文章编号：0253-374X(2022)12-1693-10 **DOI**：10. 11908/j. issn. 0253-374x. 22388

车联网环境下的隐私安全度量方法

徐小雅[1]，于海洋[1]，崔志勇[1]，王颖会[1]，王朋成[2]
（1. 北京航空航天大学 交通科学与工程学院，北京 100191；2. 北京航空航天大学 网络安全与空间学院，北京 100191）

摘要：阐释了车联网的特点、隐私特性和隐私攻击等，从身份隐私、位置隐私、数据隐私三个方面阐述了不同类型隐私在隐私保护方面的特点和需求，并梳理了针对不同隐私类型的隐私保护关键技术和方法演进；在此基础上，研究了不同类型隐私的隐私度量代表性方案、理论模型、度量难题等方面，归纳了涉及车联网环境下隐私度量的8个关键属性，并讨论了相关的32个度量指标。最后，总结了车联网隐私度量的未来研究方向及挑战。

关键词：车联网；隐私度量；隐私安全；隐私保护
中图分类号：U495 **文献标志码**：A

Privacy Metric Review of Internet of Vehicles

XU Xiaoya[1], *YU Haiyang*[1], *CUI Zhiyong*[1], *WANG Yinghui*[1], *WANG Pengcheng*[2]
(1. School of Transportation Science and Engineering, Beihang University, Beijing 100191, China; 2. School of Cyber Science and Technology, Beihang University, Beijing 100191, China)

Abstract: This paper first reviews the privacy backgrounds especially the Internet of Vehicle (IoV) characteristics, privacy features and common privacy leakage attacks. Then, it elaborates the characteristics and needs of different types of privacy in terms of identity privacy, location privacy and data privacy as well as the key privacy protection technologies and methodological evolutions for different privacy types . On this basis, the representative privacy metrics schemes, theoretical models, and metrics challenges for different types of privacy are studied in depth across multiple privacy domains. Furthermore, 8 key attributes and 32 indicators of privacy metrics in IoV are summarized. Finally, a discussion is also made on the future research directions and challenges of privacy metrics.

Key words: Internet of Vehicle(IoV); privacy metric; privacy and security; privacy protection

在车联网中，无线通信技术实现了人、车、路、云等功能实体之间高效、敏捷的数据交换与信息分享。交互数据包括车辆数据、个人数据、应用数据等。然而，大量的多元化接入用户以及网络设备也带来了日益凸显的隐私安全风险。隐私泄露关系到行车安全和生命财产安全，甚至可能会上升到国家安全。因此，隐私保护是推动车联网广泛应用的关键因素之一。

隐私度量作为评估隐私保护强度的关键方法，对推动车联网的隐私保护技术的发展有重大意义[1]。与传统数据库领域中的小规模、结构化、静态化的数据不同，车联网数据是大规模、非结构化、动态性的，其隐私保护需求和保护技术更加多样化。传统隐私度量的方法并不完全适应于车联网，因此需要对车联网隐私度量方法进行全面的研究和阐述。

针对不同隐私类型的特点，构建合理的、可靠的度量指标体系是保证车联网隐私安全的关键因素。本文首先通过分析车联网环境下的隐私安全风险和隐私需求，将车联网的隐私保护方法和理论进行分类、归纳和总结。通过梳理不同隐私类型的度量标准，总结了适用于车联网隐私的评估指标体系，并对隐私度量方法的发展方向进行了讨论。

1 车联网的隐私安全与隐私保护

1.1 车联网隐私保护需求

车联网隐私包括身份隐私、位置隐私、数据隐私

收稿日期：2022-09-08
基金项目：中国国家重点研发计划(2020YFB1600301)；国家自然科学基金青年科学家基金(52002013)；中国博士后科学基金(BX20200036, 2020M680298)
第一作者：徐小雅(1993-)，女，博士生，研究方向为智能交通系统安全理论与技术。 Email：cherie_xu@buaa. edu. cn
通信作者：王朋成(1989-)，男，副教授，工学博士，研究方向为车联网信息安全理论与技术。
Email：pcwang@buaa. edu. cn

三个重要部分[2]。车联网中的隐私风险不仅威胁到交互的信息数据,也会对车辆、用户、云端系统造成直接危害。

面向车联网中的隐私保护应考虑以下几个隐私安全需求[3-4]:①匿名性,旨在能够在数据发布环境下防止用户个人数据被泄露,同时又能保证发布数据的真实性。②有条件的隐私性,旨在当出现可问责性和匿名性冲突时[5],信任机构(如警察,交通管控等)有权透露车辆的真实身份,以实现安全的要求。③保密性,是指不将有用信息泄漏给非授权用户的特性,确保数据信息只能被授权者看到。④最小暴露值,用户在通信过程中应披露最小的信息量并且披露的用户数据应该是最低标准的,且不暴露任何额外信息。⑤不可链接性,指对同一角色或身份的两个或多个行为、两个或多个用于特征识别的特性,无法互相链接或者链接到信息主体。⑥前向保密性,当用户的真实身份或凭证被暴露时,前向保密性可以保护以前的通信信息不被链接到其身份,敏感信息不被泄露。

1.2 车联网隐私风险分类

结合车联网的隐私属性、攻击类型以及影响范围等,将现有的车联网隐私风险按照其攻击目的进行分类[6],如表1所示。

表1 车联网环境中的隐私风险

Tab. 1 Privacy risks in IoV

攻击	攻击者属性	攻击目的
窃听攻击	被动	攻击者从车联网数据中窃听机密信息,并从中提取敏感信息(如用户身份,数据位置)来识别车辆。
信息暴露	主动	攻击者获得车辆或用户的ID,定位、跟踪车辆的位置。
恶意软件	主动	恶意软件收集、分析、泄露用户信息,扰乱甚至破坏整个网络的功能。
流量分析	被动	主要针对匿名性,攻击者收集所有信息并对流量进行分析,捕捉数据包和ID。
伪装	被动	攻击者伪装成合法用户向网络注入有害信息。
重放攻击	主动	攻击者将先前截获或窃取的消息重新在网络中广播。
消息篡改攻击	主动	攻击者修改先前传递的信息。
女巫攻击	主动	攻击者伪造许多虚假身份参与通信。
节点冒充攻击	主动	攻击者冒充车联网的其他用户真实身份。
节点复制攻击	主动	真实节点的身份信息被攻击者获取,并利用截获身份在车联网中进行破坏活动。
自由搭车攻击	主动	攻击者利用其他节点的身份认证结果参与通信。

1.3 车联网隐私保护技术

1.3.1 车联网身份隐私保护技术

车联网中身份隐私的保护目标是实现任何节点都不能获得源节点和目标节点的真实身份相关信息[7-8]。针对车联网身份隐私的攻击主要包括伪装攻击[9]、女巫攻击[10]、节点复制攻击[11]、映射攻击[12]。车联网身份隐私保护技术主要包括:假名认证技术及属性凭证技术如图1所示。

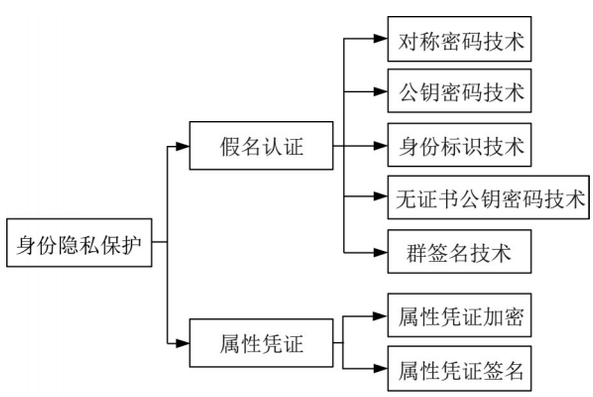

图1 车联网身份隐私保护技术

Fig. 1 Identity privacy preservation technologies in IoV

(1)假名认证

基于假名认证的隐私保护技术旨在用假名来代替现实世界的身份信息[13]。然而,只使用基础的假名认证方案时,恶意攻击者仍有可能把固定的假名与跟踪特定车辆的真实身份关联在一起[14]。

(2)属性凭证

基于属性凭证的身份隐私保护方案作为假名认证的替代方案开始受到关注[15],该方案允许用户以数据最小化的方式向验证者进行认证,并且只披露其凭证中与验证者相关的属性。与基于假名的身份隐私保护方案相比,基于属性凭证需要为有隐私保护需求的所有节点创建共享的属性,并且对资源的要求更高。不同身份隐私保护技术及其特点详见表2。

1.3.2 车联网位置隐私保护技术

目前,车联网位置隐私的保护技术主要可以分为如下4大类:

(1)加密机制

基于密码学的位置隐私保护方案通常使用加密技术来保护用户的位置。除了拥有密钥的车辆节点

表2　身份隐私保护技术及特点

Tab. 2　Identity privacy preservation technologies and features

隐私保护方案		特点
基于假名机制	对称密码技术	计算开销小，安全性不高
	公钥密码技术	安全性高，计算开销高
	身份标识技术	在复杂环境中计算时间开销大，难以实现细粒度的访问控制，存在密钥托管问题
	无证书公钥秘密技术	避免密钥托管问题，计算开销大
	群签名技术	可以保证匿名性，撤销开销大
基于属性凭证	匿名凭证认证	根据共享秘密限制对信息的访问，用户需要提供共享的秘密

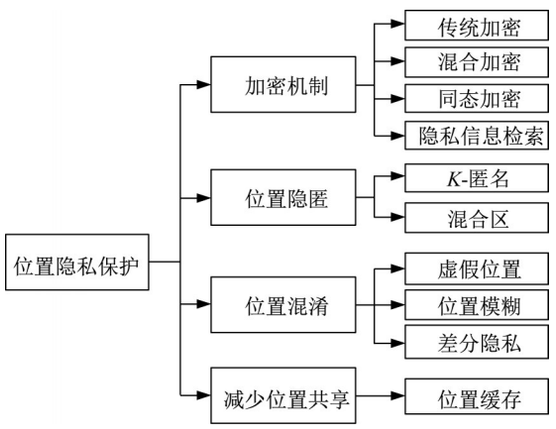

图2　车联网位置隐私保护技术

Fig.2　Location privacy preservation technologies in IoV

可以解密具体查询内容，除此之外，包括提供查询服务在内的任何第三方都无法获取具体内容。除了传统加密方案，还有3种典型的基于加密的位置隐私保护技术，分别是基于隐私信息检索(private information retrieval，PIR)的位置隐私保护技术[16]、基于同态加密(homomorphic encryption)的位置隐私保护技术[17]和混合加密方案。

(2)位置隐匿

基于位置隐匿的位置隐私保护技术旨在打破身份和位置信息之间的联系，主要分为k－匿名和混合区两类。k－匿名属于一种泛化技术，将用户所在的位置模糊成一个包含用户位置的区域，即在泛化形成的区域中，包含查询用户及其他$k-1$个用户。基于混合区的方案是通过建立一个混合区使多个车辆同时在一个区域改变假名，以混淆攻击者对新旧假名的联系，从而达到对位置信息模糊化的目标[18]。混合区的位置隐私保护方案更适用于车辆密度高且合作车辆连接紧密的场景。

(3)位置混淆

位置混淆机制的关键是通过一系列降低位置信息精度的方法模糊准确位置信息，进而达到保护位置隐私的目的。位置混淆机制主要分为添加多个虚假位置迷惑攻击者、添加扰动降低位置精度、差分隐私等几种类型[19]。

(4)减少位置共享

减少位置信息共享的位置隐私保护机制主要是通过位置缓存技术来实现。位置缓存作为一种改善隐私的方式，需要把数据预下载进行缓存，因此需要大量的存储空间。

表3　位置隐私保护技术及特点

Tab. 3 Location privacy preservation technologies and features

隐私保护方案		特点
基于加密机制	传统加密 混合加密 同态加密	信任第三方平台，容易受到串通攻击，通信开销大
	隐私信息检索	需要数据库所有者合作，计算开销大
基于位置隐匿	k-匿名 混合区	需要用户合作，且不同交通环境中不能保持相同的匿名水平
位置混淆	虚假位置 位置模糊 差分隐私	适用对位置的精度要求不高的场景，混淆技术会降低实用性
减少位置共享	位置缓存	需用户合作，数据存储开销大

1.3.3　车联网数据隐私及保护技术

数据隐私保护是通过隐私保护技术对敏感数据进行处理，从而实现数据的隐私性、保密性和不可更改性。车联网环境下的数据隐私保护技术除了传统技术外，还包括3类新兴技术，如图3所示。

(1)基于密码技术的数据隐私保护

除了传统的基于对称和非对称密码学的加密技术以外，属性基加密方案也可以保护数据隐私。这类方案让密文和密钥与属性集合和访问结构产生关联，当且仅当属性集合满足访问结构的时候，才能成功解密。区块链技术通过密码学技术能够实现数据一致存储、不可篡改、防抵赖的分布式账本。

(2)基于数据失真的数据隐私保护

基于数据失真的隐私保护技术是在数据属性不变的前提下,通过对数据添加噪声或干扰后隐藏数据的敏感信息。数据失真越大,隐私保护强度越高,但数据可用性也越低。基于数据失真的典型代表即基于差分隐私的隐私保护机制。

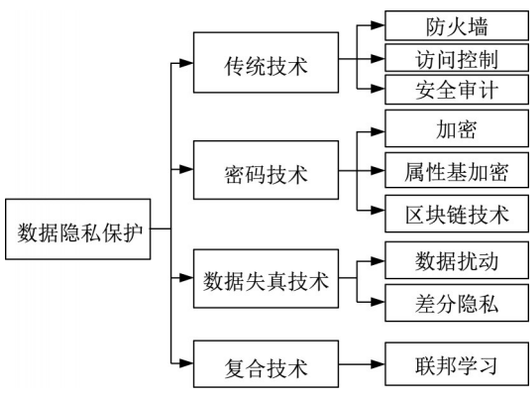

图3 车联网数据隐私保护技术

Fig. 3 Data privacy preservation technologies in IoV

(3)基于联邦学习的数据隐私保护

联邦学习本质上是一种加密的去中心化机器学习技术,结合了安全多方计算、同态加密、差分隐私等技术,使各个参与者在不透露底层数据的基础上构建学习模型。

2 车联网隐私度量方法

随着大量隐私保护方案的涌现,隐私保护技术的性能评估也是一个亟需解决的问题。隐私度量可以反映出隐私保护的强度,其目标是衡量用户隐私在一个环境中具备的隐私程度以及隐私保护技术所提供的保护程度。

2.1 车联网隐私安全度量框架

车联网中的数据有着体量大、多源化、多维度、非结构化的特性。目前,针对车联网环境的隐私度量方法缺少一个全面的、系统的隐私评估指标体系,使得选择合适的隐私指标具有一定难度。由于不同场景、不同的隐私类型,隐私度量方法所评估的内容和指标也不尽相同。因此,对车联网隐私保护评估机制进行分类时,必须考虑3个主要维度:隐私安全风险、隐私保护技术属性、隐私度量属性/指标,如图4所示。

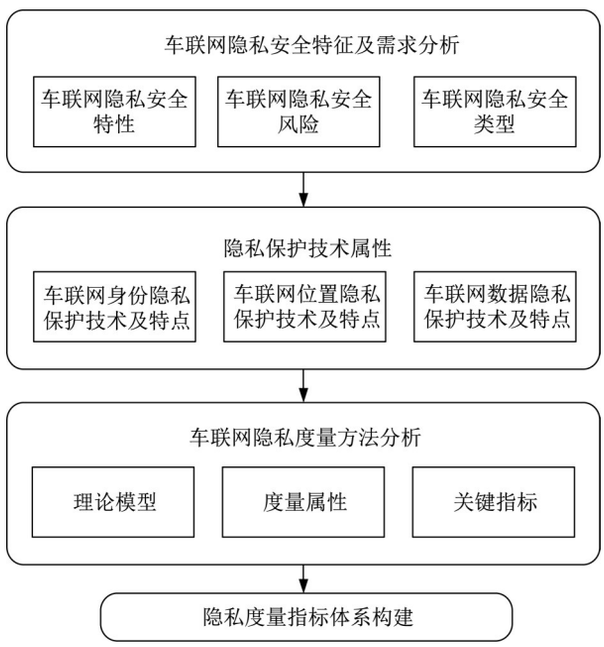

图4 车联网隐私安全度量框架

Fig. 4 Privacy metric framework in IoV

2.2 身份隐私度量

车联网中身份隐私主要是通过匿名性方案来解决[20]。根据不同的理论模型,匿名性方案的隐私度量可以归纳为5个指标:熵、匿名集大小、k-匿名、攻击者指标、隐私泄露程度。这些指标可以量化车联网匿名方案所能提供的匿名性,并且能够表示期望的隐私保护程度。

(1)熵

1984年,Shannon[21]建立了信息理论。信息理论的度量标准提供了一种实用的、相对轻量级的方法来衡量匿名系统在不同环境和不同约束条件下的匿

表4 数据隐私保护技术及特点

Tab. 4 Data privacy preservation technologies and features

隐私保护方案		特点
基于密码技术	传统加密	保留原始数据,高隐私保护,计算和通信开销大
	属性基加密	高隐私保护,串谋抵抗性,计算开销大
	同态加密	计算开销大,计算复杂,实际应用困难
	区块链	提供可靠的环境的数据共享环境,部署困难
基于数据失真	差分隐私	包含多种隐私模型,并适用于各种数据和真实的应用环境
	数据扰动	随机干扰、乱序等数据干扰是不可逆的,在数据可用性和完整性方面存在问题
基于联邦学习		保护原始数据,但由于敏感性,很难收集到足够的样本

名水平。基于信息论的度量标准包括：信息熵、最小熵/最大熵、相互熵、相对熵、条件熵、Rényi 熵、相互信息等。

信息熵被定义为一个离散的随机事件的出现概率，可以衡量与预测随机变量值相关的不确定性和信息获取及损失的大小，其表达式为

$$P_{\mathrm{E}}=h(X)=-\sum_{k\in X}p(k)\log_2 p(k) \tag{1}$$

式中：P_{E}为X的熵值；k是一个随机变量，离散随机变量X集合的每个值$X=\{k_1,\cdots,k_n\}$代表匿名集的一个成员；$p(k_i)$表示成员X是目标的估计概率。

在此基础上，许多学者基于信息熵的匿名性度量进行了延伸研究，O′Connor[22]提出了一个流量确认熵值界限的方案，通过计算熵来度量信息发送者的匿名性随时间的推移而减弱的速度。2018 年，Cui 等人[23]的方案中采用了匿名集的熵来表示车辆真实值与所有其他可能值之间关系的不确定性程度。

(2)匿名集大小

1988 年 Chaum[24]提出了匿名集的概念，在特定信息的发送者和接收者的集合中，匿名集被用来隐藏真正的发件人或收件人。匿名集的大小被看作是用户可以混入的集合的大小。随着可以混入集合大小的增加，那么被发现的几率越低，匿名的程度越高，其匿名程度P_A的表达式为

$$P_A\equiv|A_u|=\log_2 N(N\geqslant 1) \tag{2}$$

式中：u为随机用户；$|A_u|$表示u混入的用户集合，可以被看作为攻击者无法将u中区分出来的匿名集；N代表匿名集$|A_u|$中的用户数。

Chen 等人[25]提出通过基于集合理论的条件性匿名概念度量系统的匿名程度，并提出当对手从系统中获得更多可观察的输出时，系统会失去更多的匿名性结论。

(3)k—匿名

k—匿名的概念首次是由 Sweeney[26]在 2002 年提出并被用于信息发布中保护私人数据。k—匿名是衡量数据相似性的指标之一，用于表示数据集中的准标识符属性的匿名程度，其表达式为

$$P_y\equiv k \tag{3}$$

式中：k代表匿名数据集中的不可识别元组，并且k个匿名元组被识别的概率相等。匿名数据中的k值越大，攻击者越难推测出隐私信息，隐私保护强度越高。2019 年，有部分学者提出k—匿名是数据匿名的最佳概括算法[27]。

(4)攻击者指标

攻击者指标是指通过概率分析攻击者的成功率，并将隐私保护强度量化为对手任何一次尝试或多次尝试攻击的成功率，对手成功率的表达式为

$$P_S\equiv P(\mathrm{Sim}(s,s')\geqslant\tau_{\mathrm{s}})\geqslant\tau_e \tag{4}$$

式中：P_S为对手成功率；s为目标记录；s'为攻击者可以找到的相似记录；τ_{s}为相似度阈值；τ_{e}为误差阈值。因此，当对手能够找到一个与其相似度阈值为τ_{s}，误差阈值为τ_{e}时，攻击者则成功获取到隐私信息。Agrawal 和 Kesdogan[28]认为量化攻击者所需的观察次数是衡量匿名性的有效方法。

(5)隐私泄露程度指标

与对手成功率相似，对于匿名性隐私的破坏程度或者隐私的泄露程度也可以作为度量指标，其表达式为

$$P_P\equiv\tau\exists s\in St \tag{5}$$
$$P(s\in T_{x^*}|S\subseteq T_y)\geqslant\tau$$

式中：τ为阈值，当给定其先验概率后，一个属性的后验概率高于阈值τ，则会发生隐私泄露。s为一个目标，S为目标集合。当目标集S包含在随机传输信息T_{y}，并且已知s包含在内的概率。目标s包含在传输信息T_{x^*}中的概率高于阈值τ，则会发生隐私泄露问题。在此基础上，Huang[29]利用基于广义信息论的证据理论，在给定时间段内检测到达的数据包的数量，进而衡量无线移动车联网网络的匿名程度。

2.3　位置隐私度量

本节讨论在各种位置隐私评估中所使用的隐私属性和评估指标，着重介绍七个主要位置隐私度量指标：

(1)匿名集大小

Chaum[24]将匿名集定义为具有发送特定信息概率的用户集，并提出了针对不确定性，匿名集的大小是衡量匿名程度的一个很好的指标。在车联网位置隐私中匿名集大小描述了在车辆中难以区分出目标车辆的情况。这个指标的优点在于其比较简单、容易计算。

(2)熵

熵通常被用作车联网中位置隐私的精确测量方法[23]。当熵越大，车辆位置在匿名集中的混乱程度就越大，车辆的位置隐私就越安全。由于熵的取值范围取决于匿名集的元素数量，并且绝对值不能被用来比较熵值。因此，最大熵被归一化到$[0,1]$区间，并用归一化熵来表示对手的不确定性程度，其表达式为

$$P_N=\frac{H(X)}{H_0(X)} \tag{6}$$

式中:$H_0(X)$为最大熵,归一化后的熵是有界的数值范围,更适合于场景间的比较。Diaz等人[30]讨论了使用所有可能接收者分布的熵来量化隐私,并结合匿名集大小和归一化熵来提供更好的隐私保证。王彩梅等人[31]设计了一种基于信息熵的用户轨迹隐私水平计算方法,基于信息熵的角度计算用户的轨迹隐私水平。

(3)信息增益/损失

信息增益/损失指标也是一个基于信息理论的指标,衡量对手通过观察获得多少隐私信息或用户失去多少隐私信息。假设对手能获得的信息越少,那么隐私程度就越高。信息量泄露的表达式为

$$P_A\equiv|v|,\ \forall v\in V \tag{7}$$

式中:以车联网场景为例,在这一指标所度量的泄露信息量中,v表示对手能正确跟踪多少车辆,其概率很大程度上取决于一个场景中的车辆总数V。

相互信息也可以作为信息增益/损失指标量化两个随机变量之间的信息共享程度,通过计算熵和条件熵之间的差异得到,表达式为

$$P_L=I(X^*;\ Y)=H(X^*)-H(X^*|Y) \tag{8}$$

式中:X^*是数据的真实分布;Y为对手观测到被混淆的观测值。

通过计算熵的公式可以进一步得到

$$P_L=\sum_{x^*\in X^*}\sum_{y^*\in Y}p(x^*,y)\log_2\frac{p(x^*,y)}{p(x^*)p(y)} \tag{9}$$

式中:x^*是一个真实数据分布中的一个随机变量,离散随机变量X^*的每个值$X=\{x_1,\ \cdots,\ x^*\}$代表一个数据真实分布。y是观测数据分布中的随机值,离散随机变量Y的是多个y观测值的集合。

条件性隐私损失是归一化的相互信息,可以作为另一个隐私度量指标,表达式为

$$P_C\equiv 1-2^{-I(X^*;\ Y)} \tag{10}$$

式中:Y为泄露部分,X^*为由于Y泄露所损失的隐私。

(4)地理不可区分性

地理不可区分性指标是将差分隐私扩展到位置隐私场景,目的是确保用户在任何距离$d>0$时可以保证ϵd差分隐私。基于地理不可区分性的隐私度量指标P_{G-I}表示为

$$P_{G-I}\equiv d_\varsigma(\psi_{y1},\ \psi_{y2})\leqslant\epsilon d(l_1,\ l_2) \tag{11}$$

式中:d为距离;Ψ为隐私机制用于生成随机位置观测值;$d_\varsigma(\psi_{y1},\ \psi_{y2})$为随机位置观测值分布之间的距离。$l_1$和$l_2$为任意两个位置,$d(l_1,\ l_2)$为任意两个位置之间的距离,用户的隐私保护水平取决于距离d。

(5)攻击成功率

攻击成功率作为一个位置隐私指标适用于衡量攻击者成功追踪目标用户的概率,如Sholri等人[32]通过攻击者的成功概率和准确度两个指标来度量位置隐私。其中,准确度指标是指混淆区域的准确性,表达式为

$$P_A=\frac{T_a^2}{r_{\min}^2} \tag{12}$$

式中:T_a是指传感技术的最佳精度;$r_{\min}$是指为了保护位置隐私,将区域放大到满足最低用户位置服务需求的位置半径。

(6)平均混淆时间

平均混淆时间作为基于时间属性的隐私保护评估指标是用熵来衡量攻击者正确跟踪一个轨迹所需的时间,表达式为

$$P_A=H(X)_t<\tau \tag{13}$$

式中:τ是特定阈值时间;X为随机变量表示对手对匿名集中的每个成员的估计概率;$H(X)$为熵值。因此,平均混淆时间衡量的是对手的不确定性保持在混淆阈值τ以下的时间。攻击者保持不确定的时间越长,则隐私性越高。

(7)预期误差

预期误差指标可以用来衡量攻击者重建目标轨迹的成功率[33],这一指标反映了攻击者通过观察发布的位置和推断实际位置的准确度。

发布位置x'并使用攻击者可用的先验知识推断出实际位置x的准确程度。

$$P(x|x')=\frac{\pi(x)p(x|x')}{\sum_{x\in\chi}Pr(x|x')} \tag{14}$$

式中:x'为发布位置,x为实际位置。χ为用户的可能位置,先验知识通常由一组可能的用户位置χ上的先验概率分布π来获取。

通过估计位置x^*和实际位置x之间的预期偏差,计算出一个估计位置x^*为

$$x^*=a_{x^*}\min\sum_{x\in\chi}P(x|x')\|x^*-x\| \tag{15}$$

在位置隐私度量中$\|x^*-x\|$表示地点之间的地理距离,因此,预期估计误差E_p为

$$E_p=E\|x^*-x\|\sum_{x^*}P(x^*|x')\|x^*-x\| \tag{16}$$

Corser等人[34]提出可以使用综合指标,例如平

均匿名集的大小、平均距离偏差和匿名持续时间来度量隐私，来平衡不同用户之间时间和位置的偏差影响。

2.4　数据隐私度量

车联网数据隐私保护技术通常对原始数据进行干扰，以达到模糊敏感属性的目的。从隐私保护强度和数据可用性两个方面来看，可以将数据隐私度量指标分为4类：

(1)熵

在数据隐私度量可以通过熵、互信息和条件熵来度量隐私水平，其中熵和互信息的计算方法与身份隐私和位置隐私度量一致。基于条件熵的隐私度量表达式为

$$P_{\mathrm{L}}=H(X|Y)=-\sum_{i=1}^{n}\sum_{j=i}^{m}p(x_i y_j)\log_2 p(\frac{x_i}{y_j}) \quad (17)$$

式中：随机变量$X=\{x_1, \cdots, x_n\}$代表原始数据，$Y=\{y_1, \cdots, y_n\}$代表经过隐私保护处理的发布数据。当已知Y是，条件熵表示对手通过Y推测出X的平均不确定性。Begum[35]等学者提出了用联合熵来维护云数据的隐私，并利用熵值和数据库差异率被作为评估矩阵来评估隐私水平。

(2)集对分析

集对分析理论是一种解决不确定性和确定性理论的研究方法，该理论可以处理随机的、不明确的和不确定的问题[36]。这个理论是把复杂的事物作为一个集合对来分析，并探究两个集合之间的关联隐私的不确定性，其表达式如下：

$$\mu(W)=\frac{P}{S}+\frac{U}{S}i+\frac{N}{S}j \quad (18)$$

假设集合系统H由集合A和集合B组成，表示为$H=(A, B)$。两个集合合并后的特性总数为S，P属性是集合A和B的共同部分，N属性是集合A和B中的对立属性，剩余的属性$U=S-P-N$，即既不统一也不对立的不确定属性。由此可知，$\frac{P}{S}$是相似程度，$\frac{N}{S}$是对立程度，$\frac{U}{S}$是不确定性。其中i为不确定性标记，j为对立度标记，当$j=1$时，$i\in[-1, 1]$。

(3)相互信息

相互信息描述隐私泄露风险是通过攻击者在获取到隐私相关信息之前和之后，对原始数据不确定性减少的量来反映的。对原始数据不确定性减少的量越多，与之而来的隐私信息泄露的风险越大。Sankar[37]等人选取相互信息作为隐私度量指标提出了一个既能量化隐私，又能度量效用和隐私的框架，还有一系列研究[38]将相互信息作为隐私泄露的度量标准。

(4)差分隐私

基于差分隐私的隐私度量方法，隐私保护的强度主要取决于差分隐私中的ε值。分析ε的值可以反映隐私保护的强度[39]。基于差分隐私的度量指表达式为

$$P_{\mathrm{D}}\equiv\forall S\subseteq\mathrm{Range}(K): p\big(K(D_1)\in S\big) \leqslant \exp(\epsilon)\, p(K(D_2)\in S) \quad (19)$$

式中：D_1和D_2为两个最多只有一行不同的数据集，即两个数据集之间的汉明距离最大为1。S为数据查询响应集；K为随机化函数。

近似差分隐私与差分隐私的机制类似，通过允许一个额外的常数δ削弱差分隐私的隐私保证，但提升了数据发布/查询响应的效率[40]。近似差分隐私度量指标表达式为

$$P_{\delta}\equiv\forall S\subseteq\mathrm{Range}(K): p\big(K(D_1)\in S\big)\leqslant \exp(\epsilon)\, p\big(K(D_2)\in S\big)+\delta \quad (20)$$

其特殊性源于参数δ，参数的选择小于任意数据库D大小的任何多项式的逆数。当$\delta\approx\frac{1}{\|D\|}$时，将允许公布少量隐私数据，同时仍然满足差异化的隐私要求。

基于地理不可区分性的$d-\chi$隐私使用可区分度量d_χ来描述两个数据集之间的距离，而不是标准差分隐私中使用的汉明距离。$d-\chi$隐私使用可区分度量d_χ来描述两个数据集之间的距离，任意距离的数据集之间的可区分度由可区分度量d_χ决定，因此其表达式为

$$P_{d-\chi}\equiv d_P\big(K(D_1), K(D_2)\big)\leqslant d_\chi(D_1, D_2) \quad (21)$$

式中：D_1和D_2为任意距离的数据集K产生随机数据的隐私机制，$d_P\big(K(D_1), K(D_2)\big)$为随机产生两个数据集之间的距离。

3　隐私度量属性及评估指标

对车联网三个隐私类型的隐私特征和度量方法进行总结归纳后，总结出8个车联网隐私度量属性以及32个评估指标，如表5所示，以助于后续研究能够为特定的场景确定正确的隐私方案评估指标。

(1)不确定性

不确定性是指攻击者识别隐私的不确定程度[41]。基于不确定性的隐私度量是指由于攻击者无

法将其猜测建立在确定的已知信息上,因此在攻击者推测的信息中,信息的隐私程度越高,其不确定性越高。信息熵就是一个典型衡量所预测的随机变量值的不确定性。此外,类似指标还有基于匿名集大小,无关联性的程度等。

(2)信息获取/损失

信息获取或损失的指标量化了攻击者获得的隐私信息量或用户因信息泄露而损失的隐私量。假设攻击者能获得的信息越少,隐私度就越高。例如,信息损失的平均大小这一指标度量系统泄露的信息量或被泄露的用户数量。类似地还有相对熵、相互信息、条件互信息等方法。

(3)数据相似性

数据相似性指标是在大多数攻击者无法获取真实数据集的情况下,用于度量已发布或公布的数据的指标。这类方法利用数据的相似性度量隐私的大小,并完全从暴露的数据特征中得出隐私级别。这类指标的典型代表有k—匿名、l多样性(l-diversity)、m不变性(m-invariance)、t紧密(t-closeness)等。

(4)不可区分性

不可区分性指标是指攻击者区分目标的能力[41]。这类指标包括差分隐私,近似差分隐私、分布式差分隐私、分布式隐私、地理不可区分性、联合差分隐私、计算差分隐私等。

(5)攻击成功率

基于攻击者成功概率的度量可以被看作是通用的度量标准,取决于对手模型和成功的确切定义。

(6)误差

基于误差的度量量化了攻击者在创建其推测时的错误。由于度量过程需要基于真实结果的信息,所以不能由攻击者计算出来。

表5 车联网隐私评估属性及评估指标

Tab. 5 IoV privacy assessment attributes and assessment metrics

属性	理论模型	相关应用	评估指标
不确定性	信息论	位置隐私 节点隐私 个人隐私等	匿名集大小 条件熵 交叉熵 归一熵
	统计学	个人隐私 位置隐私	基于攻击类型 保护等级
信息获取/损失大小	信息论	数据隐私 数据隐私 数据隐私 位置隐私	相互信息 条件互信息 相对熵/KL散度
信息相似度	计算机技术	个人隐私 数据隐私 数据库隐私	k匿名算法 l多样性 m不变性 t紧密
不可区分性	差分隐私	位置隐私 个人隐私 数据隐私 参数保护	分布差分隐私 近似差分隐私 联合差分隐私 计算差分隐私
		位置隐私保护	地理不可区分性
	密码学	位置隐私 个人隐私 数据隐私	无关联性的程度
攻击成功率	概率论	数据库隐私 个人隐私 数据隐私	攻击者的成功率 匿名度 隐私破坏程度
误差	概率论 概率论 统计学	位置隐私 位置隐私 数据隐私	攻击者的推测估计误差 距离误差 平均平方误差
时间	统计学 统计学 统计学	通信隐私 位置隐私 位置隐私	攻击成功时间 最大追踪时间 攻击者混淆时间
精确度	统计学 统计学 概率论	数据隐私 位置隐私 位置隐私	置信区间宽度 不可观察性 混淆区域准确性

(7)时间

基于时间的度量侧重于将时间作为攻击者为破坏用户隐私所需花费的资源,这些指标通常应用于通信和定位领域。一般测量分为两种情况:直到对手成功的时间和直到对手放弃的时间。

(8)精确度

精确度指标量化了攻击者推测的精确程度,而不考虑估计的正确性。攻击者更精确的推测对应着较低的隐私性。

针对以上的隐私属性,对所应用的理论模型、隐私类别、评估指标和相关应用领域进行了总结归纳,如表5所示。

4　总结与展望

本文基于车联网环境的组成部分、数据来源以及环境特征深入研究了车联网的隐私特性和隐私需求,并将所涉及的隐私类型进行了分类与详细阐述。基于多种理论模型对其不同的隐私度量指标进行归纳、总结,构建了一个结构化的车联网隐私度量指标体系。然而,在隐私度量方面车联网还面临着很多挑战。区别于传统网络的特殊属性决定了其隐私保护方案可能囊括了多种技术并涵盖了多个隐私度量属性。因此,由于各属性之间的重要性的不同,在隐私度量过程中其隐私属性的权重比例还需要进一步研究。另外,当前面向车联网的隐私度量的研究还处于研究初期,隐私度量的研究与实际应用还需进一步研究与探索。

作者贡献声明:

徐小雅:设计论文框架,起草论文;

于海洋:指导性支持,项目管理;

崔志勇:设计研究方案,审核;

王颖会:修订与编辑论文;

王朋成:论文选题,学术指导,论文审阅。

参考文献:

[1] AMAN M N, JAVAID U, SIKDAR B. A privacy-preserving and scalable authentication protocol for the internet of vehicles [J]. IEEE Internet of Things Journal,2020, 8(2): 1123.

[2] USHA M, RAMAKRISHNAN B. A robust architecture of the OLSR protocol for channel utilization and optimized transmission using minimal multi point relay selection in VANET [J]. Wireless Personal Communications, 2019, 109 (1): 271.

[3] LU Z, QU G, LIU Z. A survey on recent advances in vehicular network security, trust, and privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(2): 760.

[4] VIJAYAKUMAR P, CHANG V, DEBORAH L J, *et al*. Computationally efficient privacy preserving anonymous mutual and batch authentication schemes for vehicular ad hoc networks [J]. Future generation computer systems, 2018, 78: 943.

[5] MANSOUR M B, SALAMA C, MOHAMED H K, *et al*. VANET security and privacy—An overview [J]. International Journal of Network Security & Its Applications, 2018, 10 (2): 13.

[6] HAFIZ M. A pattern language for developing privacy enhancing technologies[J]. Software: Practice and Experience, 2013, 43(7): 769.

[7] SUTHANA S, REDDY B M. New identity batch verification privacy scheme in VANET [J]. International Journal of Advanced Networking and Applications, 2020, 11(4): 4354.

[8] LI H, PEI L, LIAO D, *et al*. Blockchain Meets VANET: An architecture for identity and location privacy protection in VANET [J]. Peer-to-Peer Networking and Applications, 2019, 12(5): 1178.

[9] ABBAS S, FAISAL M, UR RAHMAN H , *et al*. Masquerading attacks detection in mobile ad hoc networks [J]. IEEE Access, 2018(6): 55013.

[10] GROVER J, GAUR M S, LAXMI V, *et al*. A sybil attack detection approach using neighboring vehicles in VANET[C]// Proceedings of the 4th international conference on Security of information and networks. Sydney, Australia: ACM, 2011: 151-158.

[11] SUJIHELEN L, BODDU R, MURUGAVENI S, *et al*. Node replication attack detection in distributed wireless sensor networks [J]. Wireless Communications and Mobile Computing, 2022, 2022: 11.

[12] KANG J, YU R, HUANG X, *et al*. Location privacy attacks and defenses in cloud-enabled internet of vehicles [J]. IEEE Wireless Communications, 2016, 23(5): 52.

[13] BENAROUS L, KADRI B, BITAM S, *et al*. Privacy-preserving authentication scheme for on-road on-demand refilling of pseudonym in VANET [J]. International Journal of Communication Systems, 2020, 33(10): e4087.

[14] HAIDER S, GAO D, ALI R, *et al*. A privacy conserves pseudonym acquisition scheme in vehicular communication systems [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 15536.

[15] DE FUENTES J M, GONZÁLEZ-MANZANO L, SERNA-OLVERA J, *et al*. Assessment of attribute-based credentials for privacy-preserving road traffic services in smart cities [J]. Personal and Ubiquitous Computing, 2017, 21(5): 869.

[16] YADAV V K, VERMA S, VENKATESAN S. Efficient and secure location-based services scheme in VANET [J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 13567.

[17] SUN X, YU F R, ZHANG P, *et al*. A survey on secure computation based on homomorphic encryption in vehicular ad

hoc networks[J]. Sensors, 2020, 20(15): 4253.

[18] YING B, MAKRAKIS D, HOU Z. Motivation for protecting selfish vehicles' location privacy in vehicular networks [J]. IEEE Transactions on Vehicular Technology, 2015, 64(12): 5631.

[19] HE Y, CHEN J. User location privacy protection mechanism for location-based services [J]. Digital Communications and Networks, 2021, 7(2): 264.

[20] KELLY D J, RAINES R A, GRIMAILA M R, *et al*. A survey of state-of-the-art in anonymity metrics [C]// Proceedings of the 1st ACM workshop on Network data anonymization. New York, NY, USA: Association for Computing Machinery, 2008: 31 - 40.

[21] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379.

[22] O'CONNOR L. Entropy bounds for traffic confirmation [EB/OL]. [2022-09-01]. https://eprint.iacr.org/2008/365.

[23] CUI J, WEN J, HAN S, *et al*. Efficient privacy-preserving scheme for real-time location data in vehicular ad-hoc network [J]. IEEE Internet of Things Journal, 2018, 5(5): 3491.

[24] CHAUM D. The dining cryptographers problem: Unconditional sender and recipient untraceability[J]. Journal of Cryptology, 1988, 1(1): 65.

[25] CHEN W, CAO Y, WANG H. Conditional anonymity with non-probabilistic adversary [J]. Information Sciences, 2015, 324: 32.

[26] SWEENEY L. *k*-anonymity: A model for protecting privacy [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557.

[27] RAJ A, D'SOUZA R G. Big data anonymization in cloud using *k*-anonymity algorithm using map reduce framework[J]. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 2019, 5(1): 50.

[28] AGRAWAL D, KESDOGAN D. Measuring anonymity: The disclosure attack [J]. IEEE Security & Privacy, 2003, 1(6): 27.

[29] HUANG D. On measuring anonymity for wireless mobile ad-hoc networks [C]//Proceedings. 2006 31st IEEE Conference on Local Computer Networks. [S.l.]:IEEE, 2006: 779 - 786.

[30] DIAZ C, TRONCOSO C, DANEZIS G. Does additional information always reduce anonymity?[C]//Proceedings of the 2007 ACM workshop on Privacy in Electronic society.[S.l.]: ACM, 2007: 72 - 75.

[31] 王彩梅, 郭亚军, 郭艳华. 位置服务中用户轨迹的隐私度量 [J]. 软件学报, 2012, 23(2): 9.
WANG CAIMEI, GUO YAJUN, GUO YANHUA. Privacy metric for user's trajectory in location-based services [J]. Journal of Software, 2012, 23(2): 9.

[32] SHOKRI R, FREUDIGER J, HUBAUX J P . A unified framework for location privacy[EB/OL].[2022-08-26] http://infoscience.epfl.ch/record/148708

[33] SHOKRI R, THEODORAKOPOULOS G, LE BOUDEC J Y, *et al*. Quantifying location privacy [C]//2011 IEEE symposium on security and privacy. [S.l.]:IEEE, 2011: 247 - 262.

[34] CORSER G P, FU H, BANIHANI A. Evaluating location privacy in vehicular communications and applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(9): 2658.

[35] SABIN BEGUM R, SUGUMAR R. Novel entropy-based approach for cost-effective privacy preservation of intermediate datasets in cloud[J]. Cluster Computing, 2019, 22(4): 9581.

[36] HUANG W Q, XIA J F, YU M,*et al*. Personal privacy metric based on public social network data [J]. Journal of Physics: Conference Series, 2018, 1087(3): 032007.

[37] SANKAR L, RAJAGOPALAN S R, MOHAJER S, *et al*. Smart meter privacy: A theoretical framework [J]. IEEE Transactions on Smart Grid, 2012, 4(2): 837.

[38] WU Q, TANG J, DANG S, *et al*. Data privacy and utility trade-off based on mutual information neural estimator [J]. Expert Systems with Applications, 2022, 207: 118012.

[39] WANG W, YING L, ZHANG J. On the relation between identifiability, differential privacy, and mutual-information privacy[J]. IEEE Transactions on Information Theory, 2016, 62(9): 5018.

[40] BLUM A, LIGETT K, ROTH A. A learning theory approach to noninteractive database privacy [J]. Journal of the ACM (JACM), 2013, 60(2): 1.

[41] WAGNER I, ECKHOFF D. Technical privacy metrics: A systematic survey [J]. ACM Computing Surveys, 2019, 51(3): 1.