

DOI:10.20033/j.1003-7241.(2026)01-0161-06

基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术

周晓明,王峰,弓会龙,高一洪,高梓轩

(陕西陕煤陕北矿业有限公司信息技术运维分公司,陕西榆林710000)

摘要:智慧煤炭工控网络对于煤炭的生产效率起到直接影响作用,为了解工控网络的安全状态,提出基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术。采用 Netflow 采集方式,智慧煤炭工控网络安全态势感知要素的采集,通过归一化与融合完成对初始感知要素的预处理。提取智慧煤炭工控网络安全态势特征,利用改进 IDF 算法计算特征权重,通过态势值的计算与特征匹配,得出智慧煤炭工控网络安全态势的感知结果。通过感知性能测试实验得出结论,与传统感知技术相比,优化设计技术的安全态势类型感知错误率明显降低,综合态势指标感知误差减小 0.035。为智慧煤炭工控网络的安全态势感知提供了更准确、高效的方法,为煤炭行业的智能化转型提供了强有力的技术支持。

关键词:改进 IDF 算法;智慧煤炭;工控网络;网络安全态势;态势感知

中图分类号: TP393.08 **文献标志码:** A **文章编号:** 1003-7241(2026)01-0161-06

Smart coal industry control network security situation awareness technology based on improved IDF algorithm

ZHOU Xiaoming, WANG Feng, GONG Huilong, GAO Yihong, GAO Zixuan

(Shaanxi Shaanxi Coal Shaanbei Mining Co., Ltd., Information Technology Operation and Maintenance Branch, Yulin 710000, Shaanxi, China)

Abstract: The smart coal industrial control network has a direct impact on the production efficiency of coal. In order to understand the security status of the industrial control network, a smart coal industrial control network security situation awareness technology based on the improved IDF algorithm is proposed. Using the Netflow collection method, the collection of security situational awareness elements in the smart coal industrial control network is completed through normalization and fusion to preprocess the initial perception elements. It extracts the security situation characteristics of the smart coal industrial control network, uses the improved IDF algorithm to calculate feature weights, and obtains the perception results of the security situation of the smart coal industrial control network through the calculation of situation values and feature matching. The conclusion drawn from the perception performance testing experiment is that compared with traditional perception technologies, the error rate of security situation type perception in optimized design technology is significantly reduced, and the perception error of comprehensive situation indicators is reduced by 0.035. This provides a more accurate and efficient method for the security situational awareness of smart coal industrial control networks, and provides strong technical support for the intelligent transformation of the coal industry.

Keywords: improve the IDF algorithm; smart coal; industrial control network; network security situation; situation awareness

智慧煤炭作为智慧能源的重要分支,是将新一代信息技术与煤炭产业深度融合,实现煤炭产业的数字化、网络化、自动化和智能化发展。这不仅能有效提升煤炭产业的效率和安全性,还能为整个能源行业的可持续发展提供有力支撑。智慧煤炭工控网络是智慧煤炭网络建设的重要组成部分,它基于计算机、通信、自动化和传感器等技术,实现对煤矿生产过程的全面监控和智能化管理。工控网络在智慧煤炭领域的应用,旨在提高煤矿生产的效率和安全性,降低运营成本,并为决策者提供实时、准确的数据支

持。与传统煤炭相比,智慧煤炭更加符合现代能源的建设需求,同时也会受到智慧煤炭工控网络运行状态的限制,一旦工控网络受到攻击,就会导致信息失效,将会严重影响煤矿开采工作的运行,甚至会造成停工停产。智慧煤炭工控网络的安全问题主要体现在安全威胁加剧、防御能力不足、设备安全缺失等方面,为保证智慧煤炭的生产效率,需要实时感知智慧煤炭工控网络的安全态势,因此提出智慧煤炭工控网络安全态势感知技术。

网络安全态势感知是指通过对网络环境中的各种信

收稿日期:2024-04-19;录用日期:2024-05-06

基金项目:陕西省陕煤矿业项目(2023SMHKJ-C-59)

作者简介:周晓明(1983—),男,本科,高级经济师,研究方向:煤矿自动化、信息化系统建设与维护。

引用本文:周晓明,王峰,弓会龙,等. 基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术[J]. 自动化技术与应用, 2026,45(1):161-165, 170. (ZHOU Xiaoming, WANG Feng, GONG Huilong, et al. Smart coal industry control network security situation awareness technology based on improved IDF algorithm[J]. Techniques of Automation and Applications, 2026,45(1):161-165,170.)

息进行收集、分析和处理,以实时了解网络安全状况,并及时发现和应对潜在的威胁和攻击。态势感知能够帮助组织和个人提高对网络安全事件的应对能力,通过实时监测和分析网络数据,预测和预防潜在的安全风险,从而减少网络安全事件的发生。现阶段发展较为成熟的网络安全态势感知技术主要包括。文献[1]提出的基于多源数据融合的安全态势感知技术、文献[2]提出的基于威胁情报的网络安全态势感知技术、文献[3]提出的基于深度Q学习网络的网络安全态势感知技术以及文献[4]提出的基于网络流量的网络威胁态势感知技术,其中文献[1]提出的技术采用“平台+端”的微服务架构,成功地设计和实现了煤矿安全态势感知分析平台。该平台基于多源数据的融合,能够有效地对矿井安全风险数据进行集成,并通过多协议适配技术确保数据集成的高效性。此外,建立了数据可信度量分析指标,以准确评估安全风险数据的质量。基于中性参照对象,构建了风险评价指标体系,能够对安全风险进行量化分级评价,通过建立监测数据特征图谱,实现了矿井典型异常的自动识别。借助GIS技术与空间插值技术,实现了安全态势的可视化分析。文献[2]提出技术通过比较外源威胁情报与系统内部安全事件之间的相似度,对目标系统进行威胁察觉。根据系统内部的威胁信息,生成内源威胁情报,从而更全面地了解网络安全的态势。文献[3]提出技术在分析威胁传播的基础上,构建了威胁传播-访问关系网络,以深入探究威胁的传播路径和影响范围。建立了随机博弈模型,模拟威胁行动与保护策略实施之间的博弈过程,通过求解混合策略纳什均衡问题,为应对不同的安全威胁提供了有效的策略参考。分析攻击信息、漏洞信息和防御措施的动态变化,从主机和网络两个层面进行网络态势预测。而文献[4]提出技术利用流熵算法生成态势信息,介绍基于机器学习的威胁态势理解方法,通过改进AdaBoost算法,利用上传的态势信息来判断流量中的威胁,根据威胁情报得出防御措施,得出网络安全态势的感知结果。然而上述传统感知技术在运行过程中存在明显的感知误差大的问题,为此引入改进IDF算法。

IDF算法,也就是逆向文件频率算法,是一种常用于信息检索和文本挖掘的权重计算方法。IDF算法广泛应用于文本挖掘、信息检索、自然语言处理等领域,用于提高检索和分类的准确性。改进IDF算法充分考虑词序信息和语义之间的相似度,保证文本的分类效果。利用改进IDF算法优化设计智慧煤炭工控网络安全态势感知技术,以期提升对工控网络安全态势的感知性能。

1 工控网络安全态势感知技术

现阶段较为成熟的网络安全态势感知技术主要包括,多源数据融合技术、威胁情报技术、深度Q学习网络技术以及网络流量威胁态势感知技术。这些技术均取得了一定的成果,但仍存在感知误差大的问题。为此,引入改进IDF算法,以提高感知精度。

优化设计的网络安全态势感知技术主要包括态势要素获取、态势理解和态势评估感知三部分,其中态势理解部分以改进IDF算法作为支持^[5],对网络的运行特征和权重进行量化度量,根据特征提取结果,确定当前智慧煤炭工控网络的状态、类型、感知指标和预测结果,完成对网络安全态势的感知工作。

1.1 设置安全态势感知标准

从组成结构上来看,智慧煤炭工控网络由网络节点、路由器节点、服务器节点和通信链路等部分组成,其中网络节点为煤炭自动化生产与监测设备,网络中的任意两个节点之间的信道即为网络通信链路^[6]。智慧煤炭工控网络的安全态势包括。网络攻击、网络访问控制、网络安全等多个状态,其中工控网络中的攻击态势工作原理,如图1所示。

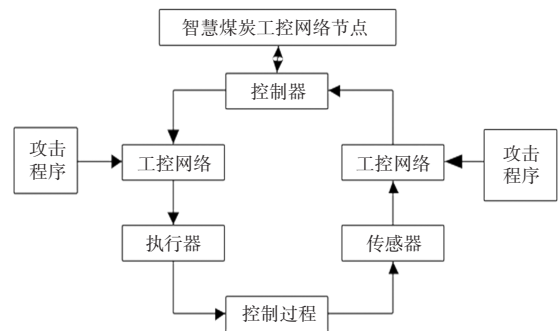


图1 智慧煤炭工控网络攻击原理图

Fig. 1 Schematic diagram of intelligent coal industrial control network attack

从图1中可以看出,智慧煤矿的控制过程在控制器的支持下进行,控制器具备接收来自不同地理位置传感器测量数据的能力,并可以通过通信网络将控制指令有效地发送到各个分布式的执行器中^[7]。然而,在工控系统运行的某个时刻 t ,系统不幸遭受了攻击者的恶意攻击。这次攻击导致在特定的运行时间 T 内,传感器发出的信号与控制器实际接收到的信号之间产生了明显偏差,即

$$\begin{cases} \tilde{y}_i(t) = \begin{cases} y_i(t), & t \notin T_{\text{attack}} \\ a_i(t), & t \in T_{\text{attack}} \end{cases} \\ \tilde{y}_i(t) \in [y_i^{\min}, y_i^{\max}] \end{cases} \quad (1)$$

式中, $\tilde{y}_i(t)$ 表示的是智慧煤炭工控网络中控制器实际接收到的信号, $y_i(t)$ 和 $a_i(t)$ 分别表示智慧煤炭传感器的信号和攻击端的攻击信号, T_{attack} 为工控网络遭受的攻击时间, y_i^{\min} 和 y_i^{\max} 对应的是智慧煤炭中传感器检测范围内的最小值和最大值^[8]。根据智慧煤炭工控网络中攻击程序的执行原理,设置工控网络在安全攻击态势下的标准特征,将其标记为 $\gamma_{b-\text{attack}}$ 。按照上述方式,可以得出其他智慧煤炭工控网络安全态势下的标准特征。

1.2 态势感知要素采集与预处理

智慧煤炭工控网络安全态势要素包括流量信息、资产信息、脆弱性信息、威胁信息以及性能信息等,其中网络流量信息采用Netflow采集方式,也就是以网络数据包交换的技术^[9]。Netflow采集技术能够显著提升网络设备间的

数据交换速度,并且具备实时统计和记录高速转发的网络数据流的能力。在 NetFlow 记录的数据中,仅包含与各个流相关的统计信息,例如端口和 IP 地址等,而不包括 MAC 地址信息。这一特点使得 Netflow 技术在处理大规模网络流量时更加高效和准确。通过采集必要的信息,Netflow 技术有助于管理员快速识别和解决网络问题,提高网络性能和安全性^[10]。在智慧煤炭工控网络流量信息采集过程中,Netflow 数据采集器首先需要选择并监听一个或多个网络接口,以便能够捕获流经这些接口的数据包,当数据包在网络接口上流动时,Netflow 数据采集器可以直接捕获数据包。捕获的数据包需要被解析,以提取出关键的流量信息,如源 IP 地址、目的 IP 地址、协议类型等,并对流量的大小、持续时间、发送频率等指标进行统计,自动生成 Netflow 记录。最终采集的智慧煤炭工控网络流量数据为

$$x_L(t) = \kappa_{\text{Netflow}} x_s(t) (1 - \chi) \quad (2)$$

式中, $x_s(t)$ 表示的是 t 时刻智慧煤炭工控网络的实际工作流量, κ_{Netflow} 为 Netflow 数据采集系数, χ 表示的是网络流量在工控网络中的损耗系数。

按照上述方式可以得出资产、威胁等其他信息的采集结果,并对初始采集数据进行归一化处理,处理结果为

$$x'_L(t) = \frac{f_{\max}(x_L(t)) - f_{\min}(x_L(t))}{x_L(t) - f_{\min}(x_L(t))} \quad (3)$$

式中, $f_{\max}()$ 和 $f_{\min}()$ 分别表示的是最大值和最小值求解函数,将初始采集网络数据代入到式(3)中,即可得出数据归一化的处理结果。以式(3)输出的归一化数据为处理对象,采用数据级融合的方式得出安全态势要素数据的融合结果为

$$x_R = f_{\text{fusion}}(x'_L(t_1), x'_L(t_2), \dots, x'_L(t_{n_c})) \quad (4)$$

式中, n_c 为采集的智慧煤炭工控网络安全态势要素数据量。

根据智慧煤炭工控网络的运行状态,完成对网络安全态势要素数据的采集与预处理工作。

1.3 提取态势感知要素特征

智慧煤炭工控网络安全态势感知要素的特征分量包括文档频率、信息增益、流量峰值、流量均值等,其中文档频率是衡量特征词在文档数据集中出现的频率。其核心思想是设定一个文档频率的范围,然后统计每个特征词的文档频率值^[11]。只有当特征词的文档频率值落在设定的范围内时,该特征词才会被保留;否则,该特征词将被剔除。信息增益基于信息熵的概念,通过计算属性出现前后的信息熵差值来衡量该属性对数据集的划分效果^[12]。信息熵用于度量数据集的混乱程度或不确定性,其值越低表示数据越有序、纯度越高。信息增益特征分量的计算公式为

$$\lambda_{\text{gain}}(x_R, A) = E(x_R) - E(x_R | A) \quad (5)$$

式中, A 表示的是信息属性, $E(x_R)$ 和 $E(x_R | A)$ 对应的是工控网络要素数据和属性划分后的子集信息熵。式(5)输

出的信息增益越大,表示该属性对数据集的划分效果越好,即该属性能够提供更多有关数据分类的信息。另外,流量峰值和流量均值特征分量的提取过程可以量化表示为

$$\begin{cases} \lambda_{\max} = f_{\max}(x_R) \\ \lambda_{\text{avg}} = \frac{\sum_{i=1}^{n_c} x_R(i)}{n_c} \end{cases} \quad (6)$$

将智慧煤炭工控网络安全态势要素的动态采集结果代入到上述公式中,即可完成对态势要素特征的采集工作。

1.4 计算工控网络特征权重

针对智慧煤炭工控网络中的文本要素,例如网络日志、安全报警日志等,采用改进 IDF 算法对网络特征分量的权重进行计算。IDF 算法的改进方式体现在 TF 融合方面,也就是与词频进行融合。改进的 IDF 算法通过结合词频和逆文档频率两个因素来衡量词的重要程度。改进 IDF 算法的工作流程如图 2 所示。

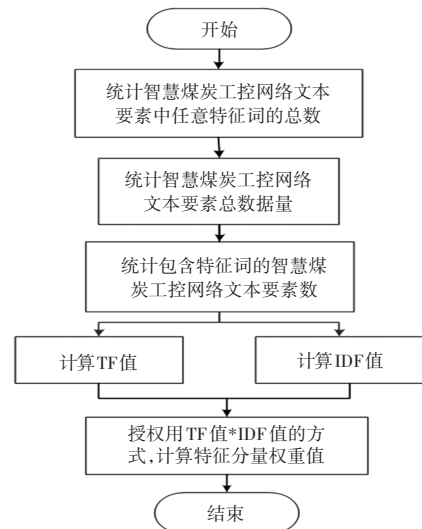


图 2 改进 IDF 算法工作流程图

Fig. 2 Workflow diagram of improving IDF algorithm

传统的 IDF 算法主要对样本数据的逆文档频率进行计算,也就是出现任意特征项文档个数的倒数,计算公式为

$$\psi_{\text{IDF}} = \lg\left(\frac{n_c}{n_r} + 0.01\right) \quad (7)$$

式中, n_r 为初始采集工控网络运行数据中出现特定特征项的次数。在此基础上,综合考虑词频和跨度权值完成对 IDF 算法的改进,并得出提取网络安全态势要素任意特征分量对应权重的计算结果为

$$\omega_i = \bar{\omega}_i \times \psi_{\text{IDF}} \times K_i \quad (8)$$

式中, $\bar{\omega}_i$ 为任意特征 i 在智慧煤炭工控网络中出现的频率, K_i 表示的是跨度权值,该参数的计算公式为

$$K_i = \frac{q_i}{Q} \quad (9)$$

式中, q_i 和 Q 分别表示特征分量出现的位置和网络要素总数。通过上述公式的联立,在改进 IDF 算法的支持下,得出工控网络特征分量对应权重的计算结果。

1.5 实现安全态势感知

1.5.1 感知智慧煤炭工控网络安全态势状态与类型

根据智慧煤炭工控网络安全态势感知要素特征及其权重的计算结果,计算网络安全态势的融合特征,融合结果为

$$\lambda = \omega_1 \lambda_f + \omega_2 \lambda_{\text{gain}}(x_R, A) + \omega_3 \lambda_{\text{max}} + \omega_4 \lambda_{\text{avg}} \quad (10)$$

式中, ω_1 、 ω_2 、 ω_3 和 ω_4 分别表示的是文档频率、信息增益、流量峰值和流量均值特征分量的权重值。最终将提取的融合权重和设置的安全态势感知标准进行匹配,得出安全态势状态的感知结果。公式(11)表示的是特征匹配过程为

$$s' = \frac{\lambda \cdot \lambda_b(i)}{\|\lambda\| \cdot \|\lambda_b(i)\|} \quad (11)$$

若式(11)的计算结果高于阈值 s_0 , 证明当前智慧煤炭工控网络的安全态势与 $\lambda_b(i)$ 一致,按照上述方式更换 $\lambda_b(i)$ 的具体取值,直至得出满足阈值条件的态势感知结果为止。

1.5.2 求解智慧煤炭工控网络安全态势感知指标

智慧煤炭工控网络安全态势指标的感知指标具体包括安全态势指数、攻击频率威胁、服务风险等,其中安全态势指数的求解结果为

$$T = \sum_{j=1}^M n_j(t) \quad (12)$$

式中, $n_j(t)$ 表示 t 时段 j 种入侵对象的入侵次数, M 为入侵的类型总数。另外,攻击威胁态势指标的求解结果可以表示为

$$f_{\text{threaten}} = \frac{f_{\text{attack}}}{1 + f_{\text{attack}}} \quad (13)$$

式中, f_{attack} 表示的是智慧煤炭工控网络的攻击频率。同理可以完成对所有安全态势感知指标的求解。设置工控网络安全态势的综合感知指标为安全风险,该指标的求解结果为

$$c = T \oplus f_{\text{threaten}} \oplus \mu_{\text{service}} \quad (14)$$

式中, μ_{service} 表示的是服务风险指标的计算结果, \oplus 表示的是耦合符号。将相关数据代入到式(14)中即可得出工控网络安全态势综合感知指标的计算结果。

智慧煤炭工控网络安全态势的预测是态势感知的内容之一,主要是根据网络当前的安全态势评估结果和影响因素,得出未来任意时刻安全态势的预测结果。

2 性能测试实验分析

为了验证优化设计基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术的态势感知性能,采用白盒测试的方式设计性能测试实验。实验基本原理为:在智慧煤炭环境中配置工控网络环境,并通过攻击场景的生成确定工控网络的实际安全态势类型和态势指标的实际取值,通过优化设计感知技术的开发与运行得出工控网络的态势感知结果,与网络态势的实际值进行比对,得出反映技术感知性能的测试结果。

2.1 配置智慧煤炭工控网络环境

此次实验选择某煤矿作为工控网络布设区域,配置的

工控网络需要完成煤炭质量检测、煤炭运输管理、煤炭配送安全保障、煤炭配煤智能化、煤炭计量标准化等工作内容,煤矿的分布面积为 80.2 km^2 ,因此配置工控网络的覆盖范围不得小于 80.2 km^2 。智慧煤炭工控网络拓扑结构的配置情况,如图3所示。

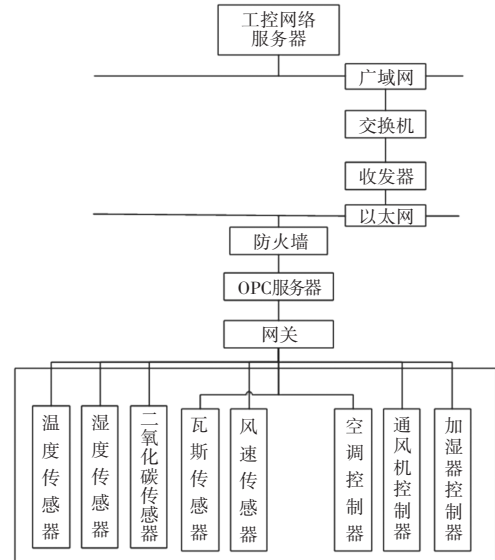


图3 智慧煤炭工控网络拓扑结构配置图

Fig. 3 Topological structure configuration of intelligent coal industry control network

所配备的智能煤炭工业控制网络服务器安装了 WindowServer2003 操作系统。在工业控制网络的关键节点上,设置了 IIS.6.0 网络服务、MSSQLserver2000 数据库服务以及 FTP 服务器。经过精确的配置,确保了 IIS 能够顺利发布各类应用,并为了保障网络安全,已关闭目标服务器主机的防火墙功能。此外,该网络系统被设定为二级等级保护标准,以满足相应的安全防护需求。除了主机设备外,在智慧煤炭环境中布设多个环境传感器和控制器作为网络节点,实现环境监测与控制功能。配置的智慧煤炭工控网络中遗留 CVE-2023-2506、CVE-2023-0057、CVE-2024-0653、CVE-2024-5112,共4项安全漏洞。

2.2 生成工控网络安全攻击场景

在配置的工业控制网络中,部署一台漏洞扫描设备作为模拟攻击源,对目标工控网络发起网络攻击。同时,在另一台服务器上部署了入侵检测系统。这三台设备通过交换机相互连接。为了实现对攻击流量的监控和检测,对交换机进行了数据流镜像配置,将漏洞扫描系统以及目标工控网络的流量镜像传输给 IDS 入侵检测系统服务器,从而生成工控网络安全的攻击场景。根据攻击场景的生成情况,确定智慧煤炭工控网络的初始安全态势,并对其进行量化标记,如表1所示。

按照上述方法可以得出作用在智慧煤矿工控网络中其他安全攻击场景的生成结果,并确定攻击场景下工控网络的实际态势情况数据。

2.3 描述性能测试实验过程

采用 MyEclipse 作为集成开发环境分别对改进 IDF

算法和优化设计的基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术进行开发,将智慧煤炭工控网络的实时运行数据输入到态势感知方法对应的运行程序中,首先输出网络特征及权重的计算结果。如图 4 所示。

表 1 工控网络安全态势初始数据表

Tab. 1 Initial data table of industrial control network security situation

工控网络场景	工控网络安全态势类型	综合安全态势感知指标	安全态势指数	攻击频率威胁
1	攻击态势	0.15	0.05	0.89
2	攻击态势	0.08	0.06	0.74
3	网络访问控制态势	0.44	0.79	0.24
4	网络访问控制态势	0.51	0.82	0.37
5	网络威胁态势	0.39	0.58	0.47
6	网络威胁态势	0.23	0.61	0.31
7	网络安全态势	0.97	0.86	0.04
8	网络安全态势	0.95	0.91	0.12

智慧煤炭工控网络安全态势感知				
态势要素	工控网络特征权重			
态势评估	特征分量	文档频率	权重值	0.10
特征分量	特征分量	信息增益	权重值	0.10
权重计算	特征分量	互信息	权重值	0.05
指标计算	特征分量	熵信息	权重值	0.05
态势预测	特征分量	流量峰值	权重值	0.15
	特征分量	流量均值	权重值	0.15
	特征分量	流量方差	权重值	0.10
	特征分量	脉冲因子	权重值	0.10
	特征分量	裕度因子	权重值	0.10
	特征分量	波形因子	权重值	0.10

图 4 智慧煤炭工控网络特征权重计算结果

Fig. 4 Characteristics and weight calculation results of smart coal industrial control network

在此基础上,得出智慧煤炭工控网络安全态势类型和指标的感知结果,如图 5 所示。

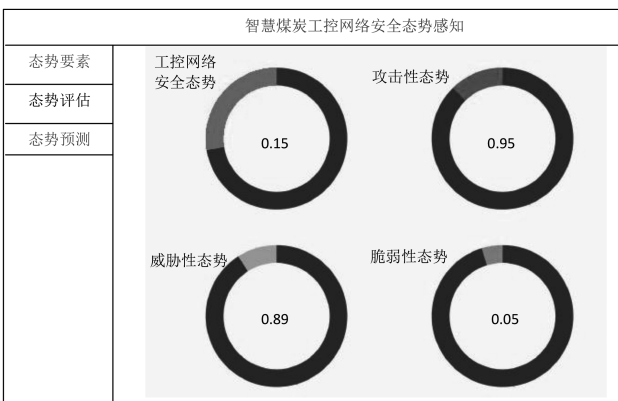


图 5 智慧煤炭工控网络安全态势感知结果

Fig. 5 Results of security situation awareness for smart coal industry control network

为体现出优化设计感知技术在性能方面的优势,设置传统的基于多源数据融合的安全态势感知技术和基于威胁情报的网络安全态势感知技术作为实验的对比技术,在相同的实验环境下完成对应的安全态势感知输出结果。

2.4 网络安全态势感知性能测试

实验设置安全态势类型感知错误率和安全态势指标感知误差作为感知技术性能的量化测试指标,其中态势类型感知错误率的测试结果

$$\eta = \frac{n_{err}}{n_{all}} \times 100\% \quad (15)$$

式中,变量 n_{err} 和 n_{all} 分别表示的是网络安全态势类型感知错误的实验次数以及实验执行的总次数。另外,安全态势指标感知误差指标的数值结果为

$$\varepsilon = |c - c_{set}| \quad (16)$$

式中, c 和 c_{set} 分别表示的是安全态势综合指标的感知值和设定值。

最终计算得出态势类型感知错误率和态势指标感知误差取值越小,证明对应技术的感知性能越优。

2.5 测试实验结果与分析

通过相关数据的统计与公式(15)的计算,得出优化设计技术对应安全态势类型感知错误率的测试对比结果,如图 6 所示。

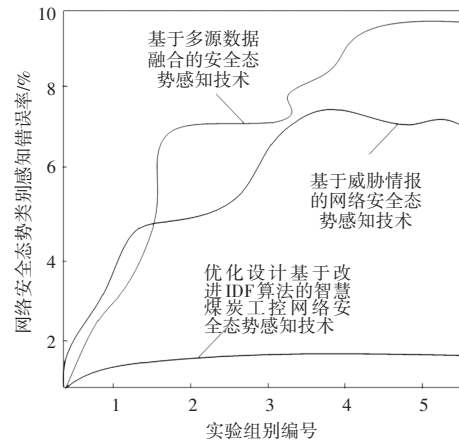


图 6 智慧煤炭工控网络安全态势类型感知错误率

Fig. 6 Error rate of perception of security situation types in smart coal industrial control network

从图 6 中可以直观地看出,与传统感知技术相比,优化设计技术得出的工控网络安全态势类型感知错误率更低。另外,安全态势指标感知误差的测试结果见表 2。

表 2 工控网络安全态势指标感知误差测试数据表

Tab. 2 Test data table for perception error of industrial control network security situation indicators

工控网络场景	基于多源数据融合的安全态势感知技术	基于威胁情报的网络安全态势感知技术	优化设计基于改进 IDF 算法的智慧煤炭工控网络安全态势感知技术
1	0.19	0.17	0.14
2	0.15	0.10	0.08
3	0.40	0.41	0.43
4	0.58	0.56	0.50
5	0.32	0.34	0.38
6	0.18	0.20	0.22
7	0.91	0.94	0.96
8	0.91	0.94	0.95

(下转第 170 页)