基于 NTA 的工业数据安全监测方法设计与应用研究*

胡冰蔚[1],洪 晟[2],王泽政[3],李鹏超[3],刘博宇[1],冼 峰[4],赵立鸣[4]

(1.国际关系学院 网络空间安全学院,北京 100091;2.北京航空航天大学 网络空间安全学院,北京 100191;
3.恒安嘉新(北京)科技股份公司,北京 100098;4.上海城投水务(集团)有限公司,上海 200002)

摘 要: 工业互联网安全监测平台趋于成熟。在工业互联网+大数据的环境下,工业数据如今面对日常运维泄露、传输违规、跨境数据安全等问题,工业数据安全监测成为工业互联网的核心关键,是态势感知、监测预警等重要功能的基础和前提。重点介绍了传统流量识别数据安全监测技术,基于 DPI、DFI 流量的工业数据监测,以及目前工业互联网数据安全监测技术应用,解决"工业互联网+大数据"场景下工业数据安全防护问题。

关键词: 工业互联网;大数据;数据安全;流量监测技术

中图分类号: TP393 **文献标识码:** A **DOI:** 10.19358/j.issn.2096-5133.2021.09.001

引用格式: 胡冰蔚,洪晟,王泽政,等. 基于 NTA 的工业数据安全监测方法设计与应用研究[J].信息技术与网络安全,2021,40(9):2-8.

Research on design and application of industrial data security monitoring method based on NTA

Hu Bingwei[1], Hong Sheng[2], Wang Zezheng[3], Li Pengchao[3], Liu Boyu[1], Xian Feng[4], Zhao Liming[4]

(1.School of Cyber Science and Engineering, University of International Relations, Beijing 100091, China;
2.School of Cyber Science and Technology, Beihang University, Beijing 100191, China;
3.Eversec Technology Corporation, Beijing 100098, China;
4.Shanghai Chengtou Water Group Corporation, Shanghai 200002, China)

Abstract: The industrial Internet security monitoring platform is becoming mature. In the environment of industrial Internet and big data, industrial data is now facing daily operation and maintenance leakage, transmission violations, cross-border data security and other issues. Industrial data security monitoring has become the core key of the industrial Internet. It is the basis and prerequisite for important functions such as situational awareness, monitoring and early warning. This article focuses on traditional traffic identification data security monitoring technology, industrial data monitoring based on DPI and DFI traffic, as well as the current industrial Internet data security monitoring technology product application, to solve the industrial data security monitoring problem in the industrial Internet and big data environment.

Key words: industrial Internet; big data; data security; network traffic monitoring technology

0 引言

工业互联网是以数字化、网络化、智能化为主要特征的新工业革命的关键基础设施[1],工业数据安全是工业互联网创新发展的基础保障。各工业企业基于网络化协同和生产化改造需求,践行"内外网"改造和"上云"业务,工业互联网系统架构从以控制系统为中心转向以工业大数据为核心[2]。大数据安全在工业互联网新业态和应用场景下愈加重要。"5G+工业互联网"、"物联网"、"云计算"、"工业互联网平台"等新信息技术的快速发展和大规模产业化应用使得工业数剧量剧增,数据作为工业互联网的基础资源和关键要素,事关企业安全运营、

*基金项目:国家重点研发计划(2019YFB1706001);上海工业控制系统安全创新功能型平台开放课题项目(TICPSH202103010-ZC)

社会经济正常运转乃至国家安全，一旦工业互联网数据泄露、被窃取或篡改，将对社会经济和国家安全造成严重的威胁[3-6]。

1 工业互联网大数据安全研究现状

国内外对于工业互联网数据安全的研究仍处在应用探索阶段，面向“工厂内网”和“工厂外网”场景的工业互联网数据安全保障大都延续采用传统的防护手段。例如，工业数据的安全交互，主要通过私有协议封装和二次加密实现数据加密传输；工业互联网平台服务和运营，依托准入认证、接口加密、漏洞扫描、代码审计、安全加固等传统网络安全措施来抵御潜在攻击。

在数据采集方面，传统的网络数据采集技术针对的是采用固定端口传输[7]。工业互联网业务应用的快速发展使得承载其服务的端口数量持续增加，更新和维护端口数据库日渐困难。不同的工业设备、工业控制系统和工业互联网平台供应商使用专用通信协议、工控协议和自定义接口，私有工控协议封装后通过 HTTP/HTTPS 传输，一些“双跨”工业互联网平台网络传输基于安全性考虑进行了二次加密，增大了工业互联网的流量分析难度。

基于传统网络结构下数据包监测技术主要对数据链路层、网络层、传输层的数据包元素监测和分析，其流量识别检测方法概括如下：

(1)基于网络端口映射的协议识别方法[8-9]。该方法凭借网络协议通信时使用的端口号建立映射关系，并据此区别各类网络协议和应用服务。其缺点是未能给所有应用及服务定义端口号，且对于端口号不固定的数据传输协议不能与之一一对接，难以形成和积累较为完备的端口映射关系。

(2)基于有效载荷分析的应用业务识别方法(Deep Packet Inspection，DPI)[10]。此类方法依托建立的应用业务特征规则库进行应用业务识别。DPI 技术在流量分析、网络信息安全和网络服务质量(Quality of Service，QoS)监测等领域应用广泛，其可对网络数据包进行内容分析，能够解析数据包内容及有效载荷，提取内容信息。完整的 DPI 技术处理流程如图 1 所示。由于网络带宽和数据的不断增加，预先建立的规则库也快速增加，导致系统计算开销过大，且存在数据泄露的安全风险。

(3)基于行为特征的流量识别方法(Deep Flow Inspection，DFI)[11]。DFI 是基于流量行为模式的技术，可以视为 DPI 体系的一部分，可分析网络中通信行为及模式，基于流量特征对网络流量进行分类识别和检测，积累广谱规则。在不同的应用类型中数据流的状态不同，表现为数据流特征的差异。此方法存储开销较大，实时性较差，致使其技术状态进展缓慢。

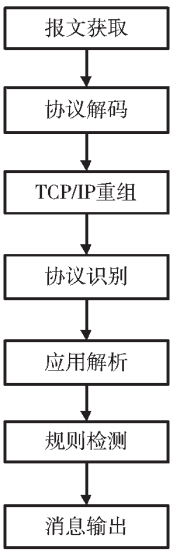


图 1 DPI 技术处理流程

(4)基于机器学习的流量分类识别方法[12-13]。具体应用可分为有监督学习的流量分类识别方法、无监督学习流量分类识别方法和半监督学习的流量分类识别方法。但从实际工程效果看，该类方法识别精确度有待提升，实时性有限，且特征值之间的关系仍待进一步的研究。

随着网络通信技术的进步加之“互联网+”模式的兴起，以数据匹配检测模式构建的监测防御体系不能应对海量数据交互带来的安全风险。在当前工业互联网安全保障探索阶段，存在工业传输中应用协议复杂、通信端口伪装、工业互联网数据跨境传输等客观现状，进行工业互联网流量安全分析的同时需要保证数据的机密性和完整性，工业互联网协议行为特征识别需要较大时间、空间开销，也存在检测实时性等问题。传统数据监测流量识别技术尚不能有效处理目前“工业互联网+大数据”环境下的诸多问题，需要探索新型工业互联网数据安全保障技术。

2 基于 NTA 的工业互联网数据监测方法

2.1 面向工业互联网的 NTA 方法

针对 DPI 和 DFI 的现存问题，本文提出了一种基于网络流量分析(Network Traffic Analysis，NTA)的工业互联网数据监测方法，该方法将 DPI、DFI、机器学习等方法以监测需求为出发点，依托“大网”监测场景，采取汇聚分流的方式，将网络中多个节点的数据进行特征分流、数据汇聚、数据预处理等操作，并将数据发送至对应的模块进行处理。该方法先汇聚全流量并依托广谱规则做流量筛选，再依托特征监测、文件还原和事件研判等技术识别潜在数据风险，在控制其计算开销和存储开销的同时，解

决原有技术缺陷。

网络流量分析技术 NTA 应用场景更为宽泛，旨在通过监控网络流量、连接和对象来识别恶意程序、攻击入侵行为迹象，本质是记录、审计和分析网络流量数据的过程。NTA 探针设备及配属产品集成高速正则事件引擎、内容结构检测引擎、异常行为识别引擎、文件还原研判引擎和全流量回溯引擎，识别流量异常，鉴别数据泄露、违规传输等可疑行为。NTA 底层技术通过 DPI[14]和 DFI[15]来解析流量，兼具规则集成和特征签名功能，能够对流量解析数据和外部数据进行建模，具备加密流量检测和威胁样本分析能力，可结合威胁情报提供威胁响应的决策支持。

由于工业应用业务类型的复杂性，仅分析报文头字段内容无法很好识别报文所承载的工业业务类型。通过识别跟踪业务协议交互过程，对报文有效载荷特征以及流量特征深度检测，检测工业互联网流量，识别工业应用业务类型，开展包解析和文件还原日渐兴起。

2.2 基于 NTA 的数据监测方法框架

基于 NTA 的底层技术是在传统基于端口识别的基础上，增加机器学习、DPI、DFI 等技术方法，即融合传统规则和机器学习的方法，探索提高工业互联网流量分析的准确性。通过将采集设备部署在骨干网、城域网、工业互联网专线、工业园区出入口、工业企业出入口等关键节点，采集流量信息，结合主动监测技术，丰富和补充各类资产对象，结合特征识别和流量还原技术增强工业流量数据的获取和分析能力。该技术适用于多工业场景下的工业互联网流量识别、协议解析和内容还原研究。

基于 NTA 的数据监测总体技术架构如图 2 所示，该架构可分为三个模块，即数据源采集模块、NTA 技术模块和工业资产管理模块。在数据源模块中，对部署网络节点如网关、交换机等设备进行流量采集。在 NTA 技术模块中，将底层技术识别能力汇聚成 NTA 技术的基础能力，实现对采集流量的特征分类，构建分析所需的流量数据集，通过广谱规则做流量筛选；然后，再对流量进行精细化分析，进行工业协议识别解析、工业互联网设备识别、工业互联网文件筛选和解析、异常流量识别等业务。在工业资产管理模块中，结合态势感知技术面向入侵检测、应急服务具体业务进行功能联动。

为了实现面向具体行业或场景的细分，本方法在 NTA 基础能力上，结合威胁情报，辅助开展工业数据安全事件分类分级评测业务。可结合不同行业要求，强化分类分级业务指导下的工业数据安全保

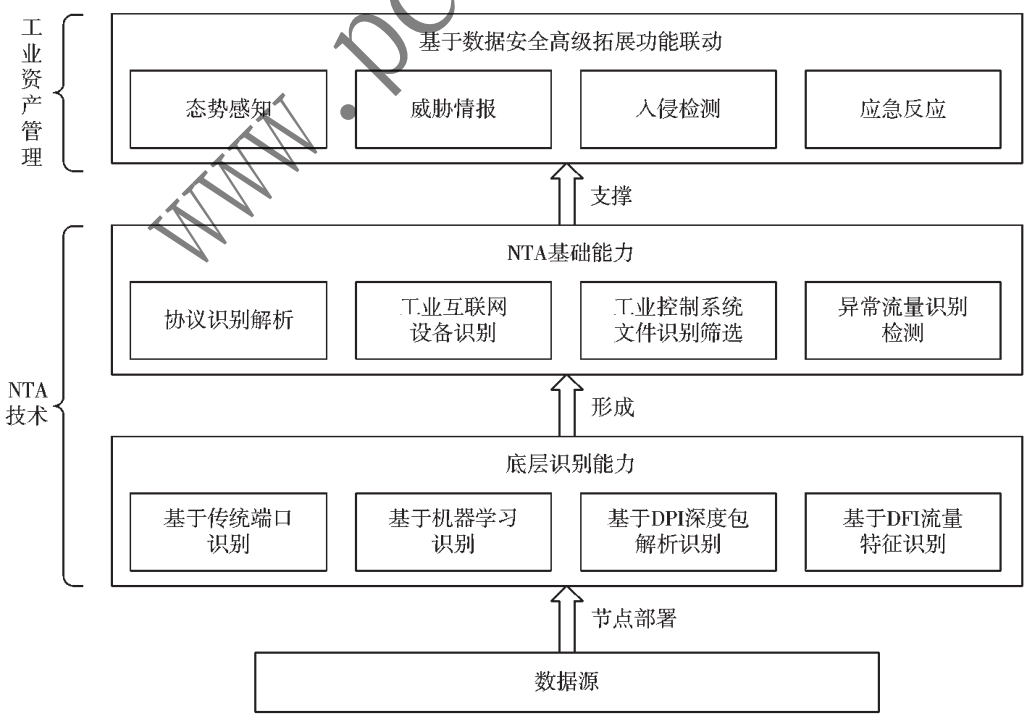


图 2 工业数据安全监测技术架构

障，从大数据安全防护视角为企业主体提供工业数据安全防护新技术手段，结合工业企业、工业互联网平台和标识解析企业已部署的安全防护手段，进一步提高数据安全防范措施。

2.2.1 工业互联网协议识别

工业互联网协议识别针对支持 S7、Modbus、BACnet 等常见工业通信协议、支持 MQTT\XMPP\COAP\UPNP\ONVIF 等物联网协议、支持 HTTP\DNS\FTP\SMTP\HTTP2 等协议识别。其中工业互联网应用层协议识别主要包括超文本传送协议(HTTP)、分布式实施数据分发服务中间件协议(DDS)、消息队列遥测传送协议(MQTT)、OPC 统一架构(OPC-UA)识别和分析。

2.2.2 工业互联网资产识别

利用工业互联网资产识别技术，可识别 2G/3G/4G/5G 移动互联网、骨干网、IDC、物联网和工业互联网专线等通信网络中的资产，包括各种工业网络设备、主机设备、安全防护设备等。该方法支持探测到资产的存活状态、资产的系统版本、设备类型、开放端口或服务、中间件、数据库等应用程序版本、组件信息、漏洞数量等。

基于流量的被动资产指纹识别，支持探测到资产的存活状态、资产的系统版本、设备类型、开放端口或服务、应用程序版本、Web 组件信息、漏洞数量等。结合工业企业备案信息，支持对未“上云”资产识别，配合云端情报联查监测相关风险。

2.2.3 基于机器学习的工业互联网敏感数据识别

为了在流量中还原数据流的样本，自动识别数据的敏感程度，本文基于机器学习方法，实现对数据包的清洗、特征识别和数据监测。

基于机器学习的流量识别方法：在不依赖于端口和规则特征的流量识别前置条件下，利用机器学习从数据流中提取特征属性，构建工业互联网安全监测业务分类模型，完成流量的识别和业务应用检测。相关特征属性主要包括数据包特征和数据流特征。

基于机器学习的风险监测方法：以传播渠道的行为检测技术为基础，基于工业互联网企业、平台、设备、业务的基础特征数据设计机器学习模型，分析异常数据传输和敏感数据泄露事件。集成联网“暴露”工业资产监测、异常流量监测、工业威胁监测技术手段，依托风险检测引擎的协同工作，可针对恶意攻击、脆弱性利用、Web 漏洞、主动数据泄露行为等场景开展检测和研判，建立多维度的特征扫描机制，构建哈希值、软件包名称、关键字、特征码、软件签名、字符串等多态特征判定和积累方法，提高对潜在风险的识别能力，实现对安全风险多态特征进行精准快速定位。

其本质是在流量中还原数据流和样本，自动识别数据的敏感度。其中文件数据根据文件的大小、类型、传输协议进行筛选，并且通过机器学习方法，对自定义文件类型进行训练生成自定义文件类型模型库，设定文件特征值，然后再根据模型库筛选出符合特征值的文件，实现对文件的过滤功能。

2.2.4 异常流量分析方法

基于对业务安全的考虑，越来越多的工业互联网应用采用 HTTPS 或 SSL/TLS 进行数据交互，根据现网统计数据，目前加密流量占比已经超过 70%，并呈现增加趋势。目前对于加密流量的识别通常采用证书或签名的方式，该种方式只能识别流量归属于哪一种应用，无法满足如恶意程序、信息安全等业务的需求。通过采用机器学习及人工智能相关技术，首先对 TLS 流量的元数据、包长和时间序列、字节分布、非加密的 TLS 头信息通过卷积神经网络 CNN 进行分类，然后将每条 TLS 流的前 N 个载荷拼接起来并转化成图像通过 CNN 进行分类，再将分类结果经过集成学习得出最终识别结果，并将识别结果反馈给学习模型。最后，采用机器学习方法，通过统计特定数量的连续字节的熵，区分不同类型的二进制流，对异常流量进行识别和持续监测。

2.3 基于 NTA 的工业数据安全监测系统设计

系统架构自下而上可大致划分为数据接入层、基础能力层、业务能力层及应用支撑层，如图3所示。

数据接入层提供基础流量、工业互联网平台企业、工业企业的数据接入，其中基础流量包括某地域的固网、移动网、互联网专线(IDC 出入口)的大流量采集能力，采集流量作为上层业务安全分析和处理的基础。

基础能力层对采集流量做汇聚和分流处理，可将网络中多个接口的原始数据进行汇聚、分流、提取、过滤、复制等操作，支持精细化、定制化分流输出给业务系统进行数据分析。搭载数据中台，实现数据入库、分析和持久化存储。集成基础规则库，积累和沉淀广谱规则和工业互联网指纹特征。

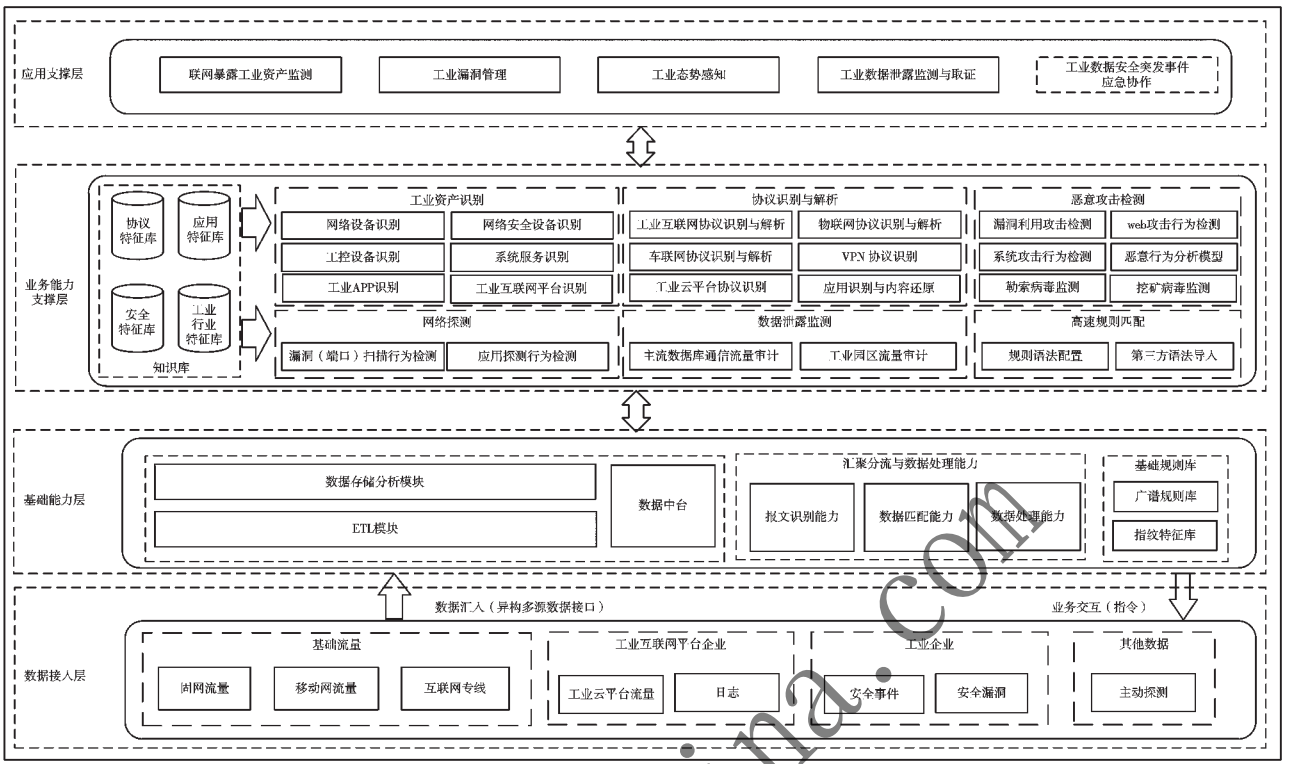


图 3 系统架构图

业务能力支撑层基于 NTA 技术，读取 IP 包载荷的内容来对 OSI 七层协议中的应用层信息进行深入分析。DPI 技术除支持对链路层、网络层、传输层及应用层工业协议进行识别与解析，还具备对物联网协议、工业互联网协议、车联网协议、VPN 协议、工业应用识别与内容还原等底层能力，提供工业资产识别、恶意代码检测、溯源取证、网络探测检测、恶意攻击检测、特殊流量检测、数据泄露检测等业务安全解析能力。NTA 引擎具备面向工业互联网全业务（含数据安全）的并行多次处理输出能力，支持自定义格式化输出给上层应用支撑系统。

应用支撑层不仅具备工业数据泄露监测功能，还具有基于联网暴露工业资产监测、工业漏洞管理功能，通过对外接口输出定制化话单，支持漏洞管理、监测预警、网络取证等工业互联网安全应用系统的分析和展示。

3 实验和结论

3.1 实验环境搭建

实验验证依托某地域现网环境，开展软硬件环境搭建和系统集成，面向本地域内的工业资产监测、风险识别、数据安全监测开展验证和分析工作。

3.1.1 硬件环境

硬件环境搭建由大数据服务器和业务系统服务器组成。服务器配置不低于 2×4 核 CPU，32 GB 内存，1 TB 硬盘，能够满足大部分网络节点数据采集和业务对接需要。

3.1.2 软件环境

软件环境操作系统采用 Cent OS 6.8；搭配第三方组件：Base－A10.1－Jdk1.8（版本：1.8.0_162），BaseA10.1－redis（版本：4.0.9），BaseA10.1－nginx（版本：1.15.6），BaseA10.1－mysql－5.7（版本：5.7）；客户端环境兼容主流浏览器：Chrome 45.0.2454.101 或以上版本，Firefox 45.1.0 或以上版本，IE10 或以上版本。

软件环境配置如表 1 所示。

表 1 软件环境配置表

环境名称	软件版本
操作系统	Cent OS 6.8
数据库	MySQL 5.7
Java 版本	JDK 1.8 以上版本
应用服务器	Tomcat 8.0

3.1.3 系统部署

系统的部署分为前台数据安全策略显示和下发模块，后台数据采集分析及监测业务模块。前台负责显示流量识别和数据安全情况，并生成合适的策略下发给用户；后台设计为统一的大数据分析平台实现对流量采集和分析，由数据采集、数据库存储、数据监测、特征库模块等功能组成，是工业数据安全监测应用的核心。

在实际部署业务中，需根据网络节点的实际情况进行部署调整，但需满足网络、硬件及软件部署所需的最低配置要求。

3.2 实验结果和应用效果

3.2.1 资产识别和安全风险监测

本文所介绍的技术已广泛应用于工业互联网安全监管领域，基于该技术体系打造的工业互联网安全监测和态势感知平台在多地实际部署应用。某地的应用效果(节选)如表 2 所示，在近期月份监测识别到设备(工控设备、物联网设备、车联网设备)106.13 万个，联网平台(工业互联网平台、物联网平台、车联网平台)516 个。3 月份监测到工业互联网企业安全漏洞 333 个，安全事件 665 万余次。

表 2 应用效果对比表

应用领域	业务类型	资产(事件)分类	数据	影响范围
工业互联网	安全风险	安全漏洞 安全事件	333 个 665 万次	装备制造、汽车、物流、家电制造等
车联网	安全事件	成功事件 攻击次数	25 次 31 479	某省 17 家单位

在车联网应用领域，以某省互联网专线监测为例，累计发现约 58 万条事件数据；其中 4 个 IP 定位到其指向同一车联网平台“GPS 定位管理平台”，发现成功事件 25 次、攻击次数 31 479 次、涉及单位 17 个。

3.2.2 工业安全事件案例

在木马远控事件中，监测到某单位的 *.*.138.14 多次访问美国纽约的 IP：199.59.242.153 的域名 ww25.stats.stuffpicks.com，该恶意域名是用于与网页木马远控端进行通信，经木马样本数据与恶意数据包对比，发现数据包内容相同，综合研判为 Trojan.unknown.a.C640 木马远控事件。

在挖矿病毒危害事件中，监测发现某单位的 IP 地址 *.*.227.26 向新加坡的 IP 地址 134.209.100.235，荷兰阿姆斯特丹的 IP 地址 142.93.137.119、178.128.242.134 进行了多次门罗币挖矿矿池登录请求，从流量包可以看出攻击者通过心跳机制 keepalive 维持与矿机的通信。

3.2.3 工业数据泄露案列

通过某省工业互联网专线流量分析，发现用户名、密码、手机号、GPS、位置信息、邮箱、智能水表、消防仪表等相关信息泄漏，涉及仪器仪表行业、消防安全、交通管理和水务等多个领域。

3.2.4 车联网跨境通联案例

基于某工业互联网专线部分流量分析，发现车联网信息泄露，包括车牌号、经纬度、sim 卡号、速度(轨迹)、车辆状态等信息。持续监测发现，某热门合资品牌智能网联汽车存在跨境通联行为，跨境访问前十位目的国家如表 3 所示。

表 3 车联网跨境通联统计表

国家	IP 数	通联次数
美国	3 059	530 420
日本	109	341 845
英国	77	241 876
德国	59	201 876
法国	39	151 563
比利时	16	142 654
泰国	11	131 876
荷兰	9	131 887
瑞士	4	121 887
巴西	4	101 876

注：以上案例进行了脱敏处理，仅用于技术研究，不涉及具体主体信息。

4 结论

工业数据安全是工业互联网企业数字化安全转型的前提，是落实工业互联网企业安全防护的重要保障。得益于 NTA 技术在工业互联网及车联网领域的拓展应用，通过主被动一体化技术面向实际防护对象的数据采集和处理，结合防护对象的实际应用场景和业务流程，集成威胁情报、工业漏洞和安全事件，开展多维关联分析，积累协议特征库、应用特征库、设备指纹库、行业特征库，支持联网“暴露”工业资产监测、工业漏洞管理、工业互联网态势感知业务。经实际应用发现，该系统能够发现联网暴露工业资产，识别其安全风险，监测各类工业数据泄露事件。

工业数据安全是落实工业互联网融合发展的核心要义，基于NTA技术开展工业数据安全监测研究是各行业主管单位进行数据安全风险监管的主要手段，是弥补各工业企业开展数据安全防护不足的有效措施。借此可支撑工业数据防护业务拓展到入侵检测、工业数据分类分级测评、工业敏感数据防护等具体场景。综合NTA技术探索工业互联网数据多点跨域环境下遭非法篡改、人为破坏、违规泄露或非法利用时的监测保障和防护处置措施，可作为落实工业数据安全分类分级管理要求、夯实企业主体责任的有效技术抓手。

参考文献

[1] 国务院关于深化“互联网+先进制造业”发展工业互联网的指导意见[EB/OL].(2017-11-27).http://www.gov.cn/zhengce/content/2017-11/27/content_5242582.htm?trs=1.

[2] YIN H, JIANG Y, LIN C, et al. Big data: transforming the design philosophy of future internet[J]. IEEE Network, 2014, 28(4): 14-19.

[3] 王晨，郭朝晖，王建民.工业大数据及其技术挑战[J].电信网技术，2017(8)：1-4.

[4] 刘阳，韩天宇，谢滨，等.基于工业互联网标识解析体系的数据共享机制[J].计算机集成制造系统，2019，25(12)：3032-3042.

[5] 门嘉平.工业大数据安全应用探讨[J].网络安全技术与应用，2019(12)：81-82.

[6] 褚健.解读《工业控制系统信息安全行动计划(2018-2020)》[J].自动化博览，2018，35(7)：54-55.

[7] 孙文.多通道数据采集系统的设计与实现[D].长沙：湖南大学，2013.

[8] 彭文峰，杜中军.应用层协议识别技术研究[J].现代计算机(专业版)，2015(8)：68-70.

[9] 张波，万丽.基于端口映射NAT网络方案分析与实施[J].软件工程师，2015，18(3)：12-13，8.

[10] 李明伟，张大方，曾彬，等.基于有效载荷分析的BT流量识别技术[J].计算机应用，2007(9)：2230-2232.

[11] 胡斌.基于混合行为特征的流量识别技术研究与应用[D].北京：北京邮电大学，2017.

[12] 孙振.基于机器学习的网络流量特征选择[J].电子测量技术，2017，40(7)：131-136.

[13] 李致远，王汝传.一种基于机器学习的P2P网络流量识别方法[J].计算机研究与发展，2011，48(12)：2253-2260.

[14] 刘胤.深度包检测的研究与设计[D].贵阳：贵州大学，2008.

[15] 吴倩.基于DPI与DFI的流量识别与控制系统的设计与实现[D].成都：电子科技大学，2013.

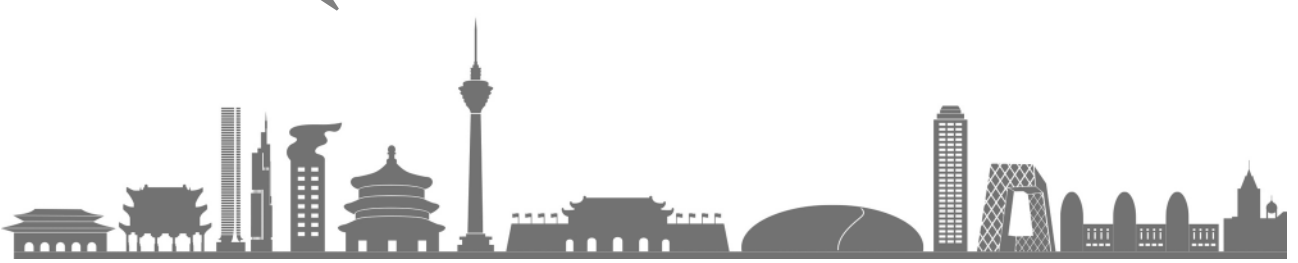
(收稿日期：2021-05-17)

作者简介：

胡冰蔚(1997-)，女，在读硕士研究生，初级工程师，主要研究方向：信息隐藏、工业互联网安全。

洪晟(1981-)，男，博士，副教授，博士生导师，主要研究方向：工业互联网安全、信息网络安全、复杂系统安全性。

王泽政(1984-)，通信作者，男，硕士，中级工程师，主要研究方向：信息安全、终端安全、工业互联网安全。E-mail：wangzezheng@eversec.cn。



版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部

中国电子信息产业集团有限公司第六研究所