

移动边缘群智感知动态隐私度量模型与评价机制

赵明烽¹, Lei Chen², 钟洋¹, 熊金波^{1,3}

(1. 福建师范大学数学与信息学院, 福建 福州 350117;

2. College of Engineering and Computing, Georgia Southern University, GA, 30458;

3. 福建省网络安全与密码技术重点实验室, 福建 福州 350117)

摘要: 移动边缘群智感知中, 用户执行感知任务采集数据所包含的隐私量是动态变化且不直观的, 数据上传亦缺乏隐私风险预警值, 提出一种动态隐私度量(DPM)模型。给出用户参与感知任务所获数据的结构化表示并转化成原始数值矩阵, 引入隐私属性偏好与时效性因素实现对该矩阵的权重叠加, 以度量数据所含隐私的动态变化, 基于权重叠加后的矩阵合理计算用户个性化隐私阈值, 并进行差分隐私处理。在此基础上, 设计一种隐私度量模型评价机制。仿真结果表明, 模型是有效且合理的, 根据所给范例, 差分隐私处理后的数据效用达到0.7, 随噪声水平增加, 隐私保护程度(PDD)可显著提升, 适应物联网移动边缘群智感知范式。

关键词: 动态隐私度量; 个性化隐私阈值; 差分隐私; 模型评价; 移动边缘群智感知

中图分类号: TP309.2

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2021016

Dynamic privacy measurement model and evaluation system for mobile edge crowdsensing

ZHAO Mingfeng¹, LEI Chen², ZHONG Yang¹, XIONG Jinbo^{1,3}

1. College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, China

2. College of Engineering and Computing, Georgia Southern University, GA 30458, USA

3. Fujian Provincial Key Laboratory of Network Security and Cryptology, Fuzhou 350117, China

Abstract: To tackle the problems of users not having intuitive cognition of the dynamic privacy changes contained in their sensing data in mobile edge crowdsensing (MECS) and lack of personalized privacy risk warning values in the data uploading stage, a dynamic privacy measurement (DPM) model was proposed. A structured representation

收稿日期: 2020-09-16; 修回日期: 2020-12-14

通信作者: 熊金波, jbxiong@fjnu.edu.cn

基金项目: 国家自然科学基金(61872088, U1905211, 61872090); 福建省自然科学基金(2019J01276); 贵州省公共大数据重点实验室开放课题(2019BDKFJJ004)

Foundation Items: The National Natural Science Foundation of China (61872088, U1905211, 61872090), The Natural Science Foundation of Fujian Province, China (2019J01276), The Guizhou Provincial Key Laboratory of Public Big Data Research Fund (2019BDKFJJ004)

论文引用格式: 赵明烽, Lei Chen, 钟洋, 等. 移动边缘群智感知动态隐私度量模型与评价机制[J]. 网络与信息安全学报, 2021, 7(1): 157-166.

ZHAO M F, LEI C, ZHONG Y, et al. Dynamic privacy measurement model and evaluation system for mobile edge crowdsensing [J]. Chinese Journal of Network and Information Security, 2021, 7(1): 157-166.

of data obtained by a user participating in a sensing task was introduced and was transformed it into a numerical matrix. Then privacy attribute preference and timeliness were presented to quantify the dynamic privacy changes of data. With this, personalized privacy thresholds of users based on the numerical matrix were reasonably calculated. Finally, differential privacy processing was performed on the numerical matrix, and a model evaluation system was designed for the proposed model. The simulation results show that the DPM model was effective and practical. According to the given example, a data utility of approximately 0.7 can be achieved, and the degree of privacy protection can be significantly improved as the noise level increases, adapting to the MECS of IoT.

Keywords: dynamic privacy measurement, personalized privacy threshold, differential privacy, model evaluation, mobile edge crowdsensing

1 引言

作为一种新颖的物联网感知模式,移动边缘群智感知(MECS)^[1]借助各式各样的智能终端^[2-3]、可穿戴设备协同完成一些泛在智慧型深度社会感知任务,帮助推动现代城市的立体功能延伸和功能耦合,高度响应了国家政府和产业界对智慧城市建设^[4-5]的号召。同时,随着第五代(5G)移动通信网络^[6-7]的到来和“万物互联”建设脚步的加快^[8],人们对移动应用服务的需求量日益剧增,由此催生的海量移动数据包含着用户众多的隐私信息。数据俨然已经渗透到各个行业和领域,成为重要的生产要素。

在实际应用中,感知用户在感知活动结束后需要上传自己的感知数据来获得相应的报酬^[9]。同时,他们面临着隐私泄露的安全隐患。感知用户对自身参与感知任务所产生数据的隐私量没有直观的认知。在采用隐私保护技术隐藏某些隐私信息之后^[10],经过处理的数据仍然会泄露隐私,例如 K-匿名技术无法抵抗同质攻击和背景知识攻击^[11]。如何确定隐私泄露的风险界限^[12],以及数据的后续效用对任务报酬有多大影响,这些都是帮助用户更好地管理自身隐私信息资产的重要因素。若不能有效地度量用户参与感知任务的数据隐私及数据效用等^[13-14],帮助用户进行更加理性的感知上传行为,将会导致用户对自身隐私泄露风险的错误认知,进而产生消极的任务参与态势,导致应用服务商获得质量低下的感知数据而无法提供优良的移动服务,以此往复形成恶性循环。

数据隐私度量较为复杂,因其在不同环境中都具有鲜明的敏感特征和属性,且在任务联动、

时空变换情况下会对数据特性产生动态影响。目前,国内外针对 MECS 网络中的隐私度量方法研究主要分为 3 种^[15]:一是利用信息熵^[16],根据隐私信息中所包含的不确定程度来度量隐私;二是基于概率统计的方法^[17],利用概率分布信息来推断隐私信息的可能性,以此度量隐私泄露的风险;三是结合集对分析理论^[18],一种将定性与定量相结合并解决确定与不确定性问题的方法。上述方法存在一定的缺陷,且考虑因素单一。例如,基于信息熵的度量本质是对信息混乱程度的界定,在对隐私风险泄露度量方面具有较好的效果,但在面临 MECS 中多源数据类型时,不同用户的数据分为敏感性或非敏感性,单一考虑信息混乱变化因素将无法适应 MECS 范式的数据;基于概率统计的方法则需要掌握一定程度的事件发生概率的先验信息,具有一定的局限性;基于结合集对分析理论的方法主要是针对两数据集之间的关联隐私信息的不确定性进行度量,因此需要具有给定数据集的前提条件,对于 MECS 中数据的实时性等问题无法进行有效度量。由此,基于数据结构化、概率论及高阶矩阵范数等理论,本文考虑多隐私属性、用户隐私偏好权重和时间等多维因素,提出一种针对 MECS 范式数据的动态隐私度量模型和评估机制。其主要贡献总结如下。

(1) 针对 MECS 服务中的用户参与感知任务所产生数据设计一种动态隐私度量模型,分析了多隐私属性、用户隐私属性偏好以及时效性等隐私度量因素,以动态度量用户参与任务的数据隐私量,并提出一种面向用户的个性化隐私阈值计算方式。

(2) 基于拉普拉斯噪声机制对所提模型进行差分隐私处理,并提出一种动态隐私度量模型评

价机制,从隐私量、数据效用以及隐私保护程度等方面对模型效果与性能进行综合评估与分析。

(3) 仿真实验结果表明,所提模型能够有效刻画用户任务数据的隐私量变化,正确反映模型的数据效用和隐私保护程度之间的关系,且能证明个性化隐私阈值计算的有效性与合理性,并为用户在上传感知数据时提供客观的隐私度量值和对应的隐私泄露风险界限,且适合 MECS 范式。

2 相关工作

关于利用信息熵的相关隐私度量方法,Shi 等^[19]针对社交网络服务中保护图形格式数据的隐私方法缺少评价标准的问题,提出了一种利用复杂网络中网络静态特征之一的网络结构熵进行隐私度量的方法,且给出了隐私度量指标(PMI)来度量图结构的隐私保护能力。在基于位置的服务中常常会因为位置信息暴露用户的家庭地址、健康状况以及购物习惯等,Shaham 等^[20]针对该问题,考虑了基于用户连续位置变换而产生的新的边信息,并提出了一种新的隐私度量方法——传递熵来研究隐私保护问题,此外,还提出了一个贪心算法来提高虚拟生成算法的传递熵性能。在开放和动态计算环境中,度量隐私损失和信任获得是一个有意义的课题,针对相关现有工作在度量过程中没有考虑到隐私信息与动态信任变化之间的关系,Gao 等^[21]提出了一种新的基于信息论的隐私度量方法,通过条件概率计算信息公开时的隐私损失熵和信任收益熵,并通过调整熵的权重,灵活地应用于不同的应用中,实验结果表明该度量方法具有显著性能,在实体之间进行隐私信息交换时,可以减少隐私损失,获得更多的信任。针对云数据间的关联性问题,张宏磊等^[22]引用条件熵提出了一种对云数据操作过程中的隐私信息泄露风险的度量方法,但只考虑了在攻击者无相关背景知识条件下的隐私度量。针对这一问题,文献[23]基于条件熵、互信息等概念提出一种拥有背景知识攻击者攻击的隐私度量方法,并通过举例位置隐私场景,构建了具体的信息熵模型及隐私保护机制和攻击者能力的度量及分析方法。针对移动服务中用户不得不向运营商披露个人信息而构成的隐私威胁,文献[24]基于信息

熵和马尔可夫链建立了移动服务隐私安全风险预估模型,给出了合理的风险度量和评估方法。

对于基于概率统计的方法,马蓉等^[25]从隐私敏感属性的角度考虑度量问题,借助公众属性和个性化属性等度量指标以及历史数据统计,对时空感知数据进行隐私度量,从而帮助选择最佳用户参与任务。Wang 等^[26]也从该角度提出了一种基于属性的统计模型,可用于基于个人可识别信息的隐私暴露度量和隐私影响评估,涉及隐私属性、隐私敏感性和属性相关性 3 个重要因素。Liu 等^[27]则从用户态度的角度对隐私进行测量,基于个体的内在和外在敏感度,提出了一种用于移动参与式感知系统的个性化隐私测量方法(PriMe),实验结果表明 PriMe 提供了合理而准确的结果,而参与者反过来也高度信任该系统。此外,根据隐私属性值的概率分布,文献[28-30]从基于贝叶斯概率的度量角度,提出了基于贝叶斯推理的度量隐私信息泄露的方法,通过分析和比较推测的信息与隐私信息之间的差异度来度量隐私信息泄露的风险,两者之间的差异度越小,隐私信息泄露风险越大。Zhang 等^[31]针对现有位置服务隐私保护机制(LPPM)缺乏隐私度量评估的问题,提出了一种基于贝叶斯条件隐私的隐私度量模型,通过关于敌手估计误差的条件隐私的一般定义来比较不同的基于位置服务(LBS)的隐私度量。为了改进对车载自组织网络(VANET)中暴露的隐私程度的度量,Han 等^[32]创建了一种以用户为中心的隐私计算系统,依赖于事件发生的概率提出了一个风险评估函数和一组决策权重来模拟决策意图,并结合信息熵知识构建了混合区自适应动态生成机制。对于一些匿名化技术忽视用户的信息自主权问题,Tesfay 等^[33]提出了一种基于组合概率数学模型和机器学习分类器的以用户为中心的隐私风险检测和度量框架,使用户能够控制自己的数据发布。

对于集对分析理论的度量方法,文献[34]基于该思想提出了一种集对分析隐私度量方法,通过对数据集之间的关联关系进行度量分析,考虑对相邻数据子集的同一度、差异度和对立度特性进行度量描述,且具体讨论了数据库隐私保护、位置隐私保护和轨迹隐私保护 3 种不同模式下隐

私保护机制的集对分析方法实例。攻击者可以利用大数据分析技术通过社交网络发布的数据发现用户的隐私,针对背景知识未定义情况下的隐私度量问题,Huang 等^[35]借鉴集对分析理论,提出了一种新的网络环境下的隐私度量方法,将个人隐私度量的结果分为确定值和不确定参数两部分,其中,确定值可以反映隐私暴露程度,不确定参数可以反映背景知识的变化对隐私暴露程度的影响,从而令隐私分析可以在不断变化的背景知识中进行动态调整。

综上所述,目前关于隐私度量的工作比较丰富。现有方案主要涉及对隐私敏感属性、用户个性化隐私偏好、数据关联性、时空状态变化等度量因素和标准的考量。然而,大多数方案均考虑单一的隐私度量标准或因素,对于适应 MECS 网络中同时具有复杂隐私属性特征、隐私属性偏好、时间变化等的感知数据隐私度量方法目前还比较匮乏。因此,从多方因素出发构建新的动态隐私度量模型对于适应 MECS 范式类型的感知数据非常重要。

3 动态隐私度量模型

3.1 感知数据的结构化表示

感知用户在确定接收到某感知任务后,首先,将分配给任务的总执行时间分成 n 等份,即将任务的执行过程划分为 n 个执行周期,分别为 $TC_1, TC_2, \dots, TC_n$ 。在第一个任务周期 TC_1 中,将感知数据存储至数据库,视为基础数据集,并根据不同传感器感知数据类型对数据集进行属性分析。将任务执行过程所感知到的数据属性记录下来,可表示为 $TA_1, TA_2, \dots, TA_k$ 。由此,给定某任务,其任务执行全周期的结构化数据形式表示如表 1 所示。其中, TC_i 表示第 i 个任务周期, TA_j 表示第 j 个感知数据属性, $Data_{i,j}$ 表示第 i 个任务周期的第 j 个数据属性。

由于任务类型的关系,数据表中可能含有非数值型数据,为了便于后续隐私敏感度的计算,采用非负数值映射方法,将数据表转化为数值矩阵。该映射定义如下。

定义 1 非负数值映射^[36]

设 X 是一个含有限个非负数值元素的集合,

f 是一个映射,如果对每一个 $x \in X$, 满足 $f(x) \in \mathbb{R}^+ \cup \{0\}$, 则 f 为非负数值映射,所有的非负数值映射构成的集合记为 F 。

依据不同数据属性自身的敏感特性,按照数据属性敏感性越高映射数值越大的原则来赋值。具体地,对于同一属性 TA_j 的多个感知数据采集值,为越敏感的类型值赋予越大的映射数值。通过选取 n 个非负数值映射 $f_1, f_2, \dots, f_n \in F$, 可以将表 1 进行数据化处理,即

$$d_{i,j} = f_j(Data_{i,j}), \quad i=1,2,\dots,n; j=1,2,\dots,k \quad (1)$$

得到结构化数据数值化处理的结果,记 $D=(d_{i,j})_{n \times k}$ 为隐私敏感数据矩阵。

表 1 任务全周期的结构化数据形式表示
Table 1 Structured data representation of the full cycle of the task

任务周期	数据属性			
	TA_1	TA_2	$\dots$	TA_k
TC_1	$Data_{1,1}$	$Data_{1,2}$	$\dots$	$Data_{1,k}$
TC_2	$Data_{2,1}$	$Data_{2,2}$	$\dots$	$Data_{2,k}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
TC_n	$Data_{n,1}$	$Data_{n,2}$	$\dots$	$Data_{n,k}$

3.2 用户的隐私属性偏好

定义 2 隐私属性偏好向量

给定任务,设 p_j 是某一用户对隐私属性 TA_j 对应数据的不愿被披露程度的权重值,由同一用户对每一隐私属性数据披露程度权重值组成的向量 $\pi=(p_1, p_2, \dots, p_k)$, 满足 $0 \leq p_j \leq 1 (j=1,2,\dots,k)$, 称 π 为这一用户的隐私属性偏好向量。

考虑参与感知任务的用户隐私的动态计算,对数据的不同属性的隐私敏感度进行等级划分显得尤为重要。感知数据在任务的连续进程中是不断变化和积累的,包括用户移动位置、时间、速度、服务请求等。对不同的用户,其某些敏感信息对于其他人可能并不在意。换句话说,每个人对自己敏感信息的偏好程度均不同。在此基础上,对于给定任务,本文将不同属性数据的隐私敏感性划分为 p 个等级,具体值为 $0,1,2,\dots,p-1$ 。同时为了增加用户对自身隐私数据保护的参与度,用户可以为每一特定感知数据属性 TA_j 选择对

应的特定值 S_{TA_j} 。其中，特定值 S_{TA_j} 的具体取值范围即上述对隐私敏感性所划分的 p 个等级，也即 $S_{TA_j} \in [0, p-1]$ 。在相同取值区间的情况下，用户自主选择的等级值更具参考性，且具备归一化的前提条件。由此，用户对特定属性 TA_j 的数据的隐私偏好度可以被定义为 $AP_degree_{TA_j}$ 。

$$\begin{aligned} AP_degree_{TA_j} &= \ln(1 + (S_{TA_j} + \beta) / p), \\ S_{TA_j} &= 0, 1, \dots, p-1 \end{aligned} \quad (2)$$

这里 $\beta(0 < \beta < 1)$ 一般取非常小的数，作用是防止偏好度参数为零，出现数据异常。

设 $\pi_{TC_i} = (p_{TC_i,1}, p_{TC_i,2}, \dots, p_{TC_i,k})$ 为某一用户第 $i(i=1, 2, \dots, n)$ 个任务周期的隐私属性偏好向量，这里假定该用户在每一任务周期对特定隐私属性的偏好度一致，则整个任务过程所构成的矩阵称为隐私属性偏好矩阵 Π ，即

$$\begin{aligned} \Pi &= (\pi_{TC_1}, \pi_{TC_2}, \dots, \pi_{TC_n})^T = \\ & (AP_degree_{TC_1, TA_j}, AP_degree_{TC_2, TA_j}, \dots, AP_degree_{TC_n, TA_j})_{n \times k} \end{aligned}$$

本文将隐私偏好对数据敏感程度的影响视为对任务感知数据矩阵的改变，因此通过对任务感知数据矩阵 D 与 Π 的点乘 $G = D \otimes \Pi = (d_{i,j} \cdot AP_degree_{i,j})_{n \times k}$ ，得到带有隐私偏好权重的隐私敏感数据矩阵 G 。

3.3 隐私属性偏好的时效性

进一步地，考虑到用户的隐私偏好会随时间迁移而变化，即隐私偏好是具有时效性的。换句话说，某一用户对某一隐私属性的重视程度会随着时间的增加而改变。为了度量隐私属性偏好的时效性，这里引入隐私属性偏好函数。

定义3 隐私属性偏好函数

设 $\varphi(t) = \delta^{\sqrt{t}}$ 是 $(0, +\infty)$ 上的函数， $\delta(\delta > 0)$ 是时间影响因子， $t(t=1, \dots, n)$ 的变化推移分别对应整个任务过程的 n 个周期，称 $\varphi(t)$ 为隐私属性偏好函数。

特别地，有3类隐私属性偏好函数。

(1) 第一类：个体对待某一隐私属性的重视程度会随时间的推移而减少，此时 $0 < \delta < 1$ 。

(2) 第二类：个体对待某一隐私属性的重视程度会随时间的推移而增加，此时 $\delta > 1$ 。

(3) 第三类：个体对待某一隐私属性的重视

程度不随时间的推移而变化，即 $\varphi(t)=1$ 为常数。

以上描述的3类函数能够简明地表示用户个体对隐私属性的重视程度随时间推移的变化情况，其他复杂的情况均可由这3类隐私属性偏好函数分段构成。类似于带隐私偏好的感知数据矩阵，本文构建带隐私偏好时效性的感知数据矩阵。设某一任务周期 TC_i 的隐私属性 TA_j 的隐私属性偏好函数为 $\Psi(t) = (\varphi_{i,j}(t))_{n \times k}$ ，则全部任务周期的每一隐私属性的偏好函数组成的矩阵为 $\Psi(t) = (\varphi_{i,j}(t))_{n \times k}$ 。因此，将带有隐私偏好权重的任务感知数据矩阵 G 与 $\Psi(t)$ 进行点乘 $Z = G \otimes \Psi(t) = (d_{i,j} AP_degree_{i,j} \varphi_{i,j}(t))_{n \times k}$ ，得到了同时具有隐私偏好权重和时效性的隐私敏感数据矩阵 Z ，其中每一项对应某一任务周期的某一隐私属性数据的隐私敏感程度值。

3.4 隐私度量模型

基于上述描述，本节给出用户参与感知任务所获数据的具体隐私度量方式。本文将对结构化数据中的隐私度量分析转换为对隐私数据矩阵的度量，便于更直接地观察。由于结构化数据矩阵的敏感属性特点均不相同，非负数值映射的值域也不同，因此在计算矩阵的隐私量时，需进行归一化处理。结构化数据在数值化后的隐私数据矩阵所构成的集合可记为 $\mathcal{D}$ 。设 $D \in \mathcal{D}$ ， D 即某一用户的隐私敏感数据矩阵。用 $|D|$ 表示敏感数据矩阵的隐私量，则有

$$|D| = \left\| \frac{D}{\max\{D\}} \right\|_F = \frac{1}{\max\{D\}} \|D\|_F \quad (3)$$

其中， $\max\{D\} = \max_i \max_j \{d_{i,j}\}$ ，矩阵 D 的 F -范

数 $\|D\|_F = \left(\sum_{i=1}^n \sum_{j=1}^k d_{i,j}^2 \right)^{\frac{1}{2}}$ 。由此，得到具有隐私偏好

权重和时效性的隐私敏感数据矩阵 Z 的隐私量为 $|Z| = \|Z / \max\{Z\}\|_F$ 。

3.5 个性化隐私阈值

在执行感知任务时，用户本地采集和存储原始感知数据，其可以选择将整个数据集或者其中的一部分上传至服务器。随着数据量的增加，用户的隐私泄露程度也随之增加。当用户选择数据的隐私敏感程度值之和达到一定程度时，将对用户

隐私造成严重威胁。因此,亟须针对用户执行感知任务这一动态过程计算其隐私阈值,帮助感知用户进行个性化的隐私数据上传。本文考虑用户最关注的数据收益,以及适应移动设备开销受限的问题,用户可以依据此阈值以在不遭受隐私泄露威胁的前提下最大化选择所能上传的数据量,从而最大化自身收益。基于所提隐私度量模型,提出式(4)来计算得到个性化隐私阈值 T 。

$$T = \sum_{i=1}^n \frac{1}{M(TC_i)} \left(\sum_{j=1}^k Z_{i,j} + q \right) \quad (4)$$

其中, $M(TC_i)$ 是任务第 i 个周期的感知数据量。首先对隐私数据矩阵的每一行进行求和,接着对每一周期的和值求期望。 $q = e^{\frac{t^2}{2\delta^2}}$ ($t=1,2,\dots,n$) 是满足正态分布的一个控制因子, δ 设置为一个较小的数,从而 q 可以适当且平滑地弥补对每个周期期望的损失值。

4 任务数据矩阵的隐私处理

差分隐私^[37]作为一种基于数据失真的隐私保护技术,即通过对敏感数据添加随机噪声使数据结果与原始数据发生一定的偏差,其具有严格的数学证明,在达到隐私保护目的的同时,也保持了一定的数据效用。

定义 4 拉普拉斯分布^[38]

假设随机变量 x , 其概率密度函数分布为

$$f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) = \frac{1}{2b} \begin{cases} \exp\left(-\frac{\mu-x}{b}\right), & x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right), & x \geq \mu \end{cases} \quad (5)$$

其中, μ 是位置参数, $b(b>0)$ 是尺度参数,称随机变量 x 服从 μ, b 的拉普拉斯分布,即 $x \sim \text{Laplace}(\mu, b)$ 。由此,可以得到其累积分布函数为

$$F(x) = \int_{-\infty}^x f(t)dt = \frac{1}{2} + \frac{1}{2} \text{sgn}(x-\mu)(1 - \exp(-\frac{|x-\mu|}{b})) \quad (6)$$

它的逆累积分布函数为

$$F^{-1}(p) = \mu - b \text{sgn}\left(p - \frac{1}{2}\right) \ln(1 - 2\left|p - \frac{1}{2}\right|) \quad (7)$$

在本文所提的隐私度量模型中的数据主要是数值型数据,因此可以借助拉普拉斯机制产生的噪声来提供差分隐私保护,其所造成的数据偏差在本文中被视为敏感数据矩阵 Z 上的偏差。具体地,通过服从均匀分布的随机数和拉普拉斯分布的逆累计分布函数来产生服从拉普拉斯分布的随机数。设 a 是均匀分布 $[-0.5, 0.5]$ 上的一个随机数,均值 $\mu=0$, 标准差为 σ , 则满足拉普拉斯分布的随机数为

$$x = \mu - \frac{\sigma}{\sqrt{2}} \text{sgn}(a) \cdot \ln(1 - 2|a|) \quad (8)$$

标准差 σ 反映了添加噪声的大小, σ 越大表示添加的噪声就越大,反之则越小。由此,通过构造相同行列数的噪声矩阵,敏感数据矩阵 Z 可变为数据加噪后的矩阵,即 $Z \rightarrow Z' = Z - |\Delta Z|$, 这里 $|\Delta Z|$ 表示 ΔZ 的绝对值,其中 $\Delta Z = (\Delta d_{i,j})_{n \times k}$, $\Delta d_{i,j}$ 为 $d_{i,j}$ 的随机噪声值。若原始隐私数据矩阵的某一元素值小于噪声矩阵对应位置的噪声值,则其加噪后变为 0。

5 模型评估机制

从上述内容可知,本文对任务感知数据进行结构化表示和数值化、隐私偏好的度量和时效性的加入,且得到了动态隐私阈值的概念和求解。接下来,本文将对模型构建评估机制,从数据效用以及隐私保护程度等评估指标,考虑模型的隐私度量效果与有效性。

5.1 数据效用

对隐私敏感数据效用的度量需要一个参考标准。由于所采集的数据属性以及类别等无法提前指定和预测,因此,采用与原始数据相比较的方法来度量丢失的信息,即将原始数据矩阵中每一数据的数据量记为 1,经过差分隐私的噪声添加处理后数据的数据量取值范围为 $[0, 1]$ 。

定义 5 矩阵 F -范数

设 $A = (a_{i,j})_{m \times n}$ 是一个矩阵,矩阵 F -范数 $\|A\|_F$ 的计算式如下。

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2 \right)^{\frac{1}{2}} \quad (9)$$

定义 6 数据效用

设 $\mathbf{Z}$ 是具有隐私偏好权重和时效性的原始隐私数据矩阵, $\mathbf{Z}'$ 是 $\mathbf{Z}$ 经过差分隐私加噪处理后的隐私数据矩阵, 且具有相同的结构, $\mathbf{Z}$ 和 $\mathbf{Z}'$ 的数据量分别用 $N(\mathbf{Z})$ 和 $N(\mathbf{Z}')$ 来表示, 则有

$$N(\mathbf{Z}) = \|\mathbf{Z} \odot \mathbf{Z}\|_F, \quad N(\mathbf{Z}') = \|\mathbf{Z}' \odot \mathbf{Z}\|_F \quad (10)$$

其中, 符号 $\odot$ 的定义如下。

设 $\mathbf{E} = (e_{i,j})_{n \times k}$ 和 $\mathbf{F} = (f_{i,j})_{n \times k}$ 是两个具有相同行和列数的矩阵, 记 $\odot: \mathbf{E} \odot \mathbf{F} = (h_{i,j})_{n \times k}$ 。

$$h_{i,j} = \begin{cases} \frac{e_{i,j}}{f_{i,j}}, & f_{i,j} \neq 0 \\ 1, & f_{i,j} = 0 \end{cases} \quad (11)$$

本文称 $N(\mathbf{Z}'|\mathbf{Z}) = N(\mathbf{Z}') / N(\mathbf{Z})$ 为矩阵 $\mathbf{Z}$ 经过加噪处理后的数据效用。容易看出, 因为 $\mathbf{Z}' = \mathbf{Z} - |\Delta \mathbf{Z}|$, 所以 $\mathbf{Z}' < \mathbf{Z}$ 恒成立, $\mathbf{Z}' \odot \mathbf{Z} \in (0, 1)$, 即 $N(\mathbf{Z}') \in (0, 1)$ 。 $N(\mathbf{Z})$ 作为原始隐私数据矩阵的数据量可以看成是单位 1, 由此 $N(\mathbf{Z}'|\mathbf{Z})$ 的值域为 $(0, 1)$, 其值越接近 1, 表明处理后的数据越接近原始隐私数据, 也即真实值。

5.2 隐私保护程度

对感知数据进行模型处理后的数据隐私保护程度 PPD 的度量也是模型评估的一个重要方面, 它与数据效用一起可以评价模型的好坏。本文通过比较采用差分隐私技术处理后的数据隐私量与个人动态隐私阈值的差值情况来衡量隐私保护程度。给定隐私数据矩阵 $\mathbf{Z}$ 和差分隐私处理后的矩阵 $\mathbf{Z}'$, 采用式(12)来进行度量。

$$\text{PPD} = \frac{\left| \frac{\|\max\{\mathbf{Z}'\} \cdot \mathbf{Z}\|_F - 1}{\|\max\{\mathbf{Z}\} \cdot \mathbf{T}\|} \right|}{\|\max\{\mathbf{Z}'\} \cdot \mathbf{Z}\|_F} \quad (12)$$

其中, T 是某一产生隐私数据矩阵 $\mathbf{Z}$ 的用户所对应的隐私阈值。PPD 的值域为 $[0, 1]$, 且与 T 成负相关。当其值越接近 1 时, 表明所提模型的隐私保护程度越高, 个人用户的动态隐私阈值计算结果也越合理。

6 性能分析

实验在 Intel®Core(TM) i7-1050U CPU@1.80 GHz, 12 GB DDR4, 1 TB hard disk 的 Windows 10 硬件平台上进行, 软件配置主要有 Python3.6、IDE PyCharm 以及 MATLAB R2018a。

数据集采用的是共享单车用户的移动行为数据, 包括使用者性别、年龄、每次骑行的起点和目的地、开始和结束时间、起点经纬度、目的地经纬度等属性。该数据集所包含的用户是具有高度移动特性的, 且所产生的数据主要依赖自身携带设备装载的传感器, 适用于 MECS 范式。

6.1 隐私量

本文随机选取了一位用户的行动数据集, 记作矩阵 $\mathbf{Z}$ 。通过模型计算得到隐私量, 从图 1 可以看出, 其隐私量随时间的推移先增长再逐步降低。在即将抵达目的地的最后时间戳部分, 隐私量有一个显著的降低, 这表明该用户在行动任务初始阶段对自身轨迹、目标地点信息具有较强的隐私概念。而在行动即将结束时, 该用户则对部分敏感属性数据的重视程度逐渐降低。此外, 很容易注意到, 用户隐私量的变化与他们对隐私属性的偏好有关, 并且短期出行的用户在感知活动过程中对隐私保护的需求更大。

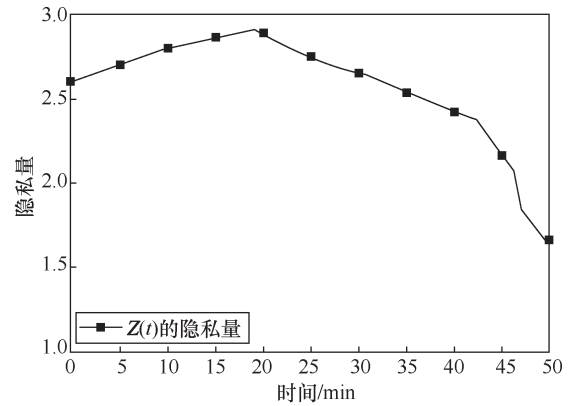


图 1 $\mathbf{Z}(t)$ 的隐私量随时间 t 的变化情况

Figure 1 Changes in the amount of privacy $\mathbf{Z}(t)$ over time t

6.2 差分隐私处理

对于矩阵 $\mathbf{Z}$, 本文设置噪声水平从 0 到 0.25 的逐步提升, σ 越大, 其噪声越大, $\mathbf{Z}$ 值根据所提数据效用和隐私保护程度的定义式进行计算。从图 2 结果可以明显看出, 所选数据在差分隐私机制的作用下, 数据效用随着噪声的增大在逐渐减低, 且与 PPD 成负相关, 符合隐私保护规律。由于 σ 的取值较低, 隐私保护程度总体较低, 但其变化趋势仍是在逐步上升的。具体而言, 随着噪声水平的逐渐增大, 根据式(9)中数据效用的定义, 处理后的模型数据的效用基本上可以保持在

0.7 以上。可以得出结论, DPM 对感知任务数据具有较好的隐私度量效果。这是因为 DPM 在测量过程中充分考虑了多个隐私属性以及不同感知用户对不同隐私属性的偏好, 并同步任务过程中相应时间戳变化的影响, 达到了动态测量的效果。

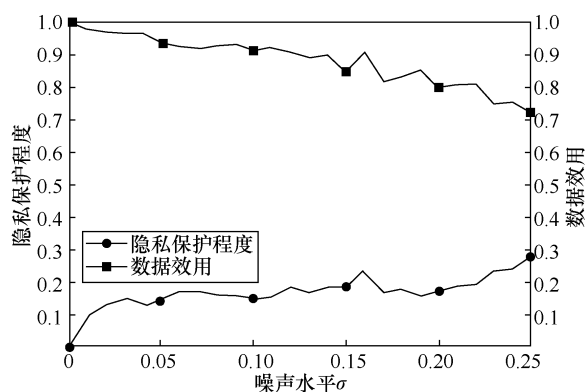


图 2 隐私保护程度与数据效用随噪声增加的变化情况
Figure 2 Changes in the degree of privacy protection and data utility as noise increases

6.3 隐私阈值合理性

本文随机选取 20 名移动感知用户的活动数据集, 根据所给隐私阈值定义式计算得到每一用户感知活动过程的个性化隐私阈值。此外, 给定 $\sigma = 0.25$, 进行统一噪声水平的差分隐私保护机制。如图 3 所示, 由于不同用户对敏感数据的偏好度不同, 在一致噪声水平的隐私保护程度也呈不规则趋势。基于局部的变化可以明显看出, 当用户隐私阈值在相对减小时, 其对应个体的隐私保护程度也在相对提高。虽然这里采用的是 20 个独立用户的数据, 但从部分用户隐私保护程度呈正比例连续上升(下降)的局部可以发现, 相应的用户的隐私阈值是呈负比例连续下降(上升)的。从多个局部变化可以总结得出, 个体隐私阈值和隐私保护程度呈负相关, 这与式(11)的定义保持一致。另外, 给定 $\sigma = 0.25$, 从图 2 可以看出感知用户的数据隐私保护度约为 0.3。由图 3 结果显示, 所选用户的数据隐私保护程度有的大于 0.3, 有的小于该值。这是因为不同的感知用户(SU, sensing user)对其隐私属性具有不同的优先权重。但是, 在式(4)给出的 q 值的影响下, 可以更好地补偿每个任务周期的隐私量的预期偏差值。 q 包含的时间因子还可以用于根据感知用户的隐私属性偏好合理地体现未来任务数据隐私度的

重要性。基于以上分析, 证明了给出的隐私阈值的计算是合理且有效的。

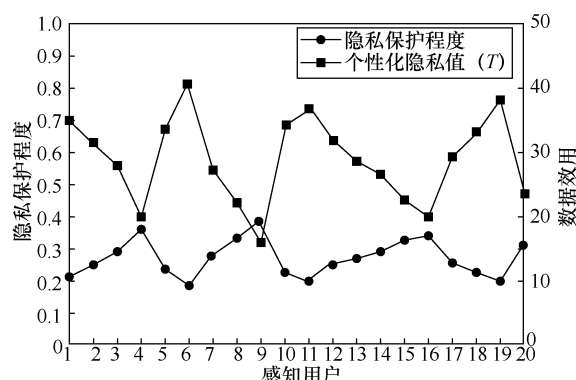


图 3 不同用户的动态隐私阈值和隐私保护程度关系
Figure 3 The relationship between the dynamic privacy threshold of different users and the degree of privacy protection

7 结束语

为了克服 MECS 范式中用户对自身感知任务隐私的认知不足, 突破现有相关工作对隐私度量因素考虑较为单一性的问题, 本文提出了一种面向感知任务数据的动态隐私度量模型。考虑多隐私属性、用户隐私属性偏好和时效性等因素, 给出了用户个性化隐私阈值计算定义式。随后, 对所提模型设计了一种有隐私量、数据效用及隐私保护程度等多指标的评估机制。通过大量的实验结果和数据分析, 验证了所提动态度量模型能够有效进行敏感数据的隐私量度量, 且客观反映出数据隐私保护效果和数据效用的变化, 解释了个性化隐私阈值计算式的合理性, 为用户感知数据的隐私动态度量和隐私阈值个性化计算提供了方案和思路。

参考文献:

- [1] XIONG J B, ZHAO M F, BHUIYAN M Z A, et al. An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT[J]. IEEE Transactions on Industrial Informatics, 2019: 1.
- [2] XIONG J, CHEN L, BHUIYAN M Z A, et al. A secure data deletion scheme for IoT devices through key derivation encryption and data analysis[J]. Future Generation Computer Systems, 2020, 111: 741-753.
- [3] YU S, WANG G, LIU X, et al. Security and privacy in the age of the smart internet of things: an overview from a networking perspective[J]. IEEE Communications Magazine, 2018, 56(9): 14-18.
- [4] QI L, HU C, ZHANG X, et al. Privacy-aware data fusion and pre-

- diction with spatial-temporal context for smart city industrial environment[J]. IEEE Transactions on Industrial Informatics, 2020: 1.
- [5] 胡宇翔, 伊鹏, 孙鹏浩, 等. 全维可定义的多模态智慧网络体系研究[J]. 通信学报, 2019, 40(8): 1-12.
HU Y X, YI P, SUN P H, et al. Research on the full-dimensional defined polymorphic smart network[J]. Journal on Communications, 2019, 40(8): 1-12.
- [6] LUO C, JI J, WANG Q, et al. Channel state information prediction for 5G wireless communications: a deep learning approach[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(1): 227-236.
- [7] 乔康, 游伟, 王领伟, 等. 基于区块链的 5G 物联网数据共享方案[J]. 网络与信息安全学报, 2020, 6(4): 45-55.
QIAO K, YOU W, WANG L W, et al. Data sharing scheme for 5G IoT based on blockchain[J]. Chinese Journal of Network and Information Security, 2020, 6(4): 45-55.
- [8] AL-GARADI M A, MOHAMED A, AL-ALI A, et al. A survey of machine and deep learning methods for internet of things (IoT) security[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1646-1685.
- [9] 熊金波, 马蓉, 牛犇, 等. 移动群智感知中基于用户联盟匹配的隐私保护激励机制[J]. 计算机研究与发展, 2018, 55(7): 1359.
XIONG J B, MA R, NIU B, et al. Privacy protection incentive mechanism based on user-Union matching in mobile crowdsensing[J]. Journal of Computer Research and Development, 2018, 55(7): 1359-1370.
- [10] ZHANG Y, WANG P, HUANG H, et al. Privacy-assured FogCS: chaotic compressive sensing for secure industrial big image data processing in fog computing[J]. IEEE Transactions on Industrial Informatics, 2020.
- [11] LIU C, TIAN Y, XIONG J, et al. Towards attack and defense views to k -Anonymous using information theory approach[J]. IEEE Access, 2019, 7: 156025-156032.
- [12] AL-ASMARI H A, SALEH M S. A conceptual framework for measuring personal privacy risks in facebook online social network[C]//International Conference on Computer and Information Sciences (ICCIS). 2019: 1-6.
- [13] ZHAO Y, WAGNER I. On the strength of privacy metrics for vehicular communication[J]. IEEE Transactions on Mobile Computing, 2018, 18(2): 390-403.
- [14] WAGNER I, ECKHOFF D. Technical privacy metrics: a systematic survey[J]. ACM Computing Surveys (CSUR), 2018, 51(3): 1-38.
- [15] 熊金波, 王敏荣, 田有亮, 等. 面向云数据的隐私度量研究进展[J]. 软件学报, 2018, 29(7): 1963-1980.
XIONG J B, WANG M S, TIAN Y L, et al. Research progress on privacy measurement for cloud data[J]. Journal of Software, 2018, 29(7): 1963-1980.
- [16] HUANG Z, YANG L, JIANG W. Uncertainty measurement with belief entropy on the interference effect in the quantum-like Bayesian networks[J]. Applied Mathematics and Computation, 2019, 347: 417-428.
- [17] SONG H, LUO T, LI J. Common criterion of privacy metrics and parameters analysis based on error probability for randomized response[J]. IEEE Access, 2019, 7: 16964-16978.
- [18] YAN F, XU K. A set pair analysis based layer of protection analysis and its application in quantitative risk assessment[J]. Journal of Loss Prevention in the Process Industries, 2018, 55: 313-319.
- [19] SHI W, HU J, YAN J, et al. A privacy measurement method using network structure entropy[C]//Proceedings of the International Conference on Networking and Network Applications (NaNA). 2017: 147-151.
- [20] SHAHAM S, DING M, LIU B, et al. Transition-entropy: a novel metric for privacy preservation in location-based services[C]//IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2019: 1-6.
- [21] GAO F, HE J, PENG S, et al. A quantifying metric for privacy protection based on Information theory[C]//Proceedings of the 2010 Third International Symposium on Intelligent Information Technology and Security Informatics. 2010: 216-220.
- [22] 张宏磊, 史玉良, 张世栋, 等. 一种基于分块混淆的动态数据隐私保护机制[J]. 计算机研究与发展, 2016, 53(11): 2454-2464.
ZHANG H L, SHI Y L, ZHANG S D, et al. A privacy protection mechanism for dynamic data based on partition-confusion[J]. Journal of Computer Research and Development, 2016, 53(11): 2454-2464.
- [23] 彭长根, 丁红发, 朱义杰, 等. 隐私保护的信息熵模型及其度量方法[J]. 软件学报, 2016, 27(8): 1891-1903.
PENG C G, DING H F, ZHU Y J, et al. Information entropy models and privacy metrics methods for privacy protection[J]. Journal of software, 2016, 27(8): 1891-1903.
- [24] ZHANG T, ZHAO K, YANG M, et al. Research on privacy security risk assessment method of mobile commerce based on information entropy and markov[J]. Wireless Communications and Mobile Computing, 2020, 2: 1-11.
- [25] 马蓉, 陈秀华, 刘慧, 等. 移动群智感知中用户隐私度量与隐私保护研究[J]. 信息网络安全, 2018, 18(8): 64-72.
MA R, CHEN X H, LIU H, et al. Research on user privacy measurement and privacy protection in mobile crowdsensing[J]. Netinfo Security, 2018, 18(8): 64-72.
- [26] WANG Y, LIU J. An attribute-based statistic model for privacy impact assessment[C]// Proceedings of the 2016 International Conference on Collaboration Technologies and Systems (CTS). 2016: 619-621.
- [27] LIU R, CAO J, VANSYCKEL S, et al. PriMe: human-centric privacy measurement based on user preferences towards data sharing in mobile participatory sensing systems[C]//Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom). 2016: 1-8.
- [28] YAMAOKA Y, ITOH K. k -presence-secrecy: practical privacy model as extension of k -anonymity[J]. IEICE Transactions on Information and Systems, 2017, 100(4): 730-740.
- [29] GKOUNTOUNA O, TERROVITIS M. Anonymizing collections of

tree-structured data[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(8): 2034-2048.

- [30] LI X Y, ZHANG C, JUNG T, et al. Graph-based privacy-preserving data publication[C]// Proceedings of the 35th Annual IEEE International Conference on Computer Communications. 2016: 1-9.
- [31] ZHANG X, GUI X, TIAN F, et al. Privacy quantification model based on the Bayes conditional risk in location-based services[J]. Tsinghua Science and Technology, 2014, 19(5): 452-462.
- [32] HAN X, TIAN D, DUAN X, et al. Optimized anonymity updating in VANET based on information and privacy joint metrics[C]//Proceedings of the 8th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications. 2018: 63-69.
- [33] TESFAY W B, SERNA-OLVERA J. Towards user-centered privacy risk detection and quantification framework[C]//Proceedings of the International Conference on New Technologies, Mobility and Security (NTMS). 2016: 1-5.
- [34] 晏燕, 郝晓弘, 王万军. 一种隐私保护度量的集对分析方法[J]. 武汉大学学报(工学版), 2015, 48(6): 883-890.
YAO Y, HAO X H, WANG W J. A set pair analysis method for privacy metric[J]. Engineering Journal of Wuhan University, 2015, 48(6): 883-890.
- [35] HUANG W Q, XIA J F, YU M, et al. Personal privacy metric based on public social network data[J]. JPhCS, 2018, 1087(3): 032007.
- [36] HE W, ZENG Y, LI G. A novel structural reliability analysis method via improved maximum entropy method based on nonlinear mapping and sparse grid numerical integration[J]. Mechanical Systems and Signal Processing, 2019, 133: 106247.
- [37] HASSAN M U, REHMANI M H, CHEN J. Differential privacy techniques for cyber physical systems: a survey[J]. IEEE Communications Surveys & Tutorials, 2019, 22(1): 746-789.
- [38] HE J, CAI L, GUAN X. Differential private noise adding mechanism and its application on consensus algorithm[J]. IEEE Transactions on Signal Processing, 2020, 68: 4069-4082.

[作者简介]



赵明烽 (1996-), 男, 江苏张家港人, 福建师范大学硕士生, 主要研究方向为移动数据安全和隐私保护。



Lei Chen (1978-), 男, 陕西西安人, 美国佐治亚州南方大学副教授, 主要研究方向为网络安全、信息安全、云计算与大数据安全等。



钟洋 (1995-), 男, 湖南湘西人, 福建师范大学硕士生, 主要研究方向为安全深度学习。



熊金波 (1981-), 男, 湖南益阳人, 福建师范大学教授、博士生导师, 主要研究方向为网联自动驾驶车辆的安全与隐私、物联网安全、大数据安全、隐私保护。

收录声明

本刊对发表的文章,拥有出版电子版、网络版版权,并拥有和其他网站交换信息的权利。本刊支付的稿酬中已经包含上述费用。

Chinese Journal of Network and Information Security has the copyright to publish electronic edition, online edition of the published articles, and has the right to exchange information with other sites. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的论文是作者独立取得的原创性研究成果,无一稿多投;论文内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,论文作者将承担全部责任。

The authors of *Chinese Journal of Network and Information Security* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Chinese Journal of Network and Information Security*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

网络 Chinese Journal of
Network and Information Security
信息安全学报



2021年2月15日出版 定价：60.00元

ISSN 2096-109X



9 772096 109219

02