文章编号:1671-4598(2024)10-0298-07 DOI:10.16526/j.cnki.11-4762/tp.2024.10.042 中图分类号:TP311 文献标识码:A

基于对抗机器学习的工业控制网络欺骗攻击行为检测系统设计

张 涛
(山西警察学院 网络安全保卫系,太原 030001)

摘要:欺骗攻击行为会干扰工业控制网络对传输信息的判断能力,从而使得风险性数据进入网络主机,造成网络安全性下降的问题;为避免上述情况的发生,设计基于对抗机器学习的工业控制网络欺骗攻击行为检测系统;设置攻击行为采集、处理、检测验证三类子模块单元,完成欺骗攻击行为检测系统的功能性模块设计;在对抗机器学习算法中定义攻击行为,并以此为基础,提取欺骗攻击行为特征,实现对攻击行为的识别;分析工业控制网络的安全风险,联合欺骗攻击行为的风险性度量条件,定义具体的检测建模标准,从而实现对工业控制网络欺骗攻击行为信息的检测;实验结果表明,设计方法的应用可以按照数据样本传输波长的差异性,将欺骗性攻击信息检测出来,且召回率测试结果在0.93~0.98之间,表明设计方法能够准确地检测出欺骗攻击行为,使工控网络的运行安全性得到了保障。

关键词:对抗机器学习;工业控制网络;欺骗攻击行为;数学表达式;行为特征;安全风险;传输波长

Design of Deception Attack Detection System for Industrial Control Networks Based on Adversarial Machine Learning

ZHANG Tao
(Department of Network Security, Shanxi Police College, Taiyuan 030001, China)

Abstract: Deceptive attack behavior can interfere with the judgment ability of industrial control networks to transmit information, causing risky data to enter network hosts and leading to a decrease in network security. To avoid the occurrence of the above situation, design an industrial control network spoofing attack behavior detection system based on adversarial machine learning. Set up three types of sub module units for attack behavior collection, processing, and detection verification, and complete the functional module design of the deception attack behavior detection system. Define attack behavior in adversarial machine learning algorithms, and based on this, extract features of deceptive attack behavior to achieve recognition of attack behavior. Analyze the security risks of industrial control networks, establish risk measurement conditions for joint deceptive attack behaviors, define specific detection modeling standards, and thus achieve the detection of information on deceptive attack behaviors in industrial control networks. The experimental results show that the application of the design method can detect deceptive attack information based on the difference in transmission wavelength of data samples, and the recall test results are between 0.93 and 0.98, indicating that the design method can accurately detect deceptive attack behavior, ensuring the operational security of industrial control networks.

Keywords: adversarial machine learning; Industrial control network; deception attack behavior; mathematical expression; behavior characteristics; security risk; transmission wavelength

0 引言

随着工业互联网的发展和智能化制造的推进,工业控制网络在工业生产中扮演着至关重要的角色。工业控制网络是适用于工业自动化系统的网络结构,其内部连接多种不同的执行器、传感器与工业设备组件,可以实现设备元件之间的协同工作与数据共享。然而,工业控制网络面临着越来越多的网络安全威胁,其中包括欺骗攻击。欺骗攻击是一种专门针对工业控制系统的恶意攻击手段,是一种常见的网络攻击方式。攻击对象通过伪造虚假信息或者冒充合法用户,诱导主机元件做出错误判断,从而使风险性数据进入网络运行环境,达到篡改数据、非法获取信息、破坏系统等目的。为了保障工业控制网络的安全,需要建立有效的欺骗攻击行为检测系统。这种系统可以通过监控网络流量、识别异常行为、分析数据包等方式,及时发现和应对潜在的欺骗攻击,并提高工业控制系统的整体安全性和可靠性。

相关国内研究如:赵嘉等人[1]利用互信息理论完成了网络数据中网络攻击行为的关键特征的选取,通过改进灰狼优化算法提出一种灰狼提升算法,并基于该算法和最小

收稿日期:2023-07-21; **修回日期**:2023-10-16。
基金项目:2022年山西省教育厅教学改革创新项目(J20221297)。
作者简介:张 涛(1980-),女,硕士,讲师。
引用格式:张 涛.基于对抗机器学习的工业控制网络欺骗攻击行为检测系统设计[J].计算机测量与控制,2024,32(10):298-304.

二乘支持向量机提出了 GWB-LSSVM 检测模型。但是在高维数据集中，计算互信息会变得非常耗时，导致检测效率下降。张宇翔等人[2]利用事件序列关联模型实现对多步攻击行为的有效检测，并通过 ATT&CK 攻击矩阵进行可视化呈现。但是随着网络环境的不断复杂化，ATT&CK 系统的检测能力已经无法满足筛选欺骗攻击行为的实际应用需求。石欣然等人[3]将待检测的网络流量数据作为训练后的 GCNBiGRU 神经网络模型的输入得到空间特征和时间特征，将时空特征相结合进行网络攻击检测。但是在网络攻击检测中，攻击行为可能存在长时间的依赖关系，模型的时空特征处理能力不足使得攻击行为无法被有效捕捉，导致检测性能的下降。吕东晓等人[4]通过实时仿真平台对稳控装置物理侧和通信侧进行监测，以获取原始数据。制定数据离散化规则，根据规则对原始数据进行离散量化处理，再对数据编码生成事件序列，接着利用 Apriori 算法对时间序列进行分析，通过已有频繁序列特征对新的攻击类别和故障点进行识别，以实现网络攻击的检测。但是 Apriori 算法的时间复杂度较高且容易受到数据维度和规模的影响。当处理大规模数据时，会导致计算效率低下，从而影响检测的实时性。

相关国外研究如：Alshehri 等人[5]提出了一种使用机器学习结合用户行为分析来检测网络攻击的新框架。该框架将用户行为建模为表示这种网络中的用户活动的事件序列。然后，将表示的序列拟合到递归神经网络模型中，以提取为单个用户绘制独特行为的特征。因此，该模型可以识别常规行为的频率，以描绘用户在网络中的行为方式。最后，递归神经网络通过将未知行为分类为规则或不规则行为来检测异常行为。但是在真实场景中，恶意网络行为的样本通常比正常行为少得多，导致训练模型时出现数据样本不平衡的情况，使模型更倾向于预测大多数样本所属的类别，而对罕见的异常行为进行辨识不足。Dai 等人[6]定义了流量行为和路由状态，并用于描述泄漏监控、隐藏流量覆盖和交易身份伪造等不同阶段的攻击特征。但是对于复杂的攻击场景，定义的特征和模型会在训练时过度拟合训练数据，导致在新数据上的泛化性能较差，影响检测的准确性。

基于现有方法存在的问题，利用对抗机器学习算法的应用优势，设计一种新型的工业控制网络欺骗攻击行为检测系统，并通过对比实验的方式，突出其应用可行性。

1　欺骗攻击行为检测系统架构设计

工业控制网络面临着越来越复杂和多样化的攻击，设计针对欺骗攻击行为的检测系统能够帮助及时发现和阻止潜在的安全威胁，从而提高网络的安全性。本文设计的欺骗攻击行为检测系统架构如图 1 所示。

根据上述系统总体框架可以看出，欺骗攻击行为检测系统的硬件设计部分包含攻击行为采集模块、攻击行为处理模块与检测验证模块，软件部分包括攻击行为识别与攻击行为检测两部分。该系统框架的硬件设计部分能够实时采集攻击行为数据，并通过软件部分对数据进行即时处理和分析，以快速准确地对攻击行为进行识别和检测，提高系统的检测性能。

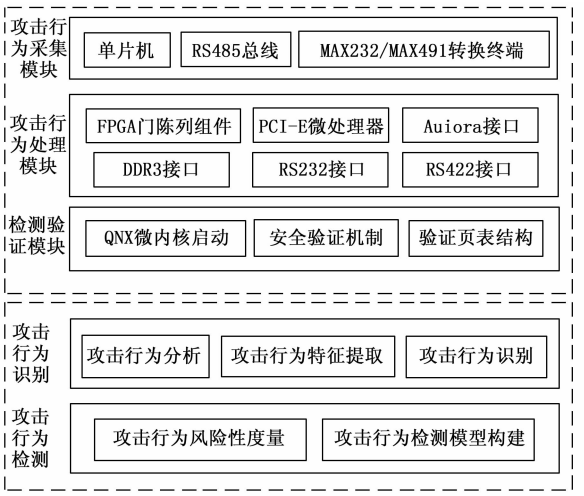

图 1　系统总体框架

2　欺骗攻击行为检测系统硬件设计

工业控制网络欺骗攻击行为检测系统的功能性模块包括攻击行为采集模块、处理模块与检测验证模块，本章节针对具体设计方法展开研究。

工业控制网络是一个庞大且复杂的系统，具有复杂的网络拓扑结构以及数据在各个节点之间的分布。攻击行为采集者需要克服物理位置分散、数据流量分割和广播等问题，使得难以高效地采集工业控制网络中各种攻击行为的数据流量和通信信息。对此本文采用 MAX232 和 MAX491 两种不同的协议文件，将风险性数据与常规传输数据进行分离，以便于后续的检测验证模块调取和利用。

工业控制网络对实时性要求较高，攻击行为的检测和应对必须在短时间内完成。因此，在对大量数据进行实时处理的过程中需要确保高效率和高准确性。而本文采用的 FPGA 门阵列设备和 PCI-E 微处理器具有并行处理能力，能够快速响应系统主机发出的数据获取指令。这种设计保证了模块能够在短时间内对大量数据进行处理和分析，从而满足工业控制网络对实时性的高要求。

攻击者采用的攻击手段和技术日益多样化和复杂化，包括但不限于恶意软件、拒绝服务攻击、社会工程、钓鱼等多种方式。这种多样性使得针对不同类型攻击行为的准确识别和验证变得具有挑战性。在本文设计中，对抗机器学习算法定义的安全验证机制是检测验证模块的核心执行部分。该机制参考了经过集中性管理后的验证页表编码形式，并考虑了 QNX 微内核中欺骗攻击行为信息的实时传输量。通过此安全验证机制，在模块执行等级保持不变的前提下，能够精准筛选出所有的欺骗性攻击行为信息，满足系统主机的实时检测需求。

本文通过设计采集模块、攻击行为处理模块与检测验证模块，有效克服欺骗攻击行为检测的技术难点，以实现高性能的攻击行为检测。

2.1 攻击行为采集模块

攻击行为采集模块对于风险性数据的采集遵循多种不同的协议文件，在数据取样与存储方面，该模块所搭载的 PC 端检测平台，可以直接驱动工业控制网络，从而在通信串口组织的配合下，实现对信息采集行为的控制。在工业控制网络中，PC 端检测平台保持开源连接状态，所以攻击行为采集模块同时支持多种不同的数据信息参量，但对抗机器学习算法规定只有具有欺骗性攻击行为的数据样本才可以作为系统主机的检测对象，所以即便是在多样化取样的情况下，系统运行过程中也不会出现错误检测数据的情况。MAX232/MAX491 转换终端，负责将攻击行为采集模块采样所得的风险性数据与常规传输数据分离开来，MAX232、MAX491 是两种不同的协议文件，基于前者检测所得的数据信息由于不符合对抗机器学习算法的定义标准，所以会被回传至 PC 端检测平台中；基于后者检测所得的数据信息是符合对抗机器学习算法所规定工业控制网络欺骗攻击行为检测需求的信息参量，可在 RS485 总线的分配作用下，按需传输至下级单片机之中，以便于检测验证模块的调取与利用。详细的攻击行为采集模块连接结构如图 2 所示。

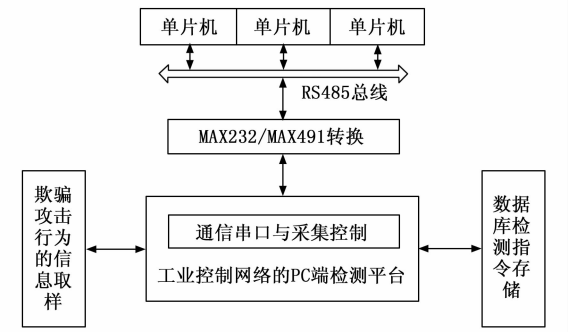

图 2 攻击行为采集模块连接结构

以精准取样工业控制网络欺骗攻击行为信息为目的，设置了攻击行为采集模块。但受到 PC 端检测平台开放性的影响，初次取样的数据样本只有得到通信串口的认证后，才能被传输至数据库主机，并于其中生成检测运行指令。

2.2 攻击行为处理模块

攻击行为处理模块的核心执行组件由 FPGA 门阵列设备和 PCI-E 微处理器共同组成。上级模块完成对欺骗攻击行为信息的取样后，会借助双向信道组织，将数据参量传输至多个不同的部件结构之中，在此过程中，攻击行为处理模块直接参与系统主机对取样所得信息参量的分类，所以该模块的运行稳定性，决定了系统主机对工业控制网络欺骗攻击行为的检测准确性。FPGA 门阵列组件与 PCI-E 微处理器之间存在检测信息的双向传输与反馈关系。工业控制网络欺骗攻击行为检测系统运行过程中，Auiora 接口、DDR3 接口接连进入相应状态，此时工业控制网络中的欺骗攻击行为信息首先进入 FPGA 门阵列组件之中，在确保对数据样本的准确分析与处理之后，PCI-E 微处理器开始向上级组件发出数据获取指令[7]。PCI-E 微处理器与工业控制网络接口、RS232 接口、RS422 接口保持并列连接关系，经过攻击行为处理模块筛选的数据信息参量，会经由这些接口组织传输至检测验证模块之中。完整的攻击行为处理模块结构模型如图 3 所示。

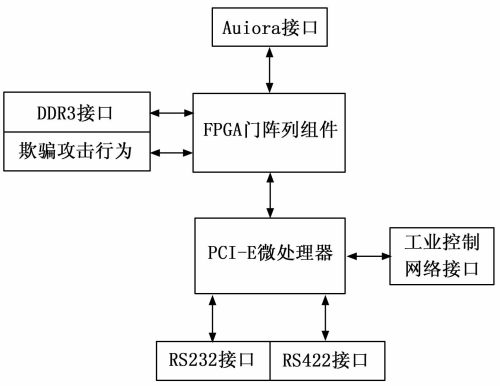

图 3 攻击行为处理模块结构模型

从执行等级的角度来看，攻击行为处理模块是攻击行为采集模块的下级功能性结构单元[8]。工业控制网络在检测欺骗攻击行为时对于数据样本准确性的要求相对较高，所以基于对抗机器学习算法对于信息参量处理标准的定义，必须直接调取单片机及数据库主机中暂存的欺骗攻击行为信息参量。

2.3 检测验证模块

在工业控制网络欺骗攻击行为检测系统中，检测验证模块的设置遵循 QNX 微内核的反传式启动模式。QNX 微内核作为检测验证模块的底层执行结构，在接收到攻击行为处理模块输出的工业控制网络欺骗攻击行为检测信息后，建立完整的验证页表结构[9]。对于系统主机而言，验证页表结构的组成形式相对较为复杂，因此只有在经过集中性管理后，其内部所存储的欺骗性攻击行为信息参量才能符合系统主机的实时检测需求。具体的检测验证模块功能框图如图 4 所示。

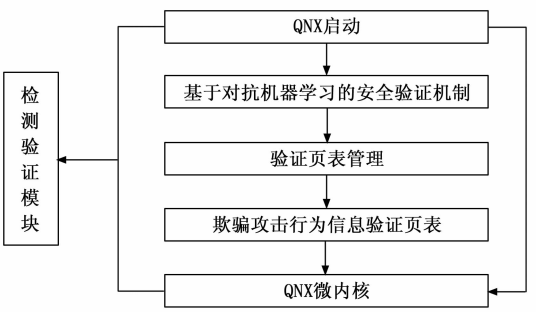

图 4 检测验证模块的功能框图

基于对抗机器学习算法定义安全验证机制，是检测系统验证模块的核心执行部分。在模块执行等级保持不变的情况下，安全验证机制的定义既要参考经主机元件管理后的验证页表编码形式，也要考虑 QNX 微内核中欺骗攻击行为信息的实时传输量能否满足系统主机的检测需求。由于

工业控制网络运行过程中，攻击对象所伪造出的欺骗行为信息类别并不固定，所以为满足多样化检测需求，安全验证机制必须对可能出现的攻击性行为进行详细列举[10]。反传式启动是 QNX 微内核特有的运行模式，对抗机器学习算法所要求的工业控制网络欺骗攻击行为检测标准不但要精准筛选出所有的测定信息，还要充分利用处理器模块输出的数据样本，而这种双向性的选择标准与 QNX 微内核的反传启动模式相符合。

3　基于对抗机器学习的攻击行为识别

工业控制网络检测系统对于欺骗攻击行为信息的处理，应根据对抗机器学习算法，提取必要的欺骗攻击行为特征，并联合相关数据样本，针对实时检测数据制定具体的识别方案。

3.1　基于对抗机器学习模型的攻击行为分析

对抗机器学习模型对于攻击方信息的定义，应以获得反馈信息为目标，在掌握攻击行为欺骗能力的同时，避免传输数据对工业控制网络的稳定运行能力造成影响。设 $\dot{k}$ 表示攻击规模定义特征，β 表示攻击目的查询系数，$\hat{l}$ 表示攻击方信息定义项，联立上述物理量，可将对抗机器学习模型计算式表示为：

$$L=\underset{\alpha\to\infty}{\operatorname{argmax}}\sum\dot{k}\cdot\beta\cdot|\hat{l}|^{2} \tag{1}$$

式中，α 表示工业控制网络中的攻击行为欺骗性等级参数。按照对抗机器学习模型，筛选工业控制网络所需检测的欺骗性攻击行为信息，既要准确查询数据样本来源，以确保检测结果的准确性，也要对信息参量的传输目的进行准确标记，从而避免系统主机对攻击行为信息的错误提取[11-12]。

攻击行为模型的通用数学表达是指按照对抗机器学习模型，求解与工业控制网络欺骗攻击行为检测信息相关的数值表达式。主要是按照机器运算的方式，将多个不相连的欺骗攻击行为信息关联起来，通过点与点之间所传递的信息，来获取工业控制网络中的数据依存关系[13]。规定 δ 表示机器运算参数，$\tilde{h}$ 表示工业控制网络欺骗攻击行为待测信息的标准赋值向量，χ 表示基于对抗机器学习模型的攻击行为信息取样系数，ε 表示随机定义项，j_{ε} 表示基于参数 ε 的攻击行为欺骗性描述参量，j'_{ε} 表示与参数 j_{ε} 相似的欺骗性描述参量，且 $j_{\varepsilon}\neq j'_{\varepsilon}\neq 0$ 的不等式取值条件恒成立。在上述物理量的支持下，联立公式（1），推导基于对抗机器学习模型的欺骗攻击行为通用数学表达式如下：

$$J=\sqrt{\delta\tilde{h}}\cdot\left(\frac{1}{\chi}L\times|j_{\varepsilon}-j'_{\varepsilon}|^{2}\right) \tag{2}$$

基于对抗机器学习模型处理攻击行为检测数据，要保证待测信息取样的相似性，通常情况下，已定义欺骗性攻击行为向量可与多个不相同的信息样本保持相似性数值关系，而实施数据处理的目的，就是将所有相似性参量全部提取出来，从而避免攻击对象所伪造出的虚假信息对工业控制网络主机的运行稳定性造成影响[14]。

3.2　欺骗攻击行为特征提取

基于上述欺骗攻击行为分析，进行欺骗攻击行为特征提取。欺骗攻击行为特征提取是工业控制网络主机分析与检测信息欺骗性攻击行为的基础。通过对欺骗攻击行为信息的特征进行抽取，检测系统主机可以构建出能够区分正常与异常数据样本的行为序列模型，且定义过程中，主机元件对每一类信息参量的取值都必须遵循对抗机器学习算法原则[15]。

由于工业控制网络对数据信息参量的取样不设置明确的门限条件，所以在攻击对象能够有效冒充合法用户的情况下，欺骗攻击行为信息极易与常规传输信息混合在一起[16]。规定 g 表示欺骗攻击行为信息的数值计量参数，其计算式如下：

$$g=\frac{\hat{h}}{\sqrt{J}}-\varphi\sum_{\iota=1}\frac{\overline{H}}{\gamma} \tag{3}$$

式中，$\hat{h}$ 表示基于对抗机器学习模型所定义的异常数据样本行为序列向量，φ 表示欺骗性攻击行为信息抽取系数，ι 表示数据样本区分参数，$\overline{H}$ 表示工业控制网络欺骗攻击行为信息的单位取样均值，γ 表示近似影响参数。

基于此，可将工业控制网络对于欺骗攻击行为特征的提取表达式定义为：

$$G=\frac{1}{|g|^{2}}\sum_{a\neq s}(d_{a}+d_{s}) \tag{4}$$

式中，a、s 表示两个不相等的门限值定义参数，d_{a} 表示基于参数 a 的欺骗攻击行为信息取样特征，d_{s} 表示基于参数 s 的欺骗攻击行为信息取样特征。对于工业控制网络主机而言，其在完成欺骗攻击行为信息的取样后，会将符合检测标准的数据样本存储于系统数据库主机之中，一方面便于验证模块对其进行针对性检测，另一方面也能够避免网络信息非标准识别行为的出现。

3.3　攻击行为识别

基于上述提取到的欺骗攻击行为特征，进行攻击行为识别处理。攻击行为检测数据识别是检测系统设计中最为关键的步骤，主要用于将正常与异常数据样本区分开来[17-18]。

设 φ 表示欺骗攻击行为信息的实时收集系数，f 表示无关虚假信息过滤参数，κ 表示待测信息编码系数，A_{0} 表示基于对抗机器学习算法所定义的欺骗攻击行为信息检测标准值，η 表示欺骗攻击行为信息的取样效率，$\tilde{q}$ 表示攻击行为信息的计数值特征，联立公式（4），可将攻击行为检测数据识别表达式定义为：

$$D=(\varphi-1)^{2}+f\sum_{\kappa=1}^{+\infty}\frac{A_{0}}{G}\log\left(\frac{\tilde{q}}{\eta}\right) \tag{5}$$

欺骗攻击行为会干扰工业控制网络对传输信息的判断能力，因此在检测攻击行为参量时，出于准确性考虑，要求一项数据识别条件只能对应一种类型的欺骗攻击行为信息参量。

4　工业控制网络欺骗攻击行为检测功能的实现

在对抗机器学习算法的基础上，分析工业控制网络中的安全性风险，再联合欺骗攻击行为的风险性度量条件，定义具体的检测建模标准，从而实现对工业控制网络欺骗

攻击行为信息的检测。

4.1 欺骗攻击行为的风险性度量

工业控制网络中，由欺骗攻击行为造成的安全风险特指恶意软件攻击事件，所涉及信息参量既可以是蠕虫、木马、病毒等恶意程序，也可以是包含漏洞的执行指令。这些恶意成分通过入侵工业控制网络体系的方式，对网络机制的运行稳定性造成严重破坏[19-20]。一旦攻击性信息累积量超过了网络主机的承载能力，工控系统还有可能停止运行，并最终导致传输数据的丢失。

在工业控制网络中，随机选择 n 个欺骗攻击行为信息，其取值满足如下表达式。

$$D(w_1,w_2,\cdots,w_n)\in[1,+\infty) \tag{6}$$

在 $w_1\neq w_2\neq\cdots\neq w_n$ 的不等式取值条件恒成立的情况下，联立公式（5），推导工业控制网络的安全风险分析表达式如下：

$$Q=\frac{1}{2|\Delta E|}\sum_{e_1=1}^{\infty}\sum_{e_2=1}^{\infty}\left(\frac{D|\dot{r}-\dot{t}|}{w_1\cdot w_2\cdot\cdots\cdot w_n}\right)^2 \tag{7}$$

式中，ΔE 表示网络主机对欺骗攻击行为信息的额定承载能力，e_1 表示风险信息的顺序传输参数，e_2 表示风险信息的逆序传输参数，$\dot{r}$ 表示欺骗攻击行为的破坏能力定义项，$\dot{t}$ 表示攻击性能力的数值定义项。定期对工业控制网络进行漏洞扫描与安全审计，能够及时发现并修复潜在的风险性问题。此外，网络系统为实现对欺骗攻击行为的精准检测，还要严格遵守安全风险分析条件，筛选所涉及传输信息，从而实现对攻击性行为的风险性等级进行有效控制。

风险性度量是指工业控制网络检测系统对于欺骗攻击行为风险性等级的评估，该项物理量的取值，决定了系统主机对于欺骗攻击行为信息的实时检测能力。对抗机器学习算法所规定的风险性度量条件涉及威胁概率与攻击影响强度[21-22]。威胁概率是指工业控制网络中发生欺骗攻击行为的可能性，常表示为 λ。攻击影响强度特指欺骗攻击行为成功后，工业控制网络主机的变化情况，常表示为 μ。在上述物理量的支持下，联立公式（7），推导欺骗攻击行为的风险性度量表达式如公式（8）所示。

$$R=\lambda\cdot\left[\frac{\mu\times(u_{\max}-u_{\min})}{\sqrt{Q}}\right] \tag{8}$$

式中，$u_{\max}$ 表示基于对抗机器学习算法所定义的网络安全性损失向量的最大取值，$u_{\min}$ 表示网络安全性损失向量的最小取值。风险性度量表达式的计算结果具有多样性特征，$R>0$ 成立时，表示检测系统对于欺骗攻击行为的取样方向与数据样本在工业控制网络中的传输方向保持一致；$R=0$ 成立时，表示工业控制网络中不存在符合对抗机器学习算法检测需求的欺骗攻击行为信息；$R<0$ 成立时，表示检测系统工业控制网络主机只有通过逆序提取传输数据的方式，才能实现对欺骗攻击行为信息的检测。

4.2 攻击行为的检测建模

对于工业控制网络的欺骗攻击行为的检测建模，联合欺骗攻击行为的风险性度量条件定义具体的检测建模标准[23-24]。设 ν 表示风险性因素组成向量，$\widetilde{O}$ 表示欺骗攻击行为信息的集中性收集特征，P 表示标准建模向量，$\vec{I}$ 表示基于对抗机器学习算法所定义的欺骗攻击行为信息检测阈值，联立公式（8），推导工业控制网络对于欺骗攻击行为的检测建模表达式定义为：

$$I=\frac{\log\left(\frac{R}{\nu^{-1}\widetilde{OP}}\right)}{\ln(\vec{I})} \tag{9}$$

从攻击检测的角度来看，工业控制网络所定义的欺骗性攻击行为必须是特定字符串，所以系统主机对于欺骗性攻击行为信息的检测也就等同于对特定传输字符串的识别与提取。

5 实验分析与研究

为突出说明基于对抗机器学习的工业控制网络欺骗攻击行为检测系统、ATT&CK 框架下的检测系统、基于少样本且不均衡的检测系统的应用差异性，设计如下对比实验。

5.1 工业控制网络设置

本次实验以工业控制网络中的欺骗攻击行为信息作为检测对象，将总主机连接在网络回路之中，设置多个分区装置，将整个工业控制网络划分成 8 个区域，每一区域就是一个单位检测环境。具体的工业控制网络接线如图 5 所示。

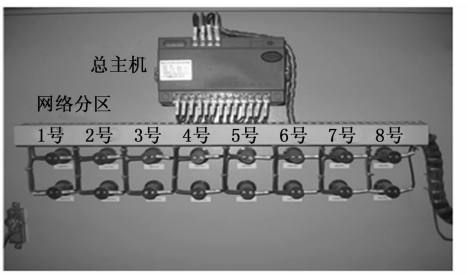

图 5 工业控制网络接线

根据上述检测环境进行工控网络数据采集，分别进行网络流量、系统日志、用户行为数据的采集，其中网络流量数据的采集速率为 1 000 数据包/秒，系统日志数据的采集速率为 500 数据包/秒，用户行为数据的采集速率为 100 数据包/秒，采集时间设定为 24 h。无关虚假信息过滤参数 $0.5<f\leqslant 0.95$，待测信息编码系数 $0.8<\kappa\leqslant 1.0$，网络主机承载能力 100 MB/s$<\Delta E\leqslant$1 GB/s。工业控制网络是较为开放的网络环境，其内部所接入设备元件必须具有独立处理各类风险性信息的能力。实验所需相关设备元件的具体型号如表 1 所示。

表 1 实验设备选型

编号	设备名称	型号
1	总主机	DT-610L-AH61 型号的中央控制器
2	分区装置	N5105 型号的软路由设备
3	处理器	RS720A-E12-RS12-C 双路主机
4	串口服务器	IPC-610L 型号的独立扩展元件
5	编码芯片	PCIE1/4 千兆口接入型主板
6	信息检测设备	G5905 型核显设备

出于公平性考虑，改用其他检测方法时，设备元件与服务器主机之间的连接关系始终保持不变。工业控制网络欺骗攻击行为检测系统如图 6 所示。

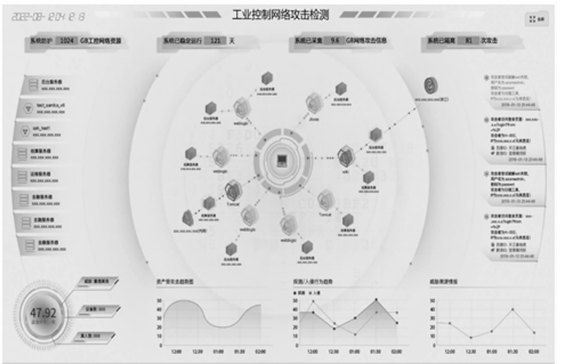

图 6　工业控制网络欺骗攻击行为检测系统图

5.2　实验指标

5.2.1　欺骗攻击行为检测

混合情况下，数据传输波长的均值水平相对较高，若将欺骗攻击行为信息、标准传输信息分离开来，数据传输波长的数值水平则会出现一定程度的下降，对于不同的检测系统而言，如果在实施检测后，欺骗攻击行为信息、标准传输信息检测波长值与其真实波长值较为相似，表示该系统能够对工业控制网络欺骗攻击行为进行准确检测，也就意味着该系统的应用能够促进工业控制网络的稳定运行。

5.2.2　召回率

召回率（Recall）是一个评估模型分类能力的指标，它衡量了模型在所有实际欺骗攻击样本中正确识别出欺骗攻击样本的比例。召回率的计算公式如下：

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

式中，TP 表示真阳性，即实际上是欺骗攻击且被模型正确识别为欺骗攻击的样本数量。FN 表示假阴性，即实际上是欺骗攻击但被模型错误地判定为非欺骗攻击的样本数量。召回率的数值范围在 0 到 1 之间，数值越高表示模型对于欺骗攻击的检测能力越强。

5.3　实验结果

5.3.1　欺骗攻击行为检测

分别将基于对抗机器学习的工业控制网络欺骗攻击行为检测系统（方法 1）、ATT&CK 框架下的检测系统（方法 2）、基于少样本且不均衡的检测系统（方法 3）的执行程序输入总主机之中，记录不同系统针对欺骗攻击行为信息、标准传输信息的波长检测结果。

图 7 反映了不同系统对欺骗攻击行为信息、标准传输信息传输波长的检测情况。

分析图 7 可知，方法 1 作用下，欺骗攻击行为信息、标准传输信息检测波长值与其真实波长之间的差值水平较低，特别是标准传输信息检测结果，检测波长曲线基本与真实波长曲线完全贴合在一起。方法 2 作用下，欺骗攻击行为信息、标准传输信息检测波长的初始值就已经远高于其真实波长值，后续实验过程中两类曲线更是出现了明显偏差。方法 3 作用下，欺骗攻击行为信息检测波长的极值水平明显下降，而标准传输信息检测波长明显上升，相较于真实波长曲线，也存在较大偏差。

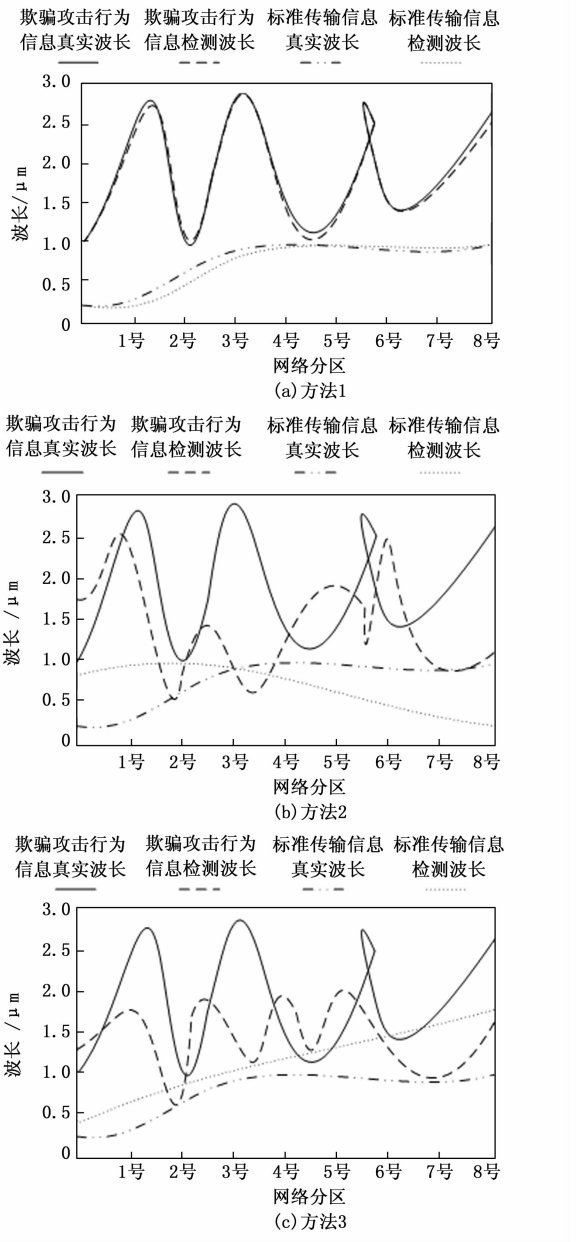

图 7　欺骗攻击行为检测

综上可知，对比方法不足以将欺骗攻击行为信息从混合数据中完全筛选出来。而设计方法可以将欺骗攻击行为信息、标准传输信息完全区分开来，保障了系统主机对欺骗攻击行为的精准检测能力，能够避免风险性数据进入网络主机，也就有效保障了工控网络的运行安全性。

5.3.2　召回率

3 种方法的召回率测试结果如图 8 所示。

分析图 8 可知，方法 1 作用下，召回率测试结果在

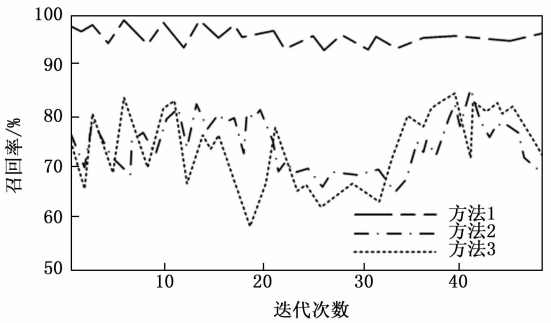

图 8 3 种方法的召回率结果

0.93～0.98 之间徘徊，表明方法 1 在欺骗攻击的识别方面表现较好，其召回率高且稳定在较高水平，能够更准确地将欺骗攻击行为识别出来，并且较少漏报。方法 2 作用下，召回率测试结果在 0.65～0.85 之间徘徊，方法 3 作用下，召回率测试结果在 0.61～0.84 之间徘徊，方法 2 和方法 3 的召回率较低且波动较大。意味着这两种对比方法对于欺骗攻击的识别能力较弱。存在一定程度的误报或漏报情况，导致无法准确地将实际上是欺骗攻击的样本正确识别出来。

6 结束语

随着工业控制网络的不断发展，网络安全问题愈加凸显出来。为了保障工业控制网络的稳定运行能力，基于对抗机器学习的工业控制网络欺骗攻击行为检测系统的设计具有重要意义。分别利用攻击行为采集模块、攻击行为处理模块、检测验证模块，对网络体系中存在的安全风险问题进行分析，在设计过程中，对抗机器学习的工业控制网络欺骗攻击行为检测系统对各类风险性信息进行了有效的训练与处理，所以网络体系中存在的各类欺骗攻击行为能够得到较为可靠的检测，而这也为工业控制网络的防御和保护提供了有力的支持。

参考文献：

[1] 赵 嘉，谷 良，吴 瑶. 基于互信息和 GWB-LSSVM 的网络攻击检测模型 [J]. 电子测量技术，2022，45 (24)：98－104.

[2] 张宇翔，韩久江，刘 建，等. ATT&CK 框架下基于事件序列关联的网络高级威胁检测系统 [J]. 计算机科学，2023，50 (s1)：710－716.

[3] 石欣然，张奇支，赵淦森，等. 一种基于少样本且不均衡的网络攻击流量检测系统 [J]. 华南师范大学学报（自然科学版），2021，53 (1)：100－108.

[4] 吕东晓，李 勇，邵 伟，等. 基于 Apriori 算法的稳控装置通信系统网络攻击检测方法 [J]. 电力信息与通信技术，2022，20 (9)：1－8.

[5] ALSHEHRI A，KHAN N，ALOWAYR A，et al. Cyberattack detection framework using machine learning and user behavior analytics [J]. Computer Systems Science & Engineering，2023，44 (2)：1679－1689.

[6] DAI Q Y，ZHANG B，XU K Y，et al. An erebus attack detection method oriented to blockchain network layer [J]. Computers，Materials & Continua，2023 (6)：5395－5431.

[7] 王嶔鑫，汪晋安，邓伟成，等. 抑制网侧交流单相接地故障下混合型 MMC 子模块过电压的谐波耦合注入策略 [J]. 电网技术，2023，47 (1)：313－322.

[8] 荣 飞，孙宗卿，徐 爽，等. 基于辅助子模块的 MMC 输出电流谐波优化控制方法 [J]. 高压电器，2023，59 (5)：154－162.

[9] 王 猛，何庆中，王 佳，等. 基于机器视觉的网络变压器模块缺陷检测系统研究 [J]. 机床与液压，2021，49 (4)：89－93.

[10] 汤彩芸，张孙杰，裴自强. 带有注意力模块的反卷积一阶检测算法研究 [J]. 小型微型计算机系统，2021，42 (6)：1199－1205.

[11] 熊 强，杨欣琦，李治文. 网络安全漏洞信息披露中多元参与主体行为策略演化博弈分析 [J]. 运筹与管理，2021，30 (7)：102－109.

[12] 彭长根，高 婷，刘惠篮，等. 面向机器学习模型的基于 PCA 的成员推理攻击 [J]. 通信学报，2022，43 (1)：149－160.

[13] 钟再敏，杨明磊，王业勤，等. 四线圈原型电机模型及其空间矢量数学表征 [J]. 电机与控制应用，2023，50 (7)：1－12.

[14] 王长健，曹雪梅，许 浩，等. 斜齿圆柱齿轮车齿加工数学模型与齿面误差分析 [J]. 河南科技大学学报（自然科学版），2022，43 (5)：15－20.

[15] 周胜利，阮琳琦，徐 睿，等. 基于关联规则特征提取的网络行为被害性识别集成优化模型 [J]. 电信科学，2023，39 (9)：129－140.

[16] 李汉伦，任建国. P2P 网络中基于特征行为检测的恶意代码传播模型 [J]. 计算机应用，2022，42 (7)：2125－2131.

[17] 郝 颖，冬 雷，王丽婕，等. 基于数学形态学去噪的光伏发电限电异常数据识别算法 [J]. 中国电机工程学报，2022，42 (21)：7843－7855.

[18] 甘 雨，郭 鹏，林立栋. 基于变分贝叶斯推断的 DPGMM 风电机组异常数据识别研究 [J]. 动力工程学报，2023，43 (7)：885－892.

[19] 王义凯，尹项根，乔 健，等. 海洋核动力平台发电机定子绕组单相接地故障风险分析与实时定位 [J]. 电力自动化设备，2022，42 (4)：178－183.

[20] 吴 坡，张江南，王 丹，等. 调度自动化主站局域网冗余可靠性风险分析 [J]. 电力系统保护与控制，2021，49 (11)：172－180.

[21] 严 钧，晏婉晨. 基于 Gumbel 分布的熵风险度量的参数估计及渐近行为 [J]. 河南师范大学学报（自然科学版），2022，50 (1)：67－72.

[22] 孙东磊，杨 思，韩学山，等. 高比例风电接入下计及时段间耦合旋转备用响应风险的动态经济调度方法 [J]. 山东大学学报（工学版），2022，52 (5)：111－122.

[23] 崔 畅，董锡超，胡 程，等. GEO 星机双基 SAR 时间同步误差对运动目标检测的影响建模与分析 [J]. 信号处理，2022，38 (1)：74－82.

[24] 张文港，张小平，李俊乐，等. 基于分段解析建模的开关磁阻电机在线转矩估算方法 [J]. 电子测量与仪器学报，2021，35 (11)：163－169.