文章编号：2096-1472(2022)-02-01-05
DOI:10.19644/j.cnki.issn2096-1472.2022.002.001

基于机器学习的牵引供电远动系统异常攻击检测技术研究

周泽岩[1]，程 鹏[2]，方付生[2]，路 涛[3]

(1.中国铁道科学研究院集团有限公司电子计算技术研究所, 北京 100081；
2.中国电子科技网络信息安全有限公司, 四川 成都 610041；
3.北京正信安盾科技有限公司, 北京 100160)
✉zhouzeyan8008@sina.com; cp975@163.com; fang4985@aliyun.com; lutao7@163.com

摘 要：聚焦铁路牵引供电远动SCADA系统，分析其通信规约和网络安全风险点，研究基于单类支持向量机算法的异常攻击检测技术，以达到检测网络异常攻击的目的。本方法先分析正常的通信数据，选取报文序列中的时间戳、源地址、目的地址、源端口、目的端口等字段，构成训练样本序列集合，再对报文序列的数据预处理形成子序列特征数据库，最后采用Python语言、Sklearn机器学习库建立单类支持向量机的行为模型，并通过仿真数据测试，验证其可有效检测网络入侵等异常行为。

关键词：牵引供电远动系统；单类支持向量机；异常检测
中图分类号：TP301.6 **文献标识码**：A

Research on Abnormal Attack Detection Technology of Traction Power Supply SCADA System based on Machine Learning

ZHOU Zeyan[1], CHENG Peng[2], FANG Fusheng[2], LU Tao[3]

(1.*Electronic Computing Technology Institute of China Academy of Railway Sciences Co., Ltd., Beijing* 100081, *China*;
2.*China Electronic Technology Cyber Security Co., Ltd., Chengdu* 610041, *China*;
3.*Beijing ZhengXin Cyber Security Technology Co., Ltd., Beijing* 100160, *China*)
✉zhouzeyan8008@sina.com; cp975@163.com; fang4985@aliyun.com; lutao7@163.com

Abstract: This paper proposes to study abnormal attack detection technology based on One-class SVM (Support Vector Machine) algorithm to achieve the purpose of detecting abnormal network attacks, through focusing on railway traction power supply SCADA (Supervisory Control and Data Acquisition System) system and analyzing its communication protocol and network security risk points. Firstly, normal communication data are analyzed, and fields in the message sequence, such as timestamp, source address, destination address, source port, destination port, are selected. Thus a training sample sequence set are formed. Then, the message sequence data are preprocessed to form a sub-sequence feature database. Finally, Python language and Sklearn machine learning library are used to establish a One-class SVM behavior model. Simulation data test verifies that the proposed technology can effectively detect abnormal behaviors such as network intrusions.

Keywords: traction power supply system; One-class SVM; abnormal detection

1 引言(Introduction)

近年来，工业信息安全整体形势不容乐观。一方面，工业控制系统安全漏洞持续增长，根据中国国家信息安全漏洞共享平台CNVD的统计数据显示，2020 年新增工业控制系统安全漏洞数已达到652 个，同比增长45.5%。另一方面，针对铁路行业的网络攻击越来越多，2020 年3 月英国火车站Wi-Fi提供商C3UK云数据库遭泄露，导致1万名英国铁路乘客的个人数据泄露。2020 年7 月，以色列铁路基础设施的150多台工业服务器遭受网络攻击，影响了28 座火车站和地铁站的运营。

铁路运输是我国主要的运输方式之一，据国家统计局

数据显示，2020 年我国铁路旅客周转量已占总周转量的42.9%，而随着IT技术在铁路原有工控领域大量应用，其信息安全风险日益增加，甚至直接影响我国交通运输的整体安全。铁路牵引供电远动系统(以下简称远动系统)作为铁路关键信息基础设施，其可靠性将对行车安全造成影响，是整个铁路安全稳定运营的核心内容。因此，有必要深入研究铁路牵引供电远动系统，分析其系统特点、通信规约和常见网络攻击行为，通过异常攻击检测技术及时发现安全隐患，增强其安全防护能力。

2 相关研究工作(Related research work)

2.1 系统介绍

远动系统是基于SCADA技术的远动监控系统，可通过远程监视和人工控制相结合的方式实现对牵引变电的控制和监视。当前远动系统多采用通用服务器，操作系统为Unix或Windows，通过网络化实现与其他系统的互联，以及对数据的采集和处理。

(1)远动系统的功能

远动系统为调度员提供远程监控与控制功能，通过下发远程遥控命令完成如开关的控制等功能，同时对远端设备进行综合监控，采集如电压、电流等数据信息，回传调度站，实现全网供电设备信息的实时采集。在变电所设备发生故障时，可及时发现断路器跳闸、故障信号等问题，及时处理事故，防止故障扩大化。

(2)远动系统的组成[1-2]

远动系统主要由一系列远端设备(被控站、远方终端RTU)和中心控制主站系统组成，其中RTU收集现场数据，并通过通信网络回送数据给主站执行端，主站显示这些采集到的数据并允许操作员执行远程控制任务，执行端负责执行相应的指令，如图1所示。

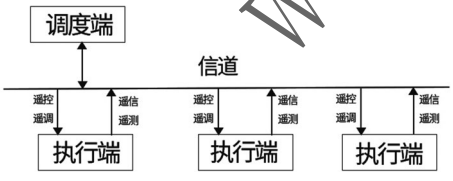

图1 铁路牵引供电远动系统组成

Fig.1 The composition of railway traction power supply SCADA system

调度端：设在控制中心内，完成对远动对象的监控、数据统计及管理等功能；

执行端：完成对远动系统的数据采集、预处理，发送、接收及输出执行等功能，一般设在牵引变电所、分区所、开闭所、AT所及V停控制站内，常规远动系统被控端为远方终端设备(Remote Terminal Unit，RTU)。

2.2 通信规约

为规范调度主站(调度端)与变电站(控制端)的通信，国际电工委员会(IEC)于1995 年发布了IEC 60870-5-101规约，此规约在我国已普遍使用。

随着通信方式从专用通道向以太网的转变，国际电工委员会(IEC)第57技术委员会(TC57)于2000 年又出版了IEC 60870-5-104：2000《远动设备及系统 第5-104部分：传输规约 采用标准传输协议集的IEC 60870-5-101网络访问》，应用层协议采用IEC 60870-5-101规约的ASDU，网络层协议为TCP/IP协议，并封装APCI传输接口以保证应用层ASDU的通信可靠性。因此，该协议具有实时性好、可靠性高的特点，能支持大流量的数据传输。

2006 年，IEC又更新了IEC 60870-5-104规约，发布了IEC 60870-5-104：2006规约。为保持与国际标准的一致性，进一步规范国内应用，全国电力系统管理及其信息交换标准化技术委员会也同步参照国际标准制定了国内标准DL/T 634.5104—2009。

在此基础上，为了规范铁路电力远动系统数据传输，铁路行业制定了《铁路供电远动PSCADA系统传输规约》，当前已在我国铁路系统中普遍采用。

2.2.1 DL/T 634.5104—2009(IEC 60870-5-104：2006)规约[3]

DL/T 634.5104—2009规约明确了通信协议结构与通信方式，其APDU(Application Protocol Data Unit，应用规约数据单元)由APCI(Application Protocol Control Information，应用规约控制信息)与ASDU(Application Service Data Unit，应用服务数据单元)构成，基于TCP协议通信，采用2404端口。其具体报文格式如图2所示。

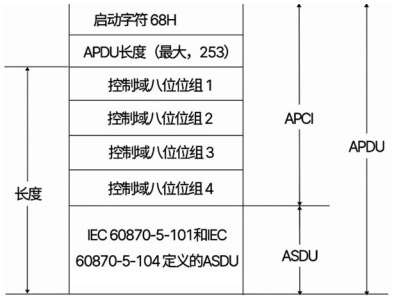

图2 规约报文结构

Fig.2 Protocol message structure

2.2.2 铁路供电远动PSCADA系统传输规约

《铁路供电远动PSCADA系统传输规约》是在DL/T 634.5104规约的基础上，针对铁路电力远动系统的应用业务，规范了铁路电力远动系统中的数据传输格式。该规约规定了铁路牵引供电设备中主站和子站(远动终端)之间以平衡方式进行数据传输的规约结构、帧格式、ASDU选取、互操作性定义、对象定义、应用功能、链路冗余、多控制站冗余切换机制。

(1)ASDU类型标识

《铁路供电远动PSCADA系统传输规约》等同采用了

DL/T 634.5104规约的协议格式和APCI的定义，从ASDU的结构出发，明确了本标准采用的ASDU类型标识，并对每一个类型都进行了定义。

①在监视方向的过程信息

1：不带时标的单点信息

3：不带时标的双点信息

5：不带时标的步位置信息

9：测量值，规一化值

11：测量值，标度化值

13：测量值，短浮点数

15：累计量

30：带时标CP56Time2a的单点信息

31：带时标CP56Time2a的双点信息

32：带时标CP56Time2a的步位置信息

34：带时标CP56Time2a的规一化测量值

35：带时标CP56Time2a的标度化测量值

36：带时标CP56Time2a的短浮点数

140：带相对时标的故障报告

141：带相对时标的事件报告

②在控制方向的过程信息

59：带时标CP56Time2a的双命令

60：带时标CP56Time2a的步调节命令

③在监视方向的系统信息

70：初始化结束

④在控制方向的系统信息

100：召唤命令

101：计数量召唤命令

103：时钟同步命令

⑤在控制方向的参数命令

145：继保装置读整定值命令

205：非继保装置读/写整定值，规一化值

206：非继保装置读/写整定值，标度化值

207：非继保装置读/写整定值，短浮点数

⑥文件传输的应用服务数据单元

120：召唤命令

121：节准备就绪

122：召唤目录，选择文件，召唤文件，召唤节

123：最后的节，最后的段

124：认可文件，认可节

125：段

126：目录，只在监视方向有效

(2)基本应用功能

在定义标识的基础上，该规约对基本应用功能进行了明确的定义，包括过程和应用层报文。基本应用功能覆盖站初始化、站(总)召唤、累计量传输、时钟同步、突发传送、命令传输、参数装载、文件传输等八个方面。

2.3 安全风险分析

对于远动系统(SCADA)的网络而言，主要的网络边界有综合信息网与控制网的边界、调度中心与子站的边界，本文主要分析这两个边界的安全风险和系统内主机的安全风险。

2.3.1 综合信息网与控制网的边界安全风险

对于综合信息网与SCADA系统控制网的网络边界，可能存在的安全隐患包括：

(1)非法访问：外部用户非法访问或内部用户访问权限之外的内容；

(2)恶意入侵：黑客通过伪造、口令破解、应用层攻击等方式，穿透访问控制机制，进入SCADA系统内部进行非法操作；

(3)恶意攻击：包括各种常规攻击和DoS/DDoS攻击；

(4)病毒和蠕虫：计算机病毒和网络蠕虫的传播和爆发，可能从综合信息网传播至控制网络中。

2.3.2 调度中心与子站的边界安全风险

对于调度中心与子站的边界，可能存在的安全风险包括：

(1)非法操作：用户试图操作权限之外的内容；

(2)非法入侵：黑客通过仿制的命令在调度中心对各子站下发非法的控制命令或操作；

(3)非正常或错误操作：调度中心的操作人员下发了非正常的操作或错误的指令；

(4)病毒和蠕虫：计算机病毒和网络蠕虫的传播和爆发，突破调度中心与子站的边界，导致其在SCADA控制网络中的所有区域内传播。

2.3.3 主机安全风险

虽然计算机应用技术在工业领域的应用带来了效率和效益，但是由于计算机网络系统的开放性和脆弱性，也使得工业网络中的操作系统及工控业务应用存在一定的安全漏洞和安全隐患，时刻威胁着企业的工业信息化发展。

同时SCADA控制网络虽然是一定程度上的物理隔离，但威胁工控系统的恶意程序仍然可能绕过隔离的网络进入工控系统，如外来人员联网访问、技术人员现场支持，以及移动存储介质接入等。

此外，主机安全风险还包括安全管理人员对工控设备权限设置不当等导致的安全问题，需防范因为管理人员的疏忽造成操作上的隐患或没有足够信息安全风险意识导致工控网络出现故障等事故。

2.4 传统安全技术局限性

当前铁路系统的连接和开放程度越来越高，所应用的技术正变得越来越具有互操作性和协调性，传统的安全技术主要基于对网络通信的实时分析，运用指纹、规则进行黑名单

式的防护控制，往往无法有效区分业务异常与安全事件，从而导致大量误报，给安全运营带来极大挑战。同时传统方式仍无法自动适应攻击者的伪装绕过，指纹攻击库的升级也严重滞后，因此对未知攻击难以察觉。

安全的本质是攻防对抗，一切的意图都需要通过行为表达，这是安全防护中最重要也是最有价值的环节，而针对行为的有效分析则是传统技术方式最欠缺的。可见，传统安全倚重旧范式，基于特征、规则和人工分析，存在安全可见性盲区，有严重的滞后效应、无力检测未知攻击、容易被绕过，以及难以适应攻防对抗的网络现实和快速变化的企业环境、外部威胁等问题。

铁路一直被普遍认为是一个特定的“安全领域”，这主要是因为铁路系统依赖专有的、独立的网络。这些网络在管理、通信和信号方面往往采用专有协议，遵循行业的特有技术规范，因此非常适合基于大数据驱动、安全分析和机器学习的安全新范式，构建铁路信息系统行为白名单的新范式，以弥补传统安全技术短板，实现从单纯强调边界防护到纵深防护的转变，逆转当前攻防不对称的情况，从海量的安全数据中识别和发现潜在的攻击和恶意行为。

远动系统是电气化铁路的重要配套设施，鉴于当前我国的远动系统功能明确、设备资产类型特定，基本遵从《铁路供电远动PSCADA系统传输规约》，因此可以首先从铁路牵引供电远动系统正常环境中获取业务数据，再根据此规约对数据进行预处理，形成数据特征库，最后通过机器学习构造异常检测模型，实时检测系统，有效发现网络隐患，减少异常攻击行为的发生。此方法是根据业务实际数据构造白名单，可有效减少网络攻击特性分析工作，同时也可以有效发现未知网络攻击，具有成本低、效率高的特点[4]。本文将对此基于机器学习的异常网络攻击检测方式进行深入研究。

3 分析与建模方法(Analysis and modeling methods)

3.1 模型技术路线

目前，基于机器学习的异常检测算法适用于业务模型相对具体的业务场景，具有分类识别准确、特征提取泛化好、误报率低等特点。机器学习中的异常检测方法种类繁多，比如神经网络[5]、支持向量机(Support Vector Machine，SVM)[6−7]等，其中，神经网络具有很强的鲁棒性、记忆能力、非线性映射能力及强大的自学习能力，但是神经网络是个“黑匣子”，无法解释自己的推理过程和推理依据，当数据集不充分时无法进行模型训练。支持向量机作为机器学习方法的一种，与其他算法相比在异常检测方面具有以下优势：

(1)作为SVM方法的理论基础，通过核函数实现高维空间的非线性映射学习。

(2)SVM目标是通过最优超平面对特征空间进行划分，最大化分类边际是其核心思想。

(3)SVM是一种有坚实理论基础的学习方法。不同于现有的统计方法，其避开了从归纳到演绎的传统过程，大大简化了分类和回归等问题[8]。

(4)SVM最终决策函数只由少数的支持向量所决定，计算复杂度取决于支持向量的数目，并非样本空间维数，进而在一定程度上避免了“维数灾难”[8]。

(5)支持向量作为SVM的训练结果，在分类决策中起决策作用。该方法能够抓住关键样本，“剔除”大量冗余样本，不仅计算简单而且具有较好的鲁棒性[8]。

综上，与其他算法相比，支持向量机方法具有更坚实的数学理论基础，可以有效地解决样本受限条件下的高维数据模型构建问题，并具有泛化能力强、收敛全局最优、维数不敏感等优点[8−9]。

尽管支持向量机的分类算法在异常检测领域优势明显，但是传统的支持向量机分类算法需要正负两类样本数据集合进行训练来建立模型，而在实际的工业控制系统环境中，往往存在异常样本数据量较少、异常样本难以收集等问题，如工控系统大多数时候处于正常运行状态，很少存在故障状态，使得故障样本数据量较少[9]。因此，针对远动系统的数据特点，选取单类支持向量机机器学习算法来建立异常检测模型。

3.2 单类支持向量机异常检测算法

单类支持向量机(One−Class Support Vector Machine，OCSVM)算法基于传统支持向量机发展而来，该算法的样本集合只需包含正常或异常一种样本数据，无须另一类样本数据。因此，对于工业控制系统环境中的异常检测问题十分适用。

单类支持向量机异常检测算法的基本思想是将低维空间的特征向量通过核函数映射到高维特征空间，在高维特征空间中构造一个超平面，使得数据对象与原点尽可能地分开[10]。因此，OCSVM算法的本质是在高维空间求解分类超平面。

在超平面求解过程中，将求解问题转化为优化问题，引入拉格朗日因子转化为对偶问题，如式(1)所示：

$$L(w,\xi,\rho,\alpha,\beta)=\frac{1}{2}\parallel w\parallel^2+\frac{1}{vl}\sum_i\xi_i-\rho-\sum_i\alpha_i\left(\left(w\phi(X_i)\right)-\rho+\xi_i\right)-\sum_i\beta_i\xi_i \tag{1}$$

根据对参数w、ξ、ρ求偏导，通过导出对偶问题和使用核技巧，便可以实现通过支持向量$\{X_i,i\in[l],\alpha_i>0\}$求得决策函数，如式(2)所示：

$$f(X)=sgn\left(\sum_i\alpha_i k\left((X_i,X)-\rho\right)\right) \tag{2}$$

由式(2)可知，当$f(X)$输出值大于0时，说明检测样本为正常样本，否则为异常样本。

3.3 单类支持向量机异常检测模型实现

基于单类支持向量机算法对远动系统流量数据进行检测的模型，具体步骤如下：

(1)流量数据采集。首先从远动系统正常环境中获取业务数据流量，在实际工作过程中单位时间内产生的数据流量具有随机性，通过固定时间间隔将报文序列划分成固定长度的子序列，选取报文序列中的时间戳、源地址、目的地址、源端口、目的端口等字段，构成训练样本序列集合。

(2)报文序列的数据预处理。对训练样本序列数据进行聚合操作，提取报文序列的前后包间隔时间，将源IP、源端口、目的IP、目的端口、包间隔时间作为子序列特征数据。

(3)单类支持向量机核函数的选择。单类支持向量机的核心思想是将低维空间的特征向量通过核函数映射到高维特征空间，其中，核函数的选择直接关系到异常检测模型的准确率和泛化能力。常见的核函数有线性核函数、多项式核函数、高斯核函数、Sigmoid核函数等。

(4)建立单类支持向量机的训练模型。采用Python语言、Sklearn机器学习库进行模型的构建与训练，并利用网格搜索和三折交叉验证的方式获取最优值参数的训练模型。以正常的远动系统报文子序列提取特征数据构成的训练集合，构造并求解最优超平面，完成异常检测模型的训练。

(5)基于单类支持向量机模型进行异常检测。将测试数据输入单类支持向量机异常检测模型进行检测，若输出结果为1，认为该数据为正常的铁路供电远动系统流量数据；否则，认为该数据为异常流量。

3.4 实验对比和结果分析

为了证明本文提出的单类支持向量机异常检测方法的有效性与可行性，选择单类支持向量机算法、传统支持向量机算法、孤立森林算法这三种算法模型进行实验对比，实验结果如表1所示。结果表明，单类支持向量机异常检测模型的分类准确率高于另外两种异常检测模型，并且误报率最低。另外，对单类支持向量机异常检测模型从不同角度进行研究，分别从滑动窗口长度和核函数选择上进行实验，结果表明滑动窗口长度为3时，分类决策达到全局最优；选择线性核函数进行映射时，模型异常检测各性能指标参数最佳。综上所述，基于单类支持向量机算法的异常检测方法在远动系统的通信流量数据异常检测中具有一定的优势。

表1 实验结果

Tab.1 Experimental result

算法	准确率/%	误报率/%
单类支持向量机算法	96.28	0.08
传统支持向量机算法	92.42	0.13
孤立森林算法	89.69	0.2

4 结论(Conclusion)

远动系统作为铁路的核心系统，一旦遭受网络攻击，必将对铁路安全运行造成直接影响。而传统的网络入侵检测方法是从网络特性和入侵特性的角度来进行检测，效率低下且不具备检测未知网络攻击的能力。

本文提出了一种基于单类支持向量机算法的异常检测方法，不需要大量的训练样本，就可以构建异常攻击检测模型。首先分析正常的通信数据，选择报文序列数据中的时间戳、源地址、目的地址、源端口、目的端口等字段组成训练样本数据集，再对报文序列的数据预处理形成子序列特征数据库，最后采用Python语言、Sklearn机器学习库建立单类支持向量机的训练模型，以实现对远动系统的异常网络攻击检测。实验表明，本方法准确率高、误报率低，适合在远动系统环境下的异常网络攻击行为检测。

参考文献(References)

[1] 方小飞.基于SCADA的铁路电力远动系统几个关键环节的设计与应用[D].长春:吉林大学,2014.

[2] 张克平.铁路牵引供电系统远动系统的设计及应用[D].南昌:南昌大学,2013.

[3] 全国电力系统管理和信息交换标准化委员会.远动设备及系统 第5-104部分:传输规约 采用标准传输协议集的IEC 60870-5-101网络访问:DL/T 634.5104—2009[S].北京:中国标准出版社,2009.

[4] ANOOP A, SREEJA M S. New genetic algorithm based intrusion detection system for SCADA[J]. International Journal of Engineering Innovations and Research, 2013, 2(2):171-175.

[5] 顾兆军,郝锦涛,周景贤.基于改进双线性卷积神经网络的恶意网络流量分类算法[J].信息网络安全,2020,20(10):67-74.

[6] 刘付民,王静咏,张治斌.组合核函数相关向量机的网络安全态势预测[J].计算机应用研究,2016,33(08):2417-2419,2424.

[7] 罗峰,胡强,侯硕,等.基于支持向量机的CAN-FD网络异常入侵检测[J].同济大学学报(自然科学版),2020,48(12):108-114.

[8] 李振刚,甘泉.改进蚁群算法优化SVM参数的网络入侵检测模型研究[J].重庆邮电大学学报(自然科学版),2014,26(06):785-789.

[9] 刘颖,王丽敏,姜建华,等.基于离群点剔除的SVM信用风险评价方法[J].吉林大学学报(理学版),2016(06):1395-1400.

[10] 尚文利,张盛山,万明,等.基于PSO-SVM的Modbus TCP通讯的异常检测方法[J].电子学报,2014,42(11):2314-2320.

作者简介：

周泽岩(1980-)，男，硕士，助理研究员.研究领域：铁路网络安全及软件测试.

程　鹏(1975-)，男，本科，工程师.研究领域：工业控制安全和网络安全.

方付生(1982-)，男，本科，工程师.研究领域：工业信息安全.

路　涛(1978-)，男，本科，工程师.研究领域：人工智能与信息安全.