arXiv:2304.07673v1 [stat.AP] 16 Apr 2023

GLHAD：面向多阶段制造系统的基于组 Lasso 的混合攻击检测与定位框架

Ahmad Kokhahi博士生，工业工程系，克莱姆森大学，电子邮箱：akokhah@clemson.edu

Dan Li助理教授，工业工程系，克莱姆森大学，电子邮箱：dli4@clemson.edu

随着工业*4.0*和数字化不断发展，对信息技术的依赖日益增加，使世界更容易受到网络攻击，尤其是那些能够操纵物理系统并损害运行数据完整性的信息物理攻击。由于攻击日益复杂，且多阶段制造系统（*MMS*）本身结构复杂，在其中检测网络攻击至关重要。攻击可能在系统中传播，影响后续阶段，使其检测难度高于单阶段系统。由于 *MMS* 中各环节之间存在复杂的相互作用，攻击定位同样至关重要。为应对这些挑战，本文提出一种基于组 *Lasso* 回归的框架，用于检测和定位 *MMS* 中的攻击。通过一个简单的线性多阶段制造系统，结果表明，与传统的基于假设检验的方法相比，所提算法在预期检测延迟和定位准确率方面表现更佳。

1 引言

随着自动化和工业物联网（IIoT）的发展，近年来制造商对网络安全的担忧显著增加 [1]。在臭名昭著的震网 [2] 事件针对伊朗核计划发起攻击后，旨在扰乱制造运营的网络攻击激增。2022年，制造业成为所有运营技术（OT）相关行业中网络攻击的首要目标，。近年来，想哭 [3] 勒索软件影响了包括台湾积体电路制造公司（TSMC）在内的多家公司，此外，针对 2014 [4] 德国一家钢铁厂的攻击也给相关国家和行业造成了巨大问题。上述攻击大多属于信息物理攻击：攻击者从网络空间侵入系统，意图扰乱物理过程。信息物理攻击的例子包括窃听、拒绝服务（DoS）攻击、隐蔽欺骗攻击、干扰攻击、密钥泄露攻击以及中间人攻击（MITM） [5]。在上述攻击中，隐蔽欺骗攻击和中间人攻击都涉及恶意数据篡改，这会影响系统的完整性，并可能损害制造业中的产品质量和系统安全。因此，本文重点关注对传感器数据完整性攻击的检测与定位。

多阶段制造系统在制造业中至关重要 [6]。这类系统由多个组件、工位或阶段构成，可建模为一系列相互连接、协同工作以生产最终产品的要素 [6, 7]。多阶段制造系统中组件与设备之间的数字连接，以及制造执行系统（MES）所采用的标准数据通信协议，使现代制造系统容易遭受网络攻击。此外，由于多阶段制造系统中各阶段相互连接，单点故障可能迅速蔓延至整个系统，并导致最终产品出现质量问题，这凸显了采取稳健的网络攻击检测和定位措施的重要性。例如，在汽车装配过程中，对加工阶段发起攻击会改变正在装配的零件的维数，导致最终产品无法发挥其应有功能，进而造成质量问题。因此，及早检测网络攻击并定位遭受攻击的阶段至关重要。

在多阶段制造系统中，各阶段相互连接，因此单点故障可能迅速蔓延至整个系统，并导致最终产品出现质量问题，这凸显了采取稳健的网络攻击检测和定位措施的重要性。例如，在汽车装配过程中，对加工阶段发起攻击会改变正在装配的零件的维数，导致最终产品无法发挥其应有功能，进而造成质量问题。因此，及早检测网络攻击并定位遭受攻击的阶段至关重要。

现有关于多阶段制造系统网络安全的研究，大多是在增材制造的背景下考察 MMS [8–11]。然而，这类方法针对增材制造而设计，无法应用于装配流程等通用 MMS。例如，在 [8], 中，每一层的打印都被视为一个阶段，并基于每层打印完成后的图像分析，开发了一种逐层变化检测方法。这类方法可推广至其他逐层制造流程。然而，目前尚缺乏基础性研究来分析网络攻击如何传播至其他阶段，以及如何利用这种传播来定位攻击。因此，我们旨在分析通用多阶段流程中的攻击传播，并开发一种可推广至多种 MMS应用的网络攻击检测算法。

本文考虑一种广义多阶段制造系统，其中每个阶段依次处理产品。每个阶段由一组传感器和一个控制器组成：传感器测量产品，控制器则根据上一阶段的传感器测量值计算控制输出，以确保该阶段的输出传感器测量值达到预期水平。由于这种控制机制，针对某个特定阶段的虚假数据注入攻击，其影响可能传播至后续阶段。也就是说，针对前一阶段的虚假数据注入可能不会对受攻击阶段造成物理影响，却会影响后续阶段。此外，当攻击者对系统有所了解时，攻击可以被设计成在该阶段无法检测到 [12]。上述因素给 MMS 中攻击的检测与定位带来了重大挑战，本文旨在解决这些问题。本文的贡献可概括如下：

1. 我们使用卡尔曼滤波器（KF）和随机状态空间模型对通用的多阶段制造过程模型进行刻画，并开展理论分析，以提取与攻击位置唯一相关的特征。
2. 我们设计了一种混合检测框架，整合了基于特征和基于异常的检测技术的优势，无需依赖攻击数据进行训练即可实现定位功能。这是基于理论分析实现的，该分析依据系统动态的领域知识提取特征。
3. 我们设计了一种基于组正则化的框架，可同时检测和定位攻击。与大多数以两阶段方式进行检测和定位的现有方法不同，基于组 Lasso 的框架使我们能够实时识别潜在虚假数据注入攻击的发生及其位置。利用这些信息，可以启动进一步调查和处理，以尽量减轻攻击对系统及其用户的影响。

2 文献综述

制造系统中的网络攻击属于信息物理攻击领域。尽管针对信息物理系统（CPS）的安全方法有很多，包括漏洞分析 [13–15], 安全物联网网络架构设计 [16, 17], 入侵检测 [18, 19], 以及具备网络攻击韧性的状态估计与控制 [20–22], ，但本文聚焦于基于过程的网络攻击检测。

制造系统中数据驱动的网络攻击检测相关文献可分为（*i*）基于特征码的方法和（*ii*）基于异常的方法。基于特征码的方法依据已知攻击运行，当模式与已知攻击匹配时便发出警报 [3, 23, 24]。换言之，这种方法类似于监督式机器学习算法：它们利用正常与受攻击数据进行训练。随后，将训练好的算法部署于真实世界数据中，以检测攻击 [25–28] 。Song 等人提出了一种在网络制造系统（CMS）中使用卷积神经网络（CNN）检测缺陷的实时攻击检测系统 [29]。Wu 等人采用机器学习算法检测网络制造系统中的信息物理攻击，并使用模拟3D打印和数控加工中的恶意攻击 [30]。Wu 等人尝试借助图像分类检测3D打印中的恶意填充缺陷。他们首先从图像中提取特征，然后应用分类算法，即朴素贝叶斯分类器和 J48决策树 [28]。

异常检测方法尝试提取系统预期行为的模式，并在识别到任何显著偏离正常模式的情况时发出警报。换言之，这些方法的工作方式类似于无监督学习算法 [26, 31–33]。Qian 等人提出了一种方案，利用论文 [32] 中介绍的基于非平行超平面的模糊分类器，检测监控与数据采集（SCADA）系统网络与物理阶段中的网络攻击。Kwon 等人提出了一种将异常检测与

特征检测算法相结合的混合方法，用于检测物理系统中的网络攻击。他们使用训练数据集中的正常数据来确定阈值，然后将训练好的模型应用于测试数据集，以评估模型的性能[26]。

本文提出的方法是特征标记与异常检测技术的混合方法。系统中每个阶段对应的特征标记描述了所有阶段的传感器测量值与注入数据之间的相关性。不过，这些特征标记并非像上述基于监督学习的算法那样通过数据驱动的方式提取，而是基于领域知识，从系统动态中推导得出。因此，本文提出的方法也被视为一种基于异常的检测算法，仅依赖正常数据，而非攻击数据。因此，在 MMS应用中，本文提出的方法更为实际，因为通常缺乏足够的攻击数据。

3 系统表示

如第1节所述，多阶段制造系统由多个阶段组成，其中阶段i 的输出是阶段$i+1$的输入。系统表示如图1所示。在每个阶段，控制动作根据输入状态的测量值确定，以控制各状态达到参考值（设定值）。在实际应用中，处理阶段前后的测量值可能在同一工位采集。不过，我们将测量值统一设定为在每个阶段之后采集，这意味着传感器测量值来自各阶段的输出状态。本文考虑的攻击是虚假数据注入攻击，即向控制器发送伪造的传感器测量值。

3.1 状态空间模型

状态空间表示已广泛用于文献中，以描述多阶段过程。[34–37]。因此，我们使用随机状态空间模型来表示一个 K 阶段 MMS。令 x_k 表示产品在阶段 k 的状态变量，其中 $\mathrm{x}_k \in \mathbb{R}^{m_k}$，$\mathrm{m}_k$表示阶段 k 的状态变量数量。令 y_k 表示阶段 k 的传感器测量值，其中 $\mathrm{y}_k \in \mathbb{R}^{n_k}$，$\mathrm{n}_k$ 表示阶段 k 的传感器数量。令 u_k表示时刻 k 的控制动作，其中 $\mathrm{u}_k \in \mathbb{R}^{p_k}$，$\mathrm{p}_k$ 表示阶段 k 控制动作的维数。w_k 和 v_k 是阶段 k 的过程噪声项和测量噪声项，它们与所有其他变量相互独立，并假设不受任何系统异常的影响。w_k 和 v_k 均服从均值为零的多元正态分布，即 $\mathrm{w}_k \sim \mathrm{K}(0,\mathrm{W}_k)$ 和 $\mathrm{v}_k \sim \mathrm{K}(0,\mathrm{V}_k)$，其中 $\mathrm{W} \in \mathbb{R}^{m_k \times}\mathrm{mk}$ 和 $\mathrm{V} \in \mathbb{R}^{n_k \times}\mathrm{nk}$ 为协方差矩阵。状态转移函数和测量函数为：

$$\mathbf{x}_k = A_k\mathbf{x}_{k-1} + B_k\mathbf{u}_k + \mathbf{w}_k, \tag{1}$$

$$\mathbf{y}_k = C_k\mathbf{x}_k + \mathbf{v}_k. \tag{2}$$

在上述方程中，$A_k \in \mathbb{R}^{m_k \times m_k-1}$、$B_k \in \mathbb{R}^{m_k \times p_k}$和$C_k \in \mathbb{R}^{n_k \times m_k}$ 分别为系统矩阵、输入矩阵和输出矩阵。我们假设每个阶段k 对应的矩阵A_k,B_k,C_k,V_k,W_k 均已知。在 (1) 中，$A_k\mathbf{x}_{k-1}$ 表示在阶段 k 对输入产品（状态变量）采取的动作。此外，$B_k\mathbf{u}_k$ 表示控制器在阶段k采取的控制动作，以确保状态变量在阶段k达到期望的状态变量（参考状态变量）。

3.2 控制器模型

我们考虑一个线性控制器，其在阶段k 的控制动作按如下方式计算：

$$\mathbf{u}_k = L_k\hat{\mathbf{x}}_{k-1|k-1} + L_{R_k}\mathbf{x}_k^r. \tag{3}$$

在上式中，$\hat{}\mathbf{x}^{k-1|k-1}$ 表示产品在前一阶段$k-1$的状态估计，而$\mathbf{x}_k^r$ 表示阶段k的状态参考值（或一般而言，控制设定参数）。线性控制器将估计状态与参考值进行线性组合，以计算控制动作。矩阵L_k 和L^{R_k} 分别是前一阶段估计状态和参考值对应的线性系数。

作为一个典型示例，线性二次高斯（LQG）控制器遵循上述公式，可根据状态估计（$\hat{}\mathrm{x}^{k|k}$）和各阶段的参考状态（x_k^r）计算控制动作。该控制器通过最小化 $\mathrm{J} = \mathbb{E}[(\mathrm{x}_K - \mathrm{x}_K^r)^T\mathrm{F}(\mathrm{x}_K - \mathrm{x}_K^r) + \Sigma_{k=0}^{K-2}(\mathrm{x}_k - \mathrm{x}_k^r)^T\mathrm{U}(\mathrm{x}_k - \mathrm{x}_k^r) + \mathrm{u}_k^T\mathrm{Zu}_k]$计算，其中 Z 和 U 是定义代价的半正定矩阵。在式（3）中，控制器的调节器为：$\mathrm{L}_k = (\mathrm{B}_k^T\mathrm{S}_{k+1}\mathrm{B}_k + \mathrm{Z})^{-1}\mathrm{B}_k\mathrm{S}_{k+1}\mathrm{A}_k$。$\mathrm{S}_k$通过以下随时间反向求解的矩阵黎卡提差分方程计算：$\mathrm{S}_k = \mathrm{A}_k^T(\mathrm{S}_{k+1} - \mathrm{S}_{k+1}\mathrm{B}_k(\mathrm{B}_k^T\mathrm{S}_{k+1}\mathrm{B}_k + \mathrm{Z})^{-1}\mathrm{B}_k\mathrm{S}_{k+1})\mathrm{A}_k + \mathrm{U}$，$\mathrm{S}_K = \mathrm{F}$。

结合状态空间模型，卡尔曼滤波器（KF）是随机线性状态空间模型 [38] 的最优状态估计器。因此，我们使用 KF 估计系统状态。计算控制器需要进行系统状态估计。卡尔曼增益（K_k）根据离散时间 KF 公式推导。KF 公式如下：

$$\hat{\mathbf{x}}_{k|k-1} = A_k\hat{\mathbf{x}}_{k-1|k-1} + B_k\mathbf{u}_k \tag{4}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k\hat{\mathbf{y}}_k \tag{5}$$

$$\hat{\mathbf{y}}_k = \mathbf{y}_k - C_k\hat{\mathbf{x}}_{k|k-1} \tag{6}$$

其中，$\hat{}\mathbf{x}^{k|k-1}$ 表示时刻$k-1$的预测系统状态，$\hat{}\mathbf{x}^{k|k}$ 表示根据时刻k的测量值$\mathbf{y}_k$更新后的状态估计，而 $\tilde{}y_k$ 表示残差

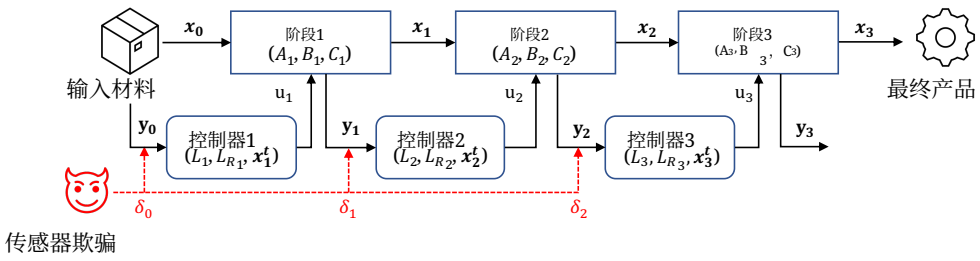


图1：系统表示：攻击向传感器注入虚假数据

时刻k的，其中 $\tilde{y}_k$ 是预测测量值与实际测量值之差，如式（6）所示。

3.3 虚假数据注入

本文考虑某一阶段遭受的虚假数据注入（FDI）攻击。此类攻击通过欺骗传感器测量值或直接向控制器发送虚假传感器数据，恶意操纵传感器测量值来实施 [39–41]。本文中，阶段k 处的虚假数据注入攻击可建模为：

$$\mathbf{y}_k^a = \mathbf{y}_k + \delta_k, \tag{7}$$

其中，δ_k 表示在阶段k由 FDI 引入传感器测量值的偏差。y_k^a 是阶段k受攻击传感器测量值的向量。

4 方法论

基于第 3 节所述的系统模型，我们提出了用于 MMS 攻击检测与定位的 GLHAD框架。我们首先进行理论分析，建立数学模型，并据此提取能够表征虚假数据注入攻击对传感器数据影响的数据特征。随后，我们利用这些表征，从状态估计残差中提取特征标记，以指示系统中是否发生攻击，以及攻击在系统中的位置（即受攻击阶段）。该问题被表述为第 4.2 节所述的组套索回归（GLR）模型；GLHAD框架则根据阶段内T^2 检验监测 GLR 系数。

4.1 理论分析

为了便于分析，我们引入由输入状态和参考值构成的增广状态变量 $\tilde{\mathbf{x}}$ 。

$$\tilde{\mathbf{x}} = [\mathbf{x}_0^T \quad \dots \quad (\mathbf{x}_K^r)^T]^T, \tag{8}$$

其中，$\mathbf{x}_0$ 表示系统的输入，x_k^r 表示阶段k的参考状态变量。注意，状态变量$\mathbf{x} = [\mathbf{x}_0^T, \mathbf{x}_1^T \quad \dots \mathbf{x}_N^T]^T$跟踪每个阶段的乘积状态，并且由于每个阶段的控制动作均基于前一阶段之后的测量值计算得出，各个$\mathbf{x}_0$ 彼此存在交叉相关性。

基于系统设置，增广状态变量是在每个阶段输入MMS的相互独立的外部输入。此处，传感器测量值 $\mathbf{y} = [\mathbf{y}_0{}^T, \mathbf{y}_1^T \quad \dots \mathbf{y}_K{}^T]^T$中，$\mathbf{y}_k$表示阶段$k$的传感器测量值，对于$k \in \{0, \dots, K\}$，可表示为 $\tilde{\mathbf{x}}$ 的线性函数，如下所示：

$$\mathbf{y} = H\tilde{\mathbf{x}} + \varepsilon, \tag{9}$$

其中H 是增广测量矩阵。具体而言，H 刻画了增广状态变量与传感器输出之间的关系。H 由$K+1 \times K+1$个子矩阵hij构成，用于刻画阶段$i-1$ 的传感器测量值与阶段$j-1$的增广状态之间的关系。我们提出命题1来定义矩阵H。

命题1 对于第*3*节所述的*MMS*，传感器测量值与状态变量之间的关系，换言之，(9) 中的矩阵H可表示为：

$$\mathbf{y} = H\tilde{\mathbf{x}} \tag{10}$$

$$H = [h_{ij}]_{(K+1)\times(K+1)} \tag{11}$$

对于 $j \geq 2$，我们有：

$$h_{ij} = \begin{cases} 0, & i = 1, .., j-1 \\ C_{i-1}B_{i-1}L_{R_{i-1}}, & i = j \\ C_i \prod_{b=j}^{i}(A_b + B_b L_b)B_1 L_{R_1}, & i = j+1, \dots, K \end{cases}$$

对于 $j = 1$，我们有：

$$h_{i1} = \begin{cases} C_0, & i = 1 \\ C_1[A_1 + B_1 L_1 K_0 C_0], & i = 2 \end{cases}$$

对于 $i \geq 3$，我们有：

$$
\begin{aligned}
h_{i1} =& C_{i-1}[\prod_{k=1}^{i-2}(A_{i-k}+B_{i-k}L_{i-k})(A_1+B_1L_1K_0C_0)-\ldots \\
& \sum_{j=1}^{i-3}[\prod_{k=1}^{j}(A_{i-k}+B_{i-k}L_{i-k})(B_{i-(j+1)}L_{i-(j+1)})\ldots \\
& \prod_{c=j+2}^{i-1}(I-K_{i-c}C_{i-c})A_{i-c}(I-K_0C_0)]-\ldots \\
& (B_{i-1}L_{i-1})\prod_{m=2}^{i-1}(I-K_{i-m}C_{i-m})A_{i-m}(I-K_0C_0)]
\end{aligned}
$$

证明见附录 A。

注释 1 命题*1*将传感器测量值形式化为基于增广状态变量 $\tilde{\mathbf{x}}$ 的$\mathbf{y}$，该变量表示系统输入。这个新系统模型的重要性在于，它有助于区分外部输入和乘积状态：前者被假定为确定性输入，不受虚假数据注入影响；后者则会受到虚假数据注入的影响。我们使用这种表示来估计增广状态变量并获得状态估计残差，从而帮助检测和定位攻击。

命题2 对于第 *3* 节所述的 *MMS*，在由向量 $\delta=[{\delta_0}^T,\delta_1^T\ldots\delta_K^T]^T$表征的虚假数据注入攻击下，其中 δ_k表示在阶段 k 注入的虚假数据，传感器测量值$\mathbf{y}$ 可表示为

$$\mathbf{y}=H\mathbf{x}+H_1\delta, \tag{12}$$

$$H_1=[\tilde{h}_{ij}]_{(K+1)\times(K+1)} \tag{13}$$

其中

$$\tilde{h}_{ij}=\begin{cases}0, & i=1,..,j-1\\ I, & i=j\\ C_{i-1}B_{i-1}L_{i-1}K_{j-1}, & i=j+1\end{cases} \tag{14}$$

对于 $i \geq j+2$:

$$
\begin{aligned}
\tilde{h}_{ij} =& C_{i-1}[\prod_{m=1}^{i-(j+1)}(A_{i-m}+B_{i-m}L_{i-m})B_jL_j+\ldots \\
& \sum_{c=1}^{i-(j+2)}[\prod_{m=1}^{c}(A_{i-m}+B_{i-m}L_{i-m})B_{i-(c+1)}L_{i-(c+1)}\cdots \\
& \prod_{b=c+2}^{i-j}(I-K_{i-b}C_{i-b})A_{i-b}]+\ldots \\
& B_{i-1}L_{i-1}\prod_{c=2}^{i-j}(I-K_{i-c}C_{i-c})A_{i-c}]K_j
\end{aligned}
$$

证明见附录 B。

注释2 在命题 *2* 中，H_1 刻画了传感器测量值与注入的虚假数据之间的关系，同时考虑了多阶段过程导致的攻击传播。方程(13) 有助于从 $\mathbf{y}$ 中提取重要特征，以准确检测虚假数据注入攻击并定位攻击源阶段。基于这种线性关系，我们将利用从状态估计残差中提取的特征构建 *GLAHD* 框架。

为识别状态估计残差中的异常模式，我们需要分析$\mathbf{y}$的方差。在式 (9) 中，由于 $\tilde{\mathbf{x}}$ 仅包含输入变量$\mathbf{x}_0$, ，而所有参考值都是确定性的，因此过程噪声和测量噪声$\mathbf{w}_k$ 和$\mathbf{v}_k$ 会对噪声项 ε产生贡献。因此，我们在命题 3 中推导 ε 的方差，以便利用 ε 中的模式识别由注入的虚假数据 δ所引入的异常模式。

命题 3 对于第 *3* 节所述的 *MMS*，分别将过程噪声和测量噪声的协方差矩阵记为 $\Sigma_x=diag(W_1,...,W_K)$ 和 $\Sigma_y=diag(V_1,...,V_K)$。式 (9) 中 ε 的协方差为：

$$\Sigma_\varepsilon=H_w\Sigma_xH_w^T+H_1\Sigma_yH_1^T, \tag{15}$$

其中 H_w 可表示为：$H_w=[h''_{ij}]_{(K+1)\times(K+1)}$。在上述表达式中，对于任意 j，有：

$$h''_{ij}=\begin{cases}0, & i=1,...,j-1\\ C_{i-1}, & i=j\\ C_{i-1}[A_{i-1}+B_{i-1}L_{i-1}K_{i-2}C_{i-2}], & i=j+1\end{cases}$$

for $i \geq j+2$:

$$
\begin{aligned}
h''_{ij} =& C_{i-1}[\prod_{m=1}^{i-(j+1)}(A_{i-m}+B_{i-m}L_{i-m})(A_j+B_jL_jK_{j-1}C_{j-1})\ldots \\
& \sum_{c=1}^{i-(j+2)}[\prod_{m=1}^{c}(A_{i-m}+B_{i-m}L_{i-m})B_{i-(c+1)}L_{i-(c+1)}\cdots \\
& \prod_{b=c+2}^{i-j}(I-K_{i-b}C_{i-b})A_{i-b}(I-K_{j-1}C_{j-1})]+\ldots \\
& B_{i-1}L_{i-1}\prod_{c=2}^{i-j}(I-K_{i-c}C_{i-c})A_{i-c}(I-K_{j-1}C_{j-1})]
\end{aligned}
$$

备注 3 命题 *3* 推导了 Σ_ε 的表达式，该表达式将用于残差ε的特征提取。通过了解正常协方差矩阵，可以识别攻击造成的数据异常模式，从而检测攻击。不同阶段的攻击造成的不同异常模式将用于定位攻击。

4.2 GLHAD框架

基于理论分析，我们提出了一种基于组 Lasso 的混合攻击检测（GLHAD）框架，该框架将系统动力学融入

组 Lasso 模型中。GLAHD 框架结合了基于特征码的方法和异常检测方法的优点，使我们能够通过基于特征码的机制识别攻击位置，而无需依赖包含带标签攻击数据的完备数据集来学习不同位置攻击的特征。相反，该框架利用理论分析结果推导出与不同攻击相关联的特征标记。

GLAHD 框架包含两个阶段。第一阶段，我们估计系统状态并分析状态估计残差，以确定检测算法的阈值。在正常情况下，系统模型遵循式（9）。增广状态变量的估计值，$\hat{\mathbf{x}}$，通过将$\mathbf{y}$ 投影到H张成的列空间上计算得出：

$$\hat{\mathbf{x}} = (H^T H)^{-1} H^T \mathbf{y},$$

状态估计残差$\mathbf{r}$计算如下：

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}},$$

其中，$\mathrm{r} = \left[\mathrm{r}_0{}^T \ \ldots \ \mathrm{r}_K{}^T\right]^T$ 和 r_k 是阶段 k 的残差。直观地说，如果 r 接近 0，我们就可以得出结论：系统未遭受攻击。因此，在第二阶段，我们可以将 GLHAD 框架应用于残差，以判断系统是否遭受攻击。此外，如果发生攻击，GLHAD 框架还能提供足够的信息，以便同时检测受攻击阶段。需要注意的是，将 y 投影到 H 张成的列空间时，H_1 的某些变化也会由 H 解释。因此，在第二阶段，我们用 R 替换 H_1。R 是 H_1 中无法由 H 解释的变化。它是 H_1 投影到 H 的列空间上的结果与 H_1 之间的差：$\mathrm{R} = \mathrm{H}_1 - \mathrm{H}(\mathrm{H}^T\mathrm{H})^{-1}\mathrm{H}^T\mathrm{H}_1$. 需要注意的是，R 矩阵可能不是满秩的。因此，我们应用主成分分析（PCA）来找出能够解释 R 方差的向量。结果如下：$\mathrm{R}' = \mathrm{PCA(R)}$。第二阶段的系统表达式如下：

$$\mathbf{r} = R'\delta \tag{16}$$

在 (16) 中，我们知道$\mathbf{r}$ 和R'的真实值。一种估计 δ 的方法是将$\mathbf{r}$ 投影到R'所张成的列空间上。鉴于只有一个阶段可能遭受攻击，我们可以得出结论，在 (13) 中，只有一个 δ_i 可以非零，$\delta_i \neq 0$。因此，我们可以得出结论，在 (13) 中，不同的 δ 可以根据其所属的阶段进行分组。换言之，参照 (13)，每个 δ_i都可以视为一个组。关于 (16)，考虑到 δ的组结构，估计 δ 的最佳方法之一是应用组 Lasso 回归算法。最初，组 Lasso 算法被定义为具有以下损失函数的正则化线性回归：

损失函数：

$$\min_{\hat{\delta}_g \in \mathbb{R}^{'d_g}} \| \sum_{g=0}^{K} [R'_g \hat{\delta}_g] - \mathbf{r} \|_2^2 + \lambda_1 \sum_{g=0}^{K} \sqrt{d_g} \| \hat{\delta}_g \|_2^2 \tag{17}$$

$$\text{s.t.} \quad \mathbf{y} = H\hat{\mathbf{x}} + \mathbf{r} \tag{18}$$

$\mathrm{R}'\mathrm{g} \in \mathbb{R}^{N \times d_g}$ 是与阶段 g 中协方差对应的残差矩阵；换言之，它仅包含与阶段 g 对应的 ˆ 列。 δ_g 是组（阶段）g 对应的回归系数。$\mathrm{r} \in \mathbb{R}^N$ 是式（16）中的残差。N 是传感器测量值的总数，$\mathrm{N} = \Sigma_{g=0}^{K} \mathrm{n}g$，其中 $\mathrm{n}g$ 是阶段 g 的传感器数量。$\mathrm{d}g$ 是阶段（组）g 的维数，λ_1 是组级正则化。在所表示的系统中，每个组代表系统的一个阶段。我们对式（16）应用组 Lasso 回归（GLR），以估计 δ。应用 GLR 后，估计得到的系数 ˆ 用于估计 r，即 $\hat{\mathrm{r}} = \mathrm{R}\delta$。最后，对每个阶段，我们基于 ˆr 应用 Hotteling T^2 检验，以判断该阶段是否遭受攻击。应用 Hotteling T^2检验时，必须已知 r 的协方差矩阵。我们知道：$\mathrm{r} = (\mathrm{I} - (\mathrm{H}^T\mathrm{H})^{-1}\mathrm{H}^T)\mathrm{y}$。因此，r 的协方差为：

$$Cov(\mathbf{r}) = (I - (H^T H)^{-1} H^T) Cov(\mathbf{y}) (I - (H^T H)^{-1} H^T)^T$$

与矩阵R相同，我们必须知道$Cov(\mathbf{r})$可能不是满秩的。因此，我们对$Cov(\mathbf{r})$进行特征值分解，以提取有意义的特征值和特征向量。

$$Eig(vectors), Eig(values) = eigen - decomposition(Cov(\mathbf{r})) \tag{19}$$

特征值决定协方差矩阵。新的协方差矩阵是一个由计算所得特征值构成的对角矩阵。其计算方式为 $Cov(\mathbf{r})' = diag(Eig(values))$。此外，由于$Cov(\mathbf{r})$发生了变化，我们必须将$\hat{\mathbf{r}}$ 映射到由式（19）中计算出的特征向量生成的列空间中。其计算方式为$\hat{\mathbf{r}}' = \hat{\mathbf{r}} Eig(vectors)$。每个阶段的主要假设是均值为零，即 $\mu = 0$。阶段k 的T^2 检验统计量计算如下：

$$t_k^2 = \hat{\mathbf{r}}'_k{}^T Cov(\mathbf{r})_k{}'^{-1} \hat{\mathbf{r}}'_k \tag{20}$$

协方差$(\mathbf{r})_k{}'^{-1}$ 是阶段k的逆协方差矩阵。对于阶段k，T^2 检验统计量的控制限为：[0,上控制限],其中上控制限 $= \chi^2_{1-\alpha, d_k}$。 α 是第一类错误，d_k 是阶段k的维度。如果多个阶段的检验统计量超出控制限，则多个阶段都可能遭受攻击。然而，由于同一时间只能有一个阶段遭受攻击，在这种情况下，将检验统计量最大的阶段视为受攻击阶段。算法的伪代码见算法 4.1：

算法 4.1 基于 GL 的攻击检测与识别
适用于多阶段线性系统

输入：H，H_1，λ_1，α，$\mathbf{y}$

1：$\hat{\mathbf{x}} \leftarrow (H^T H)^{-1} H^T \mathbf{y}$；$\hat{\mathbf{y}} \leftarrow H\hat{\mathbf{x}}$；$\mathbf{r} \leftarrow \mathbf{y} - \hat{\mathbf{y}}$

2：求解（17）

3：$k \leftarrow 0$，$pos \leftarrow 0$，$max \leftarrow 0$

4：**当** $k \neq K$ **时**

5：　$\hat{\mathbf{r}}_k \leftarrow R \quad \hat{\delta}_k$

6：　$t_k^2 \leftarrow (\hat{\mathbf{r}}_k')^T (\Sigma_\varepsilon')_k^{-1} \hat{\mathbf{r}}_k'$

7：　**如果** $t_k^2 >$上控制限 且 $t_k^2 \geq max$ **则**

8：　　最大化 $\leftarrow t_k^2$

9：　　位置 $\leftarrow k$

10：　**结束 if**

11：　$k \leftarrow k + 1$

12：**结束 while**

13：**如果** 位置 > 0 **则**

14：　**返回**pos所在的阶段遭受攻击

15：**否则**

16：**返回**没有阶段遭受攻击

17：**结束 if**

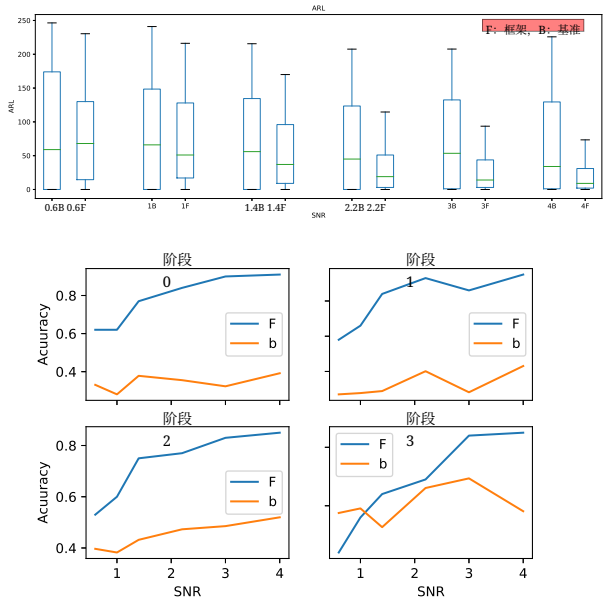


图2：数值研究结果

5 数值研究

本节将 GLHAD框架与阶段内T^2 检验进行比较。我们从标准高斯分布中随机生成矩阵（$A_{i3\times3}, B_{i3\times3}, C_{i5\times3}, i = 0,1,2,3$）中的元素。每个阶段的过程噪声和测量噪声均服从均值为零、协方差为：$W = 0.1I_3$，$V = 0.1I_5$的多元正态分布。在仿真研究中，我们首先针对系统运行的情况进行了三百次重复试验未遭受攻击。随后，我们对正常数据应用 GLHAD 框架和基准方法。接着，根据不同的信噪比（SNR）模拟系统遭受的各种传感器攻击，并比较这两种方法在攻击检测方面的性能。SNR 表示模拟攻击的幅度，定义如下：

$$SNR = \sqrt{\delta_k^T Cov(\mathbf{y})_{kk}^{-1} \delta_k}$$

协方差$(\mathbf{y})_{kk}^{-1}$ 是阶段 k 传感器的逆协方差矩阵。我们考虑：$\mathbf{x}_0 = [1,1,1]^T$、$\mathbf{x}_1^r = [-1.147, -0.726, -0.466]^T$、$\mathbf{x}_2^r = [0.239, -0.702, 0.873]^T$、$\mathbf{x}_3^r = [0.108, -0.124, -0.140]^T$。

5.1 阶段级 T^2 检验

为评估 GLAHD 的性能，我们采用阶段内 T^2 检验作为基准方法。具体而言，残差是根据测量函数得出的：

$$\mathbf{y} = C\mathbf{x} \tag{21}$$

在（21）中：$C = diag(C0,...,C_K)$。估计状态变量 $\hat{\mathbf{x}}$ 通过将$\mathbf{y}$ 投影到由C张成的列空间中计算得到：$\hat{\mathbf{x}} = (C^T C)^{-1} C^T \mathbf{y}$。计算出 $\hat{\mathbf{x}}$ 后，传感器的预测测量值 $\hat{\mathbf{y}}$ 计算如下：$\hat{\mathbf{y}} = H\hat{\mathbf{x}}$。然后，我们计算该算法的残差：$\varepsilon_b = \mathbf{y} - \hat{\mathbf{y}}$。$\varepsilon_b$ 的协方差矩阵可计算为：

$$Cov(\varepsilon_b) = (I - (C^T C)^{-1} C^T) Cov(\mathbf{y}) (I - (C^T C)^{-1} C^T)^T$$

与$Cov(\mathbf{r})$相同，我们必须知道，$Cov(\varepsilon_b)$可能不是满秩的。因此，我们对$Cov(\varepsilon_b)$进行特征值分解，以计算特征值和特征向量。

$$Eig(vectors), Eig(values) = eigen - decomposition(Cov(\varepsilon_b)) \tag{22}$$

同样，与 GLHAD框架类似，新的协方差矩阵是根据计算得到的特征值计算出的对角矩阵。其计算方式为 $\text{Cov}(\varepsilon_b)' = \text{diag}(\text{Eig(values)})$。此外，由于 $\text{Cov}(\varepsilon_b)$ 已发生变化，我们必须将 ε_b 转换到由式（22）中计算得到的特征向量生成的列空间中。其计算方式为 $\mathbf{r}' = \text{rEig(vectors)}$。然后，我们在每个阶段应用 T^2 检验，与第4.2小节类似，以检测是否有任何阶段遭受攻击。阶段 k 的 T^2 检验统计量为：$t_k^2 = \varepsilon_{b_k} \text{TCov}(\mathbf{r}'\text{b})_k^{-1} \varepsilon_{b_k}$。$\text{Cov}(\varepsilon_b)_k'^{-1}$ 是阶段 k 的 $\text{Cov}(\varepsilon_b)'$ 的逆协方差矩阵。对于阶段 k，T^2 检验统计量的控制限为：$[0, \text{UCL}]$，，其中 $\text{UCL} = \chi^2_{1-\alpha, d_k}$。 α 是第一类错误，d_k 是阶段 k 的维数。同样，与 GLHAD 框架类似，如果有多个阶段的检验统计量超出控制限，则可能有多个阶段受到攻击。

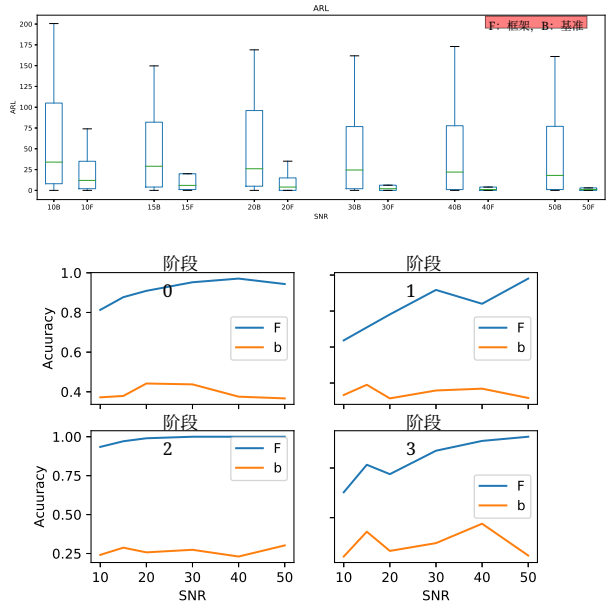


图3：案例研究结果

预测的受攻击阶段 / 真实的受攻击阶段	Zero	One	Two	三
Zero	0.76612903	0.05645161	0.08870968	0.08870968
One	0.11016949	0.83898305	0.03389831	0.01694915
Two	0.10569106	0.07317073	0.75609756	0.06504065
三	0.13138686	0.10948905	0.11678832	0.64233577

表1：GLHAD框架

预测的受攻击阶段 / 真实的受攻击阶段	Zero	One	Two	三
Zero	0.38655462	0.1092437	0.16806723	0.33613445
One	0.25694444	0.3125	0.15972222	0.27083333
Two	0.152	0.136	0.432	0.28
三	0.16541353	0.12781955	0.17293233	0.53383459

表2：基准方法

一个阶段的检验统计量超出控制限。然而，由于同一时间只能有一个阶段遭受攻击，在这种情况下，检验统计量最大的阶段被视为受攻击阶段。

预测的受攻击阶段 / 真实的受攻击阶段	Zero	One	Two	三
Zero	0.90909091	0.04545455	0.02727273	0.01818182
One	0.0859375	0.78125	0.0703125	0.0625
Two	0	0.00990099	0.99009901	0
三	0.09302326	0.08527132	0.04651163	0.7751938

表3：GLHAD框架

预测的受攻击阶段 / 真实的受攻击阶段	Zero	One	Two	三
Zero	0.44186047	0.17209302	0.11627907	0.26976744
One	0.22847682	0.31456954	0.18543046	0.27152318
Two	0.21680217	0.25474255	0.25745257	0.27100271
三	0.16019417	0.2184466	0.15533981	0.46601942

表4：基准方法

5.2 数值结果

对于每种算法，计算两个指标：(*i*) 平均运行长度（ARL）：ARL 表示一种方法在检测到失控状态之前所需的平均样本数。(*ii*) 真实定位：正确检测出受攻击阶段的样本所占百分比。对于 ARL 和真实定位，计算了两种算法在信噪比 $\in\{0.6,1,1.4,2.2,3,4\}$ 下的结果。结果如图2所示。在该图中，“F”代表 GLHAD框架，“b”代表基准方法。就 ARL 而言，除信噪比 =0.6 外，GLHAD框架检测攻击所需的样本数少于基准方法。此外，在其他四分位数下，GLHAD 的表现也优于基准方法。对于真实定位，我们逐阶段比较两种方法。例如，当阶段 1 遭受攻击时，我们分别计算 GLHAD框架和基准方法将攻击正确归因于阶段 1 的准确率。可以看出，对于阶段 k，$k \in \{0,1,2\}$，在所有信噪比取值下，GLHAD框架对攻击的定位准确率均高于基准方法。对于阶段 3，在信噪比 $=0.6$ 下，基准方法的准确率较高；但在其他所有信噪比取值下，GLHAD框架的准确率均高于基准方法。对于真实定位，我们还给出了信噪比 $=1.4$下的定位详情。详情见图1和图2。在这两幅图中，行表示

真实的受攻击阶段，列表示预测的受攻击阶段。最后，根据这两个指标的结果，我们可以得出结论：GLHAD框架优于基准方法。

5.3 案例研究

本研究中，我们在一个真实世界的多阶段装配流程案例研究中实施 GLHAD框架和基准方法，以评估所提算法的性能。为提取真实世界数据，我们使用 [7] 中的定义矩阵（A_i，B_i，$i = 1$，2，3），并且仅选取矩阵的前六行和前六列。需要注意的是，矩阵值过大可能导致协方差增大，进而使系统不稳定。为使系统稳定，我们将矩阵乘以 0.01。

接下来，我们将 GLHAD框架和基准方法应用于提取的数据。基准方法（分组假设检验）和所提出的 GLHAD框架的 ARL 如图3所示。在该图中，“F”表示 GLHAD框架，“b”表示基准方法。我们针对集合 10，15，20，30，40，50 中的信噪比值测试攻击。结果表明，GLHAD框架产生的 ARL 更低，这意味着它在所有信噪比值下均优于基准方法。此外，它在 ARL 方差方面也表现更佳。在真实定位方面，对于阶段零、一和二，GLHAD框架在所有信噪比值下的准确率均高于基准方法；但在阶段零的情况下，当信噪比值为 10，15，20 时，基准方法优于 GLHAD。

此外，我们提供了信噪比为 20 时的详细定位结果，见图 3 和图 4。在这些图中，行表示实际遭受攻击的阶段，列表示预测的遭受攻击阶段。

结果表明，与基准方法相比，GLHAD框架表现更优，说明该框架能够利用真实世界数据有效检测并定位多阶段装配过程中的网络攻击。

6 结论

本文针对多阶段制造系统提出了一种新的系统表示。我们考虑了一种适用于多阶段制造系统的一般模型，这种模型此前未曾用于攻击检测。此外，我们提出了GLHAD框架，用于检测针对传感器的网络攻击。该算法能够以较高的准确率同时检测攻击并将其定位到正确的阶段。该框架可以轻松推广至任何多阶段制造系统。未来的研究中，我们计划设计新的系统表示，涵盖多阶段制造系统中的其他类型攻击，并尝试开发用于检测这些攻击的新算法。

参考文献

[1] Mahoney, T. C., and Davis, J., 2017. 制造业网络安全：保护数字化和互联工厂。技术报告。[2] Langner, R., 2011. “震网：剖析一种网络战武器”。*IEEE Security & Privacy*，**9**(3)，第49–51页。[3] Wu, M., 2019. “网络制造系统中信息物理攻击的入侵检测”。博士论文，雪城大学。[4] Lee, R. M., Assante, M. J., and Conway, T., 2014. “德国钢铁厂网络攻击”。工业控制系统，**30**(62)，第1–15页。[5] Singh, S., Yadav, N., and Chuarasia, P. K., 2020. “信息物理系统攻击综述：问题与挑战”。载于2020年国际通信与信号处理会议（ICCSP），IEEE，第1133–1138页。[6] Liu, T., Yang, B., Li, Q., Ye, J., Song, W., and Liu, P., 2021. “多阶段制造系统（MMS）中的信息物理污点分析：案例研究”。*arXiv*预印本 *arXiv:2109.12774*。[7] Shi, J., 2006. 多阶段制造过程的变差流建模与分析。CRC出版社。[8] Al Mamun, A., Liu, C., Kan, C., and Tian, W., 2022. “通过流线视频分析和原位过程认证保障信息物理增材制造系统的安全”。制造系统杂志，**62**，第429–440页。[9] Shi, Z., Mamun, A. A., Kan, C., Tian, W., and Liu, C., 2022. “一种基于长短期记忆自编码器的增材制造信息物理攻击检测在线侧信道监测方法”。《智能制造杂志》，第1–17页。[10] Zeltmann, S. E., Gupta, N., Tsoutsos, N. G., Maniatakos, M., Rajendran, J., and Karri, R., 2016. “3D打印中的制造与安全挑战”。*JOM*，**68**(7)，第1872–1881页。[11] Liu, C., Kan, C., and Tian, W., 2020. “一种用于增材制造信息物理攻击检测的在线侧信道监测方法”。载于国际制造科学与工程会议，第84263卷，美国机械工程师学会，第V002T07A016页。[12] Liu, Y., Ning, P., and Reiter, M. K., 2011. “针对电力网状态估计的虚假数据注入攻击”。*ACM*信息与系统安全汇刊（*TISSEC*），**14**(1)，第1–33页。[13] Northern, B., Burks, T., Hatcher, M., Rogers, M., and Ulybyshev, D., 2021. “Vercasm-cps：信息物理系统的漏洞分析与网络风险评估——

信息物理系统”。*Information*，**12**(10)，第408页。[14] Zhang, Y., Jiang, T., Shi, Q., Liu, W., and Huang, S., 2022. “考虑耦合特性的信息物理系统建模与漏洞评估”。*International Journal of Electrical Power & Energy Systems*，**142**，第108321页。[15] Pan, H., Lian, H., Na, C., and Li, X., 2020. “基于社团理论的信息物理电力系统建模与漏洞分析”。*IEEE Systems Journal*，**14**(3)，第3938–3948页。[16] Pivoto, D. G., de Almeida, L. F., da Rosa Righi, R., Rodrigues, J. J., Lugli, A. B., and Alberti, A. M., 2021. “工业4.0中工业物联网应用的信息物理系统架构：文献综述”。制造系统杂志，**58**，第176–192页。[17] Patan, R., Ghantasala, G. P., Sekaran, R., Gupta, D., and Ramachandran, M., 2020. “利用灰度滤波器卷积式信息物理系统实现智慧医疗与物联网服务质量”。*Sustainable Cities and Society*，**59**，第102141页。[18] Thakur, S., Chakraborty, A., De, R., Kumar, N., and Sarkar, R., 2021. “使用通用和领域专用深度自编码器模型检测信息物理系统中的入侵”。*Computers & Electrical Engineering*，**91**，第107044页。[19] Althobaiti, M. M., Kumar, K. P. M., Gupta, D., Kumar, S., and Mansour, R. F., 2021. “基于智能认知计算的工业信息物理系统入侵检测”。*Measurement*，**186**，第110145页。[20] Kazemi, Z., Safavi, A. A., Arefi, M. M., and Naseri, F., 2021. “未知输入和传感器攻击下信息物理系统的有限时间安全动态状态估计”。*IEEE Transactions on Systems, Man, and Cybernetics: Systems*，**52**(8)，第4950–4959页。[21]Ding, D., Han, Q.-L., Ge, X., and Wang, J., 2020. “信息物理系统的安全状态估计与控制：综述”。*IEEE Transactions on Systems, Man, and Cybernetics: Systems*，**51**(1)，第176–190页。[22] Zhao, Y., Du, X., Zhou, C., and Tian, Y.-C., 2022. “执行器攻击下信息物理系统的抗饱和韧性控制”。*Information Sciences*，**608**，第1245–1260页。[23] Liao, H.-J., Lin, C.-H. R., Lin, Y.-C., and Tung, K.-Y., 2013. “入侵检测系统：全面综述”。*Journal of Network and Computer Applications*，**36**(1)，第16–24页。[24]Yaacoub, J.-P. A., Salman, O., Noura, H. N., Kaaniche, N., Chehab, A., and Malli, M., 2020. “信息物理系统安全：局限性、问——题目与未来趋势”。《微处理器与微系统》**77**，第103201页。[25] Panigrahi, R., Borah, S., Pramanik, M., Bhoi, A. K., Barsocchi, P., Nayak, S. R., 和 Alnumay, W.，2022年。“在信息物理环境中，采用混合朴素贝叶斯—决策表和多目标进化特征选择进行入侵检测”。《计算机通信》**188**，第133–144页。[26]Kwon, H.-Y., Kim, T., 和 Lee, M.-K.，2022年。“结合基于特征码和基于行为的检测方法进行高级入侵检测”。《电子学》**11**(6)，第867页。[27] Song, J., Bandaru, H., He, X., Qiu, Z., 和 Moon, Y. B.，2020。“用于增材制造实时缺陷检测的分层图像采集”。载于《ASME国际机械工程大会暨展览会》，第84492卷，美国机械工程师学会，第V02BT02A006页。[28]Wu, M., Phoha, V. V., Moon, Y. B., 和 Belman, A. K.，2016。“运用机器学习和图像分类检测3D打印过程中恶意缺陷”。载于《ASME国际机械工程大会暨展览会》，第50688卷，美国机械工程师学会，第V014T07A004页。[29]Song, J., Shukla, D., Wu, M., Phoha, V. V., 和 Moon, Y. B.，2019。“网络制造系统中用于攻击检测的物理数据审计：面向机器学习流程的区块链”。载于《ASME国际机械工程大会暨展览会》，第59384卷，美国机械工程师学会，第V02BT02A004页。[30] Wu, M., Song, Z., 和 Moon, Y. B.，2019。“运用机器学习方法检测网络制造系统中的信息物理攻击”。《智能制造杂志》**30**(3)，第1111–1123页。[31] Bhardwaj, A., Al-Turjman, F., Kumar, M., Stephan, T., 和 Mostarda, L.，2020。“捕捉不可见（CTI）：面向物联网工业控制系统的基于行为的攻击识别”。《*IEEE*开放获取》**8**，第104956–104966页。[32] Qian, J., Du, X., Chen, B., Qu, B., Zeng, K., 和 Liu, J.，2020。“流程制造业SCADA系统中的信息物理融合入侵检测方案”。《*IEEE*开放获取》**8**，第147471–147481页。[33] Abokifa, A. A., Haddad, K., Lo, C., 和 Biswas, P.，2019。“基于机器学习异常检测技术实时识别供水系统中的信息物理攻击”。《水资源规划与管理杂志》**145**(1)，第04018089页。

[34] Urbina, D. I., Giraldo, J. A., Cardenas, A. A., Tippenhauer, N. O., Valente, J., Faisal, M., Ruths, J., Candell, R., 和 Sandberg, H.，2016年。“限制隐蔽攻击对工业控制系统的影响”。载于《2016年ACM SIGSAC计算机与通信安全会议论文集》，第1092–1105页。[35]Li, D., Gebraeel, N., 和 Paynabar, K.，2020。“检测并区分监控与数据采集系统中的重放攻击和设备故障”。*IEEE*自动化科学与工程汇刊**18**(4)，第1626–1639页。[36] Mo, Y., Chabukswar, R., 和 Sinopoli, B.，2013。“检测监控与数据采集系统中的完整性攻击”。*IEEE*控制系统技术汇刊**22**(4)，第1396–1407页。[37] Van Long, D., FILLATRE, L., 和 NIKIFOROV, I.，2015。“针对网络/物理攻击对监控与数据采集系统进行序贯监测”。*IFAC*论文在线**48**(21)，第746–753页。[38] Mo, Y., 和 Sinopoli, B.，2009年。“针对重放攻击的安全控制”。载于《2009年第47届Allerton通信、控制与计算年会（Allerton）》，IEEE，第911–918页。[39] Li, B., Xiao, G., Lu, R., Deng, R., 和 Bao, H.，2019年。“关于使用分布式柔性交流输电系统设备检测电网状态估计中虚假数据注入攻击的可行性与局限性”。*IEEE*工业信息学汇刊**16**(2)，第854–864页。[40] Wang, Q., Tai, W., Tang, Y., 和 Ni, M.，2019年。“信息物理电力系统中虚假数据注入攻击综述”。*IET*信息物理系统：理论与应用**4**(2)，第101–107页。[41] Jorjani, M., Seifi, H., 和 Varjani, A. Y.，2020。“基于图论的方法：检测电力系统交流状态估计中的虚假数据注入攻击”。*IEEE*工业信息学汇刊**17**(4)，第2465–2475页。

附录 A：命题1的证明

我们首先考虑 $j \in \{2,3,...,K\}$ 的情形。我们给出 $j = 2$ 时的证明。该证明很容易推广到其他任意 $j \in \{3,...,K\}$ 的情形。h_{i2} 表示 y_{i-1} 中 x_1^r 的系数。显然：$h_{12} = 0$，$h_{22} = C_1B_1L^{R_1}$。对于 $n \geq 2$，我们使用归纳法完成证明。对于 $i = 3$，h_{32} 表示 y_2 中 x_1^T 的系数，并且有：$h_{32} = C_2[(A_2 + B_2L_2)B_1L^{R_1}$。我们假设对于任意 $i \geq 3$:

$$h_{i2} = C_{i-1}\prod_{b=2}^{i-1}(A_b + B_bL_b)B_1L_{R_1}$$

我们必须证明：$h^{(i+1)2} = C_i\prod_{b=2}^{i}(A_b + B_bL_b)B_1L^{R_1}$。根据假设，可以看出：$\gamma_{1,i-2} = \prod_{b=2}^{i-2}(A_b + B_bL_b)B_1L^{R_1}$，其中 $\gamma_{i,j}$ 是x_i^r在$hat\mathbf{x}_{j|j}$中的系数。因此，我们可以得出结论：

$$\begin{aligned}\gamma_{1,i-1} &= (I - K_{i-1}C_{i-1})(A_{i-1} + B_{i-1}L_{i-1})j_2 + K_{i-1}h_{i2} \\ &= (A_{i-1} + B_{i-1}L_{i-1})j_2 \\ &= \prod_{b=2}^{i-1}(A_b + B_bL_b)B_1L_{R_1}\end{aligned}$$

我们知道：$y_i = C_ix_i = C_iA_ix_{i-1} + C_iB_iu_i = C_iA_ix_{i-1} + C_iB_i[L_i\hat{}x_{i-1|i-1} + L^{R_i}x_i^T]$。根据假设，可以看出：$\beta_{1,i-1} = \prod_{b=2}^{i-1}(A_b + B_bL_b)B_1L^{R_1}$。$\beta_{i,j}$ 表示 x^j 中 x_i^r 的系数。因此，我们可以得出：

$$h_{(i+1)2} = C_iA_i\beta_{1,i-1} + C_iB_iL_i\gamma_{1,i-1} = C_i[\prod_{b=2}^{i}(A_b + B_bL_b)B_1L_{R_1}]$$

当 $j = 1$时，h_{i1} 表示$\mathbf{x}_0$ 在$\mathbf{y}_{i-1}$中的系数。因此，h_{11} 和h_{12} 分别表示$\mathbf{x}_0$在y_0 和y_1, 中的系数。我们使用归纳法证明$i \geq 3$。

h_{31} 表示x_0 在y_2, 中的系数，因此我们有：$h_{31} = C_2[(A_2 + B_2L_2)(A_1 + B_1L_1K_0C_0) - B_2L_2(I - K_1C_1)A_1(I - K_0C_0)]$。我们假设对于任意$i \geq 3$，有：

$$\begin{aligned}h_{i1} = C_{i-1}[&\prod_{k=1}^{i-2}(A_{i-k} + B_{i-k}L_{i-k})(A_1 + B_1L_1K_0C_0) - ... \\ &\sum_{j=1}^{i-3}[\prod_{k=1}^{j}(A_{i-k} + B_{i-k}L_{i-k})(B_{i-(j+1)}L_{i-(j+1)})... \\ &\prod_{c=j+2}^{i-1}(I - K_{i-c}C_{i-c})A_{i-c}(I - K_0C_0)] - (B_{i-1}L_{i-1})... \\ &\prod_{m=2}^{i-1}(I - K_{i-m}C_{i-m})A_{i-m}(I - K_0C_0)]\end{aligned}$$

根据假设，可以看出：

$$\begin{aligned}\theta_{i-2,i-1} = &\prod_{k=2}^{i-2}(A_{i-k} + B_{i-k}L_{i-k})(A_1 + B_1L_1K_0C_0) - ... \\ &\sum_{j=1}^{i-3}[\prod_{k=2}^{j}(A_{i-k} + B_{i-k}L_{i-k})(B_{i-(j+1)}L_{i-(j+1)})... \\ &\prod_{c=j+2}^{i-1}(I - K_{i-c}C_{i-c})A_{i-c}(I - K_0C_0)] - ... \\ &\prod_{m=2}^{i-1}(I - K_{i-m}C_{i-m})A_{i-m}(I - K_0C_0)\end{aligned}$$

其中，$\theta_{i,j}$是$\mathbf{x}_0$在y_j中的$\hat{}\mathbf{x}_{i|i}$的系数。另外，可以

看出：

$$\pi_{i-1} = \prod_{k=1}^{i-2}(A_{i-k}+B_{i-k}L_{i-k})(A_1+B_1L_1K_0C_0) - ...$$
$$sum_{j=1}^{i-3}[\prod_{k=1}^{j}(A_{i-k}+B_{i-k}L_{i-k})(B_{i-(j+1)}L_{i-(j+1)})...$$
$$\prod_{c=j+2}^{i-1}(I-K_{i-c}C_{i-c})A_{i-c}(I-K_0C_0)] - ...$$
$$(B_{i-1}L_{i-1})\prod_{m=2}^{i-1}(I-K_{i-m}C_{i-m})A_{i-m}(I-K_0C_0)$$

π_i 是$\mathbf{x}_0$ 在x_i中的系数。$\mathbf{x}_0$ 在y_i中的系数为：

$$h_{(i+1)1} = C_i[A_i\pi_{i-1}+B_iL_i((I-K_{i-1}C_{i-1})...$$
$$(A_{i-1}+B_{i-1}L_{i-1})\theta_{i-2,i-1}-K_{i-1}y_{i-1})]$$

可以看出：

$$(I-K_{i-1}C_{i-1})(A_{i-1}+B_{i-1}L_{i-1})\theta_{i-2,i-1}-K_{i-1}y_{i-1}$$
$$=(A_{i-1}+B_{i-1}L_{i-1})j_1+K_{i-1}C_{i-1}\prod_{b=1}^{i-1}A_{i-b}(I-K_{i-b-1}C_{i-b-1})$$

因此，$\mathbf{x}_0$ 在y_i 中的系数为：

$$\mathrm{h}^{(i+1)1}=\mathrm{C}_i[\mathrm{A}_i\pi_{i-1}+\mathrm{B}_i\mathrm{L}_i((\mathrm{I}-\mathrm{K}_{i-1}\mathrm{C}_{i-1})(\mathrm{A}_{i-1}+\mathrm{B}_{i-1}\mathrm{L}_{i-1})\theta_{i-2,i-1}-\mathrm{K}_{i-1}\mathrm{y}_{i-1})]=\mathrm{C}_i[\mathrm{A}_i\pi_{i-1}+\mathrm{B}_i\mathrm{L}_i((\mathrm{A}_{i-1}+\mathrm{B}_{i-1}\mathrm{L}_{i-1})\theta_{i-2,i-1}+\mathrm{K}_{i-1}\mathrm{C}_{i-1}...$$

上述陈述正是我们需要证明的内容。证明完毕。

$$\prod_{b=1}^{i-1}A_{i-b}(I-K_{i-b-1}C_{i-b-1})$$

命题2和命题3的证明

命题2和命题3可以用归纳法证明。关于命题2，对于$i=1,...,j+1$，证明显而易见。对于$i \geq j+2$，我们可以假设对于$k=a, a \geq j+2$，以下陈述成立。

$$\tilde{h}_{ij}\mathbf{r}a = C_{i-1}[\prod_{m=1}^{i-(j+1)}(A_{i-m}+B_{i-m}L_{i-m})B_jL_j + \sum_{c=1}^{i-(j+2)}[\prod_{m=1}^{c}(A_{i-m}+...$$
$$B_{i-m}L_{i-m})B_{i-(c+1)}L_{i-(c+1)}\prod_{b=c+2}^{i-j}(I-K_{i-b}C_{i-b})A_{i-b}] + ...$$
$$B_{i-1}L_{i-1}\prod_{c=2}^{i-j}(I-K_{i-c}C_{i-c})A_{i-c}]K_j$$

那么，对于$k=a+1$，我们可以证明：

$$\tilde{h}_{(i+1)j} = C_i[\prod_{m=1}^{i+1-(j+1)}(A_{i+1-m}+B_{i+1-m}L_{i+1-m})B_jL_j + ...$$
$$\sum_{c=1}^{i+1-(j+2)}[\prod_{m=1}^{c}(A_{i+1-m}+B_{i+1-m}L_{i+1-m})...$$
$$B_{i+1-(c+1)}L_{i+1-(c+1)}\prod_{b=c+2}^{i+1-j}(I-K_{i+1-b}C_{i+1-b})A_{i+1-b}] + ...$$
$$B_{i+1-1}L_{i+1-1}\prod_{c=2}^{i+1-j}(I-K_{i+1-c}C_{i+1-c})A_{i+1-c}]K_j$$

附录 B：仿真矩阵的定义

定义矩阵（A_i, B_i, C_i, $i=1, 2, 3$）如下：

$$A_1 = \begin{bmatrix}0.98451502 & 0.10019498 & 0.71348497\\ 0.14298264 & 0.6412398 & 0.90647641\\ 0.58426722 & 0.35536841 & 0.47612775\end{bmatrix}$$

$$B_1 = \begin{bmatrix}0.44096615 & 0.65555366 & 0.94144979\\ 0.78338986 & 0.9915377 & 0.04527771\\ 0.65364265 & 0.71571167 & 0.04051945\\ 0.23918787 & 0.03120722 & 0.41092476\end{bmatrix}$$

$$C_1 = \begin{bmatrix}0.29308483 & 0.25376252 & 0.27890331\\ 0.75454911 & 0.69534419 & 0.84689801\\ 0.67852479 & 0.94239412 & 0.47245499\\ 0.45955921 & 0.70151646 & 0.8589794\end{bmatrix}$$

$$A_2 = \begin{bmatrix}0.32555806 & 0.69552568 & 0.4198415\\ 0.15818161 & 0.98608914 & 0.17239575\\ 0.08682796 & 0.46574264 & 0.64864652\end{bmatrix}$$

$$B_2 = \begin{bmatrix}0.10907593 & 0.92440577 & 0.2639907\\ 0.13940577 & 0.0693751 & 0.07336545\\ 0.91392411 & 0.00977086 & 0.70578249\\ 0.11834015 & 0.82173999 & 0.30687073\end{bmatrix}$$

$$C_2 = \begin{bmatrix}0.53914991 & 0.06616444 & 0.0640871\\ 0.2704268 & 0.98044219 & 0.05198996\\ 0.8653151 & 0.23836825 & 0.53458056\\ 0.2535729 & 0.24849771 & 0.15870048\end{bmatrix}$$

$$A_3 = \begin{bmatrix}0.2825282 & 0.03752622 & 0.54049816\\ 0.72578386 & 0.68528011 & 0.71830077\\ 0.64645617 & 0.89273244 & 0.543886\end{bmatrix}$$

$$B_3 = \begin{bmatrix}0.7053606 & 0.15796312 & 0.3572694\\ 0.72379339 & 0.16706866 & 0.50119868\\ 0.66340254 & 0.80151632 & 0.24965837\\ 0.49480379 & 0.02126584 & 0.69090388\end{bmatrix}$$

$$C_3 = \begin{bmatrix}0.09041975 & 0.74105159 & 0.35007977\\ 0.76560823 & 0.96178511 & 0.02544355\\ 0.41486178 & 0.55222053 & 0.89840115\\ 0.14928482 & 0.54467456 & 0.23947464\end{bmatrix}$$