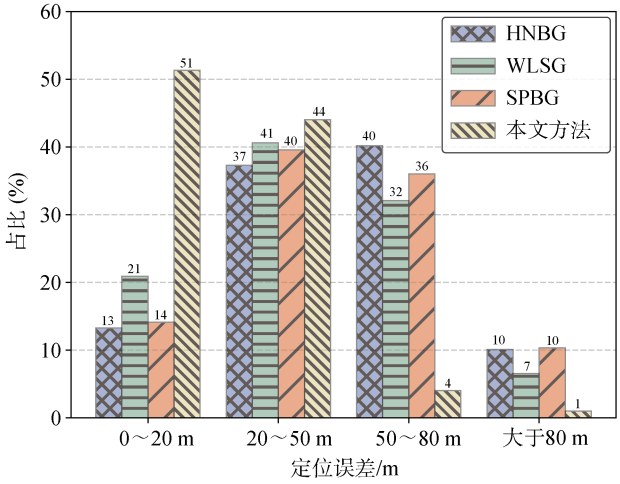

图 14　误差分布对比

Figure 14　Comparison of error distribution

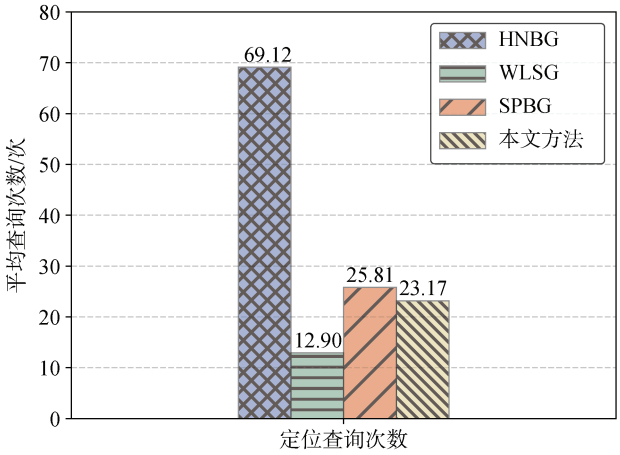

图 15　定位效率对比

Figure 15　Comparison of geolocating efficiency

由图 15 可知，本文所提方法在效率上仍旧具有较好的表现，平均查询次数为 23.16 次；相比 HNBG 与 SPBG 分别降低了 66.49%、10.23 %。WLSG 方法通过牺牲定位精度的方式提高定位效率，平均查询次数达到 12.90 次。SPBG 在定位过程中，因为需要限制目标用户的通告距离，造成额外的查询次数。HNBG 方法本身因需要在潜在区域部署大量探针以确定目标用户的初始坐标，同时应用一维数论方法精确定位目标用户时仍需频繁查询通告距离，因此需要相当多的查询次数。

综上，同现有典型 LBSD 用户定位方法相比，本文所提方法对“探探”平台用户的定位精度有明显优势，同时效率上有较好的表现。

7 结论

本文针对现有典型定位方法没有充分利用“探探”平台的特点、难以高效发现用户和实施高精度定位的问题, 提出了一种基于距离跃变的“探探”恶意用户定位方法。不同于已有的用户定位方法, 本文给出了“探探”特有的通告距离与实际距离的关系模型, 并利用该模型实现了对“探探”恶意用户的可靠定位。实验结果表明, 与现有经典方法相比, 本文提出的方法定位精度更高, 对给定的恶意用户能够进行更为有效的定位。但该方法和现有 LBSD 用户定位方法一样, 对于本身隐藏了位置或禁止被陌生人关注的用户尚无法实现定位。在未来的工作中, 我们将继续探索更为有效的 LBSD 用户定位方法, 并对其他 LBSD 服务的通告距离模型进行更深层次的刻画。

致　谢　在此向本文撰写时给予各类帮助与支持的指导老师、同学和为本文提出宝贵建议的各位评审专家表示诚挚的感谢。

参考文献

[1] Yao R X, Li H, Cao J. Overview of Privacy Preserving in Social Network[J]. *Chinese Journal of Network and Information Security*, 2016, 2(4): 33-43.
(姚瑞欣, 李晖, 曹进. 社交网络中的隐私保护研究综述[J]. *网络与信息安全学报*, 2016, 2(4): 33-43.)

[2] Zheng Y, Zhou X F. *Computing with spatial trajectories*[M]. New York: Springer Science+Business Media, LLC, 2011.: 243-276.

[3] Li J, Yan H Y, Liu Z L, et al. Location-Sharing Systems with Enhanced Privacy in Mobile Online Social Networks[J]. *IEEE Systems Journal*, 2017, 11(2): 439-448.

[4] Wang H D, Li Y, Chen Y, et al. Co-Location Social Networks: Linking the Physical World and Cyberspace[J]. *IEEE Transactions on Mobile Computing*, 2019, 18(5): 1028-1041.

[5] McGee J, Caverlee J, Cheng Z Y. Location Prediction in Social Media Based on Tie Strength[C]. *The 22nd ACM international conference on Conference on information & knowledge management - CIKM'13*, 2013: 459-468.

[6] Yuan F J, Jose J M, Guo G B, et al. Joint Geo-Spatial Preference and Pairwise Ranking for Point-of-Interest Recommendation[C]. *2016 IEEE 28th International Conference on Tools with Artificial Intelligence*, 2016: 46-53.

[7] Announces Unaudited Financial Results for the Fourth Quarter and Fiscal Year 2021. Hello Group Inc. https://ir.hellogroup.com/newsreleases/news-release-details/hello-group-inc-announces-unaudited-financial-results-fourth. March. 2022.

[8] 检察机关依法惩治危害国家安全犯罪典型案例. 最高人民检察院网上发布厅. https://www.spp.gov.cn/xwfbh/wsfbt/202204/t20220416_554500.shtml. April. 2022.

[9] 探探: 交友软件到底有多乱. 机智猫. https://new.qq.com/rain/a/20210413A092YU00. April. 2022.

[10] Wang K, Yu W, Yang S, et al. Location Inference Method in Online Social Media with Big Data[J]. *Journal of Software*, 2015, 26(11): 2951-2963.
(王凯, 余伟, 杨莎, 等. 一种大数据环境下的在线社交媒体位置推断方法[J]. *软件学报*, 2015, 26(11): 2951-2963.)

[11] Zheng X, Han J L, Sun A X. A Survey of Location Prediction on Twitter[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(9): 1652-1671.

[12] Pan X, Chen W Z, Wu L. Mobile User Location Inference Attacks Fusing with Multiple Background Knowledge in Location-Based Social Networks[J]. *Mathematics*, 2020, 8(2): 262.

[13] Luo X Y, Qiao Y Q, Li C L, et al. An Overview of Microblog User Geolocation Methods[J]. *Information Processing & Management*, 2020, 57(6): 102375.

[14] Ding Y, Peddinti S T, Ross K W. Stalking Beijing from Timbuktu: A Generic Measurement Approach for Exploiting Location-Based Social Discovery[C]. *The 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, 2014: 75-80.

[15] Li M Y, Zhu H J, Gao Z Y, et al. All Your Location Are Belong to Us: Breaking Mobile Social Networks for Automated User Location Tracking[C]. *The 15th ACM international symposium on Mobile ad hoc networking and computing*, 2014: 43-52.

[16] Cheng H N, Mao S L, Xue M H, et al. On the Impact of Location Errors on Localization Attacks in Location-Based Social Network Services[M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 343-357.

[17] Shi W Q, Luo X Y, Zhao F, et al. Geolocating a WeChat User Based on the Relation between Reported and Actual Distance[J]. *International Journal of Distributed Sensor Networks*, 2018, 14(4): 155014771877446.

[18] Shi W Q, Luo X Y, Guo J D, et al. Where Are WeChat Users: A Geolocation Method Based on User Missequence State Analysis[J]. *IEEE Transactions on Computational Social Systems*, 2021, 8(2): 319-331.

[19] Wang J L, Cheng H N, Xue M H, et al. Revisiting Localization Attacks in Mobile App People-Nearby Services[C]. *Proceedings of the 10th International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, 2017: 17-30.

[20] Shi W Q, Luo X Y, Guo J S. Social Network User Geolocating Method Based on Weighted Least Squares[J]. *Chinese Journal of Network and Information Security*, 2022, 8(3): 41-52.
(时文旗, 罗向阳, 郭家山. 基于加权最小二乘的社交网络用户定位方法[J]. *网络与信息安全学报*, 2022, 8(3): 41-52.)

[21] Xue M H, Liu Y, Ross K W, et al. I Know where You Are: Thwarting Privacy Protection in Location-Based Social Discovery Services[C]. *2015 IEEE Conference on Computer Communications Workshops*, 2015: 179-184.

[22] Peng J W, Meng Y, Xue M H, et al. Attacks and Defenses in Location-Based Social Networks: A Heuristic Number Theory Approach[C]. *2015 International Symposium on Security and Privacy in Social Networks and Big Data*, 2015: 64-71.

[23] Wang Z G, Zhu H S, Sun L M. The Concept Evolution Analysis of

Social Engineering[J]. *Journal of Cyber Security*, 2021, 6(2): 12-29.
(王作广, 朱红松, 孙利民. 社工概念演化分析[J]. *信息安全学报*, 2021, 6(2): 12-29.)
[24] Saini Y S, Sharma L, Chawla P, et al. Social Engineering Attacks[M]. Lecture Notes in Networks and Systems. Singapore: Springer Nature Singapore, 2022: 497-509.
[25] Moriya K, Fujimoto M, Arakawa Y, et al. Effective Trilateration-Based Indoor Localization Method Utilizing Active Control of Lighting Devices[J]. *Sensors and Materials*, 2020, 32(2): 625.

郭家山 于 2020 年在郑州大学软件工程专业获得学士学位。现在郑州大学网络空间安全专业攻读硕士学位。研究领域为网络空间资源测绘、网络目标定位。研究兴趣包括: 即时通信用户定位、Android 安全等。Email: silencesliver@qq.com

时文旗 河南省网络空间态势感知重点实验室任讲师。研究领域为网络空间资源测绘、网络目标定位。研究兴趣包括: 网络数据分析、即时通信用户定位。Email: shiwenqi1606@163.com

罗向阳 河南省网络空间态势感知重点实验室任教授、博士生导师。研究领域为网络空间安全、网络态势感知。研究兴趣包括: 网络实体定位、信息隐藏。Email: luoxy_ieu@sina.com

杜少勇 河南省网络空间态势感知重点实验室任讲师。研究领域为网络空间安全、网络空间安全态势感知等。研究兴趣包括: 智能移动终端安全、移动应用服务安全、网络黑产等。Email: shaoyong.du.cs@gmail.com

刘瑞婷 于 2007 年在武汉理工大学计算机应用专业获得硕士学位。现在河南省网络空间态势感知重点实验室攻读博士学位。研究领域为网络空间测绘、网络目标定位。研究兴趣包括: 网络数据分析。Email: liurt_ieu@sina.com