

异构社交网络中用户轨迹隐匿方法优化研究

张艳格

(郑州财经学院 电子商务学院,河南 郑州 450000)

摘 要:在对异构社交网络中用户轨迹进行隐匿时,当前方法大多针对用户单个位置进行轨迹隐匿,不适于复杂的异构社交网络.为此,提出一种新的基于假轨迹的异构社交网络中用户轨迹隐匿方法,通过一个例子对所提方法的基本思想进行分析.对假轨迹方法进行概述,在中心服务器系统结构上实现.给出单个位置的暴露风险、轨迹暴露风险、距离偏移度的概念和计算公式.为了避免攻击者判断出用户真实轨迹的概率,提出记忆规则进行优化.将查询消息、真实位置等参数传输至隐私保护服务器,隐私保护服务器依据用户参数要求和记忆规则产生满足条件的假位置,将含有假位置的匿名框传输至服务提供商,隐私保护服务器对返回的查询结果进行求精后传输至用户.实验结果表明,采用所提方法得到的轨迹数据有很高的可用性.

关键词: 异构社交网络;用户轨迹;隐匿;优化

中图分类号: TP309

文献标识码: A

文章编号: 1000—7180(2017)02—0132—04

User Track Hidden Method in Heterogeneous Network Optimization Research

ZHANG Yan-ge

(Electron Business College, Zhengzhou Institute of Finance and Economics Zhengzhou 450000, China)

Abstract: In the heterogeneous network users in the path of the hidden, most of the current methods for users to a single location to track hidden, not suitable for complex heterogeneous social networks. For this, put forward a new heterogeneous social networks based on false trajectory track hidden by the user, through an example analysis of the basic concept of the proposed method. Are summarized, the methods of false track in a central server system structure on the implementation. Given a single location exposure risk, trajectory exposure risk, the concept and calculation formula of distance deviation degree. In order to avoid the attacker tell users real trajectories of probability, proposed rules to optimize memory. Sending query parameters such as news, true location server to privacy protection, privacy protection server on the parameters of user requirements and the rule satisfy the condition of the false memory location, anonymous boxes containing a false position to service providers, privacy protection server to return after the refinement of the query results of transmission to the user. The experimental results show that the proposed method the trajectory data has a high availability.

Key words: heterogeneous social networks; user trajectory. the hidden; Optimization

1 引言

近年来,随着社交网络的逐渐发展,很多新型在线社交方式应运而生,单一的社会网络关系结构已无法满足人们的要求,所以学者们提出了异构社交网络^[1-3]. 异构社交网络非常复杂,其中包含多种关系与

实体,不可避免的出现一系列用户隐私安全问题^[4-5]. 因此,研究异构社交网络中用户轨迹隐匿方法具有重要意义,为保护用户隐私信息提供了重要依据^[6].

文献[7]提出一种基于位置 k 个匿名用户轨迹隐匿方法,在某一区域内,将其中一个用户与其他 $k-1$ 个用户的查询信息传输至服务器,从而实现用户轨迹隐

匿. 该方法对单个用户的轨迹隐匿保护效果不错,但不适于多个用户;文献[8]提出一种基于投影集的用户轨迹隐匿方法,该方法将部分用户轨迹看作一个准标识符对其他位置轨迹的抑制方案进行标识,构成投影集. 实现投影集方法过程很简单,但问题是投影集所处轨迹数据均不发送,导致数据利用率低;文献[9]提出一种用户轨迹隐匿模型,在同一匿名区域中,用户根据自身的隐私保护度可提出匿名要求,保证匿名范围内各用户查询消息的多样化,以提高攻击者的识别难度. 该方法能够有效实现用户轨迹隐匿,但效率低;文献[10]表明了空间变换的理论,实现用户轨迹的隐藏,此办法隐藏用户所处的地点位置,并将解锁的方法传送给用户个人,保证用户的安全性,但会造成额外的通信和计算消耗.

针对上述方法的弊端,提出基于假轨迹的异构社交网络中用户轨迹隐匿方法. 实验结果表明,采用所提方法得到的轨迹数据有很高的可用性.

2 基于假轨迹的异构社交网络用户轨迹隐匿方法优化

2.1 假轨迹方法概述

假轨迹方法是一种简单可靠的用户隐藏位置的方法,为用户提供假的位置信息数据来掩盖真是的位置信息,大大降低了用户位置被发现的概率.

为了便于分析,通过一个例子对所提方法的基本思想进行分析. 假设异构社交网络中的一个用户在 $t_0, t_1, t_2, t_3 (t_0 < t_1 < t_2 < t_3)$ 时刻对位置 A、B、C 和 D 进行查询,依据查询记录可知该用户的运动轨迹是 $A \rightarrow B \rightarrow C \rightarrow D$. 在用户发送真实查询的过程中,通过假轨迹方法传输部分虚假查询,将虚假查询与真实查询信息结合组成多条路径,使攻击这不能辨认用户的真实轨迹,实现异构社交网络中用户轨迹的隐匿. 图 1 描述的是本例中产生的全部路径,假设每条路径都从 s 点出发,点 C、F 与 I 为各路径的目的地,其他点是组成每条路径的中间点,图 1 中线段上的标注是两点之间的距离.

在异构社交网络中,假设用户每次传输的查询信息共有 $m(m > 1)$ 个位置,查询了 $n (n > 1)$ 次,则形成的总路径条数的上限是 m^n ,下限是 m . 所以,通过假轨迹方法对异构社交网络中用户轨迹进行隐匿,尽管网络中只存在一个用户,该用户仍可隐匿其真实轨迹,同时可求出其运动轨迹隐匿度的上限 m^n 与下限 m .

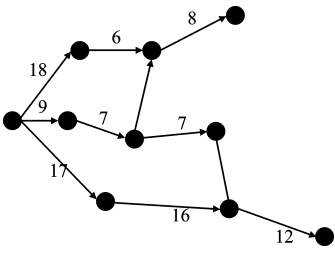


图 1 异构社交网络中用户轨迹隐匿方法示例

2.2 假轨迹方法系统结构

所提的假轨迹方法可在中心服务器系统结构上实现,如图 2 所示.

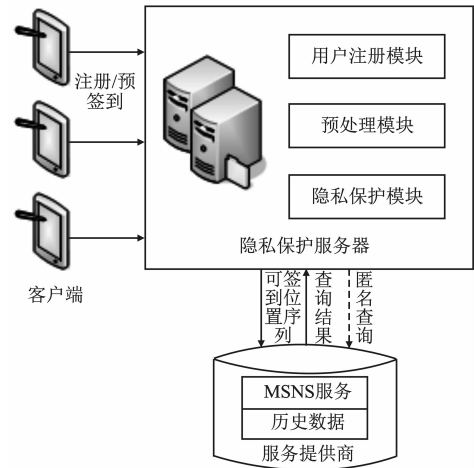


图 2 中心服务器系统结构图

分析图 2 可知,中心服务器系统结构主要包括客户端、隐私保护服务器和服务提供商. 其中,隐私保护服务器是可靠的,攻击可能源自服务提供商.

异构社交网络中用户查询过程如下:在异构社交网络中,用户在某一位置发送查询请求,将其实际位置与隐私保护参数传输至主要的隐私防护系统,隐私防护系统根据用户隐私数据生成加的地点信息,把假位置和真实位置共同传输至服务提供商. 服务提供商接收到信息后,把用户真假地理位置信息结合起来进行分析,并将分析结果发送到隐私防护系统. 隐私防护系统据个人真实情况对结果进行比较,同时将最终结果传输至用户.

2.3 准备知识

用户在平面空间中的真实位置用 (x, y) 进行描述,用户在异构社交网络中查询的消息可描述成:

$$Q = \{ID, (L_i^t, L_{d1}^t, L_{d2}^t, \dots, L_{dk}^t)\}$$
 (1)

式中, ID 用于描述异构社交网络中用户的身份数据; L_i^t 表示 t 时间点的正确位置; $(L_{d1}^t, L_{d2}^t, \dots, L_{dk}^t)$ 表示 t 时间点 k 个地点假消息 k 可以依据个人需求

设定.若用户在连续 m 个时间点进行,实际轨迹可描述成:

$$T_r = \{L_i^1, L_i^2, \dots, L_i^m\} \quad (2)$$

依据用户实际轨迹产生的假轨迹可描述成 $\{L_{dx}^1, L_{dx}^2, \dots, L_{dx}^m | x = 1, 2, \dots, k\}$.

定义 1 单独位置存在风险.用 m 表示个人实际位置 T_r 含有位置数据的总数; S_i 表示 t_i 时间的实际位置信息 T_r 和假信息 T_d 含有全部位置点; $|S_i|$ 为 S_i 的式,因此真是地点被暴露的几率为 $1/|S_i|$.某个地点的暴露风险就是从全部地点中精确得到个人用户真实地点的平均概率,公式描述如下:

$$LD = \frac{1}{m} \sum_{i=1}^m \frac{1}{|S_i|} \quad (3)$$

定义 2 位置暴露危险 TD .假设一共有 n 条真实轨迹和假轨迹,里面有 k 条轨迹和至少 1 条轨迹有交叉点,剩余的 $n-k$ 条轨迹与其他是我轨迹不存在交叉点.用 N_k 表示 k 条轨迹相交后出现是我轨迹总数.因此位置被暴露的可能性为所有可能的轨迹中获取的真实用户位置的几率,公式描述如下:

$$TD = \frac{1}{N_k + (n - k)} \quad (4)$$

定义 3 距离偏移度 DD .距离偏移度就是在第 i 个时间点上,每一条错误轨迹上的点与真实轨迹上的对应点的平均距离,用 DD 进行描述,公式如下:

$$DD = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n} \sum_{j=1}^n \text{dist}(T_r^i, T_{dj}^i) \right) \quad (5)$$

式中, $\text{dist}(T_r^i, T_{dj}^i)$ 为两个点间欧几里得距离.

2.4 用户轨迹隐匿方法优化

在异构社交网络中,当用户连续查询时,将各时间点用户位置信息连接在一起即可得到用户真实轨迹,将用户假位置连接在一起即为用户的假轨迹.若各查询点只是随机产生假位置,则攻击者判断处真实轨迹的概率将大大增大,为了避免该弊端,本节提出相应规则对其进行优化,依据所提规则产生假位置,使通过假位置产生的假轨迹和真实轨迹的相似度尽可能高,防止攻击者判断出真实轨迹.

所提规则为记忆规则:要求前后假位置连接产生的轨迹与真实轨迹之间的距离偏移度要低于 DD 阈值.

异构社交网络中用户提出查询时,将查询消息、真实位置、 $\sigma_{\min}$ 、 $\sigma_{\max}$ 、 LD 及 TD 等参数传输至隐私防护系统,隐私防护系统以个人数据条件和记忆规则生成符合要求的错误轨迹为基础,之后用存在错误地点信息的匿名框发送到服务提供商.异构社交网络里的个人位置隐藏过程如下表示:输入:用户真实

位置 (x, y) , 查询消息 q , 假位置生成范围 $(\sigma_{\min}, \sigma_{\max})$;

输出:包含假位置和真实位置的匿名区域,传输至隐私保护服务器进行查询;

第一阶段: $t = 0$ 时产生假位置,也就是产生初始点的假位置.

(1) $t = 0$ 时异构社交网络中个人实际轨迹为 $A_0 = (x_0, y_0)$;

(2) 在 A_0 周围随机产生一个假轨迹:

$$B_0 = (x'_0, y'_0) = \text{Random}(A_0) \quad (6)$$

(3) 利用下式得出实际位置与假位置之间的距离:

$$C_0 = \text{dist}(A_0, B_0) = \sqrt{(x_0 - x'_0)^2 + (y_0 - y'_0)^2} \quad (7)$$

(4) 若 C_0 符合距离范围要求,也就是 $C_0 \in (\sigma_{\min}, \sigma_{\max})$,则将其返回至假位置集中;否则,重新进行步骤(2);

(5) 重复上述步骤,得到符合要求的假位置数量.

第二阶段: $t = i, i \geq 1$ 时假位置的生成,该阶段不仅需要考虑距离范围,还需保证得到的轨迹符合 LD 和 TD 要求.

(6) t_1 时间点的实际轨迹为

$$A_1 = (x_1, y_1) \quad (8)$$

(7) 在 A_1 周围随机产生一个错位轨迹

$$B_1 = (x'_1, y'_1) = \text{Random}(A_1) \quad (9)$$

(8) 通过反复进行第一阶段的后两个步骤,直到满足 A_1 在要求范围内.

(9) $T_1 = (A_0, A_1)$ 是异构社交网络中用户前后两个真实位置产生的轨迹, $D_1 = (B_0, B_1)$ 是前后两个假位置产生的假轨迹;

(10) 依据个人实际位置 T_1 和假位置 D_1 得出 LD_1, TD_1 ;

(11) 如果位置 B_1 满足 $LD_1 \leq LD$ 且 $TD_1 \leq TD$,则返回至假位置集里;否则,重新产生符合条件的假位置 B_1 ;

(12) 反复进行上一个过程,直到生成的假轨迹数目满足用户对 DD 的要求.

(13) 依据相同的过 程 在 时 刻 $t = i (i = 2, 3, \dots, m)$ 产生符合条件的假位置.

3 实验结果分析

3.1 实验数据与环境

通过 Network-Based Generator of MoVing ob-

jects 模拟器对用户在某异构社交网络中的移动轨迹进行模拟,将得到的 12 000 条移动轨迹作为实验数据集 OLDEN. 表 1 描述的是 OLDEN 数据集的统计信息,其中, D 用于描述数据集中轨迹个数; D_{pre} 用于描述经预处理后的轨迹数量; D_{TEC} 用于描述数据集中等价类个数; Max 用于描述规模最大的等价类包含的轨迹个数; Min 用于描述规模最小的等价类包含的轨迹个数.

表 1 数据集统计信息

数据集	D	D_{pre}	D_{TEC}	Max	Min
OLDEN	12 000	28 955	152	403	64

3.2 轨迹数据可用性测试

异构社交网络中用户轨迹数据的可用性是用用户轨迹隐匿的关键,主要取决于轨迹匿名域的大小,匿名域越大,可用性越低,用户轨迹数据信息损失越高. 本节通过全部等价类轨迹匿名域之和与整个轨迹空间面积之比的均值,对轨迹数据信息损失进行衡量,进而对轨迹数据可用性进行衡量,其公式描述如下:

$$IL = \frac{1}{n_{TEC} Area} \sum_{i=1}^{n_{TEC}} S(t_i) \tag{10}$$

式中, n_{TEC} 用于描述等价类个数; $S(t_i)$ 用于描述 t_i 时刻用户轨迹集形成的匿名域面积; $Area$ 用于描述整个轨迹空间的面积.

为了确定本文方法存在的意义,对将投影集法与空间变换法进行了实验比较. 图 3 描述的是通过本文方法、投影集法和空间变换法对用户轨迹进行隐匿后,轨迹数据信息损失情况比较结果.

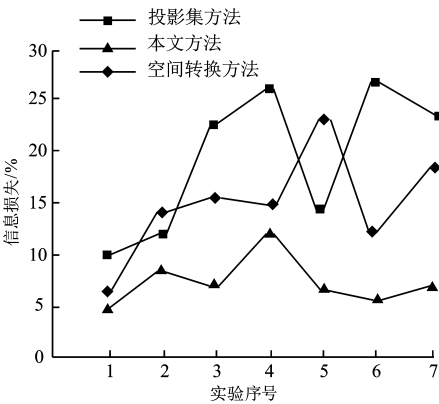


图 3 三种方法信息损失比较结果

分析图 3 可以看出,针对 7 次实验,本文方法的轨迹数据信息损失曲线一直低于投影集方法和空间转换方法,且曲线最为平稳,说明采用本文方法后得到的用户轨迹数据可用性最高.

隐匿率为用户轨迹隐匿中常用的信息损失度量方法,公式描述如下:

$$TSR = \frac{1}{n_{TEC}} \sum_{i=1}^{n_{TEC}} \frac{n_s}{|TEC|} \tag{11}$$

式中, n_{TEC} 用于描述等价类个数; n_s 用于描述隐匿的轨迹个数. 轨迹隐匿率越低,被隐匿的轨迹个数越少,轨迹数据的信息损失越低,发布数据的可用性则越高.

图 4 为通过本文方法、投影集方法和空间转换方法对用户轨迹进行隐匿后轨迹数据信息隐匿率比较结果.

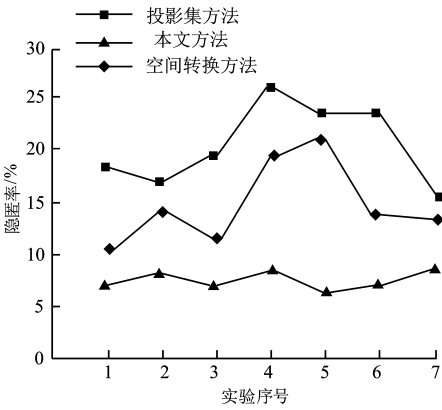


图 4 三种方法隐匿率比较结果

分析图 4 可以看出,通过与投影集方法、空间转换方法相比,本文方法的隐匿率一直是最低的,说明本文方法发布数据的可用性最高.

4 结束语

本文提出一种新的基于假轨迹的异构社交网络中用户轨迹隐匿方法,通过一个例子对所提方法的基本思想进行分析. 对假轨迹方法进行概述,在中心服务器系统结构上实现. 给出单个位置的暴露风险、轨迹暴露风险、距离偏移度的概念和计算公式. 为了避免攻击者判断出用户真实轨迹,提出记忆规则进行优化. 实验结果表明,采用所提方法得到的轨迹数据有很高的可用性.

参考文献:

[1] 施娟娟,朱晓荣,邵世祥. 一种基于用户满意度的网络接入选择和频谱分配联合优化方法[J]. 南京邮电大学学报(自然科学版),2014,34(2):78-82.
[2] 薛迪,吴礼发,李华波,等. TraDR:一种基于轨迹分解重构的移动社交网络位置预测方法[J]. 计算机科学,2016,43(3):93-98.

各类操作 14 440 次,试验中暴露的典型问题经分析可统计为以下几类:(1)部分幅角间隔无法设置相应距离圈;(2)部分数据文件卸载失败;(3)进行地图显示操作时,系统异常退出;(4)数据库异常断开连接等。

通过选取显示控制系统作为试验对象,利用提出的软硬件可靠性综合试验方法,进行软硬件可靠性综合试验,试验过程和结果表明软硬件可靠性综合试验方法考虑了软硬件综合系统的实际使用方式和工作场景,使得产品的试验方式与使用工况更加真实地反映产品实际工作状态,有效地发现影响产品可靠性的问题,保证了综合试验结果更为真实,确保了可靠性鉴定试验的质量,从而提高产品可靠性水平。

4 结束语

本文主要的贡献有:(1)针对传统硬件可靠性试验和软件可靠性测试的不足,提出一种基于剖面映射的软硬件可靠性综合试验方法,实验结果表明软硬件可靠性综合试验方法在减少测试代价的同时,实现了一次试验即验证产品的可靠性;(2)提出了软硬件可靠性综合试验剖面构造方法,该剖面构造方法能够真实反映产品的实际使用情况;(3)提供了综合试验剖面构造方法的形式化描述,该形式化描述便于研发工具原型等,从而可以逐步实现软硬件可靠性综合试验剖面构造和用例生成的自动化,实验结果表明自动化构造和用例生成是可行的。

参考文献:

[1] Jinhee Park, Hyeon-Jeong Kim, Ju-Hwan Shin, and Jongmoon Baik. An embedded software reliability

model with consideration of hardware related software failures [C]//IEEE Sixth International Conference on Software Security and Reliability. IEEE, 2012; 207-214.

[2] Choi M , Park N. Hardware-software co-reliability in field reconfigurable multi-processor-memory systems [C]// FTPDSWORKSHOP 2002. Los Angeles , 2002;1-14.

[3] Norman F. Schneidewind. Computer, network, software, and hardware engineering with applications [M]. John Wiley & Sons, 2012;315-334.

[4] Javier Cano, David Rios. Reliability forecasting in complex hardware/dofware dystems [C]//ARES 2006. Vienna, 2006; 1-5.

[5] Xiaolin Teng, Hoang Pham, Daniel R Jeske. Reliability modeling of hardware and doftwareinteractions, and its applications [J]. IEEE Transactions on Reliability, 2006, 55(4): 571-578.

[6] Bing Huang, Manuel Rodriguez, Ming Li,et al. Hardware error likelihood induced by the operation of software [J]. IEEE Transactions on Reliability, 2011, 60 (3): 622-640.

[7] Yudi Purwantoro, Stuart Bennett. Decomposition technique for integrated dependability rvaluation of hardware-dofware dystems using dtochastic activity networks[C]//The 25th EUROMICRO .Pisa ,1999;142-145.

作者简介:

封二强 男,(1984-),硕士,高级工程师.研究方向为软件可靠性工程、软件测试. E-mail:erqiangfeng@126. com.

蓝新生 男,(1988-),硕士.研究方向为软件可靠性测试、软件测试.

郑 军 男,(1969-),硕士,研究员.研究方向为软件工程.

(上接第 135 页)

[3] 姜来为,吴宣利,张乃通. 异构网络中小区间联合干扰协调算法研究[J]. 通信学报,2016,37(2):143-150.

[4] 伍杰华,朱岸青. 基于异构特征的符号社交网络关系分类[J]. 计算机应用与软件,2015,32(12):42-45.

[5] 董全,李建东,赵林靖,等. 基于效用最大的多小区异构网络调度和功率控制方法[J]. 计算机学报,2014, 37 (2):373-383.

[6] 郭洪尧,张新,贺云璐,等. 基于非合作博弈的异构融合网络非单调性选择策略研究[J]. 计算机应用研究, 2014,31(8):2450-2453.

[7] 陈万志,林澍,王丽,等. 基于用户移动轨迹的个性化健康建议推荐方法[J]. 智能系统学报,2016, 11(2):

264-271.

[8] 肖艳丽,张振宇,杨文忠. 基于 GPS 轨迹的用户移动行为挖掘算法[J]. 计算机应用与软件,2015, 32(11): 83-87.

[9] 孙小伍,马云林,袁磊. 车载主机与录像仪器相结合的新型导航产品研究[J]. 电子设计工程,2016, 24(5): 175-178.

[10] 张晓滨,杨东山,朱保华. 基于多特征相似的用户兴趣推荐[J]. 西安工程大学学报,2016,30(1):97-101.

作者简介:

张艳格 女,(1979-),硕士研究生,讲师.研究方向为网络技术应用和电子商务. E-mail:zhangyange1019@163. com.