面向 OT 跨场景的轻量化 TCA 入侵检测方法

赵筱峤 [1]
ZHAO Xiaoqiao

摘 要 针对工业控制系统操作技术（operational technology, OT）内部跨协议、跨系统、跨时间场景下入侵检测模型迁移性能下降的问题，文章提出一种基于改进迁移成分分析（transfer component analysis, TCA）的轻量化入侵检测方法。采用径向基函数（radial basis function, RBF）核提升非线性映射能力，设计周期权重最大均值差异矩阵增强周期性攻击的跨域识别，并通过降维、剪枝与 INT8 量化实现轻量化部署。实验结果表明，该方法在多个公开数据集上检测精度与先进方法相当，推理延迟仅 15.8 ms，模型体积仅 8 MB，有效平衡了检测精度与实时性，具备边缘设备部署优势。

关键词 工业控制系统；入侵检测；迁移成分分析；域自适应；轻量化模型

doi：10.3969/j.issn.1672-9528.2026.07.024

0 引言

工业控制系统（industrial control systems, ICS）是国家关键基础设施的核心组成部分。随着工业互联网发展，ICS 的入侵攻击事件频发，不仅破坏生产连续性，更可能引发安全事故，如某国电网遭到“黑能量”攻击造成大量用户停电、台积电遭勒索病毒入侵导致生产基地停摆 [1]。入侵检测系统（intrusion detection systems, IDS）成为工控安全防护的重要手段。然而，工控网络协议专有、实时性要求高、攻击样本稀少且标注成本高，同时边缘设备资源受限对模型轻量化也提出了挑战。

在当前的研究中，传统机器学习方法依赖协议先验知识，难以覆盖专有协议；深度学习方法计算量大且需要大量样本。迁移学习为解决上述问题提供了新思路，能缓解样本稀缺问题，但直接将通用网络模型应用于工控领域会因特征分布差异导致性能退化。现有研究主要分为两大类：

（1）基于分布对齐的迁移学习，如 Sun 等 [2] 提出的相关对齐（correlation alignment, CORAL）通过线性变换对齐源域和目标域的协方差矩阵，从而最小化两个域的分布差异，Long 等 [3] 提出的联合分布适配（joint distribution adaptation, JDA）同时对齐源域和目标域的边缘分布与条件分布，实现了更全面的域自适应，该类方法的优势是计算轻量、效率高，但通常采用线性假设且未针对工控周期性攻击优化；

（2）基于对抗的迁移学习，如王昊冉等 [4] 提出基于生成对抗迁移学习网络的工控系统未知攻击入侵检测方法

1. 三峡大学计算机与信息学院 湖北宜昌 443100

参考文献：

[1] 李皓皓，张进，罗亚中．基于机动目标滤波估计的航天器主动规避策略 [J]. 力学学报，2020,52(6):1560-1568.

[2] 尹聚祺，杨震，罗亚中，等．空间机动目标跟踪的改进自适应 IMM 算法 [J]. 系统工程与电子技术，2021,43(12):3658-3666.

[3] 戴定成，姚敏立，蔡宗平，等．改进的马尔可夫参数自适应 IMM 算法 [J]. 电子学报，2017,45(5):1198-1205.

[4] 许登荣，程水英，包守亮．自适应转移概率交互式多模型跟踪算法 [J]. 电子学报，2017,45(9):2113-2120.

[5] 吴志远，毛洋，周敬烈．基于变维交互式多模型算法的近区目标跟踪研究 [J]. 电子测量技术，2017,40(1):67-70.

[6] 赵兵，王桁．交互多模型 Kalman 滤波下的目标跟踪应用研究 [J]. 电子测量技术，2019,42(11):83-86.

[7] 王军，吴成富，陈怀民，等．一种改进的交互多模型算法 [J]. 计算机工程与应用，2007(13):83-85.

[8] 杨冬英，贺江鹏．针对机动目标的改进型交互多模型跟踪算法研究 [J]. 计算机工程与科学，2022,44(4):686-691.

【作者简介】

肖杰（1988—），女，四川成都人，硕士研究生，工程师，研究方向：雷达数据处理、信息融合。

（收稿日期：2025-11-12）

（generative adversarial transfer learning, GATL），通过域对抗迁移学习融合数据的差异性、共性，提高工控入侵检测模型对目标域未知攻击的检测和泛化能力，Feng 等[5]提出基于自注意力 Wasserstein 生成对抗网络的数据增强方法，提升了少数类攻击的检测性能，该类方法泛化能力强，但模型复杂、延迟高。上述方法普遍存在三个共性挑战：非线性适配能力不足、对工控周期性攻击关注不足、模型体积和延迟难以满足边缘部署需求。

针对上述问题，本文聚焦 OT 内部跨协议、跨行业场景，提出改进的迁移成分分析（TCA）入侵检测方法 TCA-IDS。通过引入时序权重 RBF 核提升非线性适配能力，设计带周期权重的（maximum mean discrepancy, MMD）矩阵，增强周期性攻击的域不变特征提取，采用降维、剪枝与 INT8 量化实现轻量化部署，并使用 XGBoost（extreme gradient boosting）作为分类器，在跨协议、跨系统、跨时间等多种工控场景中实现了检测精度的实时性的有效平衡。

1 相关理论

1.1 最大均值差异 MMD

最大均值差异 MMD 是度量两个数据分布之间差异的常用指标，被广泛用于域自适应领域。其核心思想是：将两个分布的数据映射到再生核希尔伯特空间（reproducing kernel hilbert space, RKHS）中，计算两个数据集样本均值差异来量化分布距离[6]。MMD 值越小，表示两个分布越相似。MMD 计算简单、无需密度估计，是 TCA 等域自适应方法的核心度量指标。

1.2 迁移成分分析 TCA

迁移成分分析（TCA）是一种经典域自适应学习方法。其核心思想是：学习一个特征映射，将源域和目标域数据投影到公共的再生核希尔伯特空间中，使得投影后两域的 MMD 最小化[7]。TCA 通过核矩阵和变换矩阵将原始高维特征降维至低维子空间，在降维的同时最小化源域与目标域的分布差异。TCA 的优势在于无需目标域标签即可实现域自适应，计算效率较高。然而，标准 TCA 通常采用线性核，对于工控协议间复杂的非线性分布关系适配能力有限，且未针对工控周期性攻击进行优化。为此，本文引入 RBF 核并设计周期权重 MMD 矩阵加以改进。

1.3 XGBoost

XGBoost 是 Chen 等[8]提出的集成学习算法，基于梯度提升决策树（gradient boosting decision tree, GBDT）框架进行了多项优化。其核心思想是：通过迭代训练决策树，每棵新树拟合前一棵树预测结果的残差，最终将所有树的预测结果累加作为最终输出。相较于传统 GBDT，XGBoost 引入了二阶梯度近似、列采样、并行计算等优化技术，训练效率高、泛化能力强，且支持模型剪枝，便于轻量化部署。本文采用 XGBoost 作为 TCA-IDS 的基分类器，用于对降维后的域不变特征进行分类检测。

2 TCA-IDS 方法设计

2.1 问题形式化与 MMD 度量

设源域 $D_s=\{(\boldsymbol{x}_{s,i},\boldsymbol{y}_{s,j})\}_{i=1}^{n_s}$，其中 $\boldsymbol{x}_{s,i}\in\mathbb{R}^d$ 为特征向量，$y_{s,i}\in\{0,1\}$ 为对应的攻击/正常标签（记源域标签集为 $Y_s=\{y_{s,1},y_{s,2},...,y_{s,n}\}$），目标域数据为 $D_t=\{\boldsymbol{x}_{t,j}\}_{j=1}^{n_t}$，训练阶段假设其无标签。源域与目标域特征空间相同但分布不同。采用 MMD 度量域分布差异：

$$\Delta(\boldsymbol{X}_s,\boldsymbol{X}_t)=\left\|\frac{1}{n_s}\sum_{i=1}^{n_s}\phi(\boldsymbol{x}_{s,i})-\frac{1}{n_t}\sum_{j=1}^{n_t}\phi(\boldsymbol{x}_{t,j})\right\|_{\mathcal{H}}^2 \tag{1}$$

式中：$\boldsymbol{X}_s$、$\boldsymbol{X}_t$ 为源域和目标域样本集，n_s、n_t 为样本数，$\phi(\cdot)$ 为一种非线性映射，将原始数据映射到再生核希尔伯特空间 $\mathcal{H}$，$\|\cdot\|_{\mathcal{H}}$ 为 RKHS 中的范数。

将源域和目标域样本合并为 $\boldsymbol{X}=[\boldsymbol{X}_s,\boldsymbol{X}_t]\in\mathbb{R}^{d\times n}$，$n=n_s+n_t$。定义 MMD 矩阵 $\boldsymbol{M}\in\mathbb{R}^{n\times n}$：

$$M_{ij}=\begin{cases}\dfrac{1}{n_s^2}, & \boldsymbol{x}_i,\boldsymbol{x}_j\in\boldsymbol{X}_s\\ \dfrac{1}{n_t^2}, & \boldsymbol{x}_i,\boldsymbol{x}_j\in\boldsymbol{X}_t\\ -\dfrac{1}{n_sn_t}, & \text{otherwise}\end{cases} \tag{2}$$

则 $\Delta=\mathrm{tr}(\boldsymbol{KM})$，其中 $\boldsymbol{K}$ 为核矩阵。为消除均值偏移，对核矩阵中心化：$\tilde{\boldsymbol{K}}=\boldsymbol{HKH}$，$\boldsymbol{H}=\boldsymbol{I}_n-\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$（$\boldsymbol{I}_n$ 为单位矩阵，$\mathbf{1}$ 为全列向量）。

2.2 时序权重 RBF 核

为适应工控系统可能出现的概念漂移（如设备老化、工况调整），引入时序权重 RBF 核。设样本 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的时间戳分别为 t_i 和 t_j（已归一化至 [0,1]），时序衰减因子 $w_{ij}=\exp(-\beta|t_i-t_j|)$，$\beta=0.03$ 为时序衰减率。核函数定义为：

$$k_{\mathrm{temp}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=w_{ij}\cdot k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp\left(-\beta|t_i-t_j|-\gamma\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\right) \tag{3}$$

式中：γ 为 RBF 核参数，通过 5 折交叉验证在 {0.001, 0.01, 0.1, 1, 10} 中选取；$\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2$ 为特征欧氏距离平方。该核使 MMD 更加关注近期数据，缓解工况漂移影响。

2.3 周期权重 MMD 矩阵

工控网络中针对 PROFINET 的 DoS、传感器数据伪造等攻击具有固定间隔的周期性特征。为突出周期性行为模式，定义每个样本的周期强度 $p_i=[0,1]$（基于相邻数据包时间间隔的傅里叶变换主频幅值，幅值越大周期性越强）。构建周期权重对角矩阵 $\boldsymbol{P}=\mathrm{diag}(p_1,p_2,\cdots,p_n)$，对中心化核矩阵进行加权：

$$\tilde{\boldsymbol{K}}_{\mathrm{p}}=\boldsymbol{P}\tilde{\boldsymbol{K}}\boldsymbol{P} \tag{4}$$

式中：$\tilde{\boldsymbol{K}}_{\mathrm{p}}$ 为周期加权后的核矩阵。

相应 MMD 矩阵调整为 $\boldsymbol{M}_{\mathrm{p}}=\boldsymbol{PMP}$，则 $\Delta=\mathrm{tr}(\tilde{\boldsymbol{\kappa}}_{\mathrm{p}}\boldsymbol{M}_{\mathrm{p}})$。周期强的样本（$p_i$ 较大）在核矩阵中的贡献被放大，从而迫使模型更关注周期性攻击的域不变特征提取。

2.4 学习域不变特征变换

TCA 学习变换矩阵 $\boldsymbol{W}\in\mathbb{R}^{d\times m}(m<d)$，将数据投影到低维潜在空间 $\boldsymbol{z}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}$ 中，最小化域差异的同时保留数据方差。优化目标为：

$$\min_{\boldsymbol{W}}\ \mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\tilde{\boldsymbol{M}}_{\mathrm{p}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W})+\mu\|\boldsymbol{W}\|_{\mathrm{F}}^{2}\quad \text{s.t.}\quad \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I} \tag{5}$$

式中：$\boldsymbol{X}=[\boldsymbol{X}_s,\boldsymbol{X}_t]$，$\tilde{\boldsymbol{M}}_{\mathrm{p}}=\boldsymbol{PMP}$，$\mu=0.01$ 为正则化系数，$\|\boldsymbol{W}\|_{\mathrm{F}}^{2}$为 Frobenius 范数平方，$\boldsymbol{H}$ 为中心矩阵，$\boldsymbol{I}$ 为单位矩阵。

约束条件保证投影后特征单位方差且去相关。解广义特征值问题$(\boldsymbol{X}\tilde{\boldsymbol{M}}_{\mathrm{p}}\boldsymbol{X}^{\mathrm{T}}+\mu\boldsymbol{I})w=\lambda\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}$，取前 $m=50$ 个最小特征值对应的特征向量构成 $\boldsymbol{W}$，此时方差解释率达到 95.2%。

2.5 轻量化优化

（1）特征剪枝：计算投影后源域特征 $\boldsymbol{Z}_s=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_s$ 各维度的方差贡献率，保留累计贡献率达 95% 的前 $m'=38$ 个维度，得到剪枝后的变换矩阵 $\boldsymbol{W}_{\mathrm{pruned}}\in\mathbb{R}^{d\times m'}$。投影过程公式为：

$$\boldsymbol{Z}_s^{\mathrm{pruned}}=\boldsymbol{W}_{\mathrm{pruned}}^{\mathrm{T}}\boldsymbol{X}_s,\boldsymbol{Z}_t^{\mathrm{pruned}}=\boldsymbol{W}_{\mathrm{pruned}}^{\mathrm{T}}\boldsymbol{X}_t \tag{6}$$

（2）模型剪枝与量化：在 $\boldsymbol{Z}_s^{\mathrm{pruned}}$及标签 $\boldsymbol{Y}_s$ 上训练 XGBoost（学习率 0.08，最大深度 6，迭代 120 轮）。然后对 XGBoost 进行决策树剪枝（按重要性排序保留前 50 棵）和 INT8 量化（32 位浮点转 8 位整数）。最终模型体积由 32 MB 压缩至 8 MB，推理延迟降至 15.8 ms，可部署于边缘 PLC 设备。

2.6 整体算法流程

基于 TCA 的域自适应入侵检测流程如图 1 所示。

3 实验验证与结果分析

3.1 实验设置

硬件环境：Intel i7-12700H CPU，32 GB DDR5 内存，NVIDIA RTX 3060 GPU（6 GB 显存）。西门子 S7-1200PLC 作为资源受限的边缘设备，用于验证模型在真实工控环境下的推理性能。

数据集：公开数据集 WUSTL-IIoT-2021（Modbus、EtherNet/IP 协议）；密西西比州立大学（MSU）SCADA 实验室发布的天然气管道数据集与 SWaT 水处理数据集，覆盖燃气与水务两个不同工业系统：某工控系统周一至周三正常与攻击流量数据，模拟概念漂移场景。

对比基线方法：CORAL（协方差对齐）、JDA（联合分布适应）、DAN[9]（domain adversarial neural network/ 域对抗神经网络）、Fed-IDS[10]（federated learning-based intrusion detection system/ 基于联邦学习框架的入侵检测系统）、Meta-IDS[11]（meta-learning-based intrusion detection system / 基于元学习的入侵检测系统）。

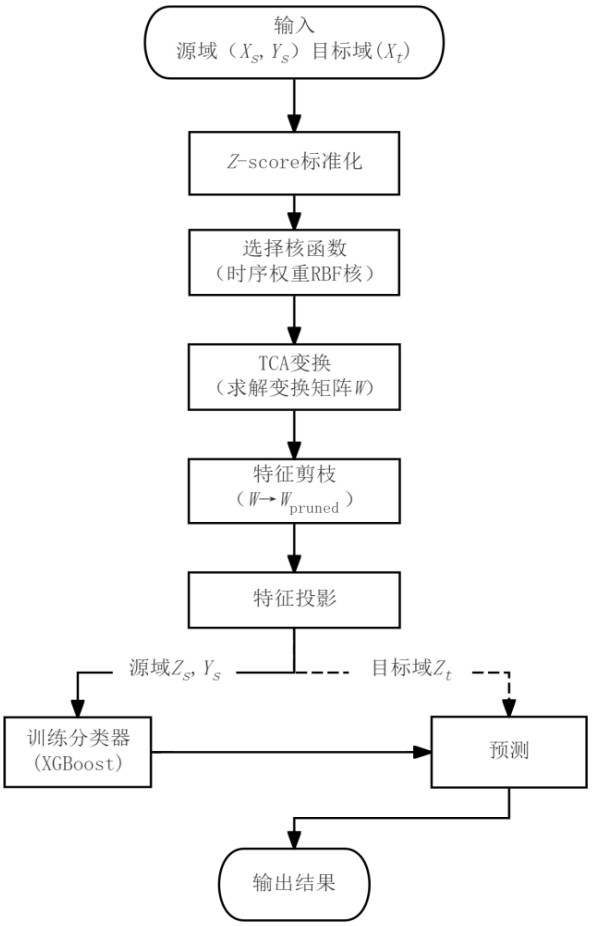

图 1 基于 TCA 的域自适应入侵检测流程图

评价指标：F_1-score（精确率与召回率的调和平均）、推理延迟（模型对单个样本进行预测的平均耗时）、模型体积。

3.2 三种跨域场景性能

根据上述数据集设计三类典型跨域场景：

（1）跨协议场景 A：Modbus→EtherNet/IP；

（2）跨系统场景 B：天然气管道 → 水处理；

（3）跨时间场景 C：周一、周二训练 → 周三测试。

实验结果如表 1 所示。

表 1 各方法跨域性能对比

方法	场景 A 的 F_1/%	场景 B 的 F_1/%	场景 C 的 F_1/%	推理延迟 /ms
CORAL	75.6±1.8	65.4±2.7	78.9±2.1	46.5
JDA	79.2±1.6	68.1±2.4	81.3±1.9	50.3
DAN	84.5±1.4	72.8±2.2	85.7±1.5	205.6
Fed-IDS	81.2±1.7	70.3±2.5	82.1±1.8	45.2
Meta-IDS	79.8±2.1	68.5±2.7	80.5±2.2	32.6
TCA-IDS(Ours)	83.9±1.2	74.5±1.9	84.2±1.6	15.8

结果分析：TCA-IDS 在三个跨域场景中均取得与最优方法 DAN 相近的检测精度（场景 A 83.9% vs 84.5%，场景 B 74.5% vs 72.8%，场景 C 84.2% vs 85.7%），且显著优于 CORAL、JDA、Fed-IDS 和 Meta-IDS；同时，TCA-IDS 的推理延迟（15.8 ms）仅为 DAN 的 1/13，远低于所有对比方法，完全满足工控边缘设备的实时性与工程部署需求。

3.3 轻量化与实时性分析

表 2 展示了 TCA-IDS 各轻量化阶段的体积与延迟变化。

表 2 模型轻量化过程相关变化

模型	体积 /MB	延迟 /ms
XGBoost 原始	32	45.2
剪枝前	12	18.6
剪枝后	8	15.8

经过特征剪枝、决策树剪枝与 INT8 量化，模型体积压缩 75%，推理延迟降低 65%，可在 S7-1200 PLC 上稳定运行。

3.4 消融实验

为验证时序权重 RBF 核与周期权重的贡献，在场景 A（跨协议）上设计三个变体进行消融实验，结果如图 2 所示。

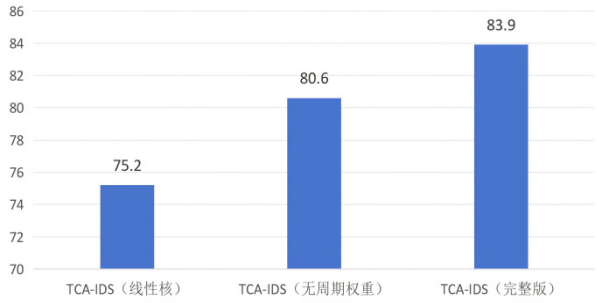

图 2 消融实验 F_1 值对比（跨协议场景）

结果分析：相比于线性核，时序权重 RBF 核贡献约 8.7 个百分点；相比于无周期权重，增加周期权重后 TCA-IDS 的 F_1 值上升 3.3 个百分点。说明本文提出的两项改进均对检测性能有显著贡献，且两者联合使用效果最佳。

4 结语

本文针对工控 OT 内部跨场景入侵检测的迁移难题，在检测精度与实时性之间取得了较好平衡，轻量化的设计使其可部署于边缘 PLC。然而，本文方法也存在一些局限，如对非周期性攻击的适配仍有限，跨系统迁移依赖共同特征维度，对完全异构系统及零样本攻击的检测能力尚需提升。未来可引入增量学习、扩展协议特征知识库，并探索半监督在线自适应策略，以增强对未知攻击的泛化能力和工程适用性。

参考文献：

[1] 东北大学“谛听”网络安全团队 . 2024 年工业控制网络安全态势白皮书 [R]. 沈阳 : 东北大学 , 2025.

[2]Sun Baochen, Feng Jiashi, Saenko K. Correlation alignment for unsupervised domain adaptation[PP/OL].(2016-12-06)[2026-06-17].https://doi.org/10.48550/arXiv.1612.01939.

[3] Long Mingsheng, Wang Jianmin, Ding Guiguang , et al. Transfer feature learning with joint distribution adaptation[C]//ICCV '13: Proceedings of the 2013 IEEE International Conference on Computer Vision. NewYork: ACM, 2013: 2200-2207.

[4] 王昊冉 , 于丹 , 杨玉丽 , 等 . 面向工控系统未知攻击的域迁移入侵检测方法 [J]. 计算机应用 ,2024,44(4):1158-1165.

[5]Feng Yuan, Si Yajie, Zhang Jianwei, et al. SA-WGAN based data enhancement method for industrial internet intrusion detection[J]. Computers, Materials & Continua, 2025, 84(3): 4431-4449.

[6]Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test[J]. Journal of Machine Learning Research, 2012, 13: 723-773.

[7]Wen Quan. Design of network intrusion detection model based on TCA[J/OL].Security and Communication Networks,2022[2026-06-17].https://www.researchgate.net/publication/361844968_Design_of_Network_Intrusion_Detection_Model_Based_on_TCA.DOI:10.1155/2022/9248853.

[8]Chen Tianqi, Guestrin C. XGBoost: a scalable tree boosting system[PP/OL].(2016-06-10)[2026-06-17].https://doi.org/10.48550/arXiv.1603.02754.

[9] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J].Journal of Machine Learning Research (JMLR), 2016, 17(1): 2030-2096.

[10]Ruzafa-Alcázar P, Fernández-Saura P, Mármol-Campos E, et al. Intrusion detection based on privacy-preserving federated learning for the industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2023, 19(2): 1145-1154.

[11] Carrión-Ojeda D, Chen Hong, Baz A E, et al. NeurIPS'22 cross-domain MetaDL competition: design and baseline results[PP/OL]. (2022-08-31)[2026-06-17].https://doi.org/10.48550/arXiv.2208.14686.

【作者简介】

赵筱峤（1999—），女，湖北宜昌人，硕士研究生，研究方向：网络安全。

（收稿日期：2026-05-15 修回日期：2026-07-06）