计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20211147
59(5):994-1014,2022

基于机器学习的工业互联网入侵检测综述

刘奇旭[1,2] 陈艳辉[1,2] 尼杰硕[1,2] 罗 成[3] 柳彩云[4] 曹雅琴[1] 谭 儒[1] 冯 云[1] 张 越[1,2]
[1](中国科学院信息工程研究所 北京 100093)
[2](中国科学院大学网络空间安全学院 北京 100049)
[3](中国信息通信研究院 北京 100191)
[4](国家工业信息安全发展研究中心 北京 100040)
(liuqixu@iie.ac.cn)

Survey on Machine Learning-Based Anomaly Detection for Industrial Internet

Liu Qixu[1,2], Chen Yanhui[1,2], Ni Jieshuo[1,2], Luo Cheng[3], Liu Caiyun[4], Cao Yaqin[1], Tan Ru[1], Feng Yun[1], and Zhang Yue[1,2]
[1](*Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093)
[2](*School of Cyber Security, University of Chinese Academy of Sciences, Beijing* 100049)
[3](*China Academy of Information and Communications Technology, Beijing* 100191)
[4](*China Industrial Control Systems Cyber Emergency Response Team, Beijing* 100040)

Abstract Machine learning has achieved great success in computer vision, natural language processing and other fields in the past few years. In recent years, machine learning technology has gradually become one of the mainstream technologies in the field of cyber-security, and many intrusion detection technologies based on machine learning have emerged in the field of the industrial Internet. Aiming at landing machine learning-based intrusion detection technology into the real industrial system network, we conduct an in-depth analysis of related work in the field. We summarize the uniqueness of machine learning-based intrusion detection in the industrial Internet and extract three research points from the workflow of intrusion detection in industrial control system (ICS). Based on the research points that different researches focus on, we divide machine learning-based intrusion detection system (IDS) in ICS into three categories: algorithm design-oriented researches, application challenges and limitations-oriented researches, and ICS attack scenario-oriented researches. The taxonomy shows the significance of different research work as well as exposes the problems existing in the research field at present. It can provide a good research direction and reference for future work. In the end, we propose two promising research directions in this field based on the latest developments in machine learning.

Key words industrial Internet; machine learning; intrusion detection; taxonomy; industrial control system (ICS)

收稿日期:2021-11-26;**修回日期**:2021-12-30

基金项目:中国科学院青年创新促进会基金项目(2019163);国家自然科学基金项目(61902396);中国科学院战略性先导科技专项(XDC02040100);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项目
This work was supported by the Foundation of the Youth Innovation Promotion Association CAS (2019163), the National Natural Science Foundation of China (61902396), the Strategic Priority Research Program of Chinese Academy of Sciences (XDC02040100), and the Project of the Key Laboratory of Network Assessment Technology at Chinese Academy of Sciences and Beijing Key Laboratory of Network Security and Protection Technology.

通信作者:张越(zhangyue@iie.ac.cn)

摘 要 过去几年中,机器学习算法在计算机视觉、自然语言处理等领域取得了巨大成功.近年来,工业互联网安全领域也涌现出许多基于机器学习技术的入侵检测工作.从工业互联网的自身特性出发,对目前该领域的相关工作进行了深入分析,总结了工业互联网入侵检测技术研究的独特性,并基于该领域中存在的3个主要研究问题提出了新的分类方法,将目前基于机器学习的互联网入侵检测技术分为面向算法设计的研究工作、面向应用限制和挑战的研究工作,以及面向不同ICS攻击场景的研究工作.该分类方法充分展现了不同研究工作的意义以及该领域目前研究工作中存在的问题,为未来的研究工作提供了很好的方向和借鉴.最后基于目前机器学习领域的最新进展,为该领域未来的发展提出了2个研究方向.

关键词 工业互联网;机器学习;入侵检测;分类法;工业控制系统

中图法分类号 TP274; TP181

工业互联网是传统工业控制系统(industrial control system, ICS)和互联网技术的结合.互联网技术在为工控系统提供便利的同时,也打破了传统的工业信息安全防护模式,不可避免地会将互联网自身固有的网络安全风险引入到工业互联网中.与传统互联网相比,工业互联网的特征更加复杂,涉及设备种类繁多,网点更加密集,协议相对脆弱,导致安全风险也就更多.

近年来,工业互联网安全事件频发,不仅给整个工业行业造成了严重的经济损失,而且造成了极其恶劣的社会影响.比如2010年破坏伊朗核设施的"震网"病毒、2014年攻陷乌克兰电网的BlackEnergy2、2017年攻击沙特天然气系统的TRITION恶意软件、2021年5月针对美国燃油运输管道商的勒索病毒攻击以及近年来频发的勒索软件攻击等.

目前工业互联网面临的安全挑战主要体现在2个方面:脆弱的终端设备和复杂的网络.

终端设备一方面漏洞频发.根据国家信息安全漏洞共享平台(China National Vulnerability Database, CNVD)最新统计,截至2021年10月,与工业控制系统相关的漏洞高达3 100个,其中高危漏洞1 432个,中危漏洞1 493个,低危漏洞175个.另一方面由于工业互联网涉及设备广泛,安全检测和监管手段不到位,导致漏洞难以及时修复.

网络方面,互联网技术的引入使得网络边界更加模糊,以隔离为主的防护体系难以满足现在的互联需求,无线网络的接入更是打破了原有系统的专网通信,追求效率的通信协议缺少相应的安全认证,这些都增加了网络被攻击者入侵的安全风险.

为了缓解安全风险,保障ICS设备和信息安全,目前工业互联网的防护工作主要利用系统的流量和设备监控数据来分析系统是否出现异常.入侵(异常)检测系统(intrusion detection system, IDS)作为工业互联网的一个安全组件,可以实时监视网络传输数据,识别安全事件,及时发现安全威胁和攻击者.传统互联网环境中,入侵检测技术一直是安全从业人员研究的热点,是保障网络安全的重要手段,但是随着黑客攻击数量的增多,传统基于规则的检测方法难以发挥作用.另外基于规则的检测方法需要专业安全人员对数据进行分析,提取特征,且难以应对未知的安全风险.随着机器学习技术在其他领域的应用和发展,比如应用于图像[1-5]和文本[6-10]等,越来越多的入侵检测方案[11]开始采用机器学习技术,由于机器学习技术具备出色的泛化能力和运算性能,以及处理大规模的数据的能力,甚至具备一定的检测未知安全风险的能力,基于机器学习的入侵检测技术成为当前主流的检测方案.

工业互联网领域借鉴传统互联网中的入侵检测技术,使得机器学习技术也开始广泛应用于工业互联网中的入侵检测系统.尽管工业互联网与传统互联网有很多相同点,使得传统互联网中的安全技术可以移植到工业互联网中,但是工业互联网也存在自己的特点.例如工业互联网中采用的是完全不同的网络协议,如Modbus,Profinet等;工业控制系统产生数据的维数高、关联性强[12];相比传统互联网系统,工业控制系统具有高实时性、资源受限、更新困难等特性[13].这些特点都增加了工业控制系统入侵检测技术的难度,研究工作应该针对工业控制系统的这些特点,提出相应的算法和模型.

近年来,随着5G技术的提出和普及,工业互联网安全越来越受到国家的重视,学术界涌现出了许多不同的基于机器学习的入侵检测技术,这些研究工作都具有不同侧重点,使用的机器学习算法也各有不同.鉴于此,本文调研了近10年机器学习技术

在工业互联网入侵检测中应用的相关研究工作.在统计中发现,2018 年以来的相关研究,相比之前的研究工作,数量和质量都有很大的提升,这与机器学习技术近几年的发展有很大关系,但同时也暴露出很多问题.本文针对这些工作进行了深入探讨和分析,总结了工业互联网入侵检测技术区别于普通互联网入侵检测技术的独特性,并从一个新的角度对这些工作进行了分类、分析和总结,以便于未来针对工业互联网安全进行更加深入的研究.

本文的主要贡献包括 3 个方面:

1) 本文调研近 10 年来工业互联网领域中基于机器学习技术的入侵检测工作,在本文的统计中,随着机器学习的发展,不仅该领域的相关工作数量在持续增加,模型的效果也在不断提升,由原先对机器学习算法的简单应用,到现在海量数据处理速度大幅提升、检测的攻击类型更加丰富,从一开始追求分类效果,到现在研究工作人员更加注重模型的落地和实用性,说明该领域经历了长足的发展,慢慢从单一到全面,已经逐渐成熟和完善.基于这些工作,本文从工业互联网自身的特点出发,对整个研究流程做了概括,总结出了 3 个主要研究点,并对每个研究点的意义做了分析和阐述.

2) 本文总结了工业互联网场景和传统互联网场景之间的差异,这些差异使得基于机器学习的入侵检测工作不能使用模型加数据的简单研究模式,而是要充分理解这些差异带来的不同.除了要考虑算法模型外,还要考虑 ICS 场景的要求和限制,比如实时性要求、计算资源限制等.由于不同的 ICS 场景面临的攻击方式存在差异,检测模型必须做出相应的适应性变化,所以工业互联网领域中基于机器学习的入侵检测方法需要关注模型设计、应用限制和挑战、独特的 ICS 场景 3 个方向.我们根据这 3 个方向将目前的研究工作分成了面向算法设计、面向应用挑战和限制,以及面向 ICS 场景 3 个类别.该分类方法不仅能够体现传统互联网中入侵检测工作和工业互联网中入侵检测工作的不同,同时也能很好地总结目前研究工作的重心,揭露工业互联网入侵检测研究工作的研究方向和存在的问题,并为以后的研究工作提供明确的方向.

3) 基于对本领域文献的深入调研,本文总结了机器学习算法在应用到工业互联网时的问题和不足,并针对这些问题做了深入分析,最后基于机器学习应用领域的最新进展工作,展望了基于机器学习的工业互联网入侵检测工作未来的发展方向,详细阐述了不同方向可能存在的研究点以及研究的重要性.

1 基础知识

1.1 工业控制系统(ICS)介绍

工业控制系统是一类工业生产使用的控制系统的总称,它包含监视控制与数据采集控制系统(supervisory control and data acquisition, SCADA)、分布式控制系统和其他常见于工业部门与关键基础设施的小型控制系统等[14].

1.1.1 工业控制系统架构

常见的工业控制系统一般由 3 层网络组成,如图 1 所示[15],从上到下分别是企业网络、控制网络和现场网络,不同网络对应不同的功能.企业网络主要包括 ERP(enterprise resource planning)系统功能单元,用于使管理者掌握和了解整个系统的运行状况和设备状态变化,实现对工艺的过程监视与控制.

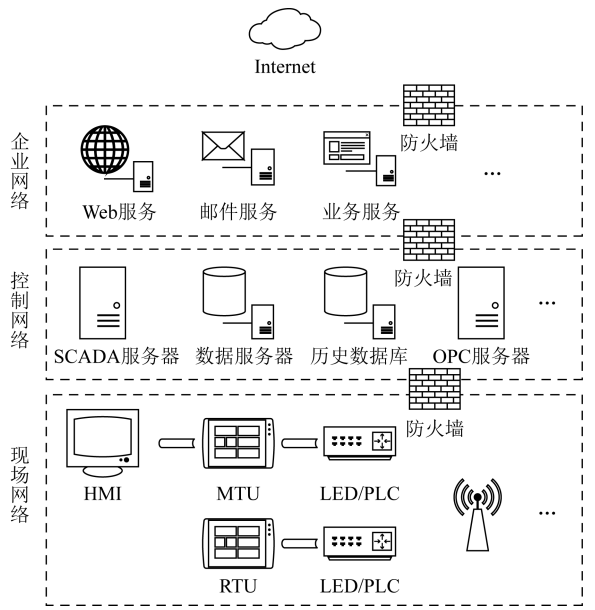


Fig. 1 Structure of industrial control system

图 1 工业控制系统框架

控制网络主要负责监视底层现场网络的行为,负责现场网络和企业网络之间的信息传递和存储.该层主要包括 SCADA 和数据服务器等设备,SCADA 可以对现场运行的设备进行监视和控制,实现对现场设备进行数据采集、设备控制、测量、信号报警以及参数调节等功能.

现场网络主要包括程序逻辑控制器(programmable logic controller, PLC)、远程终端单元(remote terminal unit, RTU)等现场设备,用于对生产过程的设备进行感知和操作.HMI 为人机界面,用于系统

和用户之间进行信息交互,向现场设备发送控制命令和查询请求.

1.1.2 工业互联网协议

工业控制通信协议是为了提高效率和可靠性而设计的,为了满足 ICS 的经济效益和运作效率,大多数的工业控制通信协议很少考虑安全性,而且大部分协议都是私有协议.本节主要介绍了 3 种常用的开放的工业控制协议:Modbus 协议、ICCP(Inter-control Center Communication Protocol),以及 DNP3(Distributed Network Protocol).这些协议由于缺少认证加密等安全措施,本身就存在着大量的安全隐患,很容易受到攻击者的攻击和利用.

Modbus[16]是历史最悠久的工业控制协议,是工业电子设备之间常用的连接方式.由于 Modbus 公开发表且无版权要求,还易于部署维护,所以是目前应用场景最广泛的工业控制协议.

Modbus 是一个 OSI 模型中应用层的协议,它可以实现设备间高效通信.通过此协议,PLC,RTU 和 SCADA 可以通信.它描述了控制器请求访问其他设备的过程,以及如何回应来自其他设备的请求、怎样侦测错误记录.它制定了通信数据的格式和内容的公共格式.Modbus 有 2 种通信传输方式,分别是 ASCII 模式和 RTU 模式.ASCII 模式在实际应用中相对较少,RTU 模式为常用模式.Modbus 在设计之初并没有考虑安全问题,缺少很多安全措施,比如认证、加密、验证等,所以很容易遭到恶意攻击.

DNP3(Distributed Network Protocol)[17]是一种应用于自动化组件之间的通信协议,常见于电力、水处理等行业.SCADA 可以使用 DNP 协议与主站、RTU、IED(intelligent electronic device)进行通信.DNP 协议提供了对数据的分片、重组、数据校验、链路控制、优先级等服务,使用大量 CRC 校验来保证数据准确性.DNP 协议同时还具有对抗恶劣环境中产生的电磁干扰的能力,但是无法有效抵抗黑客的攻击和破坏,使得该协议漏洞频出,同样也缺乏授权认证和加密等安全措施,大大增加了受到黑客攻击的风险.

ICCP (Inter-control Center Communication Protocol)[18]协议最早是美国电力科学院于 20 世纪 90 年代提出的,是目前最为流行的电力系统计算机通信的应用层规约之一.它是采用 MMS 服务和协议的一个标准化的 MMS 应用.其通信过程如图 2 所示,它采用客户端/服务器的通信方式,该方式的优点是可以提高通信效率和数据传输实时性,并且节省通信系统的开销.它的双向表和存取控制特性保证了一定的安全性,但是也十分有限,同样缺少认证和加密,容易遭受会话劫持等攻击.

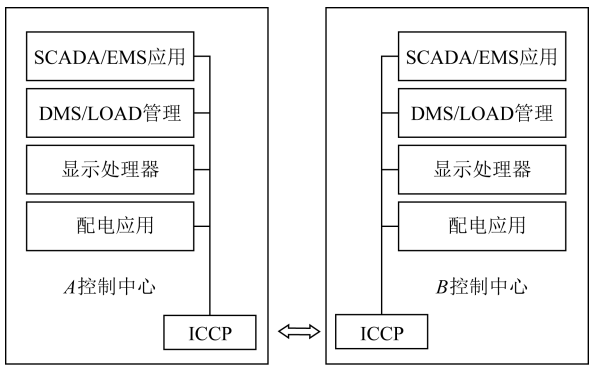


Fig. 2 ICCP protocol framework diagram

图 2 ICCP 协议框架图

1.2 工业互联网入侵检测技术的独特性

工业互联网由于历史存留问题,在建设之初就未考虑安全措施,而将安全寄托于其封闭性上,随着互联网技术的普及,这些安全问题也随之暴露出来.其独特性主要体现在 3 方面:

1) 工业互联网的独特性体现在其协议的独特性.工业互联网采用与普通 IT 网络不同的通信协议,这些协议受限于工业控制系统对实时性的要求,首要目标是提高通信效率,并非安全性,因此绝大多数协议都缺乏安全措施,导致工业互联网很容易遭受黑客攻击.

2) 工业互联网的独特性体现在其场景的多样性上.不同种类和规模的工业场景使用的设备和传感器的数量、种类都有所不同,其网络架构也存在较大差异,导致针对某一工业场景的攻击也呈现出不同的形式,即使相同攻击方式在不同工业场景也会产生不同的影响,而普通 IT 网络的场景都普遍存在相似性,攻击行为、攻击结果也都是相似的.

3) 工业互联网的独特性体现在其限制和要求上,文献[13]总结了工业互联网中独特的限制和要求,比如数据噪声多、高实时性要求、资源受限、难以重启和更新等.这些限制和要求都是普通互联网场景所不具备的.

工业互联网的独特性导致其场景下入侵检测技术也需要根据这些特性来进行研究,协议安全性差会导致遭受的攻击数量增加,从而要求检测模型处理性能要高,能够对攻击快速进行响应.场景的差异性使得检测技术要依托场景来研究,通过对 ICS 场景,以及该场景下存在的攻击类型和不同攻击产生

的不同影响进行针对性研究,才能设计出检测性能出色的检测模型.而工业场景本身存在的限制和要求也导致检测技术面临诸多限制和挑战,比如高实时性要求限制检测模型的计算复杂度,噪声数据使得对检测模型的鲁棒性要求更高.

1.3 经典机器学习算法

随着机器学习算法在其他领域的不断成功,安全领域也开始结合机器学习模型实现智能化检测来提高效率,以下主要介绍安全领域常用的一些机器学习算法.

1.3.1 聚类算法(K-means[19])

聚类算法是一种无监督学习方法,它将数据按照被明确定义的相似性划分成一定数量的集群,使同集群内部的数据之间的相似性大于不同集群的数据之间的相似性.在聚类算法中,离群值一般被认为是远离任何集群的点,作为聚类的副产物而被检测出来.

K-means 算法是入侵检测最常用的聚类算法,也是最为经典的基于划分的聚类方法.它的中心思想是,以空间中 k 个点为中心进行聚类,通过迭代的方法,逐次更新各聚类中心的值,直至得到最好的聚类结果.

1.3.2 OCSVM(One-Class SVM)

OCSVM[20]是入侵检测最常用的算法,在很难获取到离群点数据时,使用 OCSVM 对正常样本进行训练,可以将正常样本与离群点区分出来,类似于深度学习中的自编码器.

OCSVM 的优化目标为

$$\min_{\omega\in F,\xi\in\mathbf{R}^{\prime},\rho\in\mathbf{R}} \frac{1}{2}\|\omega\|^2+\frac{1}{vl}\sum_i\xi_i-\rho,$$
$$\text{s.t.}\ (\omega\cdot\varphi(x_i))\geqslant\rho-\xi_i,\ \xi_i\geqslant 0. \tag{1}$$

具体含义参见文献[20].

1.3.3 深度神经网络

深度神经网络(deep neural network, DNN)[21]是一种人工神经网络,其结构如图 3 所示,在输入层和输出层之间具有多个层,具有对复杂数据的强大拟合能力,因此被广泛应用于包括离群点检测在内的多个领域.

1.3.4 卷积神经网络

卷积神经网络(convolutional neural network, CNN)[22]是一种前馈神经网络,由一个或多个卷积层和最后的全连接层组成,在层之间也包含了池化层.

其中卷积层被用于提取数据的局部特征,池化层被用于数据降维,而全连接层与传统的神经网络一致.CNN 在图像领域具有极为出色的表现.

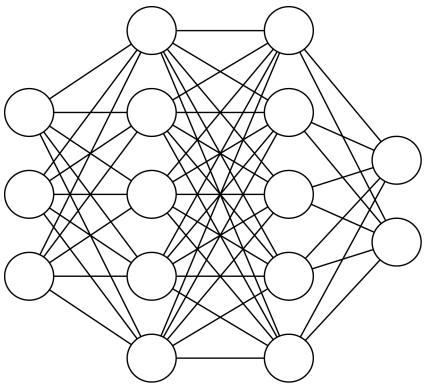


Fig. 3 Structure of DNN

图 3 常见的 DNN 结构

1.3.5 长短时记忆网络

长短时记忆网络(long short-term memory, LSTM)[23]是一种特殊的循环神经网络(recurrent neural network, RNN),不但继承了 RNN 处理序列数据的能力,而且解决了长序列训练过程中的梯度爆炸和梯度消失问题.

LSTM 在 RNN 的基础上,引入了"门"的机制,通过输入门、遗忘门和输出门,解决了长期依赖问题,使神经网络具有记忆距离较远的信息的能力.

1.3.6 深度 Q 网络

深度 Q 网络(deep Q network, DQN)[24]是一种深度强化学习方法,融合了深度神经网络与传统的 Q-Learning 强化学习方法.其核心思想是对 Q-Learning 中的 Q 值进行近似表示:

$$Q(s,a)\approx f(s,a,w), \tag{2}$$

其中 s 表示状态,a 表示行为,w 表示近似函数的参数.

DQN 的 $f(\cdot)$是深度神经网络,使用传统的 Q-Learning 算法估计的 Q 值为标签训练神经网络,来得到 Q 值的拟合,这样能够解决状态和行为数量过多的问题.

1.3.7 自编码器

自编码器(autoencoder, AE)[25]是一类在半监督学习和非监督学习中使用的人工神经网络,由一个编码器和一个解码器构成,常被应用于降维(dimensionality reduction)和异常值检测(anomaly detection).

1.4 常用数据集

通过大量的文献调研工作,本文总结出了 ICS IDS 评估常用的数据集,如表 1 所示:

Table 1 ICS IDS Datasets

表 1 工控系统的入侵检测系统数据集

数据集	良性	恶意	协议
WADI[26]	162 824	9 977	Modbus
SWaT[27]			Modbus
天然气管道数据集[28]	214 580	60 048	Modbus
EPIC[29]			Modbus
SCADA 网络数据集[30]	881 991	30 063	Modbus
UNSW-NB15[31]	93 000	164 673	IT
NSL-KDD[32]	90 503	83 206	IT

1.4.1 非公开数据集

研究人员从搭建的试验台仿真模拟得到数据或从真实的 ICS 系统中收集数据,用来评估提出的入侵检测模型的效果.

1.4.2 WADI 数据集[26]

WADI(water distribution)数据集是一个配水试验台的数据集,是通过正常运行 14 d 和异常运行 2 d 产生的(一共进行了 15 次攻击),数据来自于 123 个传感器和执行器,WADI 常用来对配水系统网络进行安全分析和评估检测.

1.4.3 SWaT 数据集[27]

SWaT(secure water treatment)是一个水处理系统试验台的数据集,通过正常运行 7 d 和在攻击场景下运行 4 d,从 51 个传感器和执行器收集得到,该数据集包含与工厂和水处理过程相关的物理属性,以及测试台上的网络流量,一共包含 946 722 个样本.

1.4.4 天然气管道数据集[28]

天然气管道数据集包含天然气管道系统试验台 SCADA 系统的网络流量和日志数据,包括正常运行的数据和遭受真实攻击的数据(如表 2 所示,包含 7 种类型攻击),该数据集使用 ARFF(attribute relationship file format)格式存储(包含 19 个特征).

Table 2 Attack Types and Description in Gas Pipeline Dataset[28]

表 2 天然气管道数据集中包含的攻击类型及其描述[28]

攻击类型	描述
NMRI	注入随机的响应包
CMRI	隐藏控制进程的真实状态
MSCI	注入恶意状态命令
MPCI	注入恶意参数命令
MFCI	注入恶意函数代码命令
DoS	拒绝服务攻击
Recon	收集网络信息、识别设备特征

1.4.5 EPIC 数据集[29]

EPIC(electric power and intelligent control)数据集是通过操作 EPIC 试验台在每个场景(共 8 个场景)下运行 30 min 产生的数据,数据包含传感器数据、执行器状态和网络流量数据.

1.4.6 SCADA 网络数据集[30]

SCADA 数据集生成自一个基于 SCADA 的小型电力网络的沙箱,通过真实的攻击工具来模拟恶意流量,数据只包含网络流量数据,该数据可以用来评估 SCADA 系统的入侵检测技术.

1.5 评估指标

为了对基于机器学习算法的 ICS IDS 进行有效评估,不仅需要有效可行的实验评估方法,还需要有衡量模型泛化能力的评价标准,也就是评估指标,我们总结了在 ICS IDS 的任务需求下常用来判别方法"好坏"的评估指标.这些指标可以分为 2 类:分类指标和应用指标.

1.5.1 分类指标

分类指标是由表 3 中的评估矩阵计算得到.

Table 3 Evaluation Matrix

表 3 评估矩阵

真实情况	预测结果	
	正例(Positive)	反例(Negative)
正例(True)	*TP*	*FN*
反例(False)	*FP*	*TN*

1) 准确率(accuracy, *ACC*)是最常见的评价指标,指正确预测的数量占总样本的比例.值越大,性能越好,其计算方法为

$$ACC=\frac{TP+TN}{TP+TN+FP+FN}. \tag{3}$$

2) 精确度(precision, *P*)反映分类器或者模型正确预测正样本精度的能力,即预测的正样本中有多少是真实的正样本.值越大,性能越好,计算为

$$P=\frac{TP}{TP+FP}. \tag{4}$$

3) 召回率(recall, *R*)反映分类器或者模型正确预测正样本全度的能力,即正样本被预测为正样本占总的正样本的比例.值越大,性能越好,计算为

$$R=\frac{TP}{TP+FN}. \tag{5}$$

4) 综合评价指标 *F*1(*F*-Measure)是精确度和

召回率的加权调和平均,用来综合考虑召回率和精确度.值越大,性能越好,计算为

$$F1=\frac{2\times P\times R}{P+R}. \tag{6}$$

5) 误报率(false positive rate, FPR)是将正常行为误判断为异常行为.由于 ICS 的高实时性,如果因为误报导致系统停止工作,会带来系统设备和资源的巨大损失,因此在 ICS IDS 中,在提高其他指标的同时也要保证误报率要尽可能低,甚至可以牺牲准确率,也不允许出现误报.值越小,性能越好,计算为

$$FPR=\frac{FP}{FP+TN}. \tag{7}$$

6) 错误率(error, ERR)与准确率相反,用来描述分类器错分的比例.值越小,性能越好,计算为

$$ERR=1-ACC=\frac{FP+FN}{TP+TN+FP+FN}. \tag{8}$$

7) 漏报率(false negative rate, FNR)反映分类器或者模型正确预测负样本纯度的能力,即正样本被预测为负样本占总的正样本的比例.值越小,性能越好,计算为

$$FNR=\frac{FN}{TP+FN}. \tag{9}$$

8) ROC 曲线是以 TPR 为纵轴、FPR 为横轴形成的曲线.ROC 曲线描述的其实是分类器性能随着分类器阈值的变化而变化的过程.

9) AUC 值为 ROC 曲线向下覆盖的面积值,其值越大,性能越好,是衡量机器学习分类器最常用的性能指标.

10) 冲突因子指数(conflict index factor, CiF)由 Gauthama Raman 等人[33]提出,作为检测准确率和误报率的平衡指标,充分考虑了这 2 个指标对模型性能的影响.CiF 可以更准确地评估检测性能.

1.5.2 基于 ICS 独特性的应用指标

基于机器学习算法的入侵检测系统的常用评估指标是分类指标,在 ICS 系统中,最重要的 2 个指标是检测率和误报率.与 IT 网络中的入侵检测系统不同,ICS 领域中的检测系统需要面临系统中存在的种种限制,比如数据维度高且噪声多、计算资源匮乏,以及对实时性要求高等,所以研究工作不仅要考虑模型的检测性能,更重要的是如何在保证检测指标不下降的同时高效地解决这些限制.在我们调研的大部分工作中,只追求高准确率和低误报率工作占比很大,但是仅凭这 2 个指标无法全面衡量复杂的工业控制系统的入侵检测系统在真正应用时的性能,所以很多研究工作使用了新的评估指标.

1) 检测时延[34]

检测延迟(time taken for detection, TTD)是指在受到攻击之后,检测到系统中异常所用的时间.对 ICS 的成功攻击可能会导致灾难性的故障,对国家经济甚至人类生命安全产生重大影响.因此,有必要尽早检测到由于攻击而产生的异常.检测延迟越低,性能越好.

2) 系统负载[35]

系统负载包括系统功耗、通信开销和处理器负载等指标.由于 ICS 中各种资源相对有限,所以 IDS 产生的负载要尽可能小,不能影响正常的系统运行.

3) 鲁棒性[36]

鲁棒性是指模型的抗干扰能力和能够适应复杂环境的能力.由于相对于实验室环境,真实的工业控制系统产生的数据包含很多噪声数据,所以模型需要具有过滤噪声或抵抗系统中噪声数据的能力.

4) 计算复杂度

由于工业控制系统计算资源相对有限,所以要求检测模型的计算复杂度应尽可能低,减少对系统其他功能的影响.

2 分类方法

本节我们介绍了不同的基于机器学习算法的 ICS IDS 的分类方法.

2.1 常见的分类方法

如图 4 所示,Mitchell 等人[37]和文献[38]对 ICS IDS 技术从检测技术和数据源 2 个角度进行了划分.根据检测技术的不同,ICS IDS 技术又可以分为基于知识的入侵检测和基于行为的入侵检测.基于知识的入侵检测技术通过特征匹配的方式,将检测到的数据与已知的攻击行为特征模式进行匹配来发现恶意行为.虽然这种检测技术误报率很低,但是这种行为只能检测到已知的攻击行为,无法检测未知

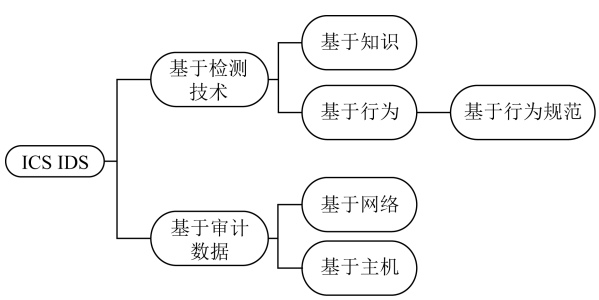


Fig. 4 A taxonomy for ICS IDS by ref [37]

图 4 文献[37]的 ICS IDS 分类方法

攻击.基于行为的入侵检测通过监测系统中的各种行为,与正常操作行为进行比较来发现不正常的行为,该技术检测效率很高,但是误报也很高,而且无法区分具体的攻击行为.这部分工作又有一部分需要通过基于行为规范的检测方法来对系统中的攻击进行更加精准的检测.

根据使用的数据来源的不同,ICS IDS 又可以分为基于网络流量的检测和基于主机的检测.基于网络流量的检测通过审计整个系统网络流量来发现网络异常现象.该方法的优势是不需要审计单个节点的网络流量和日志.但是由于无法检测每个节点的流量,合理安排检测点来反映整个系统的网络活动变得十分具有挑战性.基于主机的入侵检测主要监控系统设置、配置文件、应用程序和敏感文件,以发现系统的异常情况.

该分类方法只从数据和技术层面做了分类,没有考虑工业互联网自身具有的特性.该分类方式采用的还是传统互联网环境中入侵检测系统的分类方式.鉴于此,杨安等人[13]和 Hu 等人[39]基于对大量 ICS IDS 文献的分析和理解,考虑 ICS 本身的特点,从检测对象的角度出发提出了新的分类方法.根据 ICS 的特点,将 ICS 技术划分为三大类,分别是基于流量的检测、基于协议的检测和基于设备状态的检测.

基于流量的 IDS 依据 ICS 不同的安全区域流量特征,针对内部流量和外部流量,在不解析具体协议格式的情况下,发现异常流量.不同于传统的基于流量的检测方法,ICS 环境中设备所处的位置、功能相对固定,导致流量模式和流量特征比较稳定,所以基于流量的检测适用于 ICS 环境,可精确检测到系统中的攻击行为和异常流量.基于协议的 IDS 根据工业控制协议规范,采用协议格式和状态分析技术,对报文中的格式和协议状态进行检测,发现异常行为.基于设备状态的 IDS 根据业务逻辑和设备操作规范,通过定义正常的状态和异常的状态,根据状态转移趋势和监控操作序列等方法检测入侵行为.因为 ICS 存在大量的物理设备,而设备的状态基本上是稳定的,所以当系统存在恶意入侵时,通过检测设备状态的变化很容易发现系统中的异常,这也是不同于传统互联网环境的检测方式.

当把机器学习技术引入到工业互联网入侵检测研究中之后,工作的重点和方法都发生了改变,传统的分类方法不再适用.基于此,Wang 等人[40]提出了新的分类方法:基于数据的方法和基于工业互联网规范的方法,如图 5 所示.基于数据的方法是指将机器学习算法直接用在工业互联网数据集上,根据数据的不同又可细分为物理层数据和网络流数据.基于工业互联网规范的方法根据工业互联网的特点来区分正常和异常行为,根据使用的规范不同可以分为基于协议规范的方法、基于物理层规范的方法和基于混合规范的方法.

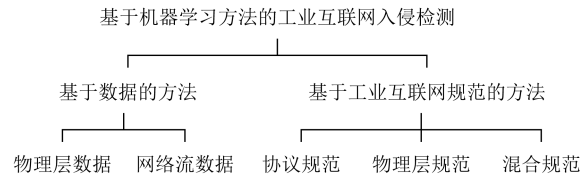


Fig. 5 A taxonomy for ICS IDS by ref [40]

图 5 文献[40]的分类方法

类似地,Gauthama Raman 等人[41]提出了以设计为中心和以数据为中心的方法.虽然这些方法都能很好地总结该领域的相关工作,但是这些分类方法专注在检测技术和数据上,视角过于局限,无法看到整个领域的发展状况,不能给以后的研究人员一个很明确的研究方向.基于此,我们需要跳出检测技术和数据的视角,从一个更加高的视角去看目前的工作,提出新的分类方法.

2.2 基于研究问题的分类方法

为了更好地对工作进行分类,我们首先对基于机器学习算法的工业互联网入侵检测技术研究过程做了总结和凝练.如图 6 所示,基于机器学习的入侵检测的研究问题如图 6 中最右侧部分所示,通常只需要关注数据预处理、特征的选择和提取,以及机器学习模型的选择 3 个问题.而图 6 中最左侧部分则是针对 ICS 攻击场景的研究问题,针对该问题的研究关注的是不同场景的架构、攻击路径以及数据特征之间的差异.而将基于机器学习的入侵检测技术应用到不同的工业互联网场景时,二者交叉就会产生图 6 中间部分所示的研究问题——如何解决应用时的限制和挑战的问题.根据研究工作中重点关注的研究问题的不同,可以分为 3 个研究点:机器学习模型的设计和选取、应用时限制和挑战的克服,以及针对不同 ICS 攻击场景的研究.我们的分类方法依据这 3 个研究点,将目前的研究工作分为面向算法设计、面向应用限制和挑战和面向特定 ICS 攻击场景 3 个类别.

面向算法设计的研究工作专注于算法的使用,创新使用机器学习算法解决检测过程中存在的问题.该类别的研究工作将 ICS IDS 看作一般的 IDS,

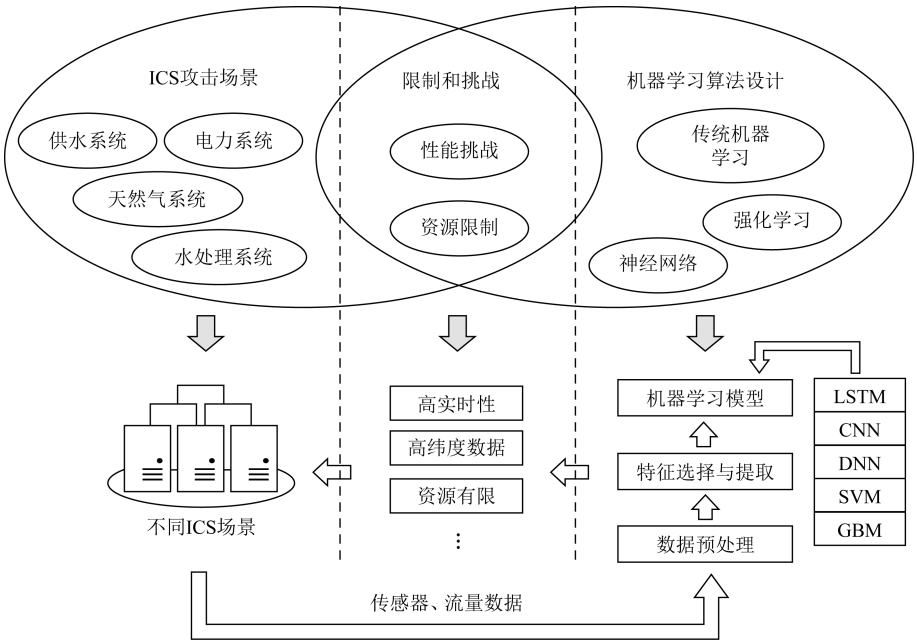


Fig. 6 Overall architecture of machine learning based ICS IDS

图 6 基于机器学习算法的 ICS IDS 整体架构

不会考虑 ICS 自身场景存在的应用限制和要求,比如速度、性能以及数据特点等,也不会考虑不同攻击场景的特点,如传感器和执行器的数量、ICS 的整体架构以及数据的采集位置等.该类别的方法一般以提高准确性、召回率、$F1$ 等通用的分类指标为主要目标.

面向应用限制和挑战的研究工作专注于将现有基于机器学习的入侵检测技术合理应用到系统中去,克服系统中的挑战和限制.比如 ICS 是一个具有高实时性的系统,一旦出现延迟就会导致突发情况,所以要求入侵检测系统计算复杂度低,消耗的计算资源要少,否则无论模型的准确率多高,都无法真正应用到真实的 ICS 当中.

攻击者在对 ICS 设施实施攻击时,往往会根据不同场景选择不同的攻击方式,比如面对智能电网系统时,常用的攻击方式就是虚假数据注入攻击.而不同攻击方式会对系统带来不同的影响,所以面向不同 ICS 攻击场景的研究专注发掘这些不同的攻击方式和攻击场景带来的不同特征,从而选择合适的方法进行检测.

3 个研究点是相辅相成的.面向算法的研究工作为其他工作提供更好的检测算法;面向应用的研究工作有利于其他工作落地到真实的系统中;面向 ICS 攻击场景的工作,更是为其他工作提供了重要的信息.研究人员通过确定自己的研究重心,可以更加明确自己的研究目标,让工作更加有针对性,这有利于整个研究领域的进步.

首先,与其他分类方法相比,本文提出的分类方法涵盖了工业控制系统入侵检测工作需要关注的 3 个重要问题,即检测模型的选择、检测模型在应用过程中所面临的限制和挑战的解决方法,以及针对不同工业控制系统攻击场景的入侵检测问题.目前其他分类方法主要聚焦在这些研究工作所使用的检测模型和数据上,而忽略了其他个重要的问题.其次,该分类方法充分体现了工业互联网入侵检测工作的特点,而其他分类方法的分类思路同普通互联网入侵检测工作分类思路是类似的.最后,该分类方法可以充分暴露目前研究工作存在的问题.比如通过该分类方法,我们可以看到,因为目前大部分工作都聚焦在新模型和新方法的设计上,模型的实验性能已经完全可以满足 ICS 的需求,而实际应用性能却不尽如人意,所以针对工业互联网场景及其限制的研究已经成为该研究领域的瓶颈.而其他分类方法很难发现该研究领域整体存在的问题,因此该分类方法相比其他较为传统的分类方法更加适合针对工业互联场景的入侵检测领域.

从图 6 看到,3 个大研究点可以细分分为更小的研究点,比如针对应用时的限制和挑战的研究,可以具体到针对工业控制系统高实时性要求的研究工

作、针对资源限制下的入侵检测研究工作等.接下来我们将按照该分类方法对目前的研究工作进行介绍和总结,发现目前工作中存在的问题并对未来的工作进行展望.

3 面向算法设计的入侵检测工作

在 ICS 中,通信模式是相对稳定的,通常是在重复相同的命令集合,这些规则模式可以被机器学习算法利用,通过数据特征构建相应的模型来区分正常行为和异常行为.面向算法设计的入侵检测相关工作大致可分为 3 个部分:数据处理、模型设计和实验评估.数据处理是指将公开数据集或者采集自模拟系统的数据转化为合适的特征向量,这个过程通常包括数据清洗、特征抽取和特征选择,选择合适的特征,可以减少数据中的噪声,提高分类效果.模型设计是指根据数据的特点选择或设计相应的分类模型对数据进行分类,合适的模型能够合理地利用数据中的信息,提高模型的性能.实验评估是为了证明模型的有效性而进行的一系列实验,通过合适的评估指标从各方面证明系统的性能.

3.1 传统机器学习算法

传统机器学习算法相比深度学习算法和强化学习算法,参数相对较少,对数据集合的大小要求较低,对计算资源的需求也小于后者,属于轻量级的算法.但是需要有应用领域知识的人从数据中提取合适的特征,传统机器学习模型算法的效果很大程度上依赖选取的特征向量.

传统的机器学习算法可以简单分为有监督学习和无监督学习,在目前很多工控系统入侵检测模型中最常用的无监督学习是 OCSVM 算法[42]和 K-means 聚类算法,由表 1 可知,工业控制系统的数据都具有较明显的数据不平衡的特点,正常数据要远大于异常数据,而 OCSVM 算法只需要一类数据就可以进行训练,所以很多工作如文献[12,43-46]都围绕着 OCSVM 算法提出不同的入侵检测模型,OCSVM 在入侵检测工作中可以看作是一种无监督学习算法,它不需要通过标记数据进行训练,在达到不错的检测效果的同时降低了检测的人工成本,但是目前基于 OCSVM 的分类器存在较高的误报率,而且无法检测到具体的攻击类型.

聚类算法是无监督学习最常用的算法,相比 OCSVM 算法,聚类算法可以检测多种攻击产生的异常,算法性能高且相对简洁.文献[47]探索了不同的聚类算法,选择了最适合对系统物理过程中产生的时间序列特征进行聚类的 K-means 算法.聚类算法非常依赖研究人员对数据的理解和特征向量的提取,当数据中存在大量噪声时,聚类算法很难得到很好的效果.

相比无监督学习算法,监督学习算法学习效率更高,入侵检测中常用的监督学习方法有 SVM[33,48-51]、贝叶斯网络[52-53]、Markov[54-57]等.SVM 同 OCSVM 类似,都只能区分异常和正常数据,缺乏精确检测攻击类型的能力,为此,文献[58]提出了基于多分类的 SVM 入侵检测方法,通过采用多个 SVM 模型结合的方法来实现精确检测多种攻击类型的能力.文献[53]使用了基于贝叶斯网络的模型来预测网络攻击对系统产生的影响,由于该方法没有考虑实际生产生活中的设备磨损和网络延迟,所以很难在真实环境中达到理想效果.Markov 过程常用来描述设备状态的变化,其算法的特点是可以发掘序列数据中的规律和特征,从而对该序列进行预测和分类,利用 Markov 算法善于处理序列数据的特点.文献[56]提出了基于隐 Markov 算法(hidden Markov model,HMM)的检测模型,该模型可以分为 2 个子系统:头部子系统和数据子系统,这 2 个子系统分别用来处理 Modbus 协议中头部和数据段的序列数据,每个子系统包含多个 HMM 分类器,当其中一个分类器检测到结果为异常时,该模型就会向 SCADA 系统发出警报,并报告异常结果,这种决策机制虽然可以提高检测率,但是会导致误报率上升.

3.2 神经网络学习算法

深度神经网络学习算法虽然结果更加精确,同时减少了特征提取的工作量,但是这些模型十分复杂和庞大,通常具有大量的参数,例如图像分类模型 Vgg19[2]就具有百万级别的参数量,这些算法不仅需要巨大的计算资源,而且要求数据集要足够大才能取得不错的分类效果.

CNN 是图像领域最常用的算法,通过卷积操作可以大幅减少模型的参数量[59-62].文献[61]将数据包序列作为输入,将数据流预处理成图像形式,通过将数据流视为一幅图像来检测正常数据包序列之间的共同特征,图像被交由 CNN 算法进行分类,进而判断出异常流量.

不同的数据结构适用的神经网络算法也不同,比如 CNN 常用于图像领域,而 RNN 及其变体 LSTM 则擅长处理序列数据,如文本等.因此根据入侵检测

工作中使用的特征,需要采用不同神经网络模型,将流量包看作一种序列数据,很多工作开始使用 RNN[63] 和 LSTM[64-68] 来处理这些序列数据.文献[64]同时使用了基于签名的包检测方法和基于 LSTM 的时间序列检测方法的 2 层检测模型,检测过程如图 7 所示.从图 7 看到,先通过基于 Bloom Filter 的检测器检测出部分异常流量,再通过基于 LSTM 的检测器检测出另一部分异常流量.但是该方法需要大量的训练集来更新模型参数才能达到不错的检测效果.

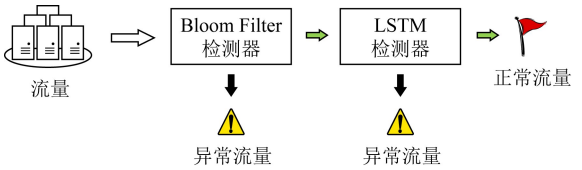


Fig. 7 Framework of the detection technique proposed in ref [64]

图 7 文献[64]中提到的检测技术整体框架

文献[65]提出了一种基于相关信息熵和 CNN-BiLSTM 的入侵检测模型,利用相关信息熵进行特征选择,减少特征的维度,然后再利用深度学习算法分类,该模型在提高准确率的同时可减少一定的噪声和计算量,但是研究人员没有对计算量这一指标进行量化分析和讨论.

在我们的调研中,AE[60,69-71] 也是工业控制系统入侵检测工作常用的算法之一.文献[70]提出了基于 LSTM-Autoencoder 的检测模型,自编码器的结构如图 8 所示,检测模型中的 LSTM 算法用于检测数据中的坏数据和缺失数据,剔除这些数据之后,再用 AE 模型去检测数据中是否存在异常行为.结合

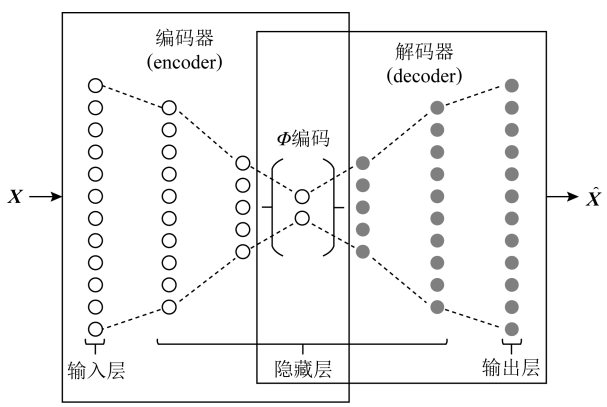


Fig. 8 Architecture of the autoencoder model in ref [70]

图 8 文献[70]中的自编码器结构

岭回归分类器的检测结果和网络日志,该模型可以定位异常事件产生的原因.

除了常见的神经网络算法,一些新型神经网络算法也被研究人员应用在工业互联网的入侵检测工作中.Demertzis 等人[72] 提出了 SOCCADF,它是一种基于进化型脉冲神经网络(evolving spiking neural network, eSNN)的入侵检测方法,只需要使用正例样本进行训练,具有检测异常行为的能力.He 等人[73] 提出基于 CDBN(conditional deep belief network)的方法实时检测智能电网系统中的虚假数据注入(false data injection, FDI)攻击,在模拟不同环境噪声实验中,该方法展现出了不错的抗干扰能力.

3.3 强化学习算法

强化学习不要求预先给定任何数据,而是通过接收环境对动作的奖励(反馈)获得学习信息并更新模型参数,强化学习模型具有独立判断决策的能力.

文献[74]和文献[75]分别将 Q-learning 和 Dyna-Q 应用到无线网络系统的欺骗检测攻击中,Dyna-Q 是 Q-Learning 利用 Dyna 架构构建的拓展版本,根据文献[75]中的实验结果,相比于 Q-Learning,Dyna-Q 具有更好的实用性和更高的学习速度.

文献[76]将深度学习算法近端策略优化(proximal policy optimization 2.0, PPO2)应用到工业互联网入侵检测工作中,相比普通 DQN 和 DDQN(double deep Q network)算法,基于 PPO2 的检测模型的准确率、召回率、精确度及 $F1$ 等指标都有提升.

4 面向应用挑战和限制的入侵检测工作

由于 ICS 本身存在诸多限制,基于机器学习算法的入侵检测技术在应用过程中会遇到很多限制和挑战,所以面向应用阶段的研究工作都是从应用时遇到的限制的角度出发,解决入侵检测技术在应用到实际 ICS 中遇到的挑战和问题.

4.1 减少检测延迟

工业控制系统本身具有实时性高、难以暂停的特点,不能因为检测的时间而导致系统停滞,所以检测延迟是入侵检测技术应用的一个重要指标,文献[77]比较了基于统计的检测技术(CUSUM 和 Bad-Data)和基于深度学习算法的检测技术(NoisePrint)的性能,结果发现基于深度学习算法的检测技术可以检测高级的攻击方式,但是其检测延迟在对实时性要求较高的 ICS 场景是无法忍受的,相比于深度学习

算法,统计算法对普通攻击检测率更高,而且具有很低的检测延迟(不超过 10 s).文献[54]在模型评估时使用检测延迟这一指标来证明其模型具有很高的实时性,可以满足真实 ICS 的需求,并提出将优化模型对资源和时间的消耗作为未来工作的重点.

4.2 降低计算复杂度

虽然很多工作都在尝试使用机器学习算法解决 ICS 中的入侵检测问题,但是大部分机器学习模型都需要大量的计算资源,而无法实际应用到真实的工业互联网场景中.传统降低计算复杂度的方法都是通过降低特征维度的方式进行,如 PCA 算法[60].而文献[66]提出了一种混合了机器学习和统计技术的检测模型,该模型具有很好的检测准确率和较低的计算复杂度,而且在真实的 ICS 中也可以有效检测网络攻击、恶意操作以及网络异常等事件.

为了降低神经网络算法带来的计算开销,文献[78-80]利用学习算法为正常的行为构造了一个模糊逻辑规则库,使用聚类方法直接从网络流量中提取模糊规则来描述不确定的事件和现象,这种方法本质上还是基于规则的方法,对计算资源的要求很低.文献[79]采用 TYPE-2 模糊逻辑,对输入数据进行模糊化,并将触发强度改成了一个范围值,解决了检测性能下降的问题.文献[80]则是在文献[79]的基础上提出了动态调节检测阈值的方法,提高了检测精度.

4.3 提高鲁棒性

在工业互联网场景中,由于设备更换、工作流程修改会引起 ICS 的行为发生变化,导致检测的效果下降,因此 Abdelaty 等人[59]提出了 AADS(adaptive anomaly detection in industrial control systems),AADS 可以使用少量的数据样本和梯度更新适应系统行为的新变化.Abdelaty 等人通过对测试数据加入噪声数据来验证 AADS 的鲁棒性,实验结果表明在噪声数据中 AADS 的 $F1$ 值保持稳定.与 Abdelaty 等人思路类似,文献[66]通过在训练时就考虑不同条件下会产生的噪声和扰动,并将其加入到训练数据中,提高模型的鲁棒性,也降低了模型误报率.

文献[81]和文献[82]考虑到控制器区域网络的不确定性,引入物理层特征,但是物理层特征的缺点是在不同环境中特征会发生变化,比如不同温度环境.文献[82]为了测试模型的稳定性,通过收集不同温度和不同时间的数据来对模型进行测试,实验表明之前的训练模型在不需要重新训练的情况下依旧可以保持相似的性能,文献[81]则是通过增量学习和减量学习的方式,增加新的学习样本和减少过时的样本来克服物理层的这种变化.

鲁棒性不仅体现在对抗噪声数据的能力方面,而且对抗样本攻击的出现也对机器学习模型的鲁棒性提出了新的要求.文献[61,83-86]中充分考虑了对抗样本攻击的影响,设计了相应的对抗攻击的实验.文献[60]提出基于一维 CNN 和 AutoEncoder 的 ICS 异常检测机制,采用一维 CNN 和 PCA 方法保证了模型不会占用太多计算资源,该模型在 WADI,BATADAL[87]和 SWaT 这 3 个数据集上表现出了出色的性能,而且检测时间也只有千分之一秒.为了证明该轻量化模型具有抵抗对抗样本攻击的能力,文献[60]还进行了对抗攻击实验,假设攻击者可以通过修改传感器的值来改变特征,可以通过梯度攻击的方式来离线生成恶意样本,实验结果表明,在数据噪声水平较高时,攻击可以绕过检测,但是无法实现攻击效果,当数据噪声水平较低时,攻击无法成功绕过检测,表明了该检测机制同时对对抗样本有很好的鲁棒性.

4.4 适应高维复杂的数据特点

相比于普通互联网中的数据,工业互联网的数据呈现出数据维度高,关联性强的特征.

文献[12]创新性地将 PU 学习(positive-unlabeled learning)应用到入侵检测系统中,针对工业系统数据维度高、关联性强的特征,通过 PU 学习的特征重要度计算方法进行特征选择,降低了特征的维度,但是 PU 学习对正例样本有更加严格的限制,现实数据很难保证正例样本和无标签样本的分布相同.

4.5 降低误报率

不仅检测延迟会影响工业控制系统的实时性,而且频繁的误报也会影响系统的实时性,对系统造成严重的后果.

为了减少误报,文献[48]通过增加训练数据的多样性来提高模型区分正常操作和攻击行为的能力,该方法不仅可以检测未知的攻击,而且在实际应用中也具有较低的延迟.在实际应用中,一旦误报产生,还可通过切换到一个相同控制器来降低误报可能带来的损失.不同于文献[48]的方法,文献[54]通过综合分析多领域的知识,提出一种基于多模型的检测方法,为了减少多模型检测方法的误报,通过引入基于 HMM 的警报分类模型,进一步区分真实

攻击和误报,具体过程如图 9 所示.图 9 中 CAD, NAD 和 AAD 分别表示不同的检测模型,实验结果表明,该方法有效减少了系统中产生的误报,并且拥有较低的检测延迟.图 9 中 λ 和 p 的下角标 f,a,n 分别表示误报(fault state)、攻击 (attack state)和正常(normal state).

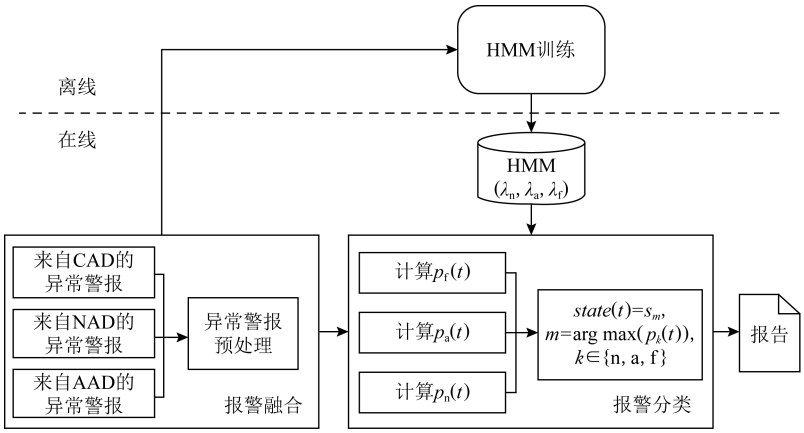


Fig. 9 Detection model proposed in ref [54]

图 9 文献[54]中的检测模型

5 面向特定 ICS 攻击场景的入侵检测工作

不同的 ICS 场景中,传感器和执行器的位置和数量都是不同的,ICS 的架构也有所不同,攻击者采取的攻击手段也会有所差异,比如在智能电网场景中虚假数据注入攻击和水处理系统中的虚假数据注入攻击的攻击路径和攻击目标是完全不一样的,对系统的影响也有很大区别,所以为了更有效地检测到特定 ICS 场景下的入侵行为,检测系统收集数据的节点和检测方法都应该是根据特定 ICS 场景而设置的,入侵检测工作要深入了解不同的 ICS 攻击场景,找到最优的探测节点来获取有效数据,结合机器学习算法,提高检测性能.

文献[73]针对智能电网系统中利用虚假数据注入攻击来窃取电力的场景提出了基于深度信念网络的检测方法.该工作对电网场景以及 FDI 攻击进行了深入调查和研究,为了更好地模拟攻击者行为,该工作对攻击者模型进行了假设,并构建了目标优化模型来模拟攻击者的行为,采用状态向量评估器(state vector estimator, SVE)和深度信念网络模型双层检测机制来检测该场景下的攻击行为,如图 10 所示,实时的测量数据首先经过 SVE 进行计算评估.当计算结果超过阈值 τ 时,模型就会判断为遭遇攻击;当小于阈值 τ 时,实时数据会传入已经训练好的 DLBI(deep-learning based identification)模型中进行分类,分类结果为 1 时判定为遭遇攻击,分类结果为 0 则为正常行为.

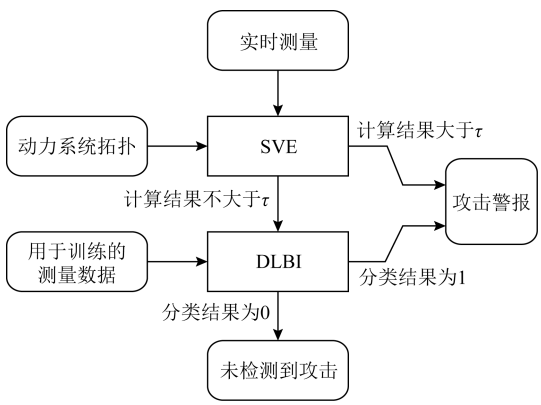


Fig. 10 Detection model proposed in ref [73]

图 10 文献[73]的检测模型

语义攻击场景近年逐渐受到了关注,语义攻击是需要攻击者对 ICS 中的协议、软件和硬件有深入了解的一种攻击类型.文献[55]针对一种特定类型语义攻击(序列攻击)场景展开深入研究,该类型攻击往往涉及序列事件,通过一系列合法的操作来破坏系统设施.针对该特点,文献[55]提出了基于离散时间 Markov 链(discrete-time Markov chains, DTMC)的检测框架,从网络流量、日志以及进程变量的变化中提取事件序列,构建 Markov 状态转移过程,并通过引入事件权重的方法,提高模型检测率的同时降低了误报率.文献[88]提出了攻击影响排名模型,采用

影响持续的时间、与正常行为之间的差距和阈值差3个指标来计算攻击方式对系统的影响程度,但是3个指标并不是独立影响的,它们之间的互相影响降低了最终评估结果的准确性.

文献[89]对多种攻击场景进行了深入研究,比如数据注入攻击、语义攻击等,从攻击原理出发,以攻击点和攻击方式为特征,对攻击策略进行了形式化描述,并对攻击产生的影响做了定量评估,最后通过基于状态转移概率的图的检测方法证明了评估方法的科学性和有效性.

Table 4 Comparisons of Research on Intrusion Detection of ICS Using Machine Learning

表4 基于机器学习的工业互联网入侵检测技术对比

来源	类别	机器学习算法	指标性能	数据或场景
文献[12]	面向应用	OCSVM PU学习	误报率、准确率、召回率、精确度以及$F1$	WADI
文献[33]	面向应用	SVM	准确率、检测率、误报率、稳定性	NSL-KDD,UNSW-NB
文献[42]	面向算法	LightGBM	准确率、精确度、召回率以及$F1$	天然气管道数据集
文献[43]	面向算法	OCSVM	准确率	流量数据
文献[44]	面向算法	OCSVM	准确率、精确度、召回率、误报率	非公开数据集
文献[45]	面向算法	贝叶斯网络	准确率、精确度、召回率、$F1$	天然气管道数据集
文献[46]	面向算法	OCSVM	检测率	非公开数据集
文献[47]	面向算法	K-means	RMSE,执行时间	流量数据
文献[48]	面向应用	SVM	反应器压强变化	非公开数据集
文献[49]	面向算法	SVM	检测率、RMSE	SWaT
文献[52]	面向算法	贝叶斯网络	精确度、运行时间	SWaT
文献[54]	面向应用	HMM	检测率(召回率)、检测时间、误报率、准确率	非公开数据集
文献[55]	面向场景	Markov算法	检测率、误报率	语义攻击
文献[59]	面向应用	CNN	精确度、召回率、$F1$以及鲁棒性	SWaT
文献[60]	面向应用	CNN,AE	精确度、召回率、$F1$、鲁棒性、检测时间	BATADAL,SWaT,WADI
文献[61]	面向算法	CNN	准确率、精确度、召回率、$F1$	非公开数据集
文献[63]	面向算法	RNN	误报率	SWaT
文献[64]	面向算法	LSTM	误报率、准确率、召回率、精确度以及$F1$	天然气管道数据集
文献[65]	面向算法	LSTM	准确率、误报率、漏报率	天然气管道数据集
文献[66]	面向应用	LSTM	R-Square、RMSE、检测时间、平均绝对比例误差(MAPE)	非公开数据集
文献[69]	面向算法	AE	准确率、精确度、召回率、$F1$	BATADAL[74]
文献[70]	面向算法	AE	准确率、精确度、召回率、$F1$、RMSE	非公开数据集
文献[72]	面向算法	eSNN	准确率、精确度、召回率、RMSE、$F1$以及ROC	天然气管道数据集
文献[73]	面向场景	CDBN	检测率、ROC	针对电网的FDI攻击
文献[76]	面向算法	强化学习	准确率、精确度、召回率记忆$F1$	天然气管道数据集
文献[77]	面向应用	NoisePrint	精确度、检测延迟	非公开数据集
文献[78]	面向应用	聚类	准确率、检测率、误报、处理时间	分公开数据
文献[89]	面向场景	概率图算法	准确率、检测率、误报率、$F1$	多种攻击场景
文献[90]	面向算法	K-means	准确率、召回率、准确度、误报率	UCI数据集
文献[91]	面向场景	奇异谱分析	检测时间、检测率、误报率以及对系统的影响	隐蔽攻击场景
文献[92]	面向应用	KNN	准确率、准确率、召回率、精确度、$F1$、灵敏度等	天然气管道数据集
文献[93]	面向算法	PNN	准确率、检测率、误报率、$F1$	SWaT
文献[94]	面向算法	SVM、随机森林	准确率、召回率、精确度、$F1$	天然气管道数据集
文献[95]	面向场景	偏度分析	计算时间和对系统的影响	隐蔽攻击场景

6 总结和展望

本节主要对目前研究工作中存在的问题以及未来的工作进行总结和展望.

6.1 当前工作存在的问题

当前的研究工作还有许多尚未很好解决的问题,这些问题的解决有利于将技术应用到现实 ICS 系统中,对未来的研究工作也具有很好的指导意义.

1) 缺乏全面的性能指标.传统的入侵检测模型一般用表 3 所示的 4 个数据(TN,FN,TP,FP)来评价模型性能,这 4 个数据可以刻画模型的准确率、精确度、召回率以及 $F1$ 值等评估指标.Gauthama Raman 等人[33]提出了冲突因子指数(CiF),作为检测准确率和误报率的权衡,这个指标可以更准确地评估检测性能.但是这些指标无法全面评估工业互联网中入侵检测模型的性能,目前的指标只专注于分类结果,并未考虑检测过程和检测环境的约束.

在 ICS 场景中,指标不应该只局限在分类指标上,研究人员应该基于 ICS 自身的特点和要求,发掘新的指标,从不同角度证明检测方法的先进性和有效性.

2) 数据问题.

① 由表 4 可知,大部分工作都是基于试验台模拟数据进行,少量工作选择在真实系统中运行测试模型效果.相比试验台模拟运行产生的数据,真实系统的数据的噪声比较多,而且攻击的种类和数量也有所不同,这就需要 IDS 模型在设计或训练的时候就要考虑到真实系统的特点,提高模型的鲁棒性.

② 随着 ICS 系统的智能化和复杂化,数据的规模和维度都会迅速增加,数据维度高会严重影响模型的处理速度,进而影响模型的实时性.针对该问题,传统的解决方法是使用聚类或降维的方式对数据进行处理,然后再使用机器学习方法进行检测分类,但是传统机器学习算法速度快,效果相对较差,不能很好地提取特征中的信息.深度神经网络算法具有强大的表征能力,可以胜任特征工程,但是深度神经网络算法训练过程复杂,需要消耗巨大的计算资源,难以达到 ICS 对实时性的要求,因此高维海量数据的处理是 ICS 领域应该着手解决的问题之一.

③ 由于异常事件的发生并非常态时间,所以系统中采集到的异常流量数据规模远远小于正常流量数据的规模,数据不平衡的问题会严重影响基于机器学习算法的 ICS IDS 的检测准确率.目前针对数据不平衡问题的研究较少,大部分工作还停留在采用欠采样、过采样等方法来缩小样本规模差距,但是这 2 种方法都会给模型带来新的问题,目前比较好的解决方式是通过 GAN 或强化学习的方式来生成异常数据,但是这方面的工作十分缺乏,所以数据不平衡问题一直是该领域存在的痛点之一.

3) 计算复杂度偏高.机器学习算法相比统计算法和规则法,计算复杂度偏高,但是 ICS 的计算资源相对有限,目前大部分工作在设计入侵检测系统时,过分注重模型分类性能,而忽略了系统计算资源的限制,导致检测模型失去实用性,研究人员需要充分考虑目标 ICS 的计算资源,设计合理的计算模型,通过数据降维等方式减少计算复杂度.

4) 概念偏移(concept drift).概念偏移是机器学习模型常见的问题之一,是指假如模型的目标变量为 y,当模型没有发生变化时,得到 y 的条件却发生了变化,导致模型不再适用的问题.按照变化的速度,概念漂移[96]又可以分为突变型、重复型、增量型以及渐变型,概念漂移可能会导致模型性能降低或失效,判断概念漂移的发生并及时更新模型对机器学习的应用的可靠性与可用性至关重要.近几年,概念漂移的检测工作已经引起了广泛的关注和参与,根据方法不同,可以分为基于错误率的检测和基于数据分布的检测.基于错误率的检测是指通过分类错误率上升时来判断是否触发进行漂移检测,如 DDM(drift detection method)[97].基于数据分布的漂移检测,通过计算数据分布的差异度来进行检测,不仅可以检测时间维度上的漂移,也可检测数据集内部的概念差异,但是这类方法通常需要消耗巨大的计算资源,比如 ITA[98].除了可以检测到模型的概念漂移,更重要的是,发生概念漂移后如何对模型进行调整,其中最简单有效的方法就是在新的数据对应关系下训练新的检测模型,但是对于 ICS 来说,设备无法暂停工作,难以进行版本更新.所以需要目前的工作专注于在 ICS 场景限制条件下的概念偏移的检测工作,以及如何在不停机的情况下实现自我更新等问题.

5) 模型评估和比较问题.由表 4 可知,大部分文献使用的数据集、评估指标各不相同,导致这些工作很难放到一起比较.大部分文献也只是展示了研究工作的优点,而忽略了工作中可能存在的问题和缺陷.所以很难发现不同模型存在的问题.另外,还存在指标本身不规范的问题,例如检测延迟这一指标,不同的研究工作使用的单位、定义以及计算方式都有所不同,而且所处的 ICS 环境也不同,这就导致虽然指标相同,但也难以放在一起比较.所以需要

进行评估指标的标准化,针对 ICS IDS 的每项性能都有唯一且确定的指标,研究人员需要针对这些繁杂的指标和数据集进行整理分析,对指标进行明确的定义,并对其计算过程进行规范.

6.2 展 望

基于目前机器学习相关工作的成果以及发展,我们对未来入侵检测工作的展望主要可以分为 2 部分:可解释机器学习的应用和对抗机器学习的应用.

6.2.1 解释学习

由于机器学习技术的不透明性,几乎所有的基于机器学习算法的入侵检测技术对于用户来说都是一个黑盒,这导致该技术在实际应用当中存在很大的不确定性.在工业控制系统等执行关键任务的系统的实际应用当中,采用机器学习技术的异常检测技术的精确度还达不到要求,所以通过解释学习的方式,增加机器学习检测过程的透明性,这对于提高基于机器学习技术的入侵检测系统的实用性具有重要的意义.比如文献[99]和文献[100]已经开始着手该方向的工作.文献[99]使用对抗的方法来解释正常和异常状态之间的差异;文献[100]专注于利用神经网络中输入特征和检测结果之间的相关性来提高透明度以增加用户信任.

6.2.2 对抗样本攻击

机器学习本身作为一个计算机系统,同样也具有安全漏洞.Szegedy 等人[101]首次提出了对抗样本攻击(adversarial example attack)的概念,通过在样本中加入微小的扰动就可以误导机器学习模型做出错误的选择.对抗攻击对于基于机器学习算法入侵检测系统来说是一个巨大的威胁,它为恶意的网络攻击者提供了强大的武器,尤其在工业互联网这种对安全相对敏感的领域,最近基于机器学习的 ICS IDS 也成了对抗攻击的对象[83-86,102-103].

在 ICS 中,基于机器学习的入侵检测模型的准确率并不是最重要的,重要的是,它不仅要提高模型对抗噪声的鲁棒性,面对对抗样本时的鲁棒性也同样重要.基于机器学习算法的入侵检测技术要在真实的 ICS 中部署,就要考虑如何防御对抗攻击的问题.图像领域研究工作者提出了一些对抗攻击的防御方法,这些方法大多可以作为借鉴参考迁移到入侵检测领域中.文献[104]阐述了一种基于生成对抗网络(generative adversarial network, GAN)框架的防御,将对抗样本转化为正常样本,降低对抗扰动的影响.文献[105]通过选择 2 个相同的模型作为教师模型和学生模型,将原始分类模型学到的信息迁移到小型网络模型中,实现了梯度遮掩.该方法可以有效抵抗一些基于梯度的小幅度扰动的对抗攻击.文献[106]提出了一种基于快速梯度符号法(fast gradient sign method, FGSM)的对抗训练方法,通过构建大量的对抗样本,将对抗样本混入训练样本中训练模型来增加模型的鲁棒性.

但是现有的对抗攻击防御措施也存在许多问题[107],比如基于识别对抗样本的防御措施存在拒绝合法样本的可能性,造成系统的可用性降低.基于对抗训练的防御措施不能从根本上加固目标,只能提高模型在遇到对抗样本时的鲁棒性,并不能完全消除安全隐患.这些问题是基于机器学习方法 ICS IDS 亟需解决的问题.

7 结束语

随着工业领域与互联网的融合,越来越多的安全问题被暴露出来.入侵检测技术逐渐成为应对这些问题的主要手段,随着机器学习技术在其他领域的成功,机器学习技术也逐渐被应用到工业互联网入侵检测系统中.但是由于工业互联网的自身特性,研究人员不仅需要了解机器学习技术,更要对 ICS 的攻击场景有深入的了解,在此基础上合理设计机器学习模型解决工业互联网入侵检测系统存在的问题和挑战.

本文调研了基于机器学习的工业互联网入侵检测技术的相关研究,分析了不同研究工作存在的优势和不足,并对这些工作中使用的数据集进行了整理,总结了传统互联网和工业互联网入侵检测工作的不同之处,指出了基于机器学习的工业互联网入侵检测技术的研究过程独特性,并提出了面向研究重点的分类方法.通过总结分析,我们发现,大部分工作过分侧重于使用新算法提高模型的准确性,而忽略了 ICS 攻击场景的特殊性以及模型应用到 ICS 中的一些困难和挑战.因此,后续工作应该在这 3 个研究方向上齐头并进,相辅相成.没有好的算法,就无法提高入侵检测系统的检测效果.忽略 ICS 场景的不同特点和应用时的条件的限制,也往往会导致模型无法在实际应用中发挥好的效果.本文最后总结了今后工作中应该改进的问题,并提出了 2 个未来非常具有前景的研究方向.

作者贡献声明:刘奇旭负责论文总体规划设计、论文的撰写和校对修改工作;陈艳辉负责工业互联网入侵检测工作相关文献的调研、阅读和整理,撰写论文以及校对最终论文格式和内容;尼杰硕负责文献的收集整理、对抗样本和解释机器学习相关知识

的调研和整理、文献引用格式的规范以及整体论文格式的校对;罗成负责工业控制系统架构、协议等相关知识的调研工作;柳彩云负责工业控制系统架构、协议等相关知识的调研工作;曹雅琴负责机器学习算法调研和论文内容的校对工作;谭儒负责评价指标的梳理及FGSM(fast gradient sign method)论文内容的校对工作;冯云负责数据集的调研和梳理以及论文内容的校对工作;张越负责已发表相关综述论文中分类方法的调研和整理,以及论文内容的校对工作.

参 考 文 献

[1] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint, arXiv:1409.1556, 2014

[3] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions [C] //Proc of 2015 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2015: 1-9

[4] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C] //Proc of the 32nd Int Conf on Machine Learning(ICML'15). New York: ACM, 2015: 448-456

[5] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C] //Proc of 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 770-778

[6] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions [C] //Proc of the 2011 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2011: 151-161

[7] Mueller J, Thyagarajan A. Siamese recurrent architectures for learning sentence similarity [C] //Proc of the 30th AAAI Conf on Artificial Intelligence(AAAI'16). Menlo Park, CA: AAAI Press, 2016: 2786-2792

[8] Peng Hao, Li Jianxin, He Yu, et al. Large-scale hierarchical text classification with recursively regularized deep graph-CNN [C] //Proc of the 2018 World Wide Web Conf (WWW'18). Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2018: 1063-1072

[9] Turc I, Chang M W, Lee K, et al. Well-read students learn better: On the importance of pre-training compact models [J]. arXiv preprint, arXiv:1908.08962, 2019

[10] Ma Xuezhe, Hovy E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1064-1074

[11] Wu S X, Banzhaf W. The use of computational intelligence in intrusion detection systems: A review [J]. Applied Soft Computing, 2010, 10(1): 1-35

[12] Lü Sicai, Zhang Ge, Zhang Yaofang, et al. A PU learning intrusion detection method for industrialcontrol system [J]. Journal of Cyber Security, 2021, 6(4): 72-89 (in Chinese)
(吕思才, 张格, 张耀方, 等. 一种面向工控系统的PU学习入侵检测方法[J]. 信息安全学报, 2021, 6(4): 72-89)

[13] Yang An, Sun Limin, Wang Xiaoshan, et al. Intrusion detection techniques for industrial control systems [J]. Journal of Computer Research and Development, 2016, 53(9): 2039-2054 (in Chinese)
(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)

[14] Stouffer K, Falco J, Scarfone K. Guide to industrial control systems (ICS) security [S]. Gaithersburg, MD: NIST Special Publication, 2007

[15] Cheminod M, Durante L, Valenzano A. Review of security issues in industrial networks [J]. IEEE Transactions on Industrial Informatics, 2012, 9(1): 277-293

[16] HopKinton. Modbus Application Protocol Specification [S]. Andover, MA: The Modbus Organization, 2006

[17] Cheng Ling. Study and application of DNP3. 0 in SCADA system [C] //Proc of 2011 Int Conf on Electronic & Mechanical Engineering and Information Technology. Piscataway, NJ: IEEE, 2011: 4563-4566

[18] Yao Heping. Power System Computer Network Communication Protocol-ICCP [J]. Automation of Electric Power Systems, 1996, 20(2): 49-53 (in Chinese)
(姚和平. 电力系统计算机网络通信协议－ICCP [J]. 电力系统自动化, 1996, 20(2): 49-53)

[19] MacQueen J. Some methods for classification and analysis of multivariate observations [C] //Proc of the 5th Berkeley Symp on Mathematical Statistics and Probability. Oakland, CA: University of California Press, 1967: 281-297

[20] Schölkopf B, Williamson R C, Smola A J, et al. Support vector method for novelty detection [C] //Proc of the 12th Int Conf on Neural Information Processing Systems(NIPS'99). Cambridge, MA: MIT press, 1999: 582-588

[21] Bengio Y. Learning deep architectures for AI [J]. Foundations and Trends® in Machine Learning, 2009, 2(1): 1-127

[22] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] //Proc of the 2015 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2015: 3431-3440

[23] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780

[24] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning [J]. arXiv preprint, arXiv:1312.5602, 2013

[25] Li S, Kawale J, Fu Y. Deep collaborative filtering via marginalized denoising auto-encoder [C] //Proc of the 24th ACM Int on Conf on Information and Knowledge Management (CIKM'15). New York: Association for Computing Machinery, 2015: 811-820

[26] Ahmed C M, Palleti V R, Mathur A P. Wadi: A water distribution testbed for research in the design of secure cyber physical systems [C] //Proc of the 3rd Int Workshop on Cyber-Physical Systems for Smart Water Networks. New York: ACM, 2017: 25-28

[27] Goh J, Adepu S, Junejo K N, et al. A dataset to support research in the design of secure water treatment systems [C] //Proc of the 11th Int Conf on Critical Information Infrastructures Security. Berlin: Springer, 2016: 88-99

[28] Morris T, Gao Wei. Industrial control system traffic data sets for intrusion detection research [C] //Proc of Int Conf on Critical Infrastructure Protection. Berlin: Springer, 2014: 65-78

[29] Ahmed C M, Kandasamy N K. A comprehensive dataset from a smart grid testbed for machine learning based CPS security research [C] //Proc of Int Workshop on Cyber-Physical Security for Critical Infrastructures Protection. Berlin: Springer, 2020: 123-135

[30] Lemay A, Fernandez J M. Providing SCADA network data sets for intrusion detection research [C] //Proc of the 9th Workshop on Cyber Security Experimentation and Test (CSET 16). Berkeley,CA: USENIX Association, 2016: 1-8

[31] Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C] //Proc of 2015 Military Communications and Information Systems Conf (MilCIS). Piscataway, NJ: IEEE, 2015: 1-6

[32] Tavallaee M, Bagheri E, Lu Wei, et al. A detailed analysis of the KDD CUP 99 data set [C] //Proc of 2009 IEEE Symp on Computational Intelligence for Security and Defense Applications. Piscataway, NJ: IEEE, 2009: 1-6

[33] Gauthama Raman M R, Somu N, Jagarapu S, et al. An efficient intrusion detection technique based on support vector machine and improved binary gravitational search algorithm [J]. Artificial Intelligence Review, 2020, 53(5): 3255-3286

[34] Striki M, Manousakis K, Kindred D, et al, Quantifying resiliency and detection latency of intrusion detection structures [C] //Proc of 2009 IEEE Military Communications Conf(MILCOM 2009). Piscataway, NJ: IEEE, 2009: 1-8

[35] Misra S, Krishna P V, Abraham K I. Energy efficient learning solution for intrusion detection in wireless sensor networks [C] //Proc of 2010 2nd Int Conf on Communication Systems and Networks (COMSNETS 2010). Piscataway, NJ: IEEE, 2010: 1-6

[36] Madani P, Vlajic N. Robustness of deep autoencoder in intrusion detection under adversarial contamination [C] // Proc of the 5th Annual Symp and Bootcamp on Hot Topics in the Science of Security (HoTSoS'18). New York: ACM, 2018: 1-8

[37] Mitchell R, Chen I R. A survey of intrusion detection techniques for cyber-physical systems [J]. ACM Computing Surveys, 2014, 46(4): Article 55

[38] Iturbe M, Garitano I, Zurutuza U, et al. Towards large-scale, heterogeneous anomaly detection systems in industrial networks: A survey of current trends [J]. Security and Communication Networks, 2017, 2017: Article ID 9150965

[39] Hu Yan, Yang An, Li Hong, et al. A survey of intrusion detection on industrial control systems [J]. International Journal of Distributed Sensor Networks, 2018, 14(8): 1-14

[40] Wang Qian, Chen He, Li Yonghui, et al. Recent advances in machine learning-based anomaly detection for industrial control networks [C] //Proc of 2019 1st Int Conf on Industrial Artificial Intelligence (IAI). Piscataway, NJ: IEEE, 2019: 1-6

[41] Gauthama Raman M R, Ahmed C M, Mathur A. Machine learning for intrusion detection in industrial control systems: Challenges and lessons from experimental evaluation [J]. Cybersecurity, 2021, 4(1): 1-12

[42] Li Lin, Shang Wenli, Yao Jun, et al. Overview of one-class support vector machine in intrusion detection of industrial control system [J]. Application Research of Computer, 2016, 33(1): 7-11 (in Chinese)
(李琳, 尚文利, 姚俊, 等. 单类支持向量机在工业控制系统入侵检测中的应用研究综述[J]. 计算机应用研究, 2016, 33(1): 7-11)

[43] Maglaras L A, Jiang Jianmin. Intrusion detection in scada systems using machine learning techniques [C] //Proc of 2014 Science and Information Conf. Piscataway, NJ: IEEE, 2014: 626-631

[44] Silva E G, Silva A S, Wickboldt J A, et al. A one-class nids for sdn-based scada systems [C] //Proc of 2016 IEEE 40th Annual Computer Software and Applications Conf (COMPSAC). Piscataway, NJ: IEEE, 2016: 303-312

[45] Ullah I, Mahmoud Q H. A hybrid model for anomaly-based intrusion detection in scada networks [C] //Proc of 2017 IEEE Int Conf on Big Data (Big Data). Piscataway, NJ: IEEE, 2017: 2160-2167

[46] Jiang Jianmin, Yasakethu L. Anomaly detection via one class SVM for protection of scada systems [C] //Proc of 2013 Int Conf on Cyber-Enabled Distributed Computing and Knowledge Discovery. Piscataway, NJ: IEEE, 2013: 82-88

[47] Kiss I, Genge B, Haller P, et al. Data clustering-based anomaly detection in industrial control systems [C] //Proc of 2014 IEEE 10th Int Conf on Intelligent Computer Communication and Processing (ICCP). Piscataway, NJ: IEEE, 2014: 275-281

[48] Keliris A, Salehghaffari H, Cairl B, et al. Machine learning-based defense against process-aware attacks on industrial control systems [C] //Proc of 2016 IEEE Int Test Conf (ITC). Piscataway, NJ: IEEE, 2016: 1-10

[49] Ahmed C M, Zhou Jianying, Mathur A P. Noise matters: Using sensor and process noise fingerprint to detect stealthy cyber attacks and authenticate sensors in CPS [C] //Proc of the 34th Annual Computer Security Applications Conf (ACSAC'18). New York: ACM, 2018: 566-581

[50] Nader P, Honeine P, Beauseroy P. Lp-norms in one-class classification for intrusion detection in scada systems [J]. IEEE Transactions on Industrial Informatics, 2014, 10(4): 2308-2317

[51] Leahy K, Hu R L, Konstantakopoulos I C, et al. Diagnosing wind turbine faults using machine learning techniques applied to operational data [C] //Proc of 2016 IEEE Int Conf on Prognostics and Health Management (ICPHM). Piscataway, NJ: IEEE, 2016: 1-8

[52] Lin Qin, Adepu S, Verwer S, et al. Tabor: A graphical model-based approach for anomaly detection in industrial control systems [C] //Proc of the 2018 on Asia Conf on Computer and Communications Security(ASIACCS'18). New York: ACM, 2018: 525-536

[53] Huang Kaixing, Zhou Chunjie, Tian Yuchu, et al. Assessing the physical impact of cyberattacks on industrial cyber-physical systems [J]. IEEE Transactions on Industrial Electronics, 2018, 65(10): 8153-8162

[54] Zhou Chunjie, Huang Shuang, Xiong Naixue, 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

[55] Caselli M, Zambon E, Kargl F. Sequence-aware intrusion detection in industrial control systems [C] //Proc of the 1st ACM Workshop on Cyber-Physical System Security (CPSS'15). New York: ACM, 2015: 13-24

[56] Stefanidis K, Voyiatzis A G. An HMM-based anomaly detection approach for SCADA systems [C] //Proc of IFIP Int Conf on Information Security Theory and Practice. Berlin: Springer, 2016: 85-99

[57] Caselli M, Zambon E, Petit J, et al. Modeling message sequences for intrusion detection in industrial control systems [C] //Proc of Int Conf on Critical Infrastructure Protection. Berlin: Springer, 2015: 49-71

[58] Luo Yaofeng. Research and design of intrusion detection method for industrial control system [D]. Hangzhou: Zhejiang University, 2013 (in chinese)
(罗耀锋. 面向工业控制系统的入侵检测方法的研究与设计[D]. 杭州: 浙江大学, 2013)

[59] Abdelaty M, Doriguzzi-Corin R, Siracusa D. AADS: A noise-robust anomaly detection framework for industrial control systems [C] //Proc of Int Conf on Information and Communications Security. Berlin: Springer, 2019: 53-70

[60] Kravchik M, Shabtai A. Efficient cyber attack detection in industrial control systems using lightweight neural networks and PCA [J/OL]. IEEE Transactions on Dependable and Secure Computing, 2021: 1-1. [2021-10-12]. https://ieeexplore.ieee.org/abstract/document/9317834

[61] Song J Y, Paul R, Yun J H, et al. CNN-based anomaly detection for packet payloads of industrial control system [J]. International Journal of Sensor Networks, 2021, 36(1): 36-49

[62] Kravchik M, Shabtai A. Detecting cyber attacks in industrial control systems using convolutional neural networks [C] // Proc of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy (CPS-SPC'18). New York: ACM, 2018: 72-83

[63] Goh J, Adepu S, Tan M, et al. Anomaly detection in cyber physical systems using recurrent neural networks [C] //Proc of 2017 IEEE 18th Int Symp on High Assurance Systems Engineering (HASE). Piscataway, NJ: IEEE, 2017: 140-145

[64] Feng Cheng, Li Tingting, Chana D. Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks [C] //Proc of 2017 47th Annual IEEE/IFIP Int Conf on Dependable Systems and Networks (DSN). Piscataway, NJ: IEEE, 2017: 261-272

[65] Shi Leyi, Zhu Hongqiang, Liu Yihao, et al. Intrusion detection of industrial control system based on correlation information entropy and CNN-BiLSTM [J]. Journal of Computer Research and Development, 2019, 56(11): 2330-2338 (in Chinese)
(石乐义, 朱红强, 刘祎豪, 等. 基于相关信息熵和CNN-BiLSTM的工业控制系统入侵检测[J]. 计算机研究与发展, 2019, 56(11): 2330-2338)

[66] Hao Weijie, Yang Tao, Yang Qiang. Hybrid statistical-machine learning for real-time anomaly detection in industrial cyber-physical systems [J/OL]. IEEE Transactions on Automation Science and Engineering, 2021 (99): 1-15. [2021-10-12]. https://ieeexplore.ieee.org/abstract/document/9424948

[67] Filonov P, Lavrentyev A, Vorontsov A. Multivariate industrial time series with cyber-attack simulation: Fault detection using an LSTM-based predictive data model [J]. arXiv preprint, arXiv:1612.06676, 2016

[68] Wu Di, Jiang Zhongkai, Xie Xiaofeng. LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT [J]. IEE Transactions on Industrial Informatics, 2020, 16(8): 5244-5253

[69] Taormina R, Galelli S. Deep-learning approach to the detection and localization of cyber-physical attacks on water distribution systems [J]. Journal of Water Resources Planning and Management, 2018, 144(10): 04018065

[70] Ahmed A, Krishnan V V, Foroutan S A, et al. Cyber physical security analytics for anomalies in transmission protection systems [J]. IEEE Transactions on Industry Applications, 2019, 55(6): 6313-6323

[71] Gauthama Raman M G, Dong Wenjie, Mathur A. Deep autoencoders as anomaly detectors: Method and case study in a distributed water treatment plant [J]. Computers & Security, 2020, 99(2): 102055

[72] Demertzis K, Iliadis L, Spartalis S. A spiking one-class anomaly detection framework for cyber-security on industrial control systems [C] //Proc of Int Conf on Engineering Applications of Neural Networks. Berlin: Springer, 2017: 122-134

[73] He Youbiao, Mendis G J, Wei Jin. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism [J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2505-2516

[74] Xiao Liang, Li Yan, Liu Guolong, et al. Spoofing detection with reinforcement learning in wireless networks [C] //Proc of 2015 IEEE Global Communications Conf (GLOBECOM). Piscataway, NJ: IEEE, 2015: 1-5

[75] Xiao Liang, Li Yan, Han Guoan, et al. PHY-layer spoofing detection with reinforcement learning in wireless networks [J]. IEEE Transactions on Vehicular Technology, 2016, 65 (12): 10037-10047

[76] Li Beibei, Song Jiarui, Du Qingyun, et al. DEL-IDS: Industrial Iot intrusion detection system based on deep reinforcement learning [J]. Computer Science, 2021, 48(7): 47-54 (in Chinese)
(李贝贝, 宋佳芮, 杜卿芸, 等. DRL-IDS: 基于深度强化学习的工业物联网入侵检测系统[J]. 计算机科学, 2021, 48 (7): 47-54)

[77] Athalye S, Ahmed C M, Zhou Jianying. A tale of two testbeds: A comparative study of attack detection techniques in CPS [C] //Proc of Int Conf on Critical Information Infrastructures Security. Berlin: Springer, 2020: 17-30

[78] Linda O, Manic M, Vollmer T, et al. Fuzzy logic based anomaly detection for embedded network security cyber sensor [C] //Proc of 2011 IEEE Symp on Computational Intelligence in Cyber Security (CICS). Piscataway, NJ: IEEE, 2013: 202-209

[79] Linda O, Manic M, Alves-Foss J, et al. Towards resilient critical infrastructures: Application of type-2 fuzzy logic in embedded network security cyber sensor [C] //Proc of 2011 4th Int Symp on Resilient Control Systems. Piscataway, NJ: IEEE, 2011: 26-32

[80] Linda O, Manic M, Vollmer T. Improving cyber-security of smart grid systems via anomaly detection and linguistic domain knowledge [C] //Proc of 2012 5th Int Symp on Resilient Control Systems. Piscataway, NJ: IEEE, 2011: 48-54

[81] Choi W, Joo K, Jo H J, et al. Voltageids: Low-level communication characteristics for automotive intrusion detection system [J]. IEEE Transactions on Information Forensics and Security, 2018, 13(8): 2114-2129

[82] Kneib M, Huth C. Scission: Signal characteristic-based sender identification and intrusion detection in automotive networks [C] //Proc of the 2018 ACM SIGSAC Conf on Computer and Communications Security (CCS'18). New York: ACM, 2018: 787-800

[83] Li Dan, Chen Dacheng, Jin Baihong, et al. Mad-GAN: Multivariate anomaly detection for time series data with generative adversarial networks [C] //Proc of Int Conf on Artificial Neural Networks. Berlin: Springer, 2019: 703-716

[84] Erba A, Taormina R, Galelli S, et al. Real-time evasion attacks with physical constraints on deep learning-based anomaly detectors in industrial control systems [J]. arXiv preprint, arXiv:1907.07487, 2019

[85] Ghafouri A, Vorobeychik Y, Koutsoukos X. Adversarial regression for detecting attacks in cyber-physical systems [C] //Proc of the 27th Int Joint Conf on Artificial Intelligence (IJCAI'18). Menlo Park, CA: AAAI, 2018: 3769-3775

[86] Feng Cheng, Li Tingting, Zhu Zhanxing. A deep learning-based framework for conducting stealthy attacks in industrial control systems [J]. arXiv preprint, arXiv:1709.06397, 2017

[87] Riccardo T, Stefano G, Nils Ole N, et al. Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks [J]. Journal of Water Resources Planning and Management, 2018, 144(8): 04018048

[88] Li Weize, Xie Lun, Deng Zulan, et al. False sequential logic attack on SCADA system and its physical impact analysis [J]. Computers & Security, 2016, 58: 149-159

[89] Xu Lijuan, Wang Bailing, Yang Meihong, et al. Multi-mode attack detection and evaluation of abnormal states for industrial control network [J]. Journal of Computer Research and Development, 2021, 58(11): 2333-2349 (in Chinese)
(徐丽娟, 王佰玲, 杨美红, 等, 工业控制网络多模式攻击检测及异常状态评估方法[J]. 计算机研究与发展, 2021, 58 (11): 2333-2349)

[90] Almalawi A, Yu Xinghuo, Tari Z, et al. An unsupervised anomaly-based detection approach for integrity attacks on scada systems [J]. Computers & Security, 2014, 46: 94-110

[91] Aoudi W, Iturbe M, Almgren M. Truth will out: Departure-based process-level detection of stealthy attacks on control systems [C] //Proc of the 2018 ACM SIGSAC Conf on Computer and Communications Security (CCS'18). New York: ACM, 2018: 817-831

[92] Khan I A, Pi D, Khan Z U, et al. HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems [J]. IEEE Access, 2019, 7: 89507-89521

[93] Gauthama Raman M G, Somu N, Mathur A P. Anomaly detection in critical infrastructure using probabilistic neural network [C] //Proc of Int Conf on Applications and Techniques in Information Security. Berlin: Springer, 2019: 129-141

[94] Anton S D D, Sinha S, Schotten H D. Anomaly-based intrusion detection in industrial data with SVM and random forests [C] //Proc of 2019 Int Conf on Software, Telecommunications and Computer Networks (SoftCOM). Piscataway, NJ: IEEE, 2019: 1-6

[95] Hu Yan, Li Hong, Luan Tom H, et al. Detecting stealthy attacks on industrial control systems using a permutation entropy-based method [J]. Future Generation Computer Systems, 2020, 108: 1230-1240

[96] Dongre P B, Malik L G. A review on real time data stream classification and adapting to various concept drift scenarios [C] //Proc of 2014 IEEE Int Advance Computing Conf (IACC). Piscataway, NJ: IEEE, 2014: 533-537

[97] Lu Ning, Lu Jie, Zhang Guangquan, et al. A concept drift-tolerant case-base editing technique [J]. Artificial Intelligence, 2016, 230: 108-133

[98] Dasu T, Krishnan S, Venkatasubramanian S, et al. An information-theoretic approach to detecting changes in multi-dimensional data streams [C] //Proc of Symp on the Interface of Statistics, Computing Science, and Applications. Fairfax, VA: Interface Foundation of North America, 2006: 1-24

[99] Marino D L, Wickramasinghe C S, Milos M. An adversarial approach for explainable AI in intrusion detection systems [C] //Proc of the 44th Annual Conf of the IEEE Industrial Electronics Society (IECON 2018). Piscataway, NJ: IEEE, 2018: 3237-3243

[100] Amarasinghe K, Milos M. Improving user trust on deep neural networks based intrusion detection systems [C] // Proc of the 44th Annual Conf of the IEEE Industrial Electronics Society(IECON 2018). Piscataway, NJ: IEEE, 2018: 3262-3268

[101] Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks [J]. arXiv preprint, arXiv: 1312.6199, 2013

[102] Kravchik M, Biggio B, Shabtai A. Poisoning attacks on cyber attack detectors for industrial control systems [C] // Proc of the 36th Annual ACM Symp on Applied Computing (SAC'21). New York: ACM, 2021: 116-125

[103] Zizzo G, Hankin C, Maffeis S, et al. Adversarial attacks on time-series intrusion detection for industrial control systems [C] //Proc of 2020 IEEE 19th Int Conf on Trust, Security and Privacy in Computing and Communications (TrustCom). Piscataway, NJ: IEEE, 2020: 899-910

[104] Samangouei P, Kabkab M, Chellappa R. Defense-GAN: Protecting classifiers against adversarial attacks using generative models [J]. arXiv preprint, arXiv: 1805.06605, 2018

[105] Papernot N, McDaniel P, Wu Xi, et al. Distillation as a defense to adversarial perturbations against deep neural networks [C] //Proc of 2016 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2016: 582-597

[106] Kurakin A, Goodfellow I, Bengio S. Adversarial machine learning at scale [J]. arXiv preprint, arXiv: 1611.01236, 2016

[107] Zhang Yuqing, Dong Ying, Liu Caiyun, et al. Situation, trends and prospects of deep learning applied to cyberspace security [J]. Journal of Computer Research and Development, 2018, 55(6): 1117-1142 (in Chinese)
(张玉清, 董颖, 柳彩云, 等. 深度学习应用于网络空间安全的现状、趋势与展望[J]. 计算机研究与发展, 2018, 55(6): 1117-1142)

Liu Qixu, born in 1984. PhD, professor, PhD supervisor. His main research interests include network attack and defense technology, cyber-attacks discovery, attribution and forensic.
刘奇旭, 1984 年生, 博士, 教授, 博士生导师. 主要研究方向为网络攻防技术、网络攻击发现和溯源取证.

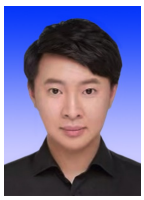
Chen Yanhui, born in 1996. PhD candidate. His main research interests include network attack and defense, malware and machine learning.
陈艳辉, 1996 年生. 博士研究生. 主要研究方向为网络攻防、恶意软件和机器学习.

Ni Jieshuo, born in 1997. Master candidate. His main research interests include Web attack and defense technology, and data analysis.
尼杰硕, 1997 年生. 硕士研究生. 主要研究方向为 Web 攻防技术、数据分析.

Luo Cheng, born in 1988. Master, engineer. His main research interests include network security, IoT security, ICS security.
罗 成, 1988 年生. 硕士, 工程师. 主要研究方向为网络安全、IoT 安全和 ICS 安全.

Liu Caiyun, born in 1992. Master, engineer. Her main research interests include data security and industrial Internet security.
柳彩云, 1992 年生. 硕士, 工程师. 主要研究方向为数据安全、工业互联网安全.

Cao Yaqin, born in 1994. Master, assistant engineer. Her main research interests include Web attacks attribution and data analytics.
曹雅琴, 1994 年生. 硕士, 助理工程师. 主要研究方向为 Web 攻击追踪溯源、安全数据分析.



Tan Ru, born in 1992. Master, assistant engineer. His main research interests include Web attacks attribution, and network attack and defense.
谭 儒, 1992 年生. 硕士, 助理工程师. 主要研究方向为 Web 攻击追踪溯源和网络攻防.

Feng Yun, born in 1993. PhD, engineer. Her main research interests include cyber security, cyber-attacks discovery, attribution and forensic.
冯 云, 1993 年生. 博士, 工程师. 主要研究方向为网络安全、网络攻击发现和溯源取证.

Zhang Yue, born in 1992. Master, assistant professor. Her main research interests include cyber security, cyber deception and attribution.
张 越, 1992 年生. 硕士, 助理研究员. 主要研究方向为网络空间安全、网络欺骗和追踪溯源.