工业互联网流量分析技术综述

刘奇旭[1,2]，肖聚鑫[1,2]，谭耀康[1,2]，王承淳[1,2]，黄昊[1,2]，张方娇[1]，尹捷[1]，刘玉岭[1,2]
（1.中国科学院信息工程研究所，北京 100085；2.中国科学院大学网络空间安全学院，北京 100049）

摘　要：为了深入理解流量分析技术在工业互联网中的应用，基于流量分析的5个主要步骤阐述了工业互联网区别于传统互联网的独特性。同时，通过调研大量相关研究工作，总结了流量预测、协议识别与逆向、工业资产指纹识别、入侵检测、加密流量识别和漏洞挖掘6个主流研究任务，并根据任务性质将其分类为面向服务质量提高和面向安全能力提升的2类应用，充分挖掘了工业互联网中的流量分析技术应用场景。最后，针对流量分析未来进一步应用于工业互联网所面临的挑战进行了讨论，并展望了潜在的研究方向。

关键词：工业互联网；工业控制系统；流量分析；机器学习
中图分类号：TP393
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024145

Survey of industrial Internet traffic analysis technology

LIU Qixu[1,2], XIAO Juxin[1,2], TAN Yaokang[1,2], WANG Chengchun[1,2], HUANG Hao[1,2], ZHANG Fangjiao[1], YIN Jie[1], LIU Yuling[1,2]
1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: To gain an in-depth awareness of the application of traffic analysis technology in the industrial Internet, the differences between the industrial Internet and the traditional Internet through the five core traffic analysis processes were illustrated. By reviewing a large number of related papers, the application of six popular were summarized in the industrial Internet, such as traffic prediction, protocol identification and reverse engineering, industrial asset fingerprinting, intrusion detection, encrypted traffic identification and vulnerability mining. Depending on the nature of the task, traffic analysis technology was classified into two types of applications, such as service quality enhancement and security capability development, allowing to thoroughly explore the application scenarios of traffic analysis technology in the industrial Internet. Finally, the challenges associated with future traffic analysis applications in the industrial Internet were examined, as well as potential development possibilities.

Keywords: industrial Internet, industrial control system, traffic analysis, machine learning

0　引言

工业互联网是互联网和新一代信息技术与工业控制系统（ICS, industrial control system）全方位深度融合所形成的产业和应用生态，是工业智能化发展的关键综合信息基础设施[1]，其核心目标在于通过连接、集成和优化工业生产过程中的物理设备、控制系统和数据流，实现生产效率、质量、灵活性和可持续性的显著提升。中国工业互联网产业规模

收稿日期：2024–03–07；修回日期：2024–07–22
通信作者：尹捷，yinjie@iie.ac.cn
基金项目：中国科学院青年创新促进会基金资助项目；国家电网有限公司科技基金资助项目（No.SG270000YXJS2311060）；中国科学院网络测评技术重点实验室和网络安全防护技术北京市重点实验室基金资助项目
Foundation Items: The Youth Innovation Promotion Association CAS, State Grid Corporation of China Technology Project (No.SG270000YXJS2311060), The Key Laboratory of Network Assessment Technology at Chinese Academy of Sciences and Beijing Key Laboratory of Network security and Protection Technology

已经超过1.2万亿元[2]，工业物联网、大数据分析、云计算和人工智能等新一代信息技术的引入，满足了不断变化的市场需求和挑战，同时也为工业互联网带来了更多不可控风险和安全性问题。从攻击伊朗核设施的“Stuxnet”和袭击乌克兰电网的“BlackEnergy”到勒索美国燃气公司的“Darkside”，工业互联网一直是黑客组织的重点攻击目标，攻击者企图通过网络攻击影响目标的工业生产和国家安全。

工业物联网、工业系统上云和标识解析等应用的大规模实施，使得工业互联网流量规模呈现爆发性增长趋势。网络流量作为贯穿于工业互联网的“血液”，具备极其重要的地位[3]。从微观上看，其携带的信息涵盖了广泛的生产控制指令和运维状态数据。从宏观上看，其存在的深层特征反映了工业生产规律及行为模式。因此，基于流量中的各类特征进行统计分析和数据挖掘，能够达到优化网络性能、加强网络设备管理、检测响应潜在威胁、发现漏洞缺陷的效果，从而提高管理者对工业互联网的生产管理质量和安全监测能力。

本文调研了近10年来流量分析技术在工业互联网的应用相关研究工作，统计发现，近5年研究工作相比之前在数量和质量上均有较大提升，如图1所示。本文主要贡献可总结为以下3个方面。

1) 本文聚焦流量分析过程的5个主要步骤，从网络流量出发比较传统互联网和工业互联网之间的差异，阐述了工业互联网流量分析的独特性。

2) 本文对流量分析在工业互联网中的应用进行了系统的调研梳理，提出了基于研究任务的分类方法，总结了6个主流研究任务，根据相应任务的性质将工业互联网流量分析技术分为面向服务质量提高和面向安全能力提升的2类应用，覆盖服务和安全2个焦点问题。

3) 本文基于现有研究工作，围绕流量数据集构建、模型可解释性和鲁棒性、对抗机器学习及其防御、资源约束和性能限制、大模型的应用5个方面，分析了流量分析在工业互联网中面临的挑战与机遇，并展望了未来研究方向。

1　基础知识

1.1　工业控制系统与工业互联网

根据ISA-95标准体系[4]，工业控制系统由信息技术（IT, information technology）和运营技术（OT, operational technology）两部分组成，涵盖了各类负责实时监控、数据分析和智能控制的网络物理系统（CPS, cyber physical system），整体架构从上到下分别为过程监控层、现场控制层和现场设备层，数据流动呈金字塔状，如图2左侧所示。

2022年，由中国牵头制定了全球首个工业互联网系统功能架构标准IEC-PAS-63441[5]，规范了工业互联网的云边端架构，明确了工业互联网和工业控制系统的关系。数据成为工业互联网架构中的

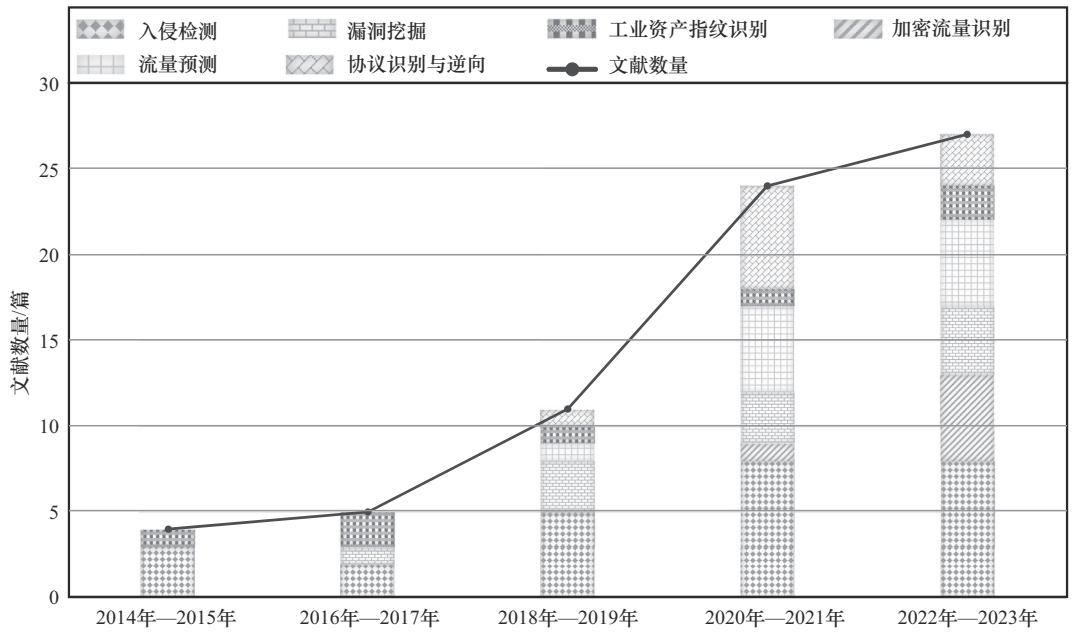

图1　现有研究工作分布

基本元素，工业控制系统的所有设备均作为资源节点接入网络，充当工业互联网的底座，使得工业数据来源更加广泛。融合工业物联网、人工智能和云计算等新一代信息技术后，传统 ISA-95 架构得到了改进[6]，新的扁平化、网状架构重新定义了制造集成方式，数据不再像金字塔状逐层流动，满足工业互联网的标准，架构如图 2 所示。

在新架构中，工业物联网的海量传感器和智能设备涌入，为 CPS 提供了数据和互联的基础，实现了物理世界和数字世界的深度融合，推动了工业自动化与智能化。但是，IT 网络和 OT 网络的融合越来越紧密，使得以往的物理隔离已不再适用于现代工业场景，这意味着整体攻击面扩大，网络安全风险成倍增长。为了保障工业互联网中 IT 网络和 OT 网络的安全隔离，需要进一步加强逻辑隔离、数据分级和访问控制等机制。另外，IT 网络和 OT 网络在安全需求和服务需求上存在差异。对于 IT 网络，其保密性优先级最高，其次是完整性和可用性，因而更侧重于防范信息层面的数据泄露和应用威胁。对于 OT 网络，其可用性优先级最高，其次是完整性和保密性，因而更侧重于防范设备层面的异常操作和物理安全。

1.2　常用算法和评价指标

统计方法广泛应用于时间序列分析、随机过程建模和预测等领域。差分自回归移动平均（ARIMA, autoregressive integrated moving average）模型[7]是一种常用于时序分析和预测的模型，差分移动平均（ARMA）是其常用于平稳时间序列的特例，仅除去了积分部分。通常，ARIMA 适用于非平稳时间序列，经过差分将序列平稳化，然后使用 ARMA 进行预测。

随着大数据的发展和计算能力的增强，机器学习技术被应用到各种复杂问题中，一部分统计方法为机器学习奠定了基础。例如，隐马尔可夫模型（HMM, hidden Markov model）[8]以马尔可夫为基础引入了隐藏状态和可观察状态的概念，是经典的机器学习算法之一。另外，传统机器学习算法还包括 K-近邻、朴素贝叶斯、决策树、支持向量机、逻辑回归和随机森林等。

为了获取流量中的深层特征表示，提升模型处理大规模和高维度流量数据的能力，卷积神经网络（CNN, conventional neural network）、循环神经网络（RNN, recurrent neural network）和长短期记忆（LSTM, long short-term memory）网络等深度学习算法相继被引入流量分析领域，产生了显著效果[9]。

上述算法被广泛应用于工业互联网流量分析中的各类研究任务，面向指定任务设计的算法和构建的模型，其效果也由不同指标来评估，主要评价指标如表 1 所示。此外，工业场景还常用训练时间和预测时间这类指标，用于评估模型的复杂度和性能。

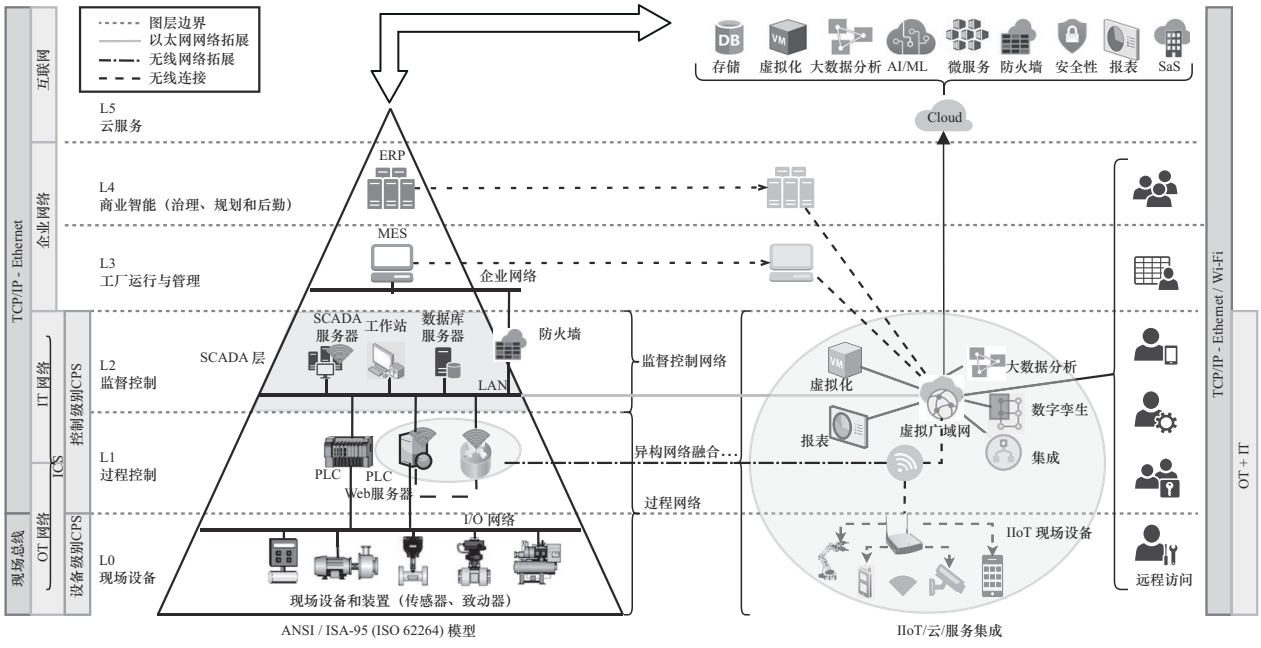

图 2　集成新一代信息技术的工业网络架构

表1 主要评价指标及其解释

主流研究任务	主要评价指标	指标解释
分类任务（入侵检测、加密流量识别、工业资产指纹识别、协议识别与逆向）	准确率	衡量正确分类样本量与总样本量之比，体现模型的分类准确率
	精确率	衡量预测正确的正样本与预测为正样本之比，体现模型的分类精确率（减少误报）
	召回率	衡量所有正样本中有多少被正确预测为正，体现模型的分类覆盖率（减少漏报）
	F1分数	衡量精确度和召回率的调和平均，体现模型准确率和覆盖率的整体能力
回归任务（流量预测）	均方根误差	衡量实际观测值与模型预测值之间的平均偏差的大小，体现模型预测准确率
	平均绝对误差	衡量实际观测值与模型预测值之间的平均平方偏差的大小，体现模型预测准确率
	均方误差	衡量实际观测值与模型预测值之间的平均绝对百分比偏差，体现模型预测准确率
	R平方	衡量模型解释了观测数据方差的比例，用于评估模型预测拟合优度
测试任务（漏洞挖掘）	覆盖率	衡量模型测试覆盖率，体现模型对测试路径的覆盖能力
	检测率	衡量模型漏洞检测率，体现模型对漏洞的检测能力

2 工业互联网流量分析的独特性

工业互联网具备传统互联网的部分特性，但不同于传统互联网。首先，数据来源不同，传统互联网中流量主要来自人类的在线活动，通过客户端与服务端应用交互产生。工业互联网流量则源自各类工业设备、传感器和控制器，覆盖生产制造中各个环节，呈现分散和多源的特点。其次，关注重点和目标对象有所不同。传统互联网流量分析主要关注用户行为，包括网站指纹、社交活动和访问习惯等，重点在于了解用户的兴趣喜好、网络习惯和消费倾向等，以便进行行为预测、内容推荐和广告投放。工业互联网则关注传感器、生产设备等实体相关数据，重点在于监控生产流程、优化生产操作、检测设备异常和保障网络安全等。

流量分析可主要分为流量采集、流量识别、流量预处理、特征工程和模型构建5个步骤[10]。基于这5个步骤，本节对传统互联网和工业互联网的差异进行了分析，如表2所示。

流量采集旨在监测记录、存储网络传输过程产生的流量，保留网络历史数据。传统互联网的网络架构成熟规范，由服务器、个人计算机和移动终端等设备组成。相对而言，工业互联网涉及各种工业设备和网络节点，网络架构复杂异构，数据规模庞大。因此，采集和处理大规模、多节点异构网络的流量数据需要定制化的方案和强大的计算存储资源。

流量识别旨在分析流量数据包特征，将其归类到不同的协议、应用和设备的过程，便于分析网络活动。工业控制系统和工业物联网是工业互联网的重要组成部分，这两部分包括各类工业控制协议和私有协议。工业控制系统中常见协议有Modbus、S7Comm、DNP3和Profinet等[11]，工业物联网中常见协议有MQTT和CoAP等[12]。相较传统网络协议，工业控制协议结构通常不公开，不遵循通用协议规范，其协议特征和通信模式的分析难度相对较大。

流量预处理旨在对流量数据进行降噪、清洗和规范化。工业互联网环境中流量数据易受干扰和噪

表2 工业互联网流量分析的独特性

流量分析阶段	传统互联网	工业互联网	主要区别
流量采集	网络成熟规范，以服务器、个人计算机和移动终端等设备为主，主要为规模庞大的用户数据	网络复杂异构，工业设备种类繁多，节点海量，存在敏感工业生产业务数据	采集方式
流量识别	网络协议规范，报文结构公开	工业控制协议通常不遵循通用规范，报文结构私有	网络协议
流量预处理	数据类型广泛，内容涵盖文本、图片、音频、视频等，加密机制多样，注重噪声数据过滤	数据类型单一，内容多为工业指令和过程数据，加密机制薄弱，注重多源数据整合与时间同步	数据类型处理方式
特征工程	数据结构灵活，偏向用户行为相关特征	数据结构规范，偏向设备状态、生产过程和控制指令相关特征	特征选择
模型构建	用于网络监控、用户画像、推荐算法等	用于设备管理、网络性能优化、故障检测等	应用场景

声影响，例如设备故障、通信错误和网络抖动等。这些干扰和噪声可能导致数据不连贯、不准确，对后续的分析产生误导。因此，数据清洗、噪声过滤、协议聚类和数据结构化等操作在保障数据质量方面具有重要作用。另外，相较传统互联网，工业互联网的多源数据整合和时间对齐较为困难。

特征工程旨在从数据集中提取符合任务需求的有效特征集合，降低特征维度并规范数据格式，提高分析效率。传统互联网流量中的数据结构灵活，内容涵盖文本、图片、音频和视频等多媒体数据，传输方式多样。相较而言，工业互联网流量的数据结构更为规范，涵盖设备状态、生产过程和控制指令相关数据。

模型构建旨在设计和构建流量分析模型，根据任务需求选用合适的算法并基于现有特征分析推理，达到任务目标。工业互联网偏向从流量中提取抽象的设备活动、网络性能状况和潜在安全威胁等，进而达到监控和预测的目的，增强管理能力和安全防护。同时，工业互联网中部署的模型应当具备大规模、多维度的数据处理能力，同时满足可观的实时性、鲁棒性和可解释性[13]。

如图 3 所示，以 Lu 等[14]在工业场景中的入侵检测研究工作为例，其流量分析思路如下：流量采集阶段，收集了来自储水箱和天然气管道 2 个工业控制系统的网络攻击流量；流量识别阶段，筛选出与 SCADA 系统相关联的工业控制协议流量；流量预处理阶段，对流量进行归一化处理和数据集平衡处理，以减少不同维度和类分布不平衡对实验结果的影响；特征工程阶段，从原始数据集中随机选择数据构建多个特征子集，形成不同的训练数据集；模型构建阶段，使用人口极值优化算法对深度信念网络的参数进行优化，构建基于不同特征子集的 PEO-DBN 模型，最后使用集成学习策略将其集成为能够检测多种攻击的 EnPEO-DBN 模型。

3　分类方法

3.1　现有综述分类方法

现有的流量分析综述主要围绕传统互联网展开研究，Abbasi 等[9]总结了深度学习在网络流量监控与分析（NTMA, network traffic monitoring and analysis）中的应用，从算法层面对流量分类、流量预测、故障管理及网络安全 4 类相关工作进行归纳，介绍了 CNN、RNN、LSTM 和 AE 等算法，并

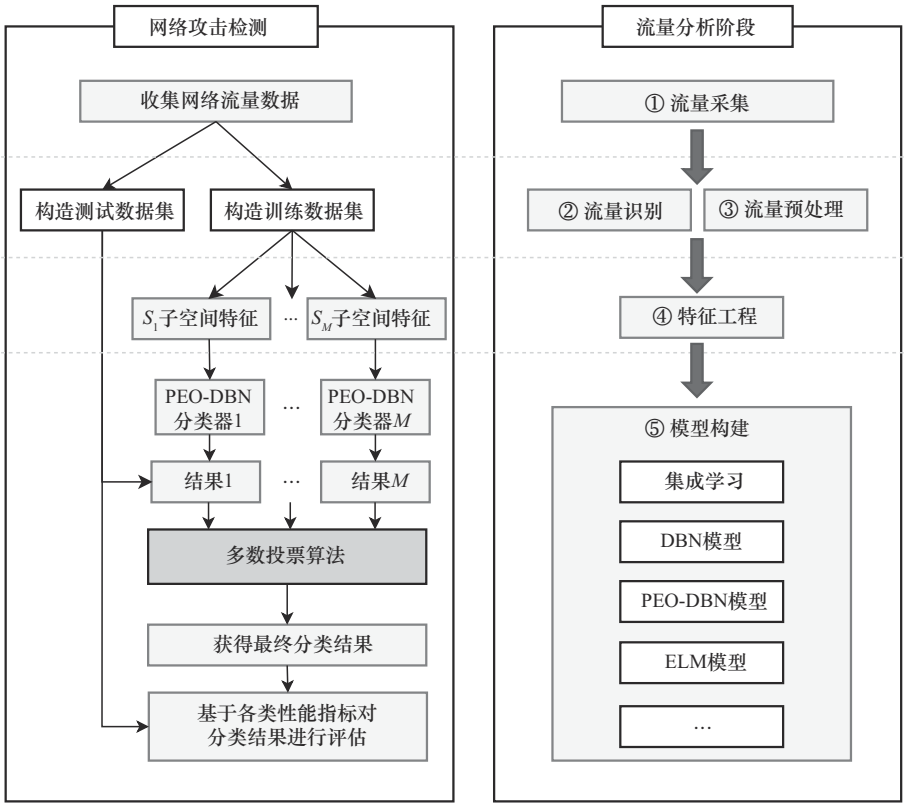

图3　工业场景流量分析案例

分析了其优缺点。随着加密机制和隐私保护的增强，流量分析变得更加困难，基于内容特征的分析方法可能失去效果。针对这种情况，Shen等[15]对机器学习驱动的加密流量分析进行了全面的调研，提取了其工作流程，包括流量收集、流量表示、流量分析方法和性能评估。然后，根据不同的分析目标将现有研究分类为网络资产识别、网络特征描述、隐私泄露检测和攻击检测。

由于工业互联网流量分析的独特性，传统互联网中流量分析侧重的场景和条件，不能完全适用于工业互联网，无法直接移植应用。相较而言，郝志强等[3]提出了工业领域网络流量安全分析技术框架，该框架虽然对工业领域网络流量安全分析关键技术进行了总结，围绕流量采集、识别、处理、分析和检测应用等方面分析了原理和方法，但是该工作的视角较为宏观、主观，对以往研究工作的调研较少，在细节上缺乏支撑。

与现有综述相比，本文的特色在于聚焦工业互联网场景，在保证研究覆盖广度的同时，对现有工作进行了细粒度的调研梳理，充分分析了流量分析技术在工业互联网中发挥的作用和面临的挑战。

3.2　基于研究任务的分类方法

本文基于中国知网、谷歌学术和DBLP等学术搜索引擎，使用“工业互联网”和“网络流量分析”相关的中英文关键词进行搜索，对索引得到的文献的摘要和关键词进行分析梳理，筛选与工业互联网流量分析技术相关的研究工作，从中提取部分词频权重较高关键词生成词云图，如图4所示。为了对以往研究进行系统的归纳和总结，本文提出基于研究任务的分类方法，通过任务性质提炼聚焦的问题。基于关键词分析和可视化词云图，本文提取了工业互联网中研究者常用流量分析驱动的6个主流研究任务，包括流量预测、协议识别与逆向、工业资产指纹识别、入侵检测、加密流量识别和漏洞挖掘。

在工业互联网架构的不同层级中，流量分析的部署需求不同，通过各类研究任务的赋能，实现对整个工业互联网的全面监控和保护。企业网络侧关注的是业务系统和用户行为，以加密流量识别为主；控制网络侧关注的是工业控制网络的事件监控和性能优化，以流量预测为主；现场网络侧关注的是工业控制设备和系统的状态感知，以协议识别与逆向、工业资产指纹识别和漏洞挖掘为主。入侵检测则贯穿架构各个层级，实现不同粒度的威胁防范。流量分析的特征选取、算法设计和评价指标会因研究任务的需求差异而不同。后文对6个主流研究任务的相关工作进行了详细归纳总结，其研究动机和研究内容表明，流量预测、协议识别与逆向、工业资产指纹识别旨在加强网络和资产的管理，更好地服务工业生产，聚焦工业互联网服务质量问题；入侵检测、加密流量识别、漏洞挖掘旨在从被动防御和主动防御双角度提升安全能力，聚焦工业互联网安全问题。

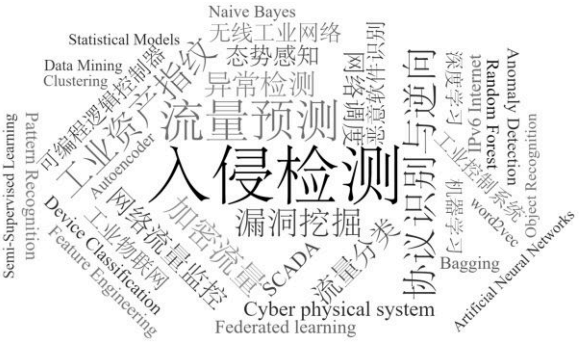

图4　相关文献关键词云图

通过深入分析工业互联网中服务质量与安全能力2个问题的关系，为本文的分类方法提供重要依据。保障工业互联网的服务质量是支撑工业生产的基础，也是提升安全能力的底座。一方面，工业互联网服务质量的提高帮助企业更好地管理和利用设备资源，达到降低成本和提高效率的目的。另一方面，工业互联网安全能力的提升保护工业资产免受潜在威胁和网络攻击的影响，保障系统稳定性和生产连续性，同时也对服务质量起到反馈作用。

综上所述，本文基于研究任务将工业互联网中的流量分析技术分类为面向服务质量提高和面向安全能力提升的2类应用，二者相互交融、共同促进工业互联网发展，如图5所示。

4　面向服务质量提高的流量分析

本节围绕流量预测、协议识别与逆向、工业资产指纹识别这3个研究任务对工业互联网中面向服务质量提高的流量分析展开叙述[16-43]，表3对该部分主要工作进行了简要归纳。

4.1　流量预测

流量预测任务旨在基于网络历史流量对未来流量的数据规模或通信模式进行预测[16]，对算法和

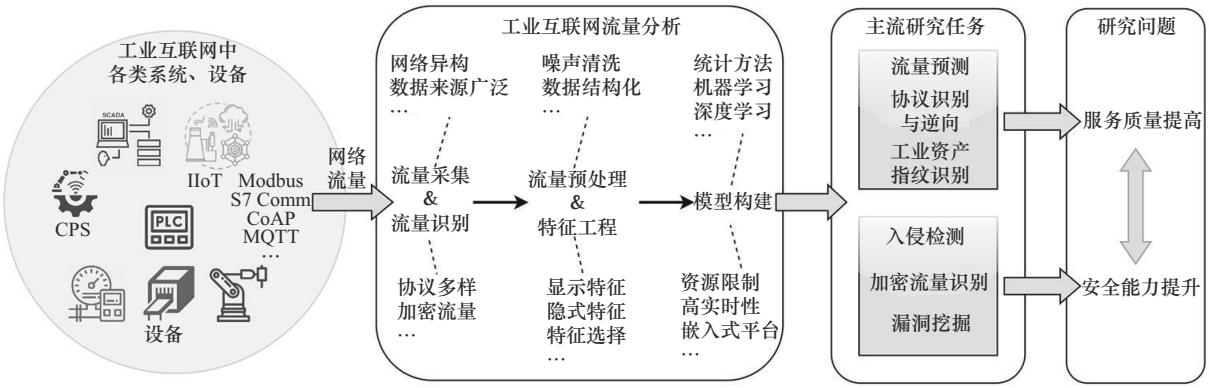

图5　流量分析过程及分类框架

表3　　**面向服务质量提高的流量分析**

文献	数据集	针对协议	针对特征	具体算法	研究任务
文献[17]	非公开	未指定	统计特征、时间特征	广义回归神经网络	流量预测
文献[18]	非公开	未指定	统计特征、时间特征	回声状态网络、岭回归	流量预测
文献[19]	Education	TCP、UDP、RTSP	统计特征、时间特征	eERBM	流量预测
文献[20]	非公开	未指定	统计特征、时空特征	Q-Learning	流量预测
文献[21]	Abilene、Geant	未指定	统计特征、时空特征	MTL、LSTM	流量预测
文献[22]	TON-IoT	未指定	时间特征	CNN、LSTM	流量预测
文献[23]	FedCSIS 2020	未指定	时间特征	图同构网络	流量预测
文献[24]	Abilene、Geant	未指定	统计特征、时空特征	自注意力机制	流量预测
文献[25]	非公开	未指定	统计特征、时间特征	多维深度神经网络	流量预测
文献[27]	非公开	Modbus、IEC 104等	协议特征	Global Voting Expert	协议识别与逆向
文献[28]	非公开	Modbus、Ethernet等	协议特征、消息序列	mean-shift	协议识别与逆向
文献[30]	非公开	S7Comm	协议特征	XGBoost	协议识别与逆向
文献[32]	非公开	Modbus、S7Comm	协议特征、时间特征	LSTM-FCN	协议识别与逆向
文献[29]	非公开	Modbus、IEC 104等	协议特征	基于熵的聚类	协议识别与逆向
文献[35]	多数据集组合	DNP3、Modbus等	协议特征	多序列比对	协议识别与逆向
文献[33]	多数据集组合	S7Comm、BACNET等	时间特征	1DCNN、LSTM等	协议识别与逆向
文献[36]	非公开	未指定	协议特征	侧信道辅助语义分析	协议识别与逆向
文献[30]	SwaT、EPIC	Ethernet、MMS	协议特征	决策树	协议识别与逆向
文献[34]	攻防平台采集	DNP3、Modbus、Fins等	协议特征	Bi-LSTM、AM、CRF	协议识别与逆向
文献[37]	非公开	ISO-Over-TCP	统计特征、时间特征	数据特征向量匹配	工业资产指纹识别
文献[38]	非公开	DNP3	时间特征	前馈人工神经网络	工业资产指纹识别
文献[39]	Shodan引擎	Modbus	协议特征	特征匹配	工业资产指纹识别
文献[40]	非公开	ARP	协议特征	特征匹配	工业资产指纹识别
文献[41]	非公开	DNP3	统计特征、协议特征	SVM、Word2VEC	工业资产指纹识别
文献[42]	非公开	DNP3、MMS、Modbus等	协议特征、载荷特征	SVM、KNN、决策树等	工业资产指纹识别
文献[43]	非公开	Modbus、S7Comm、Enip	时间特征	模式匹配	工业资产指纹识别

数据处理要求高，以便进行实时的网络优化和资源规划，可以及时发现网络异常、防范故障，控制生产成本的同时提高网络质量。目前，工业互联网的流量预测大多针对细分的工业物联网场景，且少有研究者围绕特定协议进行研究。

面对流量的非线性特性，工业互联网特定场景有相应的解决方案。在电力线通信场景，网络流量的动态特性导致其预测困难，给流量和功率调度带来了巨大挑战。Guo等[17]提出了基于广义回归神经网络（GRNN, generalized regression neural network）的电力线表面波通信流量重构方法，通过广义回归神经网络理解时变网络流量特征，实现对高

度动态的网络流量预测，该工作的不足之处是缺乏实际环境验证。Li等[18]提出基于回声状态网络（ESN, echo state network）的流量预测方法，应用ESN学习输出连接权重矩阵，并采用岭回归代替传统线性回归，有效避免了弱条件问题。Sun等[19]则改进性地提出集成神经网络模型，即增强回声状态受限玻尔兹曼机（eERBM, enhanced echo-state restricted Boltzmann machine），该模型可以应对网络流量复杂的统计特性，包括长程相关性和自相似性。

工业物联网具备海量连接、确定性时延和大带宽等特性，对于大规模数据集学习和采样的高消耗问题，Nie等[20]提出基于强化学习（RL, reinforcement learning）的方案，基于残差的自适应字典学习将网络流量投影到另一空间，降低了计算复杂度。在这之后，他们又提出面向工业物联网的多任务学习（MTL, multi-task learning）方法[21]，使用迁移学习和LSTM构建模型，利用链路负载作为附加信息提高预测精度。Wang等[22]则提出结合CNN、Bi-LSTM和注意力机制的方案，通过减少动作空间和增加动作选择约束，使用启发式策略，有效降低复杂度并缩短训练收敛时间，推进了高效网络资源分配。

工业互联网中网络流量存在时变特性，数据驱动的模型能够应对这类挑战。Wang等[23]提出了基于图同构网络的方法，根据数据吞吐量动态预测网络工作负载的未来变化，设计了具有可学习参数的预测序列后处理方法，为流量预测提供一定的可解释性。Liu等[24]提出基于自注意力机制的流量预测模型，该模型复杂度较高，存在过拟合风险，但相比其他模型，拥有更突出的时空特征和长距离依赖关系捕捉能力。

工业互联网的连接设备大多具备服务质量（QoS）需求，还存在网络架构的兼容适配问题。Zhao等[25]提出一种基于多维深度神经网络（MDNN, multi-dimensional deep neural network）的网络切片预测算法，优于ARIMA、RNN和DNN，该算法的流量预测准确率高达98.70%，有效应对异构场景的适应性需求。

4.2 协议识别与逆向

协议识别与逆向旨在理解网络中的非标准私有协议，是分析私有协议的重要手段[26]。由于工业互联网中存在复杂的工业控制系统和工业物联网设备，通常来自多个不同的供应商且具备定制的私有协议。通过流量分析对网络中的协议进行识别和聚类，在黑盒中感知设备和网络情况，可为协议模糊测试提供先验知识。

机器学习在识别复杂模式和关系方面表现出色，适合从已有流量中推理未知协议格式。Wang等[27]提出名为IPART的无监督学习的工业控制协议逆向工具，通过扩展投票专家算法推断工业协议字段边界，由统计方法推导字段类型并将消息分类到子簇中，同时推理各子簇的格式，但该工具无法推断长度可变协议字段。Shim等[28]提出基于均值漂移聚类的协议结构分析方法，该方法由流量采集、消息抽取、按大小聚类、相似度聚类、字段抽取和会话分析6个模块组成，在简洁性和覆盖率方面表现优于Netzob和AutoReEngine。Liu等[29]提出了基于熵的协议逆向方法，可用于定位协议字段边界并通过启发式方法推断类型，在常量字段和功能码推断上具备出色的精确率和召回率。

针对工业控制流量周期性和协议结构固定的特点，Wang等[30]采用渐进式多序列比对将相同载荷长度的流量进行聚类，分离消息序列的可变域和固定域后生成V-gram，通过XGBoost提取特征，有效提取语义信息。Yu等[31]提出基于决策树的方法，能够准确提取协议关键字特征，但该方法属于监督学习，对数据标注要求高，且推理耗时较长。

深度学习在处理高维、复杂数据时优于传统机器学习，特征工程工作量小，高度依赖数据集规模和质量，可以提供优秀的泛化能力。Zhao等[32]提出基于LSTM和全卷积神经网络（FCN, full convolutional neural network）的方法，根据数据包内随时间变化的值判断字段类型，推断协议格式。Zhai等[33]提出基于自编码器（AE, auto encoder）和一维卷积神经网络（1DCNN, one dimensional convolutional neural networks）的方法，通过提取数据编码与空间结构特征之间的相关性识别协议类型，但无法进一步识别载荷中的指令类型。Ning等[34]提出基于注意力机制和Bi-LSTM条件随机场融合模型的协议逆向工具PREIUD，与MSERA相比，平均精度提高了7.4%；与Discoverer相比，平均精度提高了15.4%，实现了复杂度和性能的均衡。

可解释性在工业互联网中较为重要，并非所有研究者都热衷于机器学习和深度学习技术。Ye

等[35]提出基于概率的协议逆向方法NetPlier，通过分析消息序列并引入随机变量推理指定字段是关键字的可能性，计算各个约束条件的联合概率，选择后验概率最大的字段作为关键字，实验表明，该方法优于Netzob和Discoverer。Cai等[36]提出侧信道信息辅助的工业控制协议语义挖掘框架SeMiner，创新地利用图像识别技术观察HMI交互发现潜在语义通道，发掘流量中与语义通道关联的数据包，对工业过程的行为语义进行建模。

4.3　工业资产指纹识别

工业资产指纹识别旨在对工业互联网中的系统和设备进行发现和管理，方便资产管理、统筹维护和风险控制。资产测绘可以帮助建立资产拓扑，追踪设备位置和状态；未知指纹作为管理者发现潜在风险或异常活动的线索。当设备或系统出现故障时，全面的网络资产清单可以帮助管理员迅速定位和解决问题。

工业控制系统及设备存在固定的交互模式，为指纹识别提供了基础。Peng等[37]通过分析流量数据包特征及到达时间间隔区分人机界面和可编程逻辑控制器之间的交互会话或事务模式，进而创建相应工业控制系统的行为指纹，该方法局限于相对表层的流量特征。Formby等[38]提出基于前馈神经网络的2种指纹识别方法，一种对跨层响应时间测量进而识别工业控制设备，但它只支持具备“读取”和“响应”消息的工业控制协议；另一种则通过分析工业控制设备运行时间的差异来识别设备，这受限于高分辨率和低误差的操作时间特征。

工业控制协议结构化且种类繁多，通过协议的特定字段值可以获取设备指纹信息，但这种方法通常需要高质量的先验知识。针对Modbus协议无加密和认证的特性，Keliris等[39]提出了一种利用Modbus协议特征进行指纹识别的方法，通过观察Modbus针对不同供应商的实现差异，直接读取寄存器值对知识库进行匹配。Niedermaier等[40]提出更简易的基于MAC地址的被动监控技术，构建已知MAC地址到设备的非完整映射，通过ARP协议特征可以获得设备和供应商信息，但是无法获取固件及版本，且受限于MAC地址随机化防护机制。

随着算力增长，聚类和机器学习的应用使得工业资产指纹识别面临的挑战逐渐简化。Chakraborty等[41]使用SVM针对DNP3协议实现了对工业控制与非工业控制设备的分类以及工业控制设备类型的识别。李长连等[42]提取流量数据包头部特征和负载特征，使用聚类算法对特征进行聚合，将得到的结果作为设备指纹。为减弱资源约束和性能限制的影响，Tao等[43]提出基于模式匹配的设备流量指纹库构建方案和设备识别方法，应用对角跳跃和最优切片算法，提高识别精度并降低时间开销。

在工业互联网中，流量预测能够优化网络资源的调度和分配能力，工业资产指纹识别和协议识别与逆向提升了设备资产和网络环境的感知能力。然而，目前的流量预测工作主要集中在工业物联网环境，且对工业控制网络协议的针对性还有待提高，同时也存在实时性方面的不足。相比之下，协议识别与逆向工程更关注于推断流量中网络协议的格式和字段语义，而不太关注深层次载荷特征。随着工业控制协议的机密性和安全性的提升，工业资产指纹识别也逐渐从显式特征匹配迭代为隐式特征挖掘，虽然提高了识别的准确性，但也增加了实现的难度。

5　面向安全能力提升的流量分析

本节围绕入侵检测、加密流量识别、漏洞挖掘这3个研究任务对工业互联网中面向安全能力提升的流量分析[44-87]展开叙述，表4对该部分主要工作进行了简要归纳。

5.1　入侵检测

入侵检测任务旨在保护系统和设备免受未经授权的访问、恶意攻击或异常行为威胁，提高工业互联网的安全性。入侵检测技术传统上分为异常入侵检测和误用入侵检测[44]，根据数据来源也可以分为基于主机的入侵检测和基于网络流量的入侵检测[13]，本文主要关注基于网络流量的工业互联网入侵检测研究工作。20世纪80年代，Anderson[45]提出入侵检测这一概念并对入侵行为进行了简单划分。工业互联网中的入侵检测相比其他安全性研究也发展更早，如图1所示。

在早期研究中，基于网络流量的工业互联网入侵检测大多使用统计方法或其他传统模型。文献[46]提出了基于自回归模型的方法，提取流量中过程变量和时间序列并持续跟踪过程变量更新，推导出特定变量的预测模型，有效检测针对过程控制的直接和间接攻击，但是难以检测动态变化的配置参数。Caselli等[47]提出基于序列感知的入侵检测系统，通

表4 面向安全能力提升的流量分析

文献	数据集	针对协议	特征	具体算法	研究任务
文献[46]	非公开	Modbus	过程变量、时间序列	自回归模型	入侵检测
文献[47]	非公开	Modbus	过程变量、事件序列	离散时间马尔可夫	入侵检测
文献[52]	Conpot蜜罐	Modbus	时空特征	REPTree、NB、LR等	入侵检测
文献[54]	非公开	Modbus	时间特征	神经网络、SVM等	入侵检测
文献[57]	天然气管道	未指定	时间特征	GRU	入侵检测
文献[65]	Electra等	Modbus、S7Comm等	统计特征、内容特征	TrAdaBoost、LSTM	入侵检测
文献[58]	非公开	Modbus、Fins、CIP等	协议特征、时间特征	CNN、LSTM	入侵检测
文献[64]	UNSW-NB15等	未指定	统计特征	LocalGRU	入侵检测
文献[14]	天然气管道等	未指定	未指定	PEO-DBN、LSTM	入侵检测
文献[69]	Electra	S7Comm、Modbus	协议特征	SIGSM、DNN等	入侵检测
文献[66]	SwaT、S7 Rodo	S7Comm	载荷特征	CNN	入侵检测
文献[67]	Electra	Modbus、S7Comm	统计特征、协议特征	XGBoost	入侵检测
文献[68]	天然气管道	Modbus	统计特征、载荷特征	CNN、Bi-LSTM、GRU	入侵检测
文献[48]	BUT、RTS等	IEC 104、MMS	窗口会话相对频率	确定性概率自动机	入侵检测
文献[71]	Tor、VPN等	TCP、UDP	统计特征、时间特征	CNN、RNN、Bi-LSTM	加密流量识别
文献[72]	非公开	OpenFlow、SSL	统计特征、时间特征	RNN、LSTM	加密流量识别
文献[73]	MCFP、CTU等	SSL	统计特征	改进自适应随机森林	加密流量识别
文献[76]	非公开	S7commPlus-TLS	统计特征、时间特征	聚类	加密流量识别
文献[74]	Tor、VPN等	TCP、UDP	时间序列、载荷特征	CNN、知识蒸馏	加密流量识别
文献[75]	VPN、CIC等	TCP、UDP	统计特征	Transformer、RF	加密流量识别
文献[77]	ILC 171、150	Modbus、Profinet等	协议特征	Ratcliff模式匹配	漏洞挖掘
文献[81]	非公开	Modbus	协议特征	GAN、LSTM等	漏洞挖掘
文献[82]	非公开	EtherCAT	协议特征	LSTM	漏洞挖掘
文献[80]	天然气管道等	Modbus	协议特征	RNN	漏洞挖掘
文献[83]	非公开	Modbus、EtherCAT	协议特征	Bi-LSTM、DCGAN	漏洞挖掘
文献[84]	2017 QUT_DNP3等	DNP3	协议特征	LSTM、CovGAN	漏洞挖掘
文献[87]	开源协议	IEC 104、CoAP、MQTT等	协议特征	SSTG引导包序列生成	漏洞挖掘

过离散时间马尔可夫链对流量中Modbus事件序列进行统计分析，可对类似“震网”这类不属于异常行为却能对基础设施造成破坏的操作事件进行识别，但该模型对训练集的要求较为严格。Havlena等[48]提出了一种基于确定性概率自动机的方法，对流量指定时间窗口中会话的相对频率进行语义分析，实现了较低的误报率，具备识别异常并且对其持续跟踪的能力，但该方法无法识别拒绝服务攻击和载荷数据篡改攻击。

为充分利用先验知识和已有数据，研究者尝试将多维度的数据结合并通过多个模块实现入侵检测。Zhou等[49]提出了一种基于多模块的方法，其中基于知识驱动CAD模块作为流量报文分类器，辅助区分实际攻击和故障，该方法具备良好的协议兼容性，但是实验只在模拟环境下得到验证。Zhang等[50]的主要创新之处在于将流量数据、主机系统数据和工业控制处理数据集成，提供多层网络检测，但对于能够规避传统监控方法的攻击检测仍存在局限性，其中的流量分析只起到辅助决策作用。Liu等[51]提出一种主被动结合的方法：被动检测引擎通过监测网络流量，分析非法地址操作、非法值违规等异常行为；主动检测引擎保持与PLC通信，以低中断和轮询的方式映射PLC存储空间，实时监控异常情况。该方法依赖PLC程序的分析，且无法检测使用加密私有协议的系统。

随着网络的复杂化和攻击的多样化，研究者试图用机器学习驱动入侵检测任务。Ponomarev等[52]通过流量遥测数据来识别不同设备间的交互，经比较发现减少错误修剪树（RepTree, reduce error by pruning tree）模型构建速度最快且精度较高。Terai等[53]提出基于SVM的方法，因其依赖于指定渗透场景训练数据的流量特征，欠缺检测未知攻击的能

力。Shen 等[54]通过分析层间响应时间来得到设备的运行状态，依据这种设备指纹进行实时异常监测，入侵检测和分类准确率超过 99%。Yun 等[55]使用最近邻搜索学习工业控制系统正常活动对应的流量模式，借此检测异常行为，但该方法会对新任务产生的流量造成误报。后来，研究者探索了使用多个模型集成决策的策略，例如，Hassan 等[56]提出基于随机子空间（RS, random subspace）与随机树（RT, random tree）组合的集成学习模型，当特征数量非常少时，RSRT 模型在选择最优随机特征子集方面存在局限性。

深度学习相比传统机器学习来说能够有效处理高维、时间相关的流量数据的问题。Chen 等[57]提出了基于 GRU 的入侵检测方法，其精度高于 RNN，在分析时间步长大的时间序列时，RNN 存在梯度消失或梯度爆炸问题，导致其时间记忆能力有限。Yang 等[58]提出基于寄存器状态指纹的主被动双模块入侵检测方法 iFinger，基于寄存器状态数据，使用确定性状态自动机生成设备指纹被动检测入侵。同时通过 CNN 和 LSTM 构成的流量预测方法判断工业控制系统流量上界，根据结果启发式发包探测寄存器状态并解析响应结果，该方法具有低时延、高召回率等优势，但对不改变寄存器状态的攻击检测存在局限性。Wang 等[59]首次提出基于一维卷积自编码器和支持向量数据描述（SVDD, support vector data description）的方法，与目前的基线方法相比在多数指标上得到了改进。Wang 等[60]提出基于 RNN 的电力工业控制系统异常检测方法，基于应用层载荷数据进行分析，与传统的机器算法相比有显著的优势。

为了减少特征工程的工作量，提高特征选择和的提取效率，Lu 等[14]提出基于群体极值优化（PEO, population extremum optimization）和深度信念网络（DBN, deep belief network）的方法，通过集成策略检测攻击，该方法虽然效果较好，但是 PEO 适应度评估过程耗时较多，需要考虑使用更合适的方法加速评估过程。Zhang 等[61]也采用 DBN 进行特征提取，通过 LSTM 进行特征学习解决传统方法特征表示能力弱的问题，但模型过于复杂，可能会存在过拟合的风险。Deng 等[62]使用 CNN 自主学习特征，减少了特征工程的工作量，解决了机器学习高度依赖的特征选择问题。该方法准确率高达 99.88%，比决策树、Adaboost 和 NB 准确率都高。

针对工业物联网中的雾环境特性，Zhou 等[63]提出一种基于雾计算的 DDoS 缓解处理机制，解决了设备计算能力的限制和实时响应需求，但对基于协议漏洞等类型的 DDoS 攻击的防御存在局限性。Abdel-Basset 等[64]提出了名为 Deep-IFS 的面向云雾端环境的方法，基于局部门控循环单元（LocalGRU, local gated recurrent unit），联合学习空间特征和时间特征，使用注意力机制筛选重要特征。该方法整体通过 worker 雾节点进行分布式计算，基于 IoT 端点的数据完成模型训练，将检测结果传输到云端，能够有效处理大型 IIoT 流量数据，具有较好的鲁棒性和可扩展性，但是存在数据集不平衡问题。

工业互联网入侵检测面临数据匮乏的问题，且不同协议的异常流量具有不同特征分布，传统机器学习数据满足相同分布的假设失效。Li 等[65]提出基于跨领域知识迁移的方法，通过 TrAdaBoost-LSTM 获得比单纯 LSTM 更低的错误率。Gu 等[66]将重构的 CNN 和 CB_SMOTE 数据扩展算法应用于入侵检测系统 DEIDS，采用了分类激活图结构，深度挖掘流量的潜在特征，解决了攻击样本数据不足的问题。Jiang 等[67]提出的方案集成了 DAE、SMOTE、T-Link 和 XGBoost，使用 DAE 降低数据噪声提取数据的核心特征，采用 SMOTE 和 T-Link 机制对数据进行增强，解决数据不平衡问题，最后通过 XGBoost 来避免过拟合。Chen 等[68]首次对工业控制通信网络中的不平衡数据进行分析，提出信息增强的对抗域适应方法，在数据不平衡的条件下训练跨域模型并保证较高精度。

面对机器学习和深度学习在入侵检测中的高速发展和应用，部分研究者试图对抗或者逃避检测。Gomez 等[69]提出一种对抗样本攻击方法 SIGSM，用于对抗基于 DNN 的入侵检测模型，相比 FGSM 和 BIM，SIGSM 能够生成中间网络设备理解的对抗样本流量，绕过入侵检测系统的同时保障流量对抗样本的可用性。

考虑到隐私保护和数据安全等需求，Gao 等[70]首次提出适用于工业控制系统的隐私保护型异常检测平台，有效地应对国家标准和隐私保护问题。该平台基于梯度提升决策树和分类回归树，采用多方安全计算，在敏感数据不可见的情况下达到入侵检测的目的。

5.2 加密流量识别

加密流量识别旨在监测和分析通过加密协议传输的数据流，以保障合规性和安全性。由于早年工业控制系统的网络安全要求欠缺，现今对于工业敏感数据的隐私保护要求增强，同时在计算资源的增加和深度学习的发展应用大背景下，工业互联网中的加密流量识别及分类研究才得以发展。

工业物联网中，针对传统流量分类难以区分加密数据流、无法对加密流量设计有效特征等问题，Lin等[71]提出基于TSCRNN的方案，实现对网络流量时空特征的自动提取，该方案使用CNN提取流量的抽象特征，通过堆栈Bi-LSTM和低维特征图学习时间特征，有效识别加密和匿名流量，但同样存在数据不平衡和高度依赖于标注数据等问题。Luksha等[72]提出了过滤加密流量的技术，针对不同类型流量形成训练样本，使用随机森林进行加密流量，具备一定效果。Niu等[73]提出改进的自适应随机森林算法，用于检测加密流量中的恶意软件行为，检测到新样本后实现自适应更新，以准确检测加密、漂移和不平衡流量中的恶意软件家族，该方法优于原始自适应随机森林、D2LAD和FS-Net等代表性方法。

加密流量的实时分析对模型性能要求较高，Dai等[74]提出名为CMFTC的跨模态融合模型，通过对模型进行知识蒸馏，调整蒸馏算法提高模型推理速度而不影响性能，该模型性能优于SOTA模型，且参数和性能消耗显著减少，在真实物联网设备上具有较好的推理速度。Zhao等[75]使用对抗性训练和深度聚类设计了一种无监督的方案，在预训练阶段使用自动编码器降低特征维度，省去了数据标注工作，降低了特征工程成本，具有更好的鲁棒性。

相对而言，针对传统工业控制系统协议的加密流量分析工作较少。Zhu等[76]首次对S7commPlus-TLS协议流量的敏感行为进行分析，使用聚类分离周期性流量，提取加密流量中潜在的敏感行为指纹，在性能和资源消耗上有一定优势。

5.3 漏洞挖掘

漏洞挖掘旨在发现操作系统、应用程序、网络协议中潜在的安全漏洞，通过主动发现漏洞并对其修复和防范，降低网络系统的脆弱性。由于工业互联网中各类系统应用的源码和固件获取难度高，因此漏洞挖掘主要聚焦于种类繁多的闭源协议，在协议识别与逆向的帮助下理解协议并且构造有效测试用例，实现针对性模糊测试。流量分析在此起到提供先验知识的作用，区别于其他5个研究任务。

根据为模糊测试创建输入的方式，可以大致将其分为2类。第一类是基于生成的模糊测试器从零开始创建输入，需要了解协议数据字段；第二类是基于变异的模糊测试器利用有效样本生成畸形样本输入。Niedermairer等[77]对专有协议的消息进行理解，通过Ratcliff-Obershelp模式识别算法对流量进行分析，进而确定集成开发环境和PLC之间的握手方式，为后续的模糊测试步骤提供“先验知识”，但是该方案不适于具有会话管理和加密措施的复杂协议。Fang等[78]提出可移植、模块化的模糊测试框架ICS3Fuzzer，需要人工获取协议先验知识，通过差异性分析识别字段和约束，识别协议状态并过滤无效值，建立执行路径和对应输入的State-Book映射，分析其通信范式并仿真PLC设备以进行漏洞挖掘。由于不同协议的结构具有差异性，部分方案只能针对指定协议进行模糊测试，扩展性较差，且测试用例效率较低。Ramos等[79]提出针对MQTT的模糊测试框架，流量嗅探及分析模块负责监听网络流量，对指定流量过滤和处理并生成模板，模糊测试器根据模板来确定数据包和字段，但该方法局限于MQTT协议。Lai等[80]提出基于Anti-Sample的工业控制协议漏洞挖掘测试用例生成模型，使用RNN学习协议数据单元的语义，并通过Softmax函数表示数值概率分布。该方法能够捕获3种类型的Modbus/TCP协议漏洞，相比Kitty fuzzer和Peach fuzzer具有更高的测试用例接收率和漏洞挖掘效率。

针对无状态协议和有状态协议的差异，模糊测试方案的设计有所不同。对于无状态协议，Hu等[81]结合GAN和SeqGAN，基于流量中的协议消息进行语法的学习推断，计算基础分布函数，进一步设计了自动化模糊测试框架GANFuzz，可以应用于公开或私有工业协议，测试效果优于多个已有模糊测试工具。对于有状态协议，Zhao等[82]提出模糊测试框架SeqFuzzer，不受协议种类限制，使用LSTM自动学习流量中EtherCAT协议格式，在模糊测试过程中成功检测出包括数据包注入攻击、中间人攻击和MAC地址欺骗等多种安全漏洞。

生成对抗网络在工业控制协议模糊测试中比较受欢迎，部分研究者对其进行改进以更适用于模糊测试任务。Lv 等[83]提出了基于改进深度卷积生成对抗网络（DCGAN, deep convolutional generative adversarial network）的自动化模糊测试方法，使用帧长聚类、K-means 等方法对流量中的协议消息进行分类。通过理解协议消息之间的差异，提高测试的深度。通过 Bi-LSTM 学习测试用例发送后的响应数据，得到特征优化数据变异和扩展策略，进而提供新的训练数据指导 DCGAN 生成更可靠的测试用例。Yu 等[84]提出名为 CGFuzzer 的面向 DNP3 的模糊测试框架，将覆盖引导生成对抗网络（Cov-GAN, coverage guided generative adversarial network）和树置信上限与序列生成对抗网络（Seq-GAN, sequential generative adversarial network）结合，以提高通过率和代码覆盖率，通过分析响应流量中的异常确定测试状态，实验表明 CGFuzzer 比 GANFuzz 和 SeqFuzzer 更具优势。

漏洞挖掘任务中流量分析发挥有限的作用，且高度依赖于专家经验。Bytes 等[85]提出适用于控制应用程序和工业运行时的模糊测试框架，通过网络流量自动化命令发现和状态码提取支撑模糊测试开展，能在黑盒设备上运行时执行上下文中进行控制应用程序模糊测试，但该方法存在性能评估限制、代码覆盖率受限等问题。Che 等[86]提出了基于互信息率的协议结构分析算法和基于遗传算法的模糊测试用例生成方法。通过 Smith-Waterman 算法评估测试用例与正常数据包的相似度，然后计算测试用例的适应度，筛选符合私有协议结构且具有较高适配度的测试用例，但该方法对协议交互规则的推断能力有限，针对协议数据段的分析方法还有待优化。Luo 等[87]提出名为 Bleem 的面向数据包序列的黑盒协议模糊测试器。相比现有的基于协议模型或现有报文产生测试用例的方式，该方法观察协议逻辑的方式在服务器和客户端内部实现，可以根据观察到的双方交互流量，生成协议逻辑感知的数据包序列，尽可能满足报文间字段参数依赖性，提升测试效率，实验表明，该方法相比 Peach、BooFuzz 和 Snipuzz 等协议模糊测试工具具有更高的代码覆盖率，最高提高 174.93%。

在工业互联网中，入侵检测能够监测网络流量，识别和响应潜在的安全威胁。加密流量识别技术确保即使在数据加密的情况下，也能对流量进行有效分析。漏洞挖掘则主动寻找并修复系统和设备中的安全漏洞，防止攻击者利用未知漏洞入侵。与此同时，传统机器学习和深度学习在工业互联网流量分析中得到了广泛应用，不仅提高了模型的准确性和实时性，还能够处理海量数据，上述技术的结合进一步提升了工业互联网的整体安全性。但是，工业互联网架构的不断更新迭代，使其面临着更广泛的数据来源和更多样化的攻击手段，这对流量分析的性能和数据集的质量提出了更高的要求。

6　挑战与机遇

本节对现有研究工作中存在的挑战及其应对方案进行了讨论，同时展望了未来流量分析在工业互联网中的研究方向。

1) 流量数据集的构建

现有工作的数据集大多源自早年公开发布或工业仿真平台采集，以往网络协议和攻击种类较少，网络结构简单且流量规模小，甚至部分研究中采用的数据集并非来自工业网络。工业互联网架构的迭代，使得大多按以往网络架构构建的模型过时。因此，需要对工业互联网的流量采集方案进行定制化，构建高标准、多样性丰富的数据集。

2) 模型可解释性和鲁棒性

深度学习存在不透明性，研究者关注应用效果而忽略探索其能力产生的本质[88]。工业互联网中模型的可解释性和鲁棒性至关重要，只有模型的预测和决策足够透明可靠，管理者才能理解模型的决策过程以对未知风险进行管控，保障工业生产持续稳定。

3) 对抗机器学习及其防御

机器学习广泛应用后，研究者开始关注对抗机器学习，利用模型的脆弱点逃避检测，使得模型能力削弱或失效[89]。例如，通过生成对抗网络构建嵌入微小扰动的流量样本，使得基于网络流量的入侵检测模型发生误判。工业控制协议不同于传统网络协议，一旦改动可能失去可用性，具备一定的局限性。文献[69]在保证对抗样本有效的同时逃避模型检测，攻击流量成功抵达工业设备并达到破坏效果，可见该类技术及其防御机制仍具备一定研究前景。

4) 资源约束和性能限制

工业互联网中设备具有性能低、数量多和分布广等特点，组成了一个异构资源约束网络[90]。深度学习的广泛应用，导致算力爆发性增长。现有工作在实用性上有所欠缺，尤其是流量预测、入侵检测等实时性要求高的任务，缺乏在资源约束环境的可用性实验评估。联邦学习、迁移学习和知识蒸馏等技术可以缓解资源受限的情况，降低存储和计算需求，适用于工业互联网的性能受限场景。

5) 大模型技术的应用

GPT、Claude 和 LLaMA 等大模型的发展与应用[91]，展现出令人震惊的理解和推理能力，尤其是其独有的“涌现”能力。未来，可以考虑在工业互联网中存在语义理解、上下文感知、多模态和持续学习的任务应用大模型提示和微调技术，实现流量分析的自动化、智能化和定制化。例如，通过大模型生成工业资产指纹识别规则、进行协议格式分析和语义推断、指导测试用例生成、协助工业协议模糊测试、发现潜在漏洞和缺陷；通过大模型分析工业互联网中海量网络事件和系统日志，识别异常行为和未知威胁。

7 结束语

工业互联网的发展为人们的生活带来巨大变革，同时也面临险峻的安全形势。流量分析作为网络监控和异常检测的关键技术，在工业互联网生态中发挥着重要作用。本文从流量分析的 5 个主要步骤出发，剖析了工业互联网和传统互联网的差异，阐述了工业互联网流量分析的独特性。对流量分析在工业互联网中的应用工作进行了系统梳理，归纳出 6 个主流研究任务，并将其分为提高服务质量和提升安全能力 2 个大类。最后，分析了流量分析在工业互联网中面临的挑战与机遇，并展望了未来研究方向。

参考文献：

[1] 张恒升, 赵锋, 刘东坡. 工业互联网 总体网络架构: GB/T 42021-2022[S]. 2022.
ZHANG H S, ZHAO F, LIU D B. Industrial Internet overall network architecture: GB/T 42021-2022[S]. 2022.

[2] 张云明. 数实融合–数智赋能 高质量推进新型工业化[R]. 2023.
ZHANG Y M. Digital real integration-digital intelligence empowers high quality promotion of new industrialization[R]. 2023.

[3] 郝志强, 刘冬, 王冲华. 工业领域网络流量安全分析关键技术研究[J]. 工业信息安全, 2022(3): 27-35.
HAO Z Q, LIU D, WANG C H. Research on key techniques of industrial network traffic security analysis[J]. Industry Information Security, 2022(3): 27-35.

[4] SCHOLTEN B. The road to integration: a guide to applying the I-SA-95 standard in manufacturing[M]. Durham: ISA, 2007.

[5] Functional architecture of industrial Internet system for industrial automation applications: IEC PAS 63441[S]. 2022.

[6] SVERKO M, GRBAC T G, MIKUC M. SCADA systems with focus on continuous manufacturing and steel industry: a survey on architectures, standards, challenges and industry 5.0[J]. IEEE Access, 2022, 10: 109395-109430.

[7] MOAYEDI H Z, MASNADI-SHIRAZI M A. ARIMA model for network traffic prediction and anomaly detection[C]//Proceedings of the 2008 International Symposium on Information Technology. Piscataway: IEEE Press, 2008: 1-6.

[8] EDDY S R. What is a hidden Markov model?[J]. Nature Biotechnology, 2004, 22(10): 1315-1316.

[9] ABBASI M, SHAHRAKI A, TAHERKORDI A. Deep learning for network traffic monitoring and analysis (NTMA): a survey[J]. Computer Communications, 2021, 170: 19-41.

[10] ALQUDAH N, YASEEN Q. Machine learning for traffic analysis: a review[J]. Procedia Computer Science, 2020, 170: 911-916.

[11] KOAY A M Y, KO R K L, HETTEMA H, et al. Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges[J]. Journal of Intelligent Information Systems, 2023, 60(2): 377-405.

[12] FIGUEROA-LORENZO S, AÑORGA J, ARRIZABALAGA S. A survey of IIoT protocols: a measure of vulnerability risk analysis based on CVSS[J]. ACM Computing Surveys, 2020, 53: 1-53.

[13] 刘奇旭, 陈艳辉, 尼杰硕, 等. 基于机器学习的工业互联网入侵检测综述[J]. 计算机研究与发展, 2022, 59(5): 994-1014.
LIU Q X, CHEN Y H, NI J S, et al. Survey on machine learning-based anomaly detection for industrial Internet[J]. Journal of Computer Research and Development, 2022, 59(5): 994-1014.

[14] LU K D, ZENG G Q, LUO X Z, et al. Evolutionary deep belief network for cyber-attack detection in industrial automation and control system[J]. IEEE Transactions on Industrial Informatics, 2021, 17(11): 7618-7627.

[15] SHEN M, YE K, LIU X T, et al. Machine learning-powered encrypted network traffic analysis: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 791-824.

[16] 康梦轩, 宋俊平, 范鹏飞, 等. 基于深度学习的网络流量预测研究综述[J]. 计算机工程与应用, 2021, 57(10): 1-9.
KANG M X, SONG J P, FAN P F, et al. Survey of network traffic forecast based on deep learning[J]. Computer Engineering and Applications, 2021, 57(10): 1-9.

[17] GUO Y G, CONG H Y, GUO K Y, et al. A GRNN-based traffic reconstruction algorithm to surface wave-based power line communications[C]// Proceedings of the 2019 IEEE International Conference on Industrial Internet (ICII). Piscataway: IEEE Press, 2019: 138-142.

[18] LI Q M, WANG S, LIU Y Z, et al. Traffic self-similarity analysis and application of industrial Internet[J]. Wireless Networks, 2020, 30(5): 3571-3585.

[19] SUN X C, MA S H, LI Y Q, et al. Enhanced echo-state restricted Boltzmann machines for network traffic prediction[J]. IEEE Internet of Things Journal, 2020, 7(2): 1287-1297.

[20] NIE L S, NING Z L, OBAIDAT M S, et al. A reinforcement learning-

based network traffic prediction mechanism in intelligent Internet of Things[J]. IEEE Transactions on Industrial Informatics, 2021, 17(3): 2169-2180.

[21] NIE L S, WANG X J, WANG S P, et al. Network traffic prediction in industrial Internet of things backbone networks: a multitask learning mechanism[J]. IEEE Transactions on Industrial Informatics, 2021, 17(10): 7123-7132.

[22] WANG H, BAI Y X, XIE X. Dynamic resource allocation for 5G-enabled industrial Internet of Things system with delay tolerance[C]// Proceedings of the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Piscataway: IEEE Press, 2022: 1-6.

[23] WANG R R, ZHANG Y, PENG L M, et al. Time-varying-aware network traffic prediction via deep learning in IIoT[J]. IEEE Transactions on Industrial Informatics, 2022, 18(11): 8129-8137.

[24] LIU X T, HUANG C H, ASHRAF M W A, et al. Spatiotemporal self-attention-based network traffic prediction in IIoT[J]. Wireless Communications and Mobile Computing, 2023(1): 8331642..

[25] ZHAO J H, PENG G J. Industrial Internet network slice prediction algorithm based on multidimensional and deep neural networks[C]//Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition. New York: ACM Press, 2022: 1045-1051.

[26] 李峻辰, 程光, 杨刚芹 . 基于网络流量的私有协议逆向技术综述[J]. 计算机研究与发展, 2023, 60(1): 167-190.
LI J C, CHENG G, YANG G Q. Private protocol reverse engineering based on network traffic: a survey[J]. Journal of Computer Research and Development, 2023, 60(1): 167-190.

[27] WANG X W, LV K Z, LI B. IPART: an automatic protocol reverse engineering tool based on global voting expert for industrial protocols[J]. International Journal of Parallel, Emergent and Distributed Systems, 2020, 35(3): 376-395.

[28] SHIM K S, GOO Y H, LEE M S, et al. Clustering method in protocol reverse engineering for industrial protocols[J]. International Journal of Network Management, 2020, 30(6): e2126.

[29] LIU O Y, ZHENG B, SUN W, et al. A data-driven approach for reverse engineering electric power protocols[J]. Journal of Signal Processing Systems, 2021, 93(7): 769-777.

[30] WANG R, SHI Y J, DING J K. Reverse engineering of industrial control protocol by XGBoost with V-gram[C]//Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2020: 172-176.

[31] YU C H, ZHANG Z, GAO M. An ICS traffic classification based on industrial control protocol keyword feature extraction algorithm[J]. Applied Sciences, 2022, 12(21): 11193.

[32] ZHAO R, LIU Z H. Analysis of private industrial control protocol format based on LSTM-FCN model[C]//Proceedings of the 2020 International Conference on Aviation Safety and Information Technology. New York: ACM Press, 2020: 330-335.

[33] ZHAI L, ZHENG Q H, ZHANG X, et al. Identification of private ICS protocols based on raw traffic[J]. Symmetry, 2021, 13(9): 1743.

[34] NING B W, ZONG X J, HE K, et al. PREIUD: an industrial control protocols reverse engineering tool based on unsupervised learning and deep neural network methods[J]. Symmetry, 2023, 15(3): 706.

[35] YE Y P, ZHANG Z, WANG F, et al. NetPlier: probabilistic network protocol reverse engineering from message traces[C]//Proceedings of the 2021 Network and Distributed System Security Symposium. Reston: Internet Society, 2021: 1-18.

[36] CAI J, ZHONG W J, LUO J Z. SeMiner: side-information-based semantics miner for proprietary industrial control protocols[J]. IEEE Internet of Things Journal, 2022, 9(22): 22796-22810.

[37] PENG Y, XIANG C, GAO H H, et al. Industrial control system fingerprinting and anomaly detection[C]//IFIP Advances in Information and Communication Technology. Berlin: Springer, 2015: 73-85.

[38] FORMBY D, SRINIVASAN P, LEONARD A, et al. Who's in control of your control system? device fingerprinting for cyber-physical systems[C]//Proceedings of the 2016 Network and Distributed System Security Symposium. Reston: Internet Society, 2016: 1-15.

[39] KELIRIS A, MANIATAKOS M. Remote field device fingerprinting using device-specific modbus information[C]//Proceedings of the 2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS). Piscataway: IEEE Press, 2016: 1-4.

[40] NIEDERMAIER M, HANKA T, PLAGA S, et al. Efficient passive ICS device discovery and identification by MAC address correlation[J]. arXiv Preprint, arXiv: 1904.04271, 2019.

[41] CHAKRABORTY I, KELLEY B M, GALLAGHER B. Industrial control system device classification using network traffic features and neural network embeddings[J]. Array, 2021, 12: 100081.

[42] 李长连, 余思阳, 程驰 . 基于设备流量行为的工业物联网指纹识别技术[J]. 工业信息安全, 2022(10): 47-56.
LI C L, YU S Y, CHENG C. Industrial Internet of things fingerprint identification technology based on traffic behavior drift[J]. Industry Information Security, 2022(10): 47-56.

[43] TAO J, YUAN X, ZHANG S Z, et al. Development of fingerprint identification based on device flow in industrial control system[J]. Applied Sciences, 2023, 13(2): 731.

[44] 卿斯汉, 蒋建春, 马恒太, 等 . 入侵检测技术研究综述[J]. 通信学报, 2004, 25(7): 19-29.
QING S H, JIANG J C, MA H T, et al. Research on intrusion detection techniques: a survey[J]. Journal on Communications, 2004, 25(7): 19-29.

[45] ANDERSON J P. Computer security threat monitoring and surveillance[R]. 1980.

[46] HADŽIOSMANOVIĆ D, SOMMER R, ZAMBON E, et al. Through the eye of the PLC: semantic security monitoring for industrial processes[C]//Proceedings of the 30th Annual Computer Security Applications Conference. New York: ACM Press, 2014: 126-135.

[47] CASELLI M, ZAMBON E, KARGL F. Sequence-aware intrusion detection in industrial control systems[C]//Proceedings of the 1st ACM Workshop on Cyber-Physical System Security. New York: ACM Press, 2015: 13-24.

[48] HAVLENA V, MATOUŠEK P, RYŠAVÝ O, et al. Accurate automata-based detection of cyber threats in smart grid communication[J]. IEEE Transactions on Smart Grid, 2023, 14(3): 2352-2366.

[49] ZHOU C J, HUANG S, XIONG N X, et al. Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015, 45(10): 1345-1360.

[50] ZHANG F, KODITUWAKKU H A D E, HINES J W, et al. Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data[J]. IEEE Transactions on Industrial Informatics, 2019, 15(7): 4362-4369.

[51] LIU J J, LIN X D, CHEN X, et al. ShadowPLCs: a novel scheme for remote detection of industrial process control attacks[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(3): 2054-2069.

[52] PONOMAREV S, ATKISON T. Industrial control system network intrusion detection by telemetry analysis[J]. IEEE Transactions on De-

pendable and Secure Computing, 2016, 13(2): 252-260.

[53] TERAI A, ABE S, KOJIMA S, et al. Cyber-attack detection for industrial control system monitoring with support vector machine based on communication profile[C]//Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Piscataway: IEEE Press, 2017: 132-138.

[54] SHEN C, LIU C, TAN H L, et al. Hybrid-augmented device fingerprinting for intrusion detection in industrial control system networks[J]. IEEE Wireless Communications, 2018, 25(6): 26-31.

[55] YUN J H, HWANG Y, LEE W, et al. Statistical similarity of critical infrastructure network traffic based on nearest neighbor distances[C]//International Symposium on Research in Attacks, Intrusions, and Defenses. Berlin: Springer, 2018: 577-599.

[56] HASSAN M M, GUMAEI A, HUDA S, et al. Increasing the trustworthiness in the industrial IoT networks through a reliable cyberattack detection model[J]. IEEE Transactions on Industrial Informatics, 2020, 16(9): 6154-6162.

[57] CHEN T S, LIN P, LING J. An intrusion detection method for industrial control system based on gate recurrent unit[J]. Journal of Physics: Conference Series, 2019, 1302(2): 022016.

[58] YANG K, LI Q, LIN X D, et al. iFinger: intrusion detection in industrial control systems via register-based fingerprinting[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(5): 955-967.

[59] WANG J, LI P, KONG W, et al. Unknown security attack detection of industrial control system by deep learning[J]. Mathematics, 2022, 10(16): 2872.

[60] WANG W B, ZHANG B, YU Z C, et al. Anomaly detection method of unknown protocol in power industrial control system based on RNN[C]//Proceedings of the 2022 5th International Conference on Renewable Energy and Power Engineering (REPE). Piscataway: IEEE Press, 2022: 68-72.

[61] ZHANG S J, LAI J, YAO Q G. Traffic anomaly detection model of electric power industrial control based on DBN-LSTM[C]//Proceedings of the 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). Piscataway: IEEE Press, 2021: 1902-1907.

[62] DENG H W, ZHAO Y Q, LI X W, et al. Abnormal flow monitoring of industrial control network based on neural network[C]//Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2022: 627-631.

[63] ZHOU L Y, GUO H Q, DENG G L. A fog computing based approach to DDoS mitigation in IIoT systems[J]. Computers & Security, 2019, 85: 51-62.

[64] ABDEL-BASSET M, CHANG V, HAWASH H, et al. Deep-IFS: intrusion detection approach for industrial Internet of things traffic in fog environment[J]. IEEE Transactions on Industrial Informatics, 2021, 17(11): 7704-7715.

[65] LI Y J, JI X Y, LI C G, et al. Cross-domain anomaly detection for power industrial control system[C]//Proceedings of the 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC). Piscataway: IEEE Press, 2020: 383-386.

[66] GU H R, LAI Y X, WANG Y P, et al. DEIDS: a novel intrusion detection system for industrial control systems[J]. Neural Computing and Applications, 2022, 34(12): 9793-9811.

[67] JIANG J R, CHEN Y T. Industrial control system anomaly detection and classification based on network traffic[J]. IEEE Access, 2022, 10: 41874-41888.

[68] CHEN Y L, SU S D, YU D, et al. Cross-domain industrial intrusion detection deep model trained with imbalanced data[J]. IEEE Internet of Things Journal, 2023, 10(1): 584-596.

[69] GÓMEZ Á L P, MAIMÓ L F, CELDRÁN A H, et al. Crafting adversarial samples for anomaly detectors in industrial control systems[J]. Procedia Computer Science, 2021, 184: 573-580.

[70] GAO S, CHEN J J, ZHANG B S, et al. Privacy-preserving industrial control system anomaly detection platform[J]. Security and Communication Networks, 2023(1): 7010155.

[71] LIN K D, XU X L, GAO H H. TSCRNN: a novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of IIoT[J]. Computer Networks, 2021, 190: 107974.

[72] LUKSHA I, DUY DINH T, KARELIN E, et al. Method for filtering encrypted traffic using a neural network between an industrial Internet of things system and digital twin[C]//Proceedings of the 5th International Conference on Future Networks & Distributed Systems. New York: ACM Press, 2021: 595-601.

[73] NIU Z Q, XUE J F, QU D C, et al. A novel approach based on adaptive online analysis of encrypted traffic for identifying Malware in IIoT[J]. Information Sciences, 2022, 601: 162-174.

[74] DAI J B, XU X L, GAO H H, et al. CMFTC: cross modality fusion efficient multitask encrypt traffic classification in IIoT environment[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(6): 3989-4009.

[75] ZHAO R J, HUANG Y T, DENG X W, et al. A novel traffic classifier with attention mechanism for industrial Internet of things[J]. IEEE Transactions on Industrial Informatics, 2023, 19(11): 10799-10810.

[76] ZHU Z S, SHI J Z, WANG C H, et al. MCFM: discover sensitive behavior from encrypted traffic in industrial control system[C]//Proceedings of the 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Piscataway: IEEE Press, 2022: 897-904.

[77] NIEDERMAIER M, FISCHER F, BODISCO A V. PropFuzz—an IT-security fuzzing framework for proprietary ICS protocols[C]//Proceedings of the 2017 International Conference on Applied Electronics (AE). Piscataway: IEEE Press, 2017: 1-4.

[78] FANG D L, SONG Z W, GUAN L, et al. ICS3Fuzzer: a framework for discovering protocol implementation bugs in ICS supervisory software by fuzzing[C]//Annual Computer Security Applications Conference. New York: ACM Press, 2021: 849-860.

[79] RAMOS S H, VILLALBA M T, LACUESTA R. MQTT security: a novel fuzzing approach[J]. Wireless Communications and Mobile Computing, 2018, 2018: 1-11.

[80] LAI Y X, GAO H J, LIU J. Vulnerability mining method for the modbus TCP using an anti-sample fuzzer[J]. Sensors, 2020, 20(7): 2040.

[81] HU Z C, SHI J Q, HUANG Y H, et al. GANFuzz: a GAN-based industrial network protocol fuzzing framework[C]//Proceedings of the 15th ACM International Conference on Computing Frontiers. New York: ACM Press, 2018: 138-145.

[82] ZHAO H, LI Z H, WEI H S, et al. SeqFuzzer: an industrial protocol fuzzing framework from a deep learning perspective[C]//Proceedings of the 2019 12th IEEE Conference on Software Testing, Validation and Verification (ICST). Piscataway: IEEE Press, 2019: 59-67.

[83] LV W Y, XIONG J W, SHI J Q, et al. A deep convolution generative adversarial networks based fuzzing framework for industry control pro-

tocols[J]. Journal of Intelligent Manufacturing, 2021, 32(2): 441-457.
[84] YU Z H, WANG H L, WANG D, et al. CGFuzzer: a fuzzing approach based on coverage-guided generative adversarial networks for industrial IoT protocols[J]. IEEE Internet of Things Journal, 2022, 9(21): 21607-21619.
[85] BYTES A, RAJPUT P H N, DOUMANIDIS C, et al. FieldFuzz: in situ blackbox fuzzing of proprietary industrial automation runtimes via the network[C]//Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses. New York: ACM Press, 2023: 499-512.
[86] CHE X, GENG Y Y, ZHANG G, et al. Fuzzing technology based on information theory for industrial proprietary protocol[J]. Electronics, 2023, 12(14): 3041.
[87] LUO Z, YU J, ZUO F, et al. Bleem: packet sequence oriented guzzing for protocol implementations[C]//32nd USENIX Security Symposium (USENIX Security 23). Berkeley: USENIX Association, 2023: 4481-4498.
[88] 化盈盈, 张岱墀, 葛仕明. 深度学习模型可解释性的研究进展[J]. 信息安全学报, 2020, 5(3): 1-12.
HUA Y Y, ZHANG D C, GE S M. Research progress in the interpretability of deep learning models[J]. Journal of Cyber Security, 2020, 5(3): 1-12.
[89] 刘奇旭, 王君楠, 尹捷, 等. 对抗机器学习在网络入侵检测领域的应用[J]. 通信学报, 2021, 42(11): 1-12.
LIU Q X, WANG J N, YIN J, et al. Application of adversarial machine learning in network intrusion detection[J]. Journal on Communications, 2021, 42(11): 1-12.
[90] 杨安, 孙利民, 王小山, 等. 工业控制系统入侵检测技术综述[J]. 计算机研究与发展, 2016, 53(9): 2039-2054.
YANG A, SUN L M, WANG X S, et al. Intrusion detection techniques for industrial control systems[J]. Journal of Computer Research and Development, 2016, 53(9): 2039-2054.
[91] YAO Y F, DUAN J H, XU K D, et al. A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly[J]. High-Confidence Computing, 2024, 4(2): 100211.

[作者简介]

刘奇旭（1984– ），男，江苏徐州人，博士，中国科学院信息工程研究所研究员，中国科学院大学教授，主要研究方向为网络攻防技术、网络安全评测。

肖聚鑫（1999– ），男，江西萍乡人，中国科学院大学博士生，主要研究方向为网络攻防技术、物联网安全、网络流量分析等。

谭耀康（1998– ），男，广东肇庆人，中国科学院大学硕士生，主要研究方向为Web安全、程序分析。

王承淳（2000– ），男，山东济南人，中国科学院大学博士生，主要研究方向为异常流量检测、恶意软件分析等。

黄昊（2001– ），男，江西九江人，中国科学院大学博士生，主要研究方向为Web安全、AI安全等。

张方娇（1989– ），女，山东泰安人，博士，中国科学院信息工程研究所高级工程师，主要研究方向为Web安全、溯源取证等。

尹捷（1991– ），女，重庆人，博士，中国科学院信息工程研究所工程师，主要研究方向为恶意代码对抗、僵尸网络、物联网安全等。

刘玉岭（1982– ），男，山东济阳人，博士，中国科学院信息工程研究所正高级工程师，主要研究方向为网络安全测评和等级保护。