arXiv:2603.01621v1 [cs.CR] 2 Mar 2026

# 用于工业控制系统隐蔽攻击检测的信息论数字孪生：一种闭式 KL 散度方法

*注意：https://ieeexplore.ieee.org 不会显示副标题，因此请勿使用

1st IndaKreso萨拉热窝大学刑事司法、犯罪学与安全研究系萨拉热窝，波斯尼亚和黑塞哥维那
0000-0002-5556-4669

2nd Mehran Tarif
维罗纳大学计算机科学系维罗纳，意大利
0009-0003-0951-1800

3rd FatemehMoradi伊斯兰阿扎德大学计算机工程系）伊斯法罕，伊朗
0009-0001-6077-7503

4th ImanKhazrak鲍灵格林大学计算机科学系俄亥俄州，美国
0000-0001-7087-2283

5th MostafaM Rezaee计算机科学系 鲍灵格林大学俄亥俄州，美国
0000-0002-0849-3650

6th Mohammadhossein Homaei计算机科学系 埃斯特雷马杜拉大学卡塞雷斯，西班牙
0000-0002-6108-6632

摘要—**数字孪生（DT）越来越多地用于监测和保护工业控制系统（ICS），但要检测在正常物理范围内操纵系统状态的隐蔽虚假数据注入攻击（FDIA），仍然颇具挑战。深度学习异常检测器往往会对这类微妙操纵过度泛化，而经典故障检测方法在高度相关的多变量系统中难以扩展。我们提出一种用于实时异常检测的闭环信息论数字孪生（IT-DT）框架。该框架将N4SID辨识与稳态卡尔曼滤波相结合，通过闭式KL散度量化残差分布偏移，从而捕捉均值偏差和恶意交叉协方差偏移。在SWaT和WADI数据集上的评估显示，IT-DT的F1分数分别达到0.832和0.615，且精确率优于TranAD等深度学习基线方法。计算性能分析表明，该解析方法仅需极少内存，在CPU硬件上相较于基于变换器的方法可实现约 600×的推理加速。这使该框架适用于无需GPU加速的资源受限工业边缘控制器。**

索引词—**数字孪生、信息论检测、工业控制系统、虚假数据注入攻击**

## I. 引言

工业控制系统（ICS）已不再是彼此隔离、采用气隙隔离的设施。数字孪生（DT）如今在数字化集成工业控制系统中发挥着核心作用，支持实时同步、数据交换和控制 [1]–[3]。尽管这种连接提升了维护水平和效率，却也消除了曾经保护这些系统的物理隔离措施。 [4], [5]。数字孪生如今通过持续验证实时行为与模型预测的一致性来提升安全性。 [6]。关键挑战在于检测那些隐藏在正常过程噪声中的攻击。

请在此处注明适用的资助机构。若无，请删除此句。

虚假数据注入攻击（虚假数据注入攻击）正是此类威胁的典型代表。设计精良的虚假数据注入攻击不会引发明显异常，而是以统计上看似合理的方式改变传感器读数，使残差始终处于标准检测器可接受的范围内 [7]。近期文献中的主流应对方法是训练深度重构模型（自编码器、长短期记忆网络、Transformer），当模型输出与输入不一致时便将其标记为异常 [8]。最小化重构误差会使模型学会重现统计上正常的注入数据。 [9], [10]。

在本研究中，我们从解析概率距离的角度重新审视异常检测，而非采用启发式数据重构方法。尽管工业控制系统具有非线性，稳态行为仍可通过局部线性时不变模型近似 [11], [12]。在工业环境中，安全机制必须具备严谨的数学基础，同时足够轻量，以适用于资源受限的边缘控制器。我们的主要贡献是提出一种闭环检测框架，直接量化稳态系统创新项的分布偏移，绕过深度神经网络不透明的重构特性，从而以数学上有界的方式识别隐蔽的信息物理操纵。

我们提出一种信息论数字孪生（IT-DT）框架，将子空间状态空间系统辨识（N4SID）与稳态卡尔曼滤波器相结合，以生成多变量创新残差。不同于采用传统阈值判定，我们通过 Kullback-Leibler（KL）散度的闭式高斯近似持续量化分布偏移，从而同时解析刻画均值偏差与互协方差偏移。在 SWaT 与 WADI 上的评估表明，精确率有所提升

相较于最先进的 Transformer；同时，计算性能分析表明，在仅 CPU 硬件上内存开销较低，推理加速约为 600× 。综合来看，这些结果使 IT-DT 适合部署于资源受限的工业边缘控制器。

本文其余部分组织如下。第二节回顾相关文献，并概述现有方法论空白。第三节从信息论视角构建系统模型与检测问题。第四节给出所提检测算法及阈值机制的推导。第五节报告实证评估、攻击延迟分析与计算性能分析。第六节讨论该方法的理论意义与局限性，第七节对全文进行总结。

## II. 相关工作

### A. 工业控制系统中的经典统计与状态空间方法

经典统计方法与状态空间估计器长期以来被用作工业控制系统异常检测的基线工具。主成分分析等降维方法可为过程监测构建线性子空间，并能对平稳攻击提供合理的精确率。然而，这些方法缺乏时间记忆，因而更难检测渐进式多阶段注入攻击，尤其是当攻击利用缓慢的传感器漂移时 [13]–[15]。混合统计方法与卡尔曼滤波改善了状态跟踪，但它们通常依赖固定阈值，在相关的多变量环境中可能表现不佳 [16], [17]。攻击者可利用这些静态阈值，设计隐蔽的虚假数据注入攻击，修改跨传感器相关性而非单个通道 [18]。总体而言，经典估计器对过程监测仍然有用，但其性能在高度相关的多变量系统中可能下降。

### B. 深度学习与重构式异常检测

近期的工业控制系统异常检测方法日益依赖基于深度学习的重构模型（例如自编码器、长短期记忆网络和 Transformer），在无需显式系统辨识的情况下对非线性信息物理动态进行建模 [19]–[20]。这些模型将传感器测量值映射为潜在表示，并通过重构误差检测异常 [8]。然而，这种方法在面对自适应对手时存在一定局限。隐蔽的虚假数据注入攻击通常经过精心设计，使其保持在正常统计范围内，并与自然过程噪声相似。因此，参数量较大的模型可能会将这些细微扰动泛化，并以较小误差重构被操纵的输入，从而降低检测器的灵敏度 [9], [10]。

USAD通过对抗训练降低重构误差；然而，对于偏离正常数据不超过两个标准差的攻击，它仍可能难以检测 [10], [21]。GDN 等基于图的方法会显式建模空间相关性，但会带来更高的计算成本，而且通常需要 GPU加速，与基于 Transformer 的 TranAD 类似 [22]。TranAD 通过注意力机制实现了强大的时间建模能力，但其基于重构的目标可能会降低对严格受限的隐蔽异常的敏感度 [23]。CAE-T [24] 在 SWaT 上取得了有竞争力的结果；然而，由于训练集/测试集划分方式不同（采用支持向量数据描述目标的 6:4 划分），难以直接比较；这与本文采用的无监督评估协议不同。

### C. 数字孪生与信息论安全

数字孪生可作为虚拟参考，持续生成物理过程稳态下的预期值，同时利用信息论度量量化预期行为与观测行为之间的差异 [5], [25]。尽管这种方法较纯重构模型具有优势，现有实现仍面临实际挑战。许多方法依赖启发式阈值判定或计算开销较大的非参数技术，例如核密度估计（KDE）；这些技术每个推理步骤需要执行 $\mathcal{O}(NWp)$ 次运算（其中 $N$ 为核的数量， $W$ 为窗口大小， $p$ 为传感器维度）。这种复杂度会使其难以在边缘硬件上持续运行 [26], [27]。此外，许多基于数字孪生的方法主要跟踪单变量均值偏移或简化的协方差结构，可能无法充分捕捉高级隐蔽注入攻击导致的交叉协方差变化 [28]。因此，仍需一种更高效的公式化方法，在避免大量数值计算的同时检测多元分布偏移。

### D. 研究空白与问题

综述文献揭示了信息物理系统安全领域的方法学空白。经典方法具有可解释性，但在高度相关的多元环境中往往表现不佳。同样，基于重构的深度学习模型可能会对隐蔽型虚假数据注入攻击过度泛化，从而导致更高的漏报率。现有的数字孪生框架通常依赖启发式阈值或计算开销较大的非参数估计器，这限制了其在边缘硬件上的实时部署。据我们所知，以往的研究尚未同时解决以下问题：多元分布距离的闭式测量、在没有先验物理模型的情况下使用 N4SID 进行数据驱动的系统辨识，以及适用于资源受限硬件的高效检测。

由此引出以下研究问题：

- **研究问题一：** 与传统检测器相比，通过闭式 KL 散度联合检测均值和协方差偏移，如何提升对隐蔽型虚假数据注入攻击的检测效果？这类攻击会保留边缘统计量，却改变联合分布。
- **研究问题二：** 避免潜在空间重构会如何影响相关环境中多阶段隐蔽攻击的检测性能？
- **研究问题 3：** 所提出的解析形式是否能够在不使用 GPU加速的情况下，为边缘硬件上的实时推理带来计算优势？

## III. 系统模型与问题表述

本节构建物理系统及其对应数字孪生的数学框架，并从信息论的角度表述异常检测问题。

### A. 物理系统与数字孪生建模

SWaT 和 WADI 等工业控制系统（ICS）具有固有的非线性行为。在稳态运行期间，可在特定工作点附近使用离散时间线性时不变（LTI）状态空间模型对物理过程进行准确近似。真实物理状态 $x_t \in \mathbb{R}^n$ 的演化如下：

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{1}$$

$$y_t = Cx_t + v_t + a_t \tag{2}$$

其中，$u_t \in \mathbb{R}^m$ 是已知控制输入，$y_t \in \mathbb{R}^p$ 是传感器测量值，$a_t \in \mathbb{R}^p$ 是稀疏的虚假数据注入攻击（FDIA）向量。$w_t \sim \mathcal{N}$，$Q$ 和 $v_t \sim \mathcal{N}$，$R$ 分别是均值为零的高斯过程噪声和测量噪声。

为确保实际可行性，我们不假设系统矩阵（$A$，$B$，$C$）和噪声协方差（$Q$，$R$）事先已知。在在线部署前，我们仅使用无攻击的历史训练数据，通过子空间状态空间系统辨识（N4SID）严格提取这些参数。具体而言，我们使用了 SWaT 最初连续 7 天的正常运行数据，以及 WADI 连续 14 天的数据，从而确保辨识阶段不存在异常。通过分析经验汉克尔矩阵的奇异值下降情况，确定了最优状态维数（模型阶数），得到 SWaT 的状态向量维数为 $n = 12$，WADI 的状态向量维数为 $n = 25$。为验证 LTI 线性化假设，我们对辨识得到的模型进行了残差白度检验。我们使用 Ljung-Box 检验评估了创新残差 $r_t$ 在滞后 1 至 20 阶的自相关。在正常运行条件下，两个数据集在所有滞后阶数上均得到 $p > 0.05$，这证实所辨识的 N4SID 模型能够充分捕捉稳态过程动态特性，且在无攻击条件下，残差不存在显著的序列相关。

一个真正的数字孪生需要与物理实体持续同步。我们通过采用在两个递归阶段运行的稳态卡尔曼滤波器来实现这一反馈回路。在预测阶段，估计先验状态 $\hat{x}_{t|t-1}$：

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1} + Bu_{t-1} \tag{3}$$

新息信号（残差） $r_t$，表示物理现实与 $y_t$ 与数字孪生期望值 $\hat{y}_t = C\hat{x}_{t|t-1}$，计算如下：

$$r_t = y_t - C\hat{x}_{t|t-1} \tag{4}$$

为防止数字孪生因建模误差而发生偏离，校正阶段使用卡尔曼增益 $K$ 动态更新状态：

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + Kr_t \tag{5}$$

在正常情况下（$a_t = 0$），$r_t$ 服从多元高斯分布 $P_0 \sim \mathcal{N}(0, \Sigma)$，其中 $\Sigma = CP_{t|t-1}C^\top + R$。

### B. 信息论攻击检测

传统的基于阈值的方法无法应对使 $r_t$ 保持在正常范围内的隐蔽攻击。我们通过测量参考分布 $P_0$ 与经验滑动窗口分布 $\hat{P}_{\mathcal{R}} \sim \mathcal{N}(\hat{\mu}_t, \hat{\Sigma}_t)$ 之间的信息距离来构建检测方法。

## IV. 提出一种信息论检测方法

本节详细阐述所提出的用于检测数字孪生中隐蔽异常的算法。利用经验数据计算 KL 散度的闭式高斯近似时，主要难点在于残差空间的高维性。

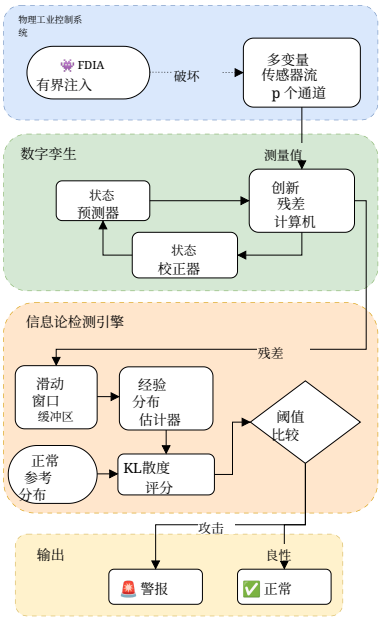


图1. 所提出的 IT-DT 框架的系统架构。物理工业控制系统层将多元传感器测量值输入数字孪生，稳态卡尔曼滤波器在其中持续生成新息残差。信息论检测引擎通过与预先计算的正常参考分布进行闭式 KL 散度计算，量化分布偏移，并在超过阈值时触发警报。

整体系统架构如图1所示，由三个相互连接的层组成：物理工业控制系统、数字孪生状态估计器和信息论检测引擎。

### A. 经验分布估计

给定大小为 $W$ 的滑动窗口 $\mathcal{R}_t = \{r_{t-W+1}, \ldots, r_t\}$，通过最大似然估计计算经验均值 $\hat{\mu}_t$ 和协方差 $\Sigma_t$。对于两个数据集，均通过在验证集上对 $W \in \{30, 60, 120, 240\}$ 进行网格搜索，将窗口大小 $W$ 选定为 $W = 60$，并在固定虚警率 $\alpha = 0.01$ 的条件下优化阈值 $\tau^*$。敏感性分析证实，在所评估的范围内，检测性能保持稳定，F1 分数变化不超过 $\pm 5\%$，表明该超参数具有较强的稳健性。为确保求逆过程中的数值稳定性，我们采用 Tikhonov 正则化 $\hat{\Sigma}_t \leftarrow \hat{\Sigma}_t + \epsilon I$，其中 $\epsilon = 10^{-4}$。

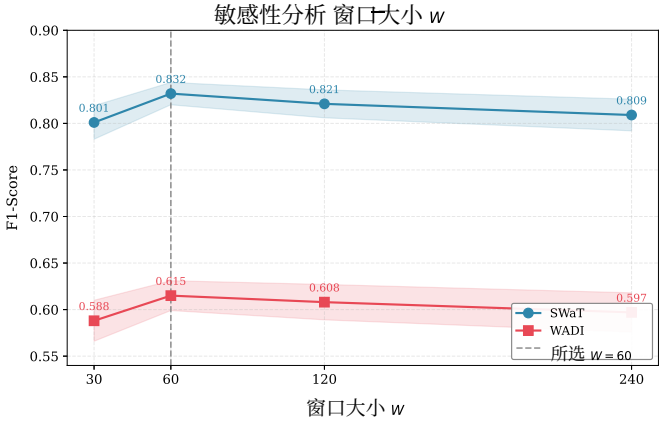


图 2. SWaT 和 WADI 数据集上 F1 分数对窗口大小 $W$ 的敏感性分析。阴影区域表示在 $K = 10$ 个按时间顺序划分的验证折上的一个标准差。性能在所评估范围内保持稳定，F1变化不超过 $\pm 5\%$，证实了对该超参数的鲁棒性。$W = 60$ 被选为最优值。

如图 2 所示，在 $W \in \{30, 60, 120, 240\}$中，检测性能保持稳定，F1变化不超过 $\pm 5\%$，证实了对该超参数选择的鲁棒性。

### B. KL散度解析计算

KL散度的闭式高斯近似通过以下闭式解析表达式计算：

$$D_{KL}(\hat{P}_{\mathcal{R}} \parallel P_0) = \frac{1}{2}\Big[\mathrm{tr}(\Sigma^{-1}\hat{\Sigma}_t) - p + \hat{\mu}_t^{\top}\Sigma^{-1}\hat{\mu}_t + \ln\left(\frac{\det(\Sigma)}{\det(\hat{\Sigma}_t)}\right)\Big] \tag{6}$$

### C. 静态分位数阈值机制

为满足奈曼–皮尔逊条件，根据由 $M$ 个样本组成的无攻击验证集离线确定决策阈值 $\tau^*$，以满足目标虚警率（$\alpha = 0.01$）。基于经验累积分布：

$$\tau^* = \inf\left\{\tau \in \mathbb{R} : \frac{1}{M}\sum_{i=1}^{M}\mathbb{I}(D_{KL}^{(i)} \le \tau) \ge 1 - \alpha\right\} \tag{7}$$

**算法 1**闭环信息论异常检测

**要求：** 窗口 $W$、矩阵 $A$, $B$, $C$, $K$, $\Sigma$、阈值 $\tau^*$

通过 N4SID 初始化 $\hat{x}_{0|0}$，$\mathcal{R} = \emptyset$
**对于**每个时间步 $t = 1, 2, \ldots$ **执行**
  **预测：** $\hat{x}_{t|t-1} \leftarrow A\hat{x}_{t-1|t-1} + Bu_{t-1}$
  **测量：** $r_t \leftarrow y_t - C\hat{x}_{t|t-1}$
  **校正：** $\hat{x}_{t|t} \leftarrow \hat{x}_{t|t-1} + Kr_t$
  更新缓冲区 $\mathcal{R}$，使用 $r_t$，保持大小为 $W$
  **如果** $|\mathcal{R}| == W$ **则**
    计算$\hat{\mu}_t$ 并进行正则化 $\hat{\Sigma}_t$
    $d_t \leftarrow D_{KL}(\mathcal{N}(\hat{\mu}_t, \hat{\Sigma}_t) \parallel \mathcal{N}(0, \Sigma))$，使用 (6)
    **如果** $d_t > \tau^*$ **则**
      **触发警报**
    **结束如果**
  **结束如果**
**结束循环**

## V. 实验评估

本节对所提出的框架进行评估。我们评估检测准确率、隐蔽鲁棒性和计算可行性。

如图1所示，检测流程通过第IV节详述的闭式评分机制，完全在仅 CPU 硬件上运行。

### A. 实验*1*：检测效能与统计显著性

表 I 给出了评估指标。为严格计算 F1 分数提升的统计显著性，我们将测试数据集划分为 $K = 10$ 个按时间顺序排列、互不重叠的子集。随后对这些子集的成对F1分数执行威尔科克森符号秩检验，结果表明IT-DT相对于TranAD的提升具有统计显著性（$p < 0.01$）。

WADI 上相对较低的 F1 分数（0.615，而 SWaT 上为 0.832）可归因于两个数据集特有的因素。首先，WADI 仅包含 15 个攻击片段，其中有三个被明确记录为隐蔽的低幅值注入，其误差界限窄于正常过程噪声的一个标准差，从根本上限制了任何无监督检测器的召回率。其次，已确认 WADI 残差存在轻微的非高斯性（$p = 0.04$，Mardia 检验，见第 VI 节），这会在式 (6) 中引入分布近似误差，导致精确率有所下降。尽管存在这些数据集特有的限制，IT-DT 在 WADI 上的 F1 分数和精确率仍高于所有其他受评方法。

尽管TranAD在SWaT数据集上取得了略高的召回率（0.772，对比0.766），但其精确率显著更低。值得注意的是，主成分分析取得了具有竞争力的精确率（0.852），但其召回率（0.528）证实了其时间盲视：它能够正确识别稳态偏差，却会漏掉需要时间记忆才能检测的渐进式多阶段攻击窗口。在真实的工业控制系统中，漏检十分危险，而过多的误报则会造成警报疲劳。所提出的IT-DT框架相对于TranAD，以可忽略的0.6%召回率下降，换取了2.5%的精确率大幅提升。这种精确率与召回率平衡表明，IT-DT为关键基础设施的实际部署提供了有利的权衡。

表 I 异常检测性能（Wilcoxon $p < 0.01$）

| 方法 | SWaT | | | WADI | | |
|---|---|---|---|---|---|---|
| | *Pre* | *Rec* | *F1* | *Pre* | *Rec* | *F1* |
| 主成分分析 [13] | 0.852 | 0.528 | 0.651 | 0.551 | 0.306 | 0.395 |
| USAD [21] | 0.871 | 0.731 | 0.795 | 0.682 | 0.450 | 0.542 |
| GDN [22] | 0.892 | 0.742 | 0.810 | 0.710 | 0.475 | 0.570 |
| TranAD [23] | 0.885 | 0.772 | 0.825 | 0.751 | 0.501 | 0.602 |
| IT-DT（我们的方法） | **0.910** | **0.766** | **0.832** | **0.785** | **0.505** | **0.615** |

## B. 实验 2：隐蔽攻击识别

为确保选择具有代表性并避免挑选偏差，我们基于两项客观标准将攻击 #41确定为典型的隐蔽攻击基准：（i）它是SWaT数据集中持续时间最长的多阶段攻击（>8分钟），以及（ii）它在先前工作中已被用作标准隐蔽攻击基准 [22]，[23]。我们明确分离了SWaT中的攻击 #41（一种隐蔽的多阶段水位操纵）。如表 II所量化，IT-DT触发警报的时间显著早于深度学习基线，从而主动防止物理损害。

表 II SWaT 攻击 #41 的检测延迟量化

| 方法 | 检测延迟（秒） | *95%* 置信区间[a] |
|---|---|---|
| USAD [21] | 112.5 | ±4.2 |
| TranAD [23] | 88.0 | ±3.1 |
| IT-DT（我们的方法） | **43.5** | **±1.8** |

[a]IT-DT 的结果基于 100 次自助法迭代得出 $\tau^*$，深度学习基线的结果则基于多次随机权重初始化得出。

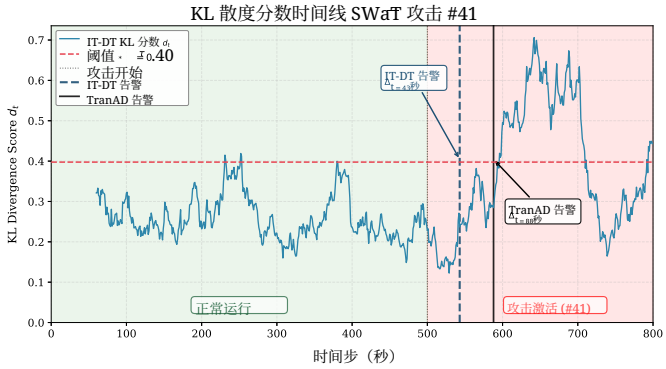


图 3. SWaT 攻击 #41 的 KL散度分数 $d_t$ 随时间的变化。红色阴影区域表示攻击活动窗口。攻击开始后 43.5 秒，IT-DT 越过阈值 $\tau^*$；而 TranAD 在 88.0 秒后才越过阈值。这是因为前者监测联合分布偏移，而非累积的重构误差。

图 3 展示了攻击 #41 期间 KL 散度得分 $d_t$ 的时间演化，直观地证实了 IT-DT 相对于 TranAD 更早产生警报。

## C. 实验 3：计算硬件性能分析

为确保计算公平性，我们在两种范式下对延迟进行了基准测试：GPU 加速和仅使用 CPU（Intel i9-14900K）。

表 III 每推理步的公平计算时延

| 方法 | 图形处理器上的时延（*ms*） | *CPU* 上的时延（*ms*） |
|---|---|---|
| USAD [21] | 2.15 | 18.40 |
| GDN [22] | 5.80 | 42.15 |
| TranAD [23] | 8.45 | 75.30 |
| IT-DT（我们的方法） | N/A | **0.12** |

在仅使用 CPU 运行以模拟工业边缘控制器时，TranAD 每步需要 75.30 毫秒。我们的方法仅需 0.12 毫秒即可完成闭式计算，在同等硬件上实现了 $600\times$ 加速比。需要注意的是，IT-DT 在 GPU 上标记为 N/A：所提出的解析方法仅依赖低维 $p\times p$ 矩阵求逆。对于如此小的矩阵，通过 PCIe 总线在 CPU 和 GPU 之间传输内存所产生的开销远远超过并行化带来的收益，因此这种特定算法结构无需 GPU 加速。

# VI. 讨论

实证评估表明，与最先进的深度学习架构和经典控制算法相比，所提出的信息论数字孪生（IT-DT）具有明显优势。本节将回答第二节中的研究问题，并概述我们方法的固有限制。

## A. 研究问题一：联合均值与交叉协方差量化

研究问题一探讨了基于联合 KL 散度的量化，能否扩展对有界虚假数据注入攻击的检测边界；这类攻击在破坏联合分布的同时保留边缘统计量。正如第三节所述，通过 (2) 中的 $a_t$ 注入的隐蔽虚假数据注入攻击，可能改变 $\hat{}P_{\mathcal{R}}$ 的跨传感器协方差结构，同时保留逐通道均值，使单变量阈值判定在结构上无法察觉。闭式表达式 (6) 同时计算均值偏差项 $\hat{}\mu_t^\top \Sigma^{-1} \hat{}\mu_t$、交叉协方差项 $\mathrm{tr}(\Sigma^{-1}\,\hat{}\Sigma_t)$ 和对数行列式比，因此即使在 $\hat{}\mu_t \approx 0$ 的情况下，也能确保产生可检测信号。表 I 中的实证结果证实了这一理论特性：在 WADI ($p = 127$) 上，IT-DT 的精确率为 0.785，高于 TranAD 的 0.751；密集的跨传感器相关结构为基于协方差的检测带来了最显著的差异化优势。

## B. 研究问题二：绕过潜在空间重构

研究问题二探讨了消除重构范式在多大程度上能够改善精确率-召回率权衡。如第二节 B 部分所述，重构模型会对隐蔽的虚假数据注入攻击过度泛化，因为有界注入在统计上与训练数据相似，从而抑制了重构误差信号。IT-DT 完全绕过了这一过程：卡尔曼校正 (5) 直接生成新息残差 $r_t$，并通过 (6) 以解析方式进行检测，无需任何潜在编码。其效果在表 I 中清晰可见：IT-DT 在 SWaT 上取得了 F1 $= 0.832$，而 TranAD 为 0.825；同时，精确率显著提升（$p < 0.01$,

Wilcoxon）。最重要的是，表 II 显示，对于隐蔽的 SWaT 攻击 #41，IT-DT 在 43.5 s 时触发，而 TranAD 则在 88.0 s 时触发；这直接源于监测残差分布偏移，而非累积的潜在重构误差。

### *C.* 研究问题 *3*：边缘硬件上的计算可行性

研究问题 3 探讨了该解析公式是否能在仅 CPU 硬件上带来显著的计算优势。如算法 1 所详述，在线推理仅需进行低维矩阵向量运算，其中 $\Sigma^{-1}$ 预先离线计算，使每步复杂度降至 $\mathcal{O}(p^2)$。相比之下，TranAD 每步需要进行 $\mathcal{O}(W^2 \cdot d)$ 次注意力运算。表 III 证实了这一实际效果：IT-DT 在英特尔 i9-14900K CPU 上的执行时间为 0.12 ms，而 TranAD 为 75.30 ms，达到报告的 $600\times$ 加速比，因此无需依赖图形处理器，该框架便能在资源受限的工业边缘控制器上切实运行。

### *D.* 局限性

一项主要局限是高斯残差假设。Mardia 多元正态性检验在 SWaT 上得到 $p = 0.08$ （未拒绝正态性假设），而在 WADI 上得到 $p = 0.04$ ，表明存在轻微的非高斯性。Tikhonov 正则化 $\epsilon I$可部分缓解这一偏差；然而，严重的运行模式转换可能导致暂时性假阳性，因此未来工作需要探索非参数分布跟踪。

## VII. 结论

本文提出了一种信息论数字孪生，该方法利用通过 N4SID 识别的最优状态估计器和解析式 Kullback-Leibler 散度。我们以直接量化残差分布取代重构式学习。通过在 SWaT 和 WADI 上进行严格的实证与统计测试，所提出的架构取得了 0.832 的 F1 分数；与最先进的 Transformer 相比，在仅使用 CPU 的边缘部署中，推理时间减少了约$600\times$。未来的工作将以自适应非参数跟踪取代高斯假设，并将 N4SID 扩展至时变系统。

## 参考文献


[1] M. Homaei, O. Mogoll´on-Guti´errez, J. C. Sancho, M. ´Avila 和 A. Caro, 《基于人工智能的数字孪生及其在网络安全中的应用综述》，《人工智能评论》，第 57 卷，第 8 期，2024年7月。[2] A. Aghazadeh Ardebili、M. Zappatore、A. Ramadan、A. Longo 和 A. Ficarella, 《智能能源系统数字孪生：关于赋能因素、设计、管理和计算挑战的系统文献综述》，《能源信息学》，2024年。[3] J. Fitzgerald、C. Gomes 和 P. Larsen, 《数字孪生工程》。施普林格，2024年。[4] A. Khazraei、S. Hallyburton、Q. Gao、Y. Wang 和 M. Paji´c, 《信息物理系统的基于学习的脆弱性分析》，载于第 *13* 届 *ACM/IEEE* 信息物理系统国际会议论文集，*ICCPS 2022*，2022年。[5] S. Krishnaveni、T. Chen、M. Sathiyanarayanan 和 B. Amutha, 《Cy- berdefender：面向基于数字孪生的工业信息物理系统的集成智能防御框架》，《集群计算》，2024年。

[6] A. Aghazadeh Ardebili、C. Martella、A. Longo、C. Rucco、F. Izzi 和 A. Ficarella, 《物联网驱动的韧性监测：信息物理系统案例研究》，《应用科学》（瑞士），2025年。[7] S. Otoom, 《信息物理系统数字孪生的风险审计：系统综述》，《网络安全与风险审计期刊》，2025年。[8] S. Mohammed、M. Jit Singh、A. Al-Jumaily、M. Isla m、M. Islam、A. Alenezi 和 M. Soliman, 《用于检测智能电网虚假数据注入的双混合入侵检测系统》，《PLoS ONE》，2025年。[9] H. Ahmad、G. Mustafa、M. Gulzar、I. Ahmed 和 M. Khalid, 《用于检测虚假数据注入攻击下配电网异常的人工智能赋能框架》，《人工智能评论》，2025年。[10] J. Sweeten、A. Takiddin、M. Ismail、S. Refaat 和 R. Atat, 《基于信息物理图神经网络的智能电网入侵检测》，载于《2023 年 IEEE 智能电网通信、控制与计算技术国际会议论文集》（SmartGridComm 2023），2023年。[11] S. Skogestad 和 I. Postlethwaite, 《多变量反馈控制：分析与设计》，第 2 版。Wiley，2005年。[12] S. X. Ding, 《基于模型的故障诊断技术：设计方案、算法与工具》。施普林格，2008年。[13] J. Shlens, 《主成分分析教程》，2014年。[14] W. Hao、T. Yang 和 Q. Yang, 《面向工业信息物理系统实时异常检测的混合统计机器学习》，《IEEE 自动化科学与工程汇刊》，2023年。[15] T. Ghosh、S. Bagui、S. Bagui、M. Kadzis 和 J. Bare, 《利用熵和基于分类的分析检测控制系统中 Modbus over TCP 的异常》，《网络安全与隐私期刊》，2023年。[16] Y. Liu、W. Deng、Z. Liu 和 F. Zeng, 《利用偏差网络和特征选择对工业控制系统进行半监督攻击检测》，《超级计算期刊》，2024年。[17] M. Homaei、M. Tarif、P. Rodriguez、A. Caro Lindo 和 M. ´Avila, 《面向水系统信息物理安全的因果数字孪生：稳健异常检测框架》，《机器学习与应用》，2026年。[18] P. Yao、X. Wang、Z. Zhang、B. Yan、Q. Yang 和 W. Wang, 《用于工业信息物理系统跨层影响评估的统计知识与博弈论集成模型》，《先进工程信息学》，2024年。[19] T. Tabassum、O. Toker 和 M. Khalghani, 《利用自编码器神经网络检测逆变型微电网中的信息物理异常》，《应用能源》，2024年。[20] S. Sun、K. Haque、X. Huo、A. Sahu、A. Goulart 和 K. Davis, 《利用长短期记忆自编码器检测信息物理系统异常：中间人（mitm）攻击案例研究》，载于《2025 年 IEEE 德克萨斯电力与能源会议》（TPEC），2025年。[21] J. Audibert、P. Michiardi、F. Guyard、S. Marti 和 M. A. Zuluaga, 《USAD：多元时间序列无监督异常检测》，载于《第 26 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集》，KDD '20 系列。ACM，2020年8月，第 3395–3404 页。[22] A. Deng 和 B. Hooi, 《基于图神经网络的多元时间序列异常检测》，《AAAI 人工智能会议论文集》，第 35 卷，第 5 期，第 4027–4035 页，2021年5月。[23] S. Tuli、G. Casale 和 N. R. Jennings, 《TranAD：用于多元时间序列数据异常检测的深度 Transformer网络》，《VLDB基金会论文集》，第 15 卷，第 6 期，第 1201–1214 页，2022年2月。[24] W. Shang、J. Qiu、H. Shi、S. Wang、L. Ding 和 Y. Xiao, 《一种面向工业控制系统的高效异常检测方法：深度卷积自编码Transformer网络》，《国际智能系统期刊》，第 2024 卷，第 1–18 页，2024年5月。[25] T. Zhukabayeva、L. Zholshiyeva、N. Karabayev、S. Khan 和 N. Al-nazzawi, 《工业物联网与边缘计算集成的网络安全解决方案：挑战、威胁与未来方向》，《传感器》，2025年。[26] R. Gupta、B. Tian、Y. Wang 和 K. Nahrstedt, 《Twin-adapt：面向物联网驱动智能实验室中数字孪生赋能的在线异常分类的持续学习》，《未来互联网》，2024年。[27] S. Otoom, 《信息物理系统数字孪生的风险审计：系统综述》，《网络安全与风险审计期刊》，2025年。[28] J. Huang、W. Yang、D. Ho、F. Li 和 Y. Tang, 《重放攻击下分布式一致性滤波的安全性分析》，《IEEE 控制论汇刊》，2024年。