DOI:10.13878/j.cnki.jnuist.20240425002

施继婷[1]　曾维昊[1]　张骞允[1]　刘凯歌[1]　秦志金[2]　李树锋[3]

语义通信安全研究综述

摘要

随着人工智能技术与无线通信领域的深度融合,作为一种新兴的通信模式,语义通信聚焦于语义层面的信息传输与交互,凭借其独特优势,显著提升了通信的精确性和可靠性.在低时延、高流通密度的通信应用场景中,语义通信技术突破了传统基于经典信息论的语法通信,为无线通信领域提供了新范式,拓宽了现代通信技术的应用范畴.目前,语义通信技术的发展尚处于起步阶段,其在应用过程中面临的安全问题尚未得到系统梳理和全面分析.为进一步推动语义通信技术的发展与应用,首先对语义通信系统中存在的各类安全威胁进行了分类阐述;然后,详细介绍了语义通信系统中模型安全和数据安全的研究现状;最后,总结了语义通信安全研究所面临的挑战,并对未来发展趋势进行了展望.

关键词

语义通信;数据安全;隐私保护;无线通信安全

中图分类号 TN918

文献标志码 A

收稿日期 2024-04-25

资助项目 媒体融合与传播国家重点实验室(中国传媒大学)开放课题(SKLMCC2023KF001);中国科协青年人才托举工程项目(2021QNRC001)

作者简介

施继婷,女,博士生,研究方向为无线通信安全、语义通信安全、隐私保护等.shijiting@buaa.edu.cn

张骞允(通信作者),女,博士,副教授,研究方向为无线网络安全技术研究,包括物理层安全技术、电磁频谱高效感知与利用、无线通信理论与硬件系统设计等.zhangqianyun@buaa.edu.cn

1 北京航空航天大学 网络空间安全学院,北京,100191

2 清华大学 信息科学技术学院,北京,100084

3 中国传媒大学 信息与通信工程学院,北京,100024

0 引言

作为现代信息论的基础理论,香农定律指导了无线通信系统的发展,构建了信息传输的基础理论体系[1],并将通信划分为3个层次:1)语法层:通信符号如何准确地传输;2)语义层:传输的符号如何精确地表达期望的含义;3)语用层:接收的信息如何以期望的方式影响行为[2].随着近几十年来无线通信系统的日益发展,尤其是6G技术的飞速发展,无线通信的系统容量日益接近香农极限.与此同时,万物互联时代,在人工智能(Artificial Intelligence,AI)技术的辅助下,通信的目的逐渐转向对具体语义信息的精确传达.

语义通信(Semantic Communication,SC)技术作为一种全新的通信范式,结合自然语言处理、机器学习和人工智能等技术,聚焦于传输语义信息而非通信符号本身,对传输信息进行语义层面的分析和处理,使得无线通信系统能够适应更加复杂和多变的应用场景,引发了学术界和工业界的广泛关注,并得到了飞速发展[3-4].目前,语义通信相关研究主要集中在利用深度学习(Deep Learning,DL)技术进行语义特征的提取与传输[5].随着智能化技术的迅猛发展,语义通信在通信效率、准确率及鲁棒性方面凸显出显著优势,其不仅大幅提升了无线通信系统的传输效率与能力,而且在复杂多变的场景中展现出卓越的抗干扰性能,在智能应用领域和下一代通信场景或恶劣环境通信[6]中具备巨大的应用潜力.然而,无线通信信道的开放性和易接入特性,为语义通信系统带来了与传统无线通信系统相似的安全挑战.除此之外,随着人工智能技术的快速发展,基于深度神经网络的语义通信架构也面临着深度学习技术自身存在的威胁带来的安全风险.

如图1所示,一个端到端深度学习语义通信系统包含语义编解码器、信道编解码器和收发端知识库等模块,其各个模块在无线信道传输和网络模型架构更新过程中面临着多种安全攻击的威胁.典型地,由于无线通信信道暴露在外部环境中,信号传输过程中容易受到窃听攻击,从而导致信息泄露、篡改等[7].在训练和使用过程中,深度网络架构容易受到对抗性扰动、模型窃取、数据投毒等攻击[8],导致网络性能下降,甚至可能泄露敏感信息.因此,为了保障语义通信技术的进一步发展,语义通信系统中的安全问题不容忽视,语义通信安全技术也逐渐成为研究热点.

语义通信安全研究的整体架构如图2所示.本文围绕着语义通信

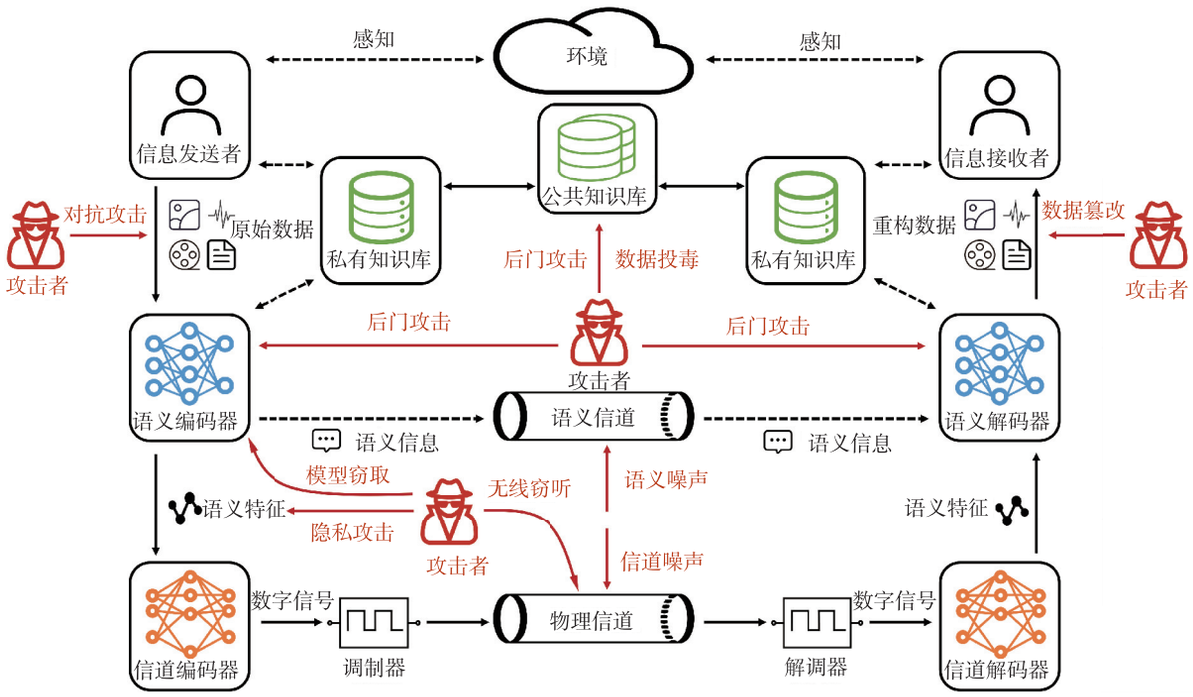

图 1　语义通信系统架构及安全攻击示意图

Fig. 1　Semantic communication system architecture and the encountered security attacks

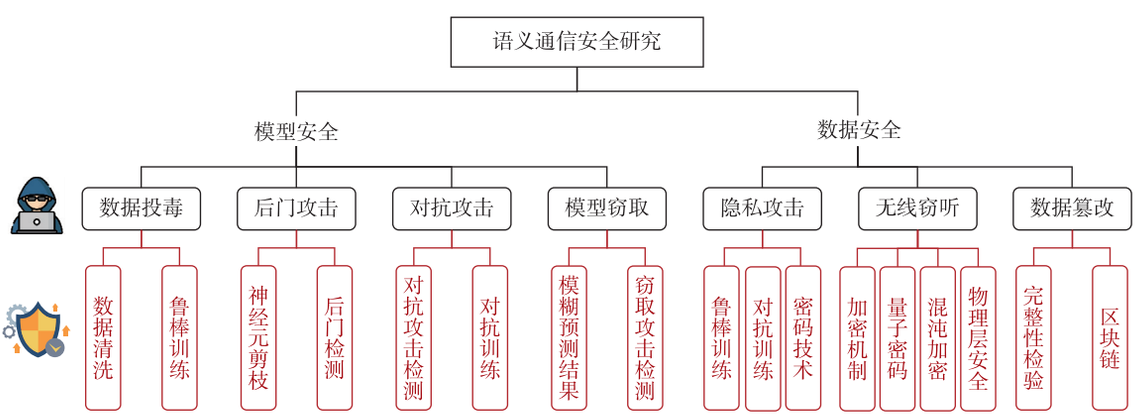

图 2　语义通信安全研究整体架构

Fig. 2　Overall architecture for semantic communication security research

系统中的模型安全和数据安全的主要攻击手段和防御策略,对现有面向语义通信的安全技术研究进行全面综述,并基于现有研究进展,总结语义通信安全研究发展所面临的挑战,展望语义通信安全研究的未来发展趋势.

1　语义通信模型安全研究现状

人工智能技术的迅猛发展,极大地推动了 AI 驱动的语义通信系统的构建,为其发展奠定了坚实的基础.在语义通信系统中,模型架构的设计和优化是实现系统功能的核心,也是语义通信的关键模块.高性能的网络模型是确保语义通信系统实现高效性、准确性和鲁棒性的关键.因此,网络模型也成为攻击者的主要攻击目标,模型的安全性对整个系统的安全性能具有至关重要的影响.

1.1　模型安全攻击

在深度学习架构中,所研究的核心问题在于确保模型的架构、权重参数或数据集的安全性,防止其被非授权的用户获取或利用.当前,模型安全攻击手段日益多样化,针对网络模型的典型攻击手段包括数据投毒、后门攻击、对抗攻击和模型窃取等,这些

攻击手段不仅在模型训练、测试、部署等多个阶段影响网络输出准确性,而且会严重威胁基于深度学习的语义通信系统的稳定性与可靠性.

1.1.1 数据投毒

语义通信系统中,语义信息的提取与准确重构至关重要.通信过程中,合法通信双方所提供的数据的质量和所依赖的网络架构的有效性,将直接影响到数据处理的准确性和可靠性.然而,在数据收集阶段,合法用户所使用的数据集或知识库可能存在遭受投毒攻击(poisoning attack)的风险.

数据投毒攻击通过在数据集、知识库中注入恶意数据,破坏原始训练数据的分布,这不仅会对网络模型的预测精度造成影响,还会影响网络模型的可用性和数据的完整性,从而威胁整个语义通信系统的安全性[9-10].常见的攻击方式包括:

1)基于标签翻转(Label Flipping,LP):攻击者直接篡改目标类别训练数据的标签信息 y,从而破坏数据样本与标签之间的正常对应关系,以达到混淆模型判别结果的目的,从而削弱或破坏网络模型的可用性.在分类问题中的标签翻转攻击可以表示如下:

$$\mathrm{LP}(x;y)=(x;y_{\mathrm{p}}),\ y_{\mathrm{p}}\in Y-\{y\}. \tag{1}$$

其中:Y 为原始标签集合,攻击者可随机或有选择地选取中毒样本标签 y_{p} 以达到攻击目标.

2)基于双层优化:攻击者将投毒问题转化为最优化问题后,通过优化目标函数生成对目标模型的训练产生负面影响的中毒样本,可表示为

$$x_{\mathrm{p}}^{*}=\underset{x_{\mathrm{p}}}{\operatorname{argmax}}\mathcal{L}_1(f_{\theta *}(x_{\mathrm{val}}),y_{\mathrm{val}}),$$
$$\text{s.t.}\quad \theta^{*}=\underset{\theta}{\operatorname{argmin}}\mathcal{L}_2(f_{\theta}(x_{\mathrm{p}}\cup x_{\mathrm{tra}}),y_{\mathrm{p}}\cup y_{\mathrm{tra}}). \tag{2}$$

其中:$\{x_{\mathrm{p}},y_{\mathrm{p}}\}$ 为中毒数据;$\{x_{\mathrm{tra}},y_{\mathrm{tra}}\}$,$\{x_{\mathrm{val}},y_{\mathrm{val}}\}$ 分别为训练数据集和测试数据集.双层优化问题通过内层优化在中毒数据集$\{x_{\mathrm{p}}\cup x_{\mathrm{tra}}\}$上更新网络模型参数 θ,并利用外层优化获得中毒数据样本 x_{p}^{*},使参数集为 θ^{*} 的模型 f 在测试集上的损失函数最大化[11].

1.1.2 后门攻击

在模型训练阶段,数据集或预训练模型的开放性会导致网络模型极易遭受后门攻击(backdoor attack).攻击者通过在数据集或模型中嵌入触发器,使模型在处理特定数据时激活触发器以操纵模型的预测结果,而不对正常样本的结果产生影响,如下所示:

$$f_{\theta}(x)=\begin{cases}y_{\mathrm{p}}, & x\in\{x_{\mathrm{p}}\},\\ y_{\mathrm{val}}, & \text{其他}.\end{cases} \tag{3}$$

其中:$\{x_{\mathrm{p}}\}$ 为带有触发器的数据集.这类攻击仅在处理触发器时对模型结果产生影响,具有很强的隐蔽性,导致网络架构面临着严重的安全隐患.

后门攻击通常被认为是一种特殊的数据投毒攻击[12].传统数据投毒攻击主要旨在通过污染训练数据来削弱模型的泛化能力,导致模型的可用性遭到破坏;而后门攻击目的则是通过植入的触发器来进一步操纵模型的预测结果,影响用户最终决策.语义通信过程中,Zhou 等[13]提出一种语义符号后门攻击范式,其通过在训练数据集中植入特定的触发样本以更改发送方数据集中的语义符号和标签,导致接收方在处理到触发样本时产生敌手预期的错误输出.

1.1.3 对抗攻击

对抗攻击(adversarial attack)是指在模型推理阶段,攻击者在正常样本中添加微小扰动以干扰网络模型的正确预测结果,对测试样本的非随机扰动甚至可以任意地操纵神经网络的输出[14].一般而言,如果敌手掌握目标模型的结构与参数,其可利用原始数据集最大化生成与原始数据相似的对抗样本,实现对网络模型的干扰攻击.

在语义通信过程中,对抗攻击者会在数据样本或无线信道内添加语义对抗扰动,这些对抗攻击样本的存在会显著降低合法接收方的解码准确性,对语义通信的合法通信方造成不利影响.常见攻击可以分为有目标的攻击(targeted attack)和无目标的攻击(non-targeted attack)[15],分别如式(4)和式(5)所示:

$$\min_{r} f_{\theta}(x+r)=y_{\mathrm{t}}, \tag{4}$$

$$\min_{r} f_{\theta}(x+r)\neq y. \tag{5}$$

其中:r 为对抗扰动;$(x+r)$ 为所构造的对抗样本;y_{t} 为敌手目标预测结果.

1.1.4 模型窃取

模型窃取攻击(Model Extraction Attack,MEA)是指在缺乏目标模型训练数据和算法细节等先验知识的情况下,攻击者通过输入特定的查询数据并观测模型输出,推断模型的内部网络架构与参数的攻击手段.

随着机器学习模型的应用程序编程接口(Application Programming Interface,API)的日益普及,使用者利用个人数据集对模型进行定制化训练.然而,模型的开放性使得攻击者能够利用 API 公开访问网络模型,并基于神经网络的泛化能力产生大量的输入

输出结果 $\{x,y\}$ 以逼近原始模型 f_θ[16],这类攻击不仅侵犯了服务提供商的合法权益,还可能直接危及服务的安全性和可靠性.更严重地,一旦网络模型被成功窃取,攻击者能够基于已有网络架构部署白盒模型,并进行进一步的攻击,如数据污染或模型欺骗等,进一步加剧对原始系统的安全威胁[17-18].

1.2 模型安全防御手段

1.2.1 防数据投毒

为抵御数据投毒攻击,常见的防御方案包括数据清洗和鲁棒训练,其中:数据清洗通过在训练数据集中检测和识别中毒数据,以有效减少数据库中的中毒样本,来保证训练过程中数据的可靠性和模型的有效性;鲁棒训练则通过提高模型的鲁棒性来削弱恶意数据对模型性能的影响,提高模型的泛化能力,以避免模型被恶意攻击.

典型地,Steinhardt 等[19]讨论了基于固定防御和数据依赖防御的数据清洗防御策略,其通过检查完整数据集并删除异常值的方式构造可行集合,并仅在可行集上训练模型,而且通过球面距离和平面距离的两类优化问题估计网络模型损失上界,以增强模型对攻击的防御能力.针对标签数据投毒攻击,Peri 等[20]提出使用基于 k-最近邻(k-Nearest Neighbor,kNN)的防御手段,用于标识和移除异常数据,并通过消融实验确定合适的 k 值,以确保防御策略的有效性.

在鲁棒训练的防御方案中,Jagielski 等[21]提出一种面向回归模型的防御手段,通过迭代估计回归参数,并在每次迭代中利用修剪损失函数来移除具有大残差的数据点,以在多轮迭代后隔离中毒数据,实现了鲁棒性回归模型.Chen 等[22]提出了 De-Pois 系统,利用生成对抗网络(Generative Adversarial Network,GAN)来增强训练数据的多样性,并通过学习中毒样本和干净样本之间的预测差异,使得系统能够在无先验知识的条件下识别中毒样本,实现了通用的数据投毒防御手段.

1.2.2 防后门攻击

在网络模型的安全性防护中,有效检测和防御后门攻击至关重要,模型剪枝和后门检测技术是两种常用的防御策略.模型剪枝技术通过移除神经元等方式微调或重构遭受攻击的网络模型,从而显著降低后门攻击带来的潜在风险;后门检测技术则通过识别训练数据集中潜在的触发器样本,从而确定是否存在后门攻击.

在剪枝和微调策略的基础上,Liu 等[23]实现了细粒度剪枝的防御手段,其通过移除在干净样本上激活水平低的神经元来防止其被后门攻击者利用,从而降低后门攻击成功率.类似地,基于后门模型处理触发样本所需修改量更小的事实,Wang 等[24]通过异常检测算法和逆向工程技术识别网络模型中具有已知触发器的训练样本,还分别讨论了利用输入滤波器移除触发器、利用神经元剪枝重构模型和利用遗忘技术提高模型鲁棒性等方案,提升了模型抵御后门攻击的有效性.

针对语义通信过程中潜在的后门攻击问题,Zhou 等[13]设计了 U 形训练架构,使得通信双方的数据集保持相互独立,以防止攻击者对任一方数据集的单独投毒;同时,提出一种剪枝算法用于识别和剪除与后门相关的神经元,在不损害重建精度的前提下消除模型中潜在的后门.

1.2.3 防对抗攻击

为降低对抗攻击对网络模型造成的安全性风险,检测对抗攻击和基于对抗训练的防御方案被提出并得到广泛研究.典型地,以 softmax 分类模型为例,Lee 等[25]基于马氏距离来衡量置信度分数并估计异常样本与正常样本之间的差异,构建了检测异常值和对抗样本的通用防御框架.在基于对抗训练的防御策略中,网络模型使用者通过在训练集中增加含有对抗扰动的样本数据,使得网络模型具有更高的鲁棒性,从而降低对抗攻击的有效性.

在语义通信中,现有防对抗攻击的策略多基于对抗训练实现,以增强网络模型和语义通信系统对对抗样本的鲁棒性.面向语义对抗性扰动的存在,考虑到无线通信系统中信道损伤造成的影响,Hu 等[26]建立了样本相关和样本无关的语义噪声模型,提出一种带有权重扰动的对抗训练方法来对抗语义噪声,并开发了一种 VQ-VAE 模型和离散码本以适配数字通信系统.类似地,Nan 等[27]建立了一种物理层攻击者模型,其利用物理层对抗扰动生成器产生面向语义的、可控的扰动以更好地模拟真实物理环境的效果,并通过对抗训练方法增强语义通信系统对攻击的抵御能力.基于生成对抗网络 GAN,Tang 等[28]设计了一个能够在语义干扰存在的情况下准确地恢复语义信息的鲁棒接收器,通过交替优化干扰器和接收器,提高语义通信系统的安全性.

1.2.4 防模型窃取

针对模型窃取攻击,模糊预测结果的防御策略

通过使网络模型返回混淆过后的预测输出,来降低攻击者窃取模型参数和网络结构的准确率.以超参数窃取攻击为例,Wang 等[29]提出通过对参数进行舍入处理的操作,使攻击者获得的模型参数与原始参数之间存在一定的差异,增加攻击者准确估计模型参数的难度.类似地,聚焦于黑盒攻击下的模型窃取攻击,Li 等[30]利用物理不可克隆函数(Physical Unclonable Function,PUF)实现了对深度神经网络输出结果的模糊,确保合法用户能够准确恢复输出结果,而攻击者无法获得有关 PUF 响应的有效信息且难以从混淆输出中提取有效模型信息.Lee 等[31]在模型最后的激活层中添加欺骗性扰动以最大化被窃取模型的损失并保持原有模型的准确率,在保证用户能够获得有用信息的前提下,显著限制攻击者窃取模型的能力.

检测模型窃取攻击也是一种有效的防御手段.Juuti 等[32]提出一种通用的检测模型窃取攻击的防御系统 PRADA,该系统不依赖于模型的具体实现或训练数据,而是通过判断查询样本之间的距离分布是否偏离正态分布来发现异常 API 查询行为.Jiang 等[33]结合对抗训练、恶意查询检测、自适应查询响应和所有权验证的方式,构建了检测 MEA 的综合防御架构,其中,恶意查询检测在最大 softmax 概率度量的基础上,引入温度参数来度量异常查询与良性查询之间的距离,提高了检测方案的准确率,有效抵御了模型窃取攻击.

2 语义通信数据安全研究现状

无线通信网络中的数据安全问题一直是无线安全领域中备受关注的研究热点,通信数据量的迅猛增加和交互延迟的持续降低,对无线网络通信数据的安全交互和可靠流通提出了更高的要求,传输过程中的安全形势也愈发严峻.与传统语法层面的无线通信相比,语义通信架构下,合法通信方根据传输任务提取所需要的语义特征信息,在模型部署和通信阶段,交互数据不仅涵盖了数据集和知识库等基础信息,还可能蕴含用户潜在的敏感数据信息,面临着严峻的数据安全风险.

2.1 数据安全攻击

在语义通信过程中,网络模型的训练与部署、特征信息的传输与交互都不可避免地涉及大量用户数据的处理和交换,然而,网络架构和无线信道的开放性为恶意攻击者提供了潜在的攻击途径.具体而言,在网络模型部署和无线通信的数据传输环节,恶意攻击者可通过隐私攻击、无线窃听、数据篡改等攻击手段对数据的机密性和完整性构成威胁,不仅侵害了用户的隐私权益,也对敏感信息的安全保护构成了实质性的挑战.

2.1.1 隐私攻击

隐私攻击(privacy attack)一般发生在模型推理阶段,常见攻击方式包含模型逆向攻击和推理攻击.模型逆向攻击(Model Inversion Attack,MIA)是指攻击者通过分析网络模型的输出,利用反向数据流来逆向推断训练数据集的攻击手段[34].语义通信系统中,Chen 等[35]提出一种 MIA 与窃听攻击相结合的攻击方式,攻击者首先截获无线信道内传输的语义信息,再通过模型逆向尝试重建原始消息,导致用户的隐私信息泄露,传输数据的机密性遭到破坏.

推理攻击一般包含成员推理攻击和属性推理攻击.成员推理攻击的目的是确定隐私的训练数据集中是否包含某条或某些特定的个体数据记录,而属性推理攻击用于获取训练数据集的统计属性或分布特征,导致数据的完整性和机密性遭到破坏[36].模型逆向攻击和推理攻击的区别在于,模型逆向攻击是从网络模型的输出中逆向获取输入数据集,导致整个数据集的机密性受到严重威胁,而推理攻击一般仅获取数据集的一般特性,而非原始数据.

2.1.2 无线窃听

语义通信系统中除模型训练、部署阶段可能遭遇的安全威胁外,系统无线传输的特性同样带来了一系列的安全挑战.无线信道固有的开放性和易接入性,使得在信道内以明文形式传输的语义特征信息极易遭受窃听攻击,无线信道内的窃听装置利用电磁波在空间自由传输的特性,调整设备接收频率即可截获信道内传输的信息.

如图 3 所示,在信号传输过程中,由于窃听装置不主动发射信号而仅处于接收模式,合法通信双方往往难以察觉窃听攻击的存在.这种隐蔽性使得用户敏感信息面临泄露的风险,进而对无线通信系统在数据机密性保护方面构成极大的安全风险.在无线通信系统中,攻击者利用窃听手段非法获取敏感信息后,可能进一步实施如假冒、伪造和篡改等更具威胁性的攻击,这些行为将对合法通信用户的数据安全造成严重威胁.

2.1.3 数据篡改

数据篡改是指攻击者通过非法手段获取对数据

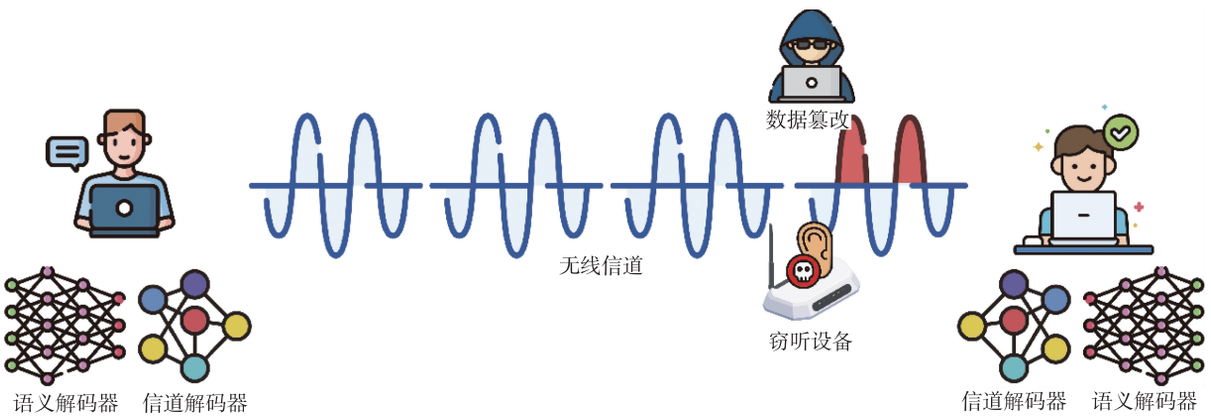

图 3　语义通信系统无线窃听和数据篡改攻击

Fig. 3　Wireless eavesdropping and data tampering attacks on semantic communication systems

的访问权限,并未经授权地修改或更改原始数据内容[37],这类攻击不仅会破坏数据的完整性和可靠性,使数据失去原有的真实性和准确性,还可能导致基于数据的深度神经网络的输出结果出现偏差,误导用户做出错误的决策,对系统的可靠性和用户利益造成严重的危害.

在语义通信系统中,包括数据收集、模型训练、模型推理以及无线传输在内的各个环节均面临着数据被篡改的风险,例如,模型安全中的数据投毒、后门攻击等也属于数据篡改攻击的范畴.这些攻击对合法用户之间交互的数据带来了极大的安全风险,严重威胁着用户敏感数据和隐私信息的机密性、完整性和系统的可靠性.如图 3 所示,特别是在无线传输过程中,篡改攻击不仅会破坏传输信息的完整性和真实性,还会直接干扰用户接收到的最终输出结果,从而对系统的整体可用性产生不利影响.

2.2　数据安全防御手段

2.2.1　防隐私攻击

为防止网络模型遭受隐私攻击而导致的用户隐私泄露,已有研究通过提高模型的鲁棒性来实现对隐私数据的安全保护.基于互信息正则化的防御机制,Wang 等[38]在损失函数中引入正则化项,来最小化模型输入输出之间的互信息,以削弱网络模型输入输出之间的依赖关系,而且建立了适用于不同模型的正则化近似方法,从而有效抵御推理攻击.针对黑盒模型下的模型逆向攻击,Wen 等[39]设计并建立了攻击者网络模型,并利用该模型的梯度信息来计算可以添加到目标模型输出上的对抗性噪声向量,从而在确保目标模型预测准确率的同时,最大化逆向攻击的误差,限制模型逆向攻击者的能力.类似地,Gong 等[40]提出一种利用生成对抗网络的防御手段,构造了逆向公共样本和逆向私有样本,并向合法输出标签中添加误导特征来抵御多种隐私攻击.

与此同时,基于同态加密、安全多方计算、差分隐私等典型密码技术的机器学习隐私安全保护架构也已有研究,在密码算法和安全协议的支撑下,深度学习网络能够在不泄露参与方隐私的同时实现模型的训练和部署,并获得预期结果,其中,同态加密支持对密文进行运算并保证结果与明文运算结果一致,安全多方计算技术允许多参与方在保护各自隐私的前提下获得函数结果,而差分隐私通过在数据集上添加噪声实现对个体数据的保护.在安全多方计算攻击模型中,Zheng 等[41]提出一种在不共享原始数据的情况下安全训练网络模型的多方协作系统,并结合同态加密和零知识证明相关技术,实现了对参与方数据隐私的保护.针对边端协同场景下的隐私攻击,Li 等[42]不仅基于分割学习原理实现了对真实输出标签的隐藏,而且通过在传输数据中加入拉普拉斯分布的随机噪声实现了差分隐私机制,在不泄露有用信息的前提下增加了数据的随机性,提供了敏感信息的机密性保护.

语义通信系统中,对所提出的模型逆向窃听攻击,Chen 等[35]实现了基于随机排列和替换的防御方法,该方案旨在防止攻击者有效地重建原始信息,从而提高传输数据的机密性和网络系统的安全性.为抵御模型逆向攻击,Wang 等[43]提出一种基于信息瓶颈理论和对抗学习的语义通信系统,通过对抗学习训练编码器,产生抗攻击的语义特征,以增强系统的隐私保护能力.针对通信节点的知识库差异导致的逆向攻击,Cheng 等[44]提出使用知识映射和消歧

来减少收发两端的知识差异,并使用路径切断方法来防止敏感数据遭到泄露.

2.2.2 防无线窃听

为防止传输的语义特征数据遭受窃听攻击,常用的关键技术之一是加密.合法通信方对传输内容进行加密处理,只有合法接收方才可以解密信息,防止信息被非授权人员获取或访问,为传输数据提供了机密性保护,进而保障了无线通信系统的传输安全.基于公钥密码体制,Tung 等[45]将计算难题转化为复合噪声,实现了联合信源-信道编码的图像安全语义通信方案.在对抗训练网络支持加密通信的研究基础上,Luo 等[46]设计了面向文本信息传输的隐私保护加密语义通信系统,以确保在加密和未加密模式下通信的准确性,并防止攻击者重建原始语义信息;Zhang 等[47]提出了兼顾语义和安全性的图像传输系统,通过在损失函数中引入窃听重构误差,达到传输效率和隐私保护的平衡.类似地,Xu 等[48]在图像传输过程中引入保护模块和去保护模块,实现了对传输数据的机密性保护.对抗加密方案模型如图 4 所示,通过设计合法通信双方和攻击者的损失函数,使合法接收方能够对消息进行准确重构,并限制攻击者重构能力以提高无线通信系统的安全性.

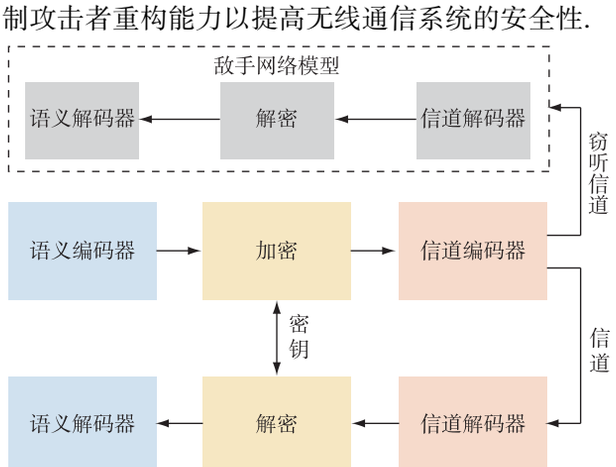

图 4 加密语义通信对抗训练网络示意图

Fig. 4 Adversarial training networks in encrypted semantic communications

除了上述加密方案,Khalid 等[49]在元宇宙背景下,基于量子机器学习技术将数据的语义信息编码为量子态,并通过量子匿名通信协议进行传输,实现用户间的安全可靠交互.类似地,结合量子密钥分发(Quantum Key Distribution,QKD)和语义信息通信系统架构,Kaewpuang 等[50]提出在边缘设备间利用 QKD 构建满足语义信息传输安全需求的密钥,实现面向语义通信的 QKD 协作管理网络.与此同时,基于混沌加密的安全语义通信方案也逐渐得到关注和研究,混沌加密方案以结构复杂且高随机性的混沌序列作为密钥,保障传输内容的机密性[51].在正交频分复用无源光网络(Orthogonal Frequency Division Multiplexed Passive Optical Network,OFDM-PON)下,Ma 等[52]在提取传输信息中的语义特征后,使用厄农映射产生混沌序列,加密 16QAM 的 OFDM 信号,以解决语义光通信系统中的安全问题.

以信息论为基础,物理层安全方案利用物理介质的固有特征或信道噪声的随机性和不确定性,实现了物理层密钥协商和对非法窃听的有效干扰[53-54].聚焦于物理层密钥协商技术,Zhao 等[55]探究了语义通信过程中特有的随机现象——语义漂移,并提出了物理层语义密钥生成方案.类似地,Qin 等[56]直接利用双语评估替补(Bilingual Evaluation Understudy,BLEU)分数的随机性生成物理层语义密钥,实现了加密和子载波级的数据混淆.结合图像语义传输系统,Rong 等[57]量化图像中包含的语义信息量,并构建了细粒度的 OFDM 子载波分配方案,实现了自适应传输和物理层加密方案.结合 Wyner 窃听模型和语义通信架构,Du 等[58]提出了面向语义物联网场景的安全性能指标,重新定义了语义保密中断概率等评估指标以量化语义物联网系统的安全性.Mu 等[59]提出一个语义辅助的保密传输框架,合法通信双方在窃听信道模型下,传输语义流以干扰恶意节点对比特流的窃听,增强无线信道传输的安全性.结合同时发射和反射的可重构智能表面(Simultaneous Transmitting and Reflecting Reconfigurable Intelligent Surface,STAR-RIS),Wang 等[60]利用波束成形技术增强了基站和目标用户之间的语义通信信号传输,并对窃听用户造成干扰.

2.2.3 防数据篡改

为了抵御数据篡改攻击并确保数据的完整性,常用的防御措施包括完整性校验机制,这些机制依赖于哈希算法和数字签名等技术来确保数据及其来源的真实性.哈希算法利用其单向性和抗碰撞性,生成数据的唯一消息摘要(哈希值),并通过比较数据的摘要是否发生变化来迅速检测数据篡改;而数字签名技术则基于公钥密码体制,通过私钥签名和公钥验证的机制,确保了传输数据的完整性和消息源身份的可信性.与此同时,随着区块链技术的不断进步,其所固有的去中心化、非篡改、可公开和匿名等

特性,不仅为解决数据篡改、溯源困难等安全问题提供了有效手段,还为构建安全、可靠且透明的数据交互系统引入了新的机制.

在端到端语义通信系统架构下,为防止数据被恶意篡改,Liu 等[61]提出的 SemProtector 构造了可生成语义签名的插拔模块,为在线语义通信系统所传输的语义特征信息提供了完整性保护.除此之外,在分布式语义通信系统中,聚焦于 Web3.0 中的高效安全信息交互,Lin 等[62-63]提出利用零知识证明、区块链和智能合约等技术,确保数据的一致性和不可篡改性,其中,零知识证明允许在不获取原数据的前提下验证数据的真实性,区块链和智能合约本身的防篡改机制也限制了恶意节点的行为.同样地,为了在元宇宙背景下提供安全的生成式人工智能(Artificial Intelligence Generated Content,AIGC)服务,Lin 等[64]构建了一个基于区块链的语义通信架构,其利用区块链技术来建立去中心化的信任机制,确保数据的真实性,并防止数据被恶意篡改.

3 语义通信安全研究面临的挑战

尽管现有研究已经提出了各类方法以抵御语义通信场景下存在的安全威胁,但是面向语义通信的安全技术研究仍然面临着一系列亟待解决的挑战,包括安全性评估、技术创新和复杂应用场景等带来的挑战.未来语义通信安全领域的研究将深入挖掘并结合语义通信场景的内在特性,并通过与网络安全领域的研究成果相融合,确保无线通信系统的安全性与稳定性,为语义通信系统建立更为坚固的安全防线.

3.1 语义通信的安全性评估的挑战

目前,面向语义通信系统的安全研究在架构和协议层面不断取得显著进展,但其在安全性分析上仍面临着严峻挑战,特别是缺乏精准的安全性分析与评估方法,给系统的安全性带来了很大的不确定性.尽管部分工作对方案的安全性进行了一定的讨论与评估,但大多依赖于实验验证,即通过实验证明攻击者无法有效重构语义信息等方法来验证安全性,缺乏一套全面和客观的评估指标和标准.此外,这种验证方法通常局限于已知的攻击模式和测试条件,而在实际应用中可能存在未知的攻击方式,导致原防御方案的安全性难以得到有效保证.因此,需综合考虑多方面因素,包括算法的复杂性、抗攻击的能力、隐私保护的强度以及对外部环境的依赖等,构建全面衡量安全性水平的评估框架.

与此同时,语义通信系统中各个编解码模块大多基于深度神经网络架构实现,这些网络模型包含大量参数和复杂层级结构,以便通过训练学习获得预测结果,并准确地处理数据.然而,深度神经网络目前通常被认为是黑盒模型,在预测和使用时,模型内部的工作机制和原理难以被理解和解释.深度学习的不可解释性使得基于数学证明的理论分析方法,在验证语义通信安全性上存在困难,不能直接分析系统、协议或算法安全性.因此,基于深度学习的语义通信系统的安全研究,需要从网络架构的特性出发,设计与神经网络相适应的安全性评估及分析方法.

另外,从性能评估角度出发,目前提出的语义通信安全方案往往都是基于网络架构的,其安全性与网络架构的设计密切相关,涉及物理层、语义层、应用层等多层次多组件的信息交互与协同.在评估基于网络架构的语义通信安全方案的性能时,需要考虑到各个层次和组件之间的相互作用,以及安全策略的实现方式.与传统的依赖于特定安全设备或算法以提供安全性的方案相比,语义通信安全技术的安全性能评估更为困难,需要针对其特殊性和复杂性在评估方法和度量标准上进行适当调整,以确保语义通信安全方案性能评估结果的准确性和可信度.

3.2 语义通信安全技术层面的挑战

语义通信作为一种新兴的通信范式,其理论框架和技术体系尚处在不断发展和完善的过程中.目前,研究工作主要聚焦于技术层面的深入探索和应用领域的广泛拓展.与此同时,尽管面向语义通信系统的安全和隐私保护技术现正处在快速发展的阶段,但仍存在诸多技术层面上的挑战亟待解决.当前,部分研究工作主要是基于语义通信的系统架构,面向新兴应用场景,在概念层面上提出了可供参考的模型架构,然而,这些研究目前尚缺乏具体的、可实施的技术细节,在实际应用中难以有效抵御安全威胁.此外,攻击者能力和攻击手段的不断变化,也给语义通信系统的安全策略提出了更高的要求.因此,在语义通信架构下,灵活智能的安全技术仍然需要进一步的研究与探索,并构建切实可行的技术路线细节.

语义通信系统与安全技术相结合,会不可避免地带来一定程度上的性能损失.除此之外,安全模块

的引入将会进一步提升原有通信网络模型的复杂度,而以提升安全性为目的的交互数据量的增长也可能会引入额外的信令与通信开销,会对通信性能造成一定的影响,尤其是在低延时、高流量密度的应用场景中,这种影响尤为显著,给通信需求的权衡带来了更复杂的挑战.因此,在推进语义通信系统与安全技术融合的过程中,要从语义通信系统的实际应用需求出发,探究解决通信安全与通信性能之间矛盾的有效策略,以实现安全策略与通信需求的有效平衡.

与此同时,目前语义通信安全研究往往未能充分探讨语义通信系统自身跨层通信的结构特点,多数研究仅聚焦于上层或物理层等单一层面的安全保障,跨层语义通信的安全研究尚处在相对匮乏的状态,且不同域之间的信息交互和数据共享也需要建立跨域信任和管理机制.尽管已有研究着眼于语义通信系统的加密技术、物理层安全技术及隐蔽通信技术等方面,但对与特定应用场景相适配的密码算法或安全协议的深入研究仍显不足,而且许多现有安全架构尚未与语义通信系统实现有效融合,在一定程度上影响了安全策略的实施效果.因此,为构建更为完善的语义通信安全架构体系,需要综合考虑并充分利用如差分隐私、同态加密、多方安全计算等多种安全技术,并建立有效的跨域信任机制,包括身份验证、授权和安全审计等以增强语义通信系统的安全防护能力,从而确保无线通信环境的安全性与可靠性.

3.3 语义通信安全应用层面的挑战

随着智能技术的迅猛发展,语义通信的应用场景和领域正不断拓展.然而,面向实际应用的语义通信模型普遍存在架构各异的问题,尚未建立完善且统一的语义通信技术标准体系,这也导致在应用层面上不断出现更为复杂的安全挑战.

从信息源的角度来看,语义通信传输的并非仅限于文本、语音、图像、视频等单一模态的数据,而是多种模态数据的融合.然而,目前大部分安全技术研究主要面向图像、文本等单模态数据,多模态应用场景下的安全技术研究稍显不足.由于不同模态数据的处理与安全传输往往依赖于不同的安全策略,在部署安全语义通信系统时,必须考虑多模态数据之间的差异性与相似性,从而构建安全高效的语义通信范式.

从应用场景的角度出发,语义通信可广泛应用于多种智能领域和下一代通信场景中.由于不同应用场景下的语义通信架构可能存在较大差异,安全性的需求也不尽相同.这种差异会导致安全技术和策略难以简单迁移,在适用性上存在一定的局限性.因此,在面向语义通信的安全技术研究过程中,必须综合考虑具有普适性的通用安全策略和面向特定场景的专用安全策略,以确保安全研究的全面性与实用性.

在语义通信过程中,用户的个人信息与隐私数据的安全至关重要.在语义通信的发展进程中,制定相关政策与法律法规,建立合法合规的技术标准,是不可或缺的可靠前提,这些举措也能够进一步推动语义通信系统的持续发展,为语义通信安全提供保障.

4 结束语

目前,面向语义通信中的安全问题及其应对策略的研究仍处在萌芽阶段,如何有效保障语义通信系统的模型安全和传输数据安全仍是一个亟待解决的热点问题,且尚未得到深入的研究与分析.本文主要围绕语义通信系统的安全研究展开综述,基于语义通信网络架构的独特性,深入剖析了模型训练、推理、部署以及无线传输等多个阶段中语义通信系统所面临的安全威胁,并概述了各类攻击及其防御手段的研究现状.最后,总结了目前语义通信安全研究在安全性评估、技术创新和实际应用等多个层面所面临的挑战,并对未来语义通信系统下安全研究的发展方向进行了展望.

参考文献

References

[1] Shannon C E. A mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27(3): 379-423

[2] Weaver W. The mathematical theory of communication [M]. Urbana-Champaign: University of Illinois Press, 1963

[3] Strinati E C, Barbarossa S. 6G networks: beyond Shannon towards semantic and goal-oriented communications [J]. Computer Networks, 2021, 190: 107930

[4] Xie H Q, Qin Z J, Li G Y, et al. Deep learning enabled semantic communication systems [J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2675

[5] Qin Z J, Tao X M, Lu J H, et al. Semantic communications: principles and challenges [J]. arXiv e-Print, 2022, arXiv: 2201.01389

[6] 王衍虎,郭帅帅.基于大语言模型的语义通信:现状,

挑战与展望[J].移动通信,2024,48(2):16-21
WANG Yanhu, GUO Shuaishuai. Large language model-based semantic communications: status, challenges, and prospects [J]. Mobile Communications, 2024, 48 (2): 16-21

[7] Zou Y L,Zhu J,Wang X B,et al.A survey on wireless security:technical challenges,recent advances,and future trends[J].Proceedings of the IEEE,2016,104(9):1727-1765

[8] He Y Z, Meng G Z, Chen K, et al. Towards security threats of deep learning systems: a survey [J]. IEEE Transactions on Software Engineering, 2022, 48 (5): 1743-1770

[9] Xia G M,Chen J,Yu C D,et al.Poisoning attacks in federated learning:a survey[J].IEEE Access,2023,11:10708-10722

[10] 姜文博.机器学习中的安全与隐私保护技术研究[D].成都:电子科技大学,2023
JIANG Wenbo. Research on security and privacy-preservation techniques in machine learning[D].Chengdu:University of Electronic Science and Technology of China,2023

[11] Mei S K,Zhu X J.Using machine teaching to identify optimal training-set attacks on machine learners[J].Proceedings of the AAAI Conference on Artificial Intelligence,2015,29(1):2871-2877

[12] Chen X Y,Liu C,Li B,et al.Targeted backdoor attacks on deep learning systems using data poisoning[J].arXiv e-Print,2017,arXiv:1712.05526

[13] Zhou Y H,Hu R,Qian Y.Backdoor attacks and defenses on semantic-symbol reconstruction in semantic communications[J].arXiv e-Print,2024,arXiv:2404.13279

[14] Kang J W,He J Y,Du H Y,et al.Adversarial attacks and defenses for semantic communication in vehicular metaverses[J].IEEE Wireless Communications,2023,30(4):48-55

[15] Sagduyu Y E,Erpek T,Ulukus S,et al.Is semantic communication secure? A tale of multi-domain adversarial attacks[J].IEEE Communications Magazine,2023,61(11):50-55

[16] Tramèr F, Zhang F, Juels A, et al. Stealing machine learning models via prediction APIs[C]//25th USENIX Security Symposium (USENIX Security 16). August 10-12,2016,Austin,TX,USA.USENIX,2016:601-618

[17] Papernot N, McDaniel P, Goodfellow I, et al. Practical black-box attacks against machine learning [C]//Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security. April 2 - 6, 2017, Abu Dhabi, United Arab Emirates. ACM, 2017: 506-519

[18] 任奎,孟泉润,闫守琨,等.人工智能模型数据泄露的攻击与防御研究综述[J].网络与信息安全学报,2021,7(1):1-10
REN Kui,MENG Quanrun,YAN Shoukun,et al.Survey of artificial intelligence data security and privacy protection [J]. Chinese Journal of Network and Information Security,2021,7(1):1-10

[19] Steinhardt J, Koh P W W, Liang P S. Certified defenses for data poisoning attacks[C]//Advances in Neural Information Processing Systems (NIPS2017). December 4-9,2017, Long Beach, CA, USA. NIPS Foundation, 2017: 3520 - 3532

[20] Peri N, Gupta N, Huang W R, et al. Deep k-NN defense against clean-label data poisoning attacks [C]// European Conference on Computer Vision. August 23-28,2020,Online.Cham:Springer,2020:55-70

[21] Jagielski M, Oprea A, Biggio B, et al. Manipulating machine learning: poisoning attacks and countermeasures for regression learning[C]//2018 IEEE Symposium on Security and Privacy (SP). May 21-23,2018,San Francisco,CA,USA.IEEE,2018:19-35

[22] Chen J,Zhang X X,Zhang R,et al.De-Pois:an attack-agnostic defense against data poisoning attacks[J].IEEE Transactions on Information Forensics and Security,2021,16:3412-3425

[23] Liu K, Dolan-Gavitt B, Garg S. Fine-pruning: defending against backdooring attacks on deep neural networks [C]//International Symposium on Research in Attacks, Intrusions, and Defenses. September 10-12, 2018, Heraklion, Greece. Cham: Springer, 2018: 273-294

[24] Wang B L,Yao Y S,Shan S,et al.Neural cleanse:identifying and mitigating backdoor attacks in neural networks [C]//2019 IEEE Symposium on Security and Privacy (SP). May 19-23,2019,San Francisco,CA,USA.IEEE,2019:707-723

[25] Lee K,Lee K,Lee H,et al.A simple unified framework for detecting out-of-distribution samples and adversarial attacks[J].arXiv e-Print,2018,arXiv:1807.03888

[26] Hu Q Y,Zhang G Y,Qin Z J,et al.Robust semantic communications with masked VQ-VAE enabled codebook [J]. IEEE Transactions on Wireless Communications, 2023,22(12):8707-8722

[27] Nan G S, Li Z C, Zhai J L, et al. Physical-layer adversarial robustness for deep learning-based semantic communications[J].IEEE Journal on Selected Areas in Communications,2023,41(8):2592-2608

[28] Tang R,Gao D H,Yang M X,et al.GAN-inspired intelligent jamming and anti-jamming strategy for semantic communication systems [C]//2023 IEEE International Conference on Communications Workshops (ICC Workshops). May 28-June 1,2023,Rome,Italy.IEEE,2023:1623-1628

[29] Wang B H, Gong N Z. Stealing hyperparameters in machine learning [C]//2018 IEEE Symposium on Security and Privacy (SP). May 20-24,2018,San Francisco,CA,USA.IEEE,2018:36-52

[30] Li D W,Liu D,Guo Y,et al.Defending against model extraction attacks with physical unclonable function[J].Information Sciences,2023,628:196-207

[31] Lee T, Edwards B, Molloy I, et al. Defending against neural network model stealing attacks using deceptive

perturbations[C]//2019 IEEE Security and Privacy Workshops (SPW). May 23, 2019, San Francisco, CA, USA. IEEE, 2019:43-49
[32] Juuti M, Szyller S, Marchal S, et al. PRADA: protecting against DNN model stealing attacks[C]//2019 IEEE European Symposium on Security and Privacy. June 17-19, 2019, Stockholm, Sweden. IEEE, 2019:512-527
[33] Jiang W B, Li H W, Xu G W, et al. A comprehensive defense framework against model extraction attacks[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(2):685-700
[34] Fang H, Qiu Y X, Yu H Y, et al. Privacy leakage on DNNs: a survey of model inversion attacks and defenses[J]. arXiv e-Print, 2024, arXiv:2402.04013
[35] Chen Y H, Yang Q Q, Shi Z G, et al. The model inversion eavesdropping attack in semantic communication systems[C]//2023 IEEE Global Communications Conference. December 4-8, 2023, Kuala Lumpur, Malaysia. IEEE, 2023:5171-5177
[36] Jegorova M, Kaul C, Mayor C, et al. Survey: leakage and privacy at inference time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7):9090-9108
[37] 林飞,刘佳宁,焦强.大数据背景下信息安全问题探析[J].计算机技术与发展,2024,34(8):1-8
LIN Fei, LIU Jianing, JIAO Qiang. Exploration of information security issues in the context of big data[J]. Computer Technology and Development, 2024, 34(8):1-8
[38] Wang T H, Zhang Y H, Jia R X. Improving robustness to model inversion attacks via mutual information regularization[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(13):11666-11673
[39] Wen J, Yiu S M, Hui L C K. Defending against model inversion attack by adversarial examples[C]//2021 IEEE International Conference on Cyber Security and Resilience (CSR). July 26-28, 2021, Rhodes, Greece. IEEE, 2021:551-556
[40] Gong X L, Wang Z Y, Li S K, et al. A GAN-based defense framework against model inversion attacks[J]. IEEE Transactions on Information Forensics and Security, 2023, 18:4475-4487
[41] Zheng W T, Popa R A, Gonzalez J E, et al. Helen: maliciously secure coopetitive learning for linear models[C]//2019 IEEE Symposium on Security and Privacy (SP). May 19-23, 2019, San Francisco, CA, USA. IEEE, 2019:724-738
[42] Li J Y, Liao G C, Chen L, et al. Roulette: a semantic privacy-preserving device-edge collaborative inference framework for deep learning classification tasks[J]. IEEE Transactions on Mobile Computing, 2024, 23(5):5494-5510
[43] Wang Y H, Guo S S, Deng Y Q, et al. Privacy-preserving task-oriented semantic communications against model inversion attacks[J]. IEEE Transactions on Wireless Communications, 2024, PP(99):1-10
[44] Cheng S Q, Zhang X F, Sun Y, et al. Knowledge discrepancy oriented privacy preserving for semantic communication[J]. IEEE Transactions on Vehicular Technology, 2024, PP(99):1-10
[45] Tung T Y, Gündüz D. Deep joint source-channel and encryption coding: secure semantic communications[C]//IEEE International Conference on Communications. May 28-June 1, 2023, Rome, Italy. IEEE, 2023:5620-5625
[46] Luo X L, Chen Z Y, Tao M X, et al. Encrypted semantic communication using adversarial training for privacy preserving[J]. IEEE Communications Letters, 2023, 27(6):1486-1490
[47] Zhang M J, Li Y, Zhang Z Z, et al. Wireless image transmission with semantic and security awareness[J]. IEEE Wireless Communications Letters, 2023, 12(8):1389-1393
[48] Xu J L, Ai B, Chen W, et al. Deep joint source-channel coding for image transmission with visual protection[J]. IEEE Transactions on Cognitive Communications and Networking, 2023, 9(6):1399-1411
[49] Khalid U, Ulum M S, Farooq A, et al. Quantum semantic communications for metaverse: principles and challenges[J]. IEEE Wireless Communications, 2023, 30(4):26-36
[50] Kaewpuang R, Xu M R, Lim W Y B, et al. Cooperative resource management in quantum key distribution (QKD) networks for semantic communication[J]. IEEE Internet of Things Journal, 2024, 11(3):4454-4469
[51] Matthews R. On the derivation of a "chaotic" encryption algorithm[J]. Cryptologia, 1989, 13(1):29-42
[52] Ma Y L, Ren J X, Liu B, et al. Secure semantic optical communication scheme based on the multi-head attention mechanism[J]. Optics Letters, 2023, 48(16):4408-4411
[53] Shannon C E. Communication theory of secrecy systems[J]. The Bell System Technical Journal, 1949, 28(4):656-715
[54] Maurer U M. Secret key agreement by public discussion from commoninformation[J]. IEEE Transactions on Information Theory, 1993, 39(3):733-742
[55] Zhao R, Qin Q, Xu N Y, et al. SemKey: boosting secret key generation for RIS-assisted semantic communication systems[C]//2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). September 26-29, 2022, London, UK. IEEE, 2022:1-5
[56] Qin Q, Rong Y K, Nan G S, et al. Securing semantic communications with physical-layer semantic encryption and obfuscation[C]//IEEE International Conference on Communications. May 28-June 1, 2023, Rome, Italy. IEEE, 2023:5608-5613
[57] Rong Y, Nan G, Zhang M, et al. Semantic entropy can simultaneously benefit transmission efficiency and channel security of wireless semantic communications[J]. arXiv e-Print, 2024, arXiv:2402.02950
[58] Du H Y, Wang J C, Niyato D, et al. Rethinking wireless

communication security in semantic Internet of Things [J]. IEEE Wireless Communications, 2023, 30(3): 36-43
[59] Mu X D, Liu Y W. Semantic communication-assisted physical layer security over fading wiretap channels[J]. arXiv e-Print,2024,arXiv:2402.14581
[60] Wang Y,Yang W,Guan P,et al.Star-RIS-assisted privacy protection in semantic communication system[J].IEEE Transactions on Vehicular Technology,2024:1-6. DOI:10.1109/TVT.2024.3383824
[61] Liu X H,Nan G S,Cui Q M,et al.SemProtector:a unified framework for semantic protection in deep learning-based semantic communication systems [J]. IEEE Communications Magazine,2023,61(11):56-62
[62] Lin Y J,Gao Z P,Tu Y F,et al.A blockchain-based semantic exchange framework for web 3.0 toward participatory economy [J]. IEEE Communications Magazine, 2023,61(8):94-100
[63] Lin Y J,Gao Z P,Du H Y,et al.A unified blockchain-semantic framework for wireless edge intelligence enabled web 3.0[J].IEEE Wireless Communications,2024,31(2):126-133
[64] Lin Y J,Du H Y,Niyato D,et al.Blockchain-aided secure semantic communication for AI-generated content in metaverse[J].IEEE Open Journal of the Computer Society,2023,4:72-83

Semantic communication security:a survey

SHI Jiting[1] ZENG Weihao[1] ZHANG Qianyun[1] LIU Kaige[1] QIN Zhijin[2] LI Shufeng[3]
1 School of Cyber Science and Technology,Beihang University,Beijing 100191,China
2 School of Information Science and Technology,Tsinghua University,Beijing 100084,China
3 School of Information and Communication Engineering,Communication University of China,Beijing 100024,China

Abstract With the deep integration of artificial intelligence technology and wireless communication,semantic communication technology emerges as a vital mode focusing on semantic-level information transmission and interaction, thereby significantly enhancing communication accuracy and reliability.In the scenarios of low latency and high traffic density communication applications, semantic communication technology surpasses traditional syntactic-level communication grounded in classical information theory,presenting a new paradigm in wireless communication and expanding the application scope of modern communication technology.However,the development of semantic communication technology is still in its infancy,and the security issues it faces in the application process have not been thoroughly researched and comprehensively analyzed.To advance the development and implementation of semantic communication technology, this paper first provides an overview of various security threats in semantic communication systems;then,it details the research status of model security and data security in semantic communication systems;finally,it summarizes the challenges faced by semantic communication security research and outlooks the future trends.

Key words semantic communication (SC);data security;privacy preservation;wireless communication security