

# 基于动态异构冗余架构的信息物理系统隐私保护弹性控制

牛玉坤<sup>1</sup>, 贺磊<sup>1</sup>, 何川<sup>2,3</sup>, 韩晓鹏<sup>1</sup>, 曹植纲<sup>1</sup>, 周鼎<sup>1\*</sup>

1. 紫金山实验室, 南京 211111

2. 中国电力科学研究院, 南京 210003

3. 东南大学网络空间安全学院, 南京 210096

\* 通信作者. E-mail: zhou dinghawk@163.com, zhouding@pmlabs.com.cn

收稿日期: 2025-01-10; 修回日期: 2025-05-17; 接受日期: 2025-06-23; 网络出版日期: 2026-01-08

国家重点研发计划 (批准号: 2022YFB3104300) 和江苏省自然科学基金 (批准号: BK20240292) 资助项目

**摘要** 本文旨在解决信息物理系统 (cyber-physical system, CPS) 在遭受虚假数据注入与窃听协同攻击时的安全控制挑战. 针对现有研究在面对复杂多样的协同攻击时需叠加多重防御技术导致的系统复杂度增加、防御机制间相互干扰和资源效率低下等问题, 本文提出了一种信息物理融合的主被动协同防御架构. 该架构以动态异构冗余 (dynamic heterogeneous redundancy, DHR) 为基座, 融合机密交互协议在信息域抵御虚假数据注入与窃听协同攻击. 设计的基于椭圆曲线加密的机密交互协议在保障数据传输安全的同时满足 CPS 实时性需求; 提出的基于历史可信度与异构度的 DHR 调度算法平衡了检测可靠性、异构度最大化和系统稳定性. 为应对 DHR 架构在实际应用中面临的异构资源受限、底层同源风险和切换频率约束等挑战, 在物理域设计了攻击检测滤波器作为补充. 构建了主动与被动相结合的防御控制算法, 主动防御通过控制器动态切换提高系统不可预测性, 被动防御则依靠攻击检测滤波器及时识别受攻击控制器. 理论分析表明, 即使面临未知攻击, 闭环系统仍能保持稳定. 多种 CPS 场景的仿真实验证明, 所提弹性控制方法增强了系统抵御虚假数据注入与窃听协同攻击的能力, 验证了该方法的有效性.

**关键词** 信息物理系统, 动态异构冗余, 主动和被动防御, 虚假数据注入攻击, 窃听攻击

## 1 引言

信息物理系统 (cyber-physical system, CPS) 是现代基础设施不可或缺的一部分, 它将计算过程与物理操作无缝融合<sup>[1]</sup>, 并广泛应用于交通、医疗、能源和制造等关键领域<sup>[2]</sup>. 随着 CPS 的不断演进与规模扩大, 针对 CPS 的安全威胁与防护机制成为了学术界和产业界关注的焦点. 攻击者针对 CPS 的攻击可能造成基础设施损坏、数据泄露, 甚至危及公共安全等灾难性后果. 特别是虚假数据注入攻击和窃听攻击, 成为了 CPS 面临的主要威胁: 虚假数据注入攻击主要威胁系统的完整性, 它可能导致控制决策的恶化, 从而影响系统的稳定性和安全性; 窃听攻击主要威胁系统的机密性, 泄漏系统状态参数、控制指令等机密信息<sup>[1]</sup>.

引用格式: 牛玉坤, 贺磊, 何川, 等. 基于动态异构冗余架构的信息物理系统隐私保护弹性控制. 中国科学: 信息科学, 2026, 56: 164–184, doi: 10.1360/SSI-2025-0010

Niu Y K, He L, He C, et al. Privacy-preserving resilience control for cyber-physical systems based on the dynamic heterogeneous redundancy architecture. *Sci Sin Inform*, 2026, 56: 164–184, doi: 10.1360/SSI-2025-0010

本文特别聚焦于 CPS 控制闭环中控制器因漏洞或后门等被攻破而注入虚假数据的场景<sup>[3]</sup>, 这类攻击不仅威胁控制器本身安全, 还会通过控制回路影响整个系统的稳定性和安全性. 值得注意的是, 虚假数据注入 (false data injection, FDI) 攻击可以根据其对系统知识的依赖程度分为两类: 一类是需要获知系统准确参数才能实施的最优 FDI 攻击, 另一类是不需要精确系统参数即可发起的 FDI 攻击. Reda 等人<sup>[4]</sup> 对智能电网中的虚假数据注入攻击进行分类研究, 指出不同类型的攻击需要攻击者掌握不同程度的系统信息, 从完整模型信息到部分拓扑信息不等. 针对这种分类, Han 等人<sup>[5]</sup> 进一步将基于信息获取程度的 FDI 攻击细分为多种类型: 零动态攻击 (zero dynamic attack) 需要完整的系统模型知识, 才能对控制器发起开环式隐蔽虚假数据注入攻击; 隐蔽攻击 (covert attack) 同时针对控制器和传感器, 需要全面的模型知识以设计合适的攻击向量, 抵消输入向量对测量的影响; 而随机或偏置注入攻击 (random or bias injection attack) 和重放攻击 (replay attack) 则需要较少的系统知识. Xin 等人<sup>[6]</sup> 的研究也证实, 不同类型的 FDI 攻击对系统模型知识的依赖度存在显著差异, 这直接影响攻击的实施难度和隐蔽性.

对于需要系统知识的 FDI 攻击, 窃听攻击往往作为其前置条件, 构成完整的攻击链. Cash 等人<sup>[7]</sup> 在针对建筑自动化系统的攻击研究中发现, 成功实施 FDI 攻击的前提是先通过窃听获取系统的详细通信协议和参数信息. Wang 等人<sup>[8]</sup> 进一步指出, 窃听攻击是发起组合攻击的基础, 高级攻击者通过监控通信信道上传的关键信息, 能够设计并实施更具针对性和破坏力的攻击. 这种窃听与 FDI 攻击的协同关系使得传统的单一防御机制难以有效应对复杂攻击场景. 因此, 防御窃听攻击成为阻断高级 FDI 攻击的重要组成部分. 而对于第 2 类不依赖精确系统知识的 FDI 攻击, 他们通常通过数据驱动或者利用协议漏洞, 仅基于有限观测数据即可发起有效攻击. 这类攻击虽然不需要完整的系统模型知识, 但往往难以实现最优攻击效果. 针对此类攻击, 需要构建专门的 FDI 检测与防御机制.

为应对 CPS 中的窃听攻击, 研究者已提出多种防御方法, 主要可分为以下几类: 数据加密技术<sup>[9~13]</sup>、噪声注入方法<sup>[14]</sup>、差分隐私技术<sup>[15~17]</sup>, 以及传输策略优化<sup>[18,19]</sup> 等. 这些防御方法在实际应用中各具特色但也面临不同挑战: 密码学加密方法虽能提供理论上较强的安全保障, 但其显著的计算开销使其在资源受限的 CPS 环境中应用受限; 噪声注入和差分隐私机制在实现简便性方面具有优势, 但如 Jin 等人<sup>[19]</sup> 所指出, 这类方法不可避免地降低数据精度, 难以满足高精度控制系统的严格要求; 传输调度策略虽试图在隐私保护与状态估计精度间取得平衡, 但仍未能完全解决保密性与控制性能间的根本矛盾. 值得注意的是, 当前主流隐私保护方案大多采用“先牺牲准确性实现隐私保护, 再证明控制协议/状态估计在不准确值下的收敛性”的方式. 这种方式导致隐私保护机制与特定控制协议形成高度耦合, 严重限制了解决方案的通用适用性. 在复杂多变的实际应用场景中, 为每一种新控制协议重复证明其在不精确数据条件下的收敛性既费时又缺乏实用性<sup>[20]</sup>. 近年来, 不透明性作为一种新兴概念引起研究者关注. 如果攻击者无法通过观察区分系统的秘密状态与非秘密状态, 则称该系统具有不透明性. Yin 和 Li<sup>[21]</sup> 指出, 实现不透明性的核心是在不同系统路径上构造相同的可观测决策序列, 使攻击者难以推断系统的真实状态. 然而, 这类方法需要精密设计系统路径和决策历史模式, 实现复杂度高且在大规模分布式系统中扩展性受限.

面对日益复杂的虚假数据注入攻击, 研究人员已经构建了多层次、全方位的防御体系<sup>[22,23]</sup>. Lian 等人<sup>[24]</sup> 系统地将这些防御策略分为两大类: 检测机制和响应机制. 检测机制主要负责识别潜在的攻击行为, 而响应机制则致力于提高系统鲁棒性, 确保系统在攻击存在时仍能维持基本功能. 在检测机制方面, 主要包括知识驱动方法和数据驱动方法<sup>[24]</sup>. 知识驱动方法 (或称模型基础检测) 依赖于对系统模型的深入分析, 通过比较传感器测量值与系统分析模型生成残差, 并将残差与预设阈值比较以识别潜在攻击<sup>[25,26]</sup>. 数据驱动方法则通过历史数据识别正常操作与潜在威胁, 包括各类机器学习技术、网络编码技术等<sup>[6,27,28]</sup>. 虽然这些检测方法在识别攻击方面取得了显著进展, 但面对高级隐蔽攻击时, 单纯依靠检测往往难以应对.

在响应机制方面, 冗余技术作为抵御虚假数据注入攻击的关键策略, 在 CPS 安全防御中扮演着至关重要的角色<sup>[3]</sup>. 现代 CPS 通过在通信、软件和硬件层面实施精细的冗余机制, 显著提升了系统的抗攻击能力. 例如, 在工业控制系统中, 可编程逻辑控制器 (programmable logic controller, PLC) 层的 ControlLogix 冗余系统和编译器层的控制器开发系统 (controller development system, CODESYS) 冗余工具包已成为常见的防御解决方案<sup>[29,30]</sup>.

这些冗余策略的核心思想是: 当一个子系统遭到攻击或发生故障时, 可以通过其他正常工作的冗余子系统迅速接管, 从而将潜在的攻击影响降到最低. 其主要实现方式包括: 实时交叉验证、投票机制, 以及设置备用系统, 确保即使部分系统受到攻击, 整体系统仍能保持稳定和可靠.

为进一步增强防御能力, 研究人员引入了移动目标防御 (moving target defense, MTD) 和异构冗余等高级防御技术<sup>[31,32]</sup>. 这些创新方法通过动态调整系统参数、通信协议和配置, 大大增加了攻击者的攻击难度. 特别是动态异构冗余 (dynamic heterogeneous redundancy, DHR) 架构, 通过闭环反馈和随机调度, 能够在不同的计算和通信层实现多样化、不可预测的防御策略<sup>[33~35]</sup>. DHR 架构的核心思想是通过部署具有功能等价但实现方式不同的异构控制器, 将基于漏洞和后门的攻击转换为控制器输出结果的差模或共模问题. 所谓差模问题, 指异构控制器在面对相同攻击时由于实现差异而产生不同输出; 而共模问题, 指所有控制器可能同时受到影响产生相同错误. 通过监控控制器输出结果的一致性并进行裁决, DHR 架构能够检测控制器异常并抑制未知攻击的影响. 与传统静态防御相比, DHR 不仅能显著提高系统对虚假数据注入攻击的抵抗力, 还能通过动态切换和异构配置, 有效降低攻击者预测和渗透系统的可能性.

综上, 现有研究多针对特定攻击类型单独设计防御方法. 在面对 CPS 中复杂多样的协同攻击行为时, 这种方法不得通过叠加多重防御技术应对, 导致系统复杂度增加、防御机制间相互干扰和资源效率低下.

针对上述挑战, 本文面向信息物理系统, 提出基于动态异构冗余架构的信息物理系统隐私保护弹性控制方法, 以应对虚假数据注入和窃听协同攻击. 本文的主要贡献如下:

- 提出了融合机密交互协议的 DHR 架构, 在信息域同时抵御虚假数据注入与窃听协同攻击, 通过控制器动态切换实现主动防御, 提高系统不可预测性, 有效增加攻击者的攻击成本;
- 设计了基于椭圆曲线加密的机密交互协议, 在保障数据传输安全的同时满足 CPS 实时性需求; 提出了基于历史可信度与异构度的 DHR 调度算法, 动态平衡检测可靠性、异构度最大化和系统稳定性;
- 针对 DHR 架构在 CPS 应用中面临的异构资源受限、底层同源风险和切换频率约束等挑战, 设计了物理域攻击检测滤波器和主被动结合的防御控制算法, 有效融合信息物理域防御方法, 通过严谨的理论证明与不同 CPS 场景下的仿真实验验证了所提弹性控制方法的有效性.

本文结构安排如下: 第 2 节给出问题描述与 DHR 安全策略; 第 3 节首先构建了机密交互协议, 然后构建了最优控制器, 并提出攻击检测滤波器及 DHR 调度算法; 第 4 节提出了主动与被动结合的防御控制算法; 第 5 节中给出仿真和性能评估结果; 最后在第 6 节给出总结.

## 2 准备工作及问题描述

### 2.1 符号

符号  $R^n$  表示  $n$  维欧几里得 (Euclidean) 空间. 对于矩阵  $A \in R^{n \times n}$ ,  $A > 0$ ,  $A \geq 0$  分别表示  $A$  为正定矩阵和正半定矩阵. 符号  $\|\cdot\|$  表示向量的欧几里得范数, 对矩阵则表示 Frobenius 范数. 符号  $\lambda(\cdot)$  表示矩阵的特征值. 上标  $(\cdot)^T$  和  $(\cdot)^{-1}$  分别表示矩阵的转置与逆. 符号  $\text{diag}(\cdot)$  表示对角矩阵. 此外,  $\text{card}(K)$  表示集合  $K$  的基数. 最后,  $\text{Supp}(\cdot)$  表示向量的支撑集, 即向量中非零元素的个数.

### 2.2 CPS 及攻击建模

首先, 将 CPS 建模为以下线性时不变系统<sup>[36~38]</sup>:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad t \geq t_0, \\ y(t) &= Cx(t), \end{aligned} \quad (1)$$

其中  $x(t) \in R^n$  表示状态向量,  $u(t) \in R^m$  表示控制输入,  $y(t) \in R^l$  表示测量输出,  $A \in R^{n \times n}$ ,  $B \in R^{n \times m}$  和  $C \in R^{l \times n}$  分别表示系统参数矩阵、输入矩阵以及输出矩阵. 其中  $(A, B)$  可控,  $(A, C)$  可观.

定义  $u_i(t)$  为与第  $i$  个执行器对应的控制信号. 系统 (1) 可重写为

$$\dot{x}(t) = Ax(t) + \sum_{i=1}^m \beta_i u_i(t), \quad (2)$$

其中  $\beta_i$  为矩阵  $B$  的第  $i$  列向量.

记  $\Upsilon$  为所有列向量  $\beta_i$  的可能组合,  $i \in \{1, 2, \dots, m\}$ . 每种组合对应一个输入矩阵  $B_l$ , 其中  $l$  取值于  $\{1, 2, \dots, 2^m\}$ . 定义候选执行器集合  $\tilde{B}_\Upsilon$  为使系统 (1) 完全可控的所有执行器组合. 当系统由可控对  $(A, B_l)$  决定时, 其动态根据给定规则改变, 形式如下:

$$\tilde{B}_\Upsilon = \{B_l \in \Upsilon : \text{rank}(B_l, AB_l, \dots, A^{n-1}B_l) = n\}. \quad (3)$$

进一步, 系统 (1) 可重写为

$$\dot{x}(t) = Ax(t) + B_l u_l(t), \quad l \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}. \quad (4)$$

由于网络层易遭受恶意行为, 攻击者可能篡改发送到执行器的控制信号. 此外, 执行器本身也可能因存在未知漏洞或后门而被攻击者所控制. 在此背景下, 考虑对执行器的虚假数据注入攻击. 当对手成功攻陷执行器时, 受攻击后的系统 (1) 重写为

$$\dot{x}(t) = Ax(t) + B_l u_l(t) + B_{l,a} u_{l,a}(t), \quad (5)$$

其中  $l \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}$ ,  $u_{l,a}$  为对手控制的攻击信号.

结合文献 [39, 40], 给出如下合理假设.

**假设1** 攻击者无法同时攻陷所有执行器, 且其攻击信号的能量是有限的.

### 2.3 椭圆曲线集成加密方案

椭圆曲线集成加密方案 (elliptic curve integrated encryption scheme, ECIES) 是一种混合加密方法, 它结合了公钥密码学的优势与对称密钥系统的操作效率<sup>[41]</sup>. 值得注意的是, 目前尚无已知的通用亚指数算法能够解决椭圆曲线离散对数问题, 这使得椭圆曲线密码学系统能够在较短的密钥长度下安全运行, 从而提供与传统公钥系统 (如 RSA 算法) 相当或更高的安全级别. ECIES 基于椭圆曲线密码学原理, 并被认可为美国国家标准学会 (American National Standards Institute, ANSI) X9.63 标准的一部分<sup>[42]</sup>. 尽管它与 ElGamal 椭圆曲线加密有相似之处, 但 ECIES 通过引入消息认证机制而有所区别. 这一附加机制不仅确保了传输数据的完整性, 还确保了其真实性, 从而比其他方案更有效地增强了对敏感信息的保护<sup>[43]</sup>.

给定一个椭圆群  $E_p(a, b)$  及其生成点  $G$ , 用户的私钥是一个随机数  $d \leftarrow Z_q$ , 对应的公钥是  $\text{pub} = d \cdot G$ . 假设发送方和接收方的私钥分别为  $d_s$  和  $d_r$ , 对应的公钥为  $\text{pub}_s = d_s \cdot G$  和  $\text{pub}_r = d_r \cdot G$ . ECIES 包含如下基本功能.

- **密钥协商 KA()**. 发送方与接收方交换彼此的公钥. 随后, 发送方计算  $(k_{\text{MAC}}^{s,r}, k_{\text{ENC}}^{s,r}) = \text{KDF}(d_s \cdot \text{pub}_r || \text{para})$ . 密钥  $k_{\text{MAC}}^{s,r}$  用于加密双方之间通信消息的消息认证码 (message authentication code, MAC),  $k_{\text{ENC}}^{s,r}$  用于加密双方之间通信的消息. 函数  $\text{KDF}(\cdot)$  是一个密钥派生函数, 它从密钥材料和一些可选参数  $\text{para}$  (如发送方公钥  $\text{pub}_s$  的二进制表示) 中生成一组密钥; 接收方以相同的可选参数计算  $(k_{\text{MAC}}^{s,r}, k_{\text{ENC}}^{s,r}) = \text{KDF}(d_r \cdot \text{pub}_s || \text{para})$ . 由于  $d_s \cdot \text{pub}_r = d_s \cdot d_r \cdot G = d_r \cdot d_s \cdot G = d_r \cdot \text{pub}_s$ , 发送方和接收方将生成相同的  $(k_{\text{MAC}}^{s,r}, k_{\text{ENC}}^{s,r})$ .

- **加密 Enc()**. 为了加密消息  $m$  并发送给接收方, 发送方分别计算密文  $c = \text{Enc}(k_{\text{ENC}}^{s,r}, m)$  和  $\text{tag} = \text{Enc}(k_{\text{MAC}}^{s,r}, H(m))$ , 随后发送方将消息  $(\text{pub}_s || \text{tag} || c)$  发送给接收方.

- **解密 Dec()**. 在接收到消息  $(\text{pub}_s || \text{tag} || c)$  后, 接收方计算  $m' = \text{Dec}(k_{\text{ENC}}^{s,r}, c)$  并验证  $H(m') = \text{Dec}(k_{\text{MAC}}^{s,r}, \text{tag})$  是否成立. 如果成立, 则  $m' = m$ .

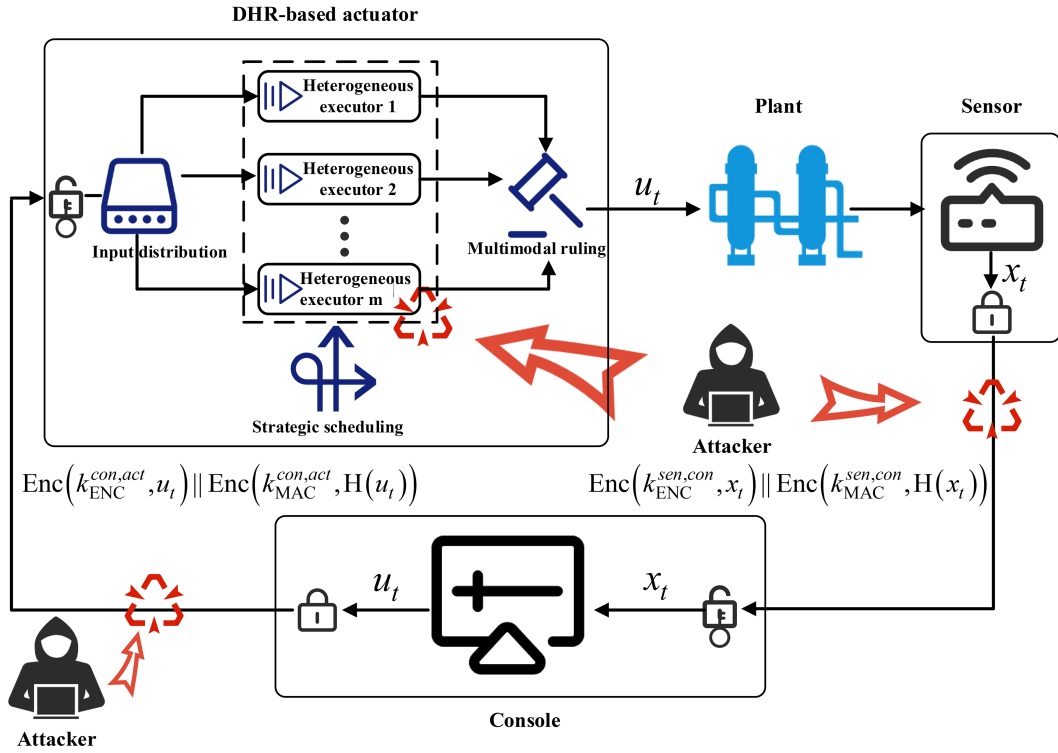


图 1 (网络版彩图) 基于 DHR 的系统架构。  
Figure 1 (Color online) System architecture with DHR.

## 2.4 DHR 安全方案

在本小节中,我们将基于 DHR 原理设计一种框架,以消除攻击者对系统发起的攻击。

结合 CPS 系统示意图与 DHR 机制,本文给出了由异构执行器组成的 DHR 系统示意图,见图 1。输入分配模块接收来自策略调度模块的指令,并将控制输入序列分配到多个异构执行器上。虽然单个执行器或许就能对系统实现有效控制,但通过冗余执行器的协作,系统的可控制性与安全性都会得到提高。这些执行器皆属于由式(4)描述的执行器集合。多模裁决模块根据规则对执行器输出向量进行合规性检测,并将满足策略要求的输出给被控系统。若出现异常结果,多模裁决模块则激活策略调度模块。策略调度模块一旦被激活,就会使当前执行器下线并及时进行修复,然后根据策略参数对执行器进行重组或重新分配。同时,当在一段时间间隔内未检测到异常时,也可以随机切换执行器,以降低多个执行器同时受到攻击的风险。在不改变服务功能的前提下,通过多模裁决与策略调度反馈机制,可以实现执行器运行环境的动态可重构和可恢复,进而向攻击者引入不确定性,同时确保系统功能的及时恢复。

注意到在同一时刻同样精确地攻击多个异构执行器并产生相同的攻击信号,是一种极小概率事件<sup>[33]</sup>,且 3 个异构执行器的 DHR 架构已具备足够高的可靠性。考虑到可靠性与成本之间的权衡,为了在攻击场景下设计 DHR 策略,我们考虑从集合  $\tilde{B}_r$  中同时选取不超过 3 个执行器,构成一个切换控制系统。通过合理的切换规则控制  $B_l, l \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}$  的激活组合,由式(4)描述的 DHR 架构就能在系统表层形成动态随机化与异构执行器,从而大幅增加攻击者的攻击难度。

## 3 隐私保护的弹性控制策略设计和稳定性分析

### 3.1 机密交互协议

我们考虑攻击者可能针对系统状态及控制指令在传输过程中的窃听攻击,即考虑系统状态数据从传感器传

递到控制台的过程,以及控制指令从控制台传递到执行器的过程.尽管控制台也存在被攻破的风险,作为系统中的核心组件,其物理和逻辑安全性通常比传感器和执行器更为重视,防护措施相对较强,且控制台的安全加固工作与本文工作正交,因此本文暂不考虑控制台被攻破的可能性.

本文所提机密交互协议基于 ECIES 加密框架,我们假设传感器、控制台、执行器均具备加解密功能,实际可以通过内置加密模块或附加加密装置来实现.如图 1 所示,机密交互协议流程如下.

(1) 传感器与控制台之间执行密钥协商,并获得两者之间的共享密钥对  $(k_{\text{ENC}}^{\text{sen,con}}, k_{\text{MAC}}^{\text{sen,con}})$ .类似地,控制台与基于 DHR 架构的执行器之间执行密钥协商并获得共享密钥对  $(k_{\text{ENC}}^{\text{con,act}}, k_{\text{MAC}}^{\text{con,act}})$ .

(2) 在第  $t$  时刻,传感器测量系统状态信息  $x_t$ ,并将加密结果  $(c_t || \text{tag}_t)$  传输给控制台,其中  $c_t = \text{Enc}(k_{\text{ENC}}^{\text{sen,con}}, x_t)$ ,  $\text{tag}_t = \text{Enc}(k_{\text{MAC}}^{\text{sen,con}}, H(x_t))$ .

(3) 控制台执行解密操作  $\text{Dec}(c_t || \text{tag}_t)$  并获得  $x_t$ ,然后根据控制算法生成控制指令  $u_t$ ,并采用控制台与基于 DHR 架构的执行器之间的共享密钥加密控制指令,最后将加密结果  $(c'_t || \text{tag}'_t)$  传输给基于 DHR 架构的执行器,其中  $c'_t = \text{Enc}(k_{\text{ENC}}^{\text{con,act}}, u_t)$ ,  $\text{tag}'_t = \text{Enc}(k_{\text{MAC}}^{\text{con,act}}, H(u_t))$ .

(4) 基于 DHR 架构的执行器执行解密操作  $\text{Dec}(c'_t || \text{tag}'_t)$  并获得  $u_t$ ,然后根据控制指令  $u_t$  操控被控系统.

需要注意的是,本文所提机密交互协议具有通用性和模块化特性,允许系统根据实际需求灵活替换为其他更轻量化的交互机制,同时保持数据的完整准确性.

### 3.2 最优控制器设计

对于每个可控模式  $B_l$ ,  $l \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}$ ,采用最优控制方法<sup>[44]</sup>,建立如下性能指标:

$$J_l = \min_{u_i} \int_{t_0}^{\infty} (x^T(t)Q_l x(t) + u_i^T(t)R_l u_i(t)) dt, \quad (6)$$

其中  $Q_l \geq 0, R_l > 0$ ,并且对于任意初始条件  $x(t_0)$  上述积分均适用.

针对每个执行模式的最优控制器可表示为

$$u_l(t) = -K_l x(t) = -R_l^{-1} B_l^T P_l x(t), \quad (7)$$

其中对称矩阵  $P_l$  满足以下代数黎卡提方程:

$$A^T P_l + P_l A + Q_l - P_l B_l R_l^{-1} B_l^T P_l = 0.$$

为了便于分析所提出 DHR 框架的稳定性,我们把闭环系统 (4) 和 (7) 表示为具有不同工作模式的切换系统.由文献 [45] 可知,对于每个切换信号,如果切换间隔时间足够长,那么系统可以稳定.首先,定义切换信号  $\vartheta(t) = l$  来表示某个主要控制器在不同时刻的启用情况,其中  $l \in 1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)$ .接着,为给出保持稳定性所需的条件,我们引入平均驻留时间的定义.

**定义1** ([45]) 对于切换信号  $\vartheta(t)$  和任意时间区间  $[t_1, t_2]$ ,其中  $0 < t_1 < t_2$ ,定义  $N_\vartheta(t_1, t_2)$  为该区间  $[t_1, t_2]$  内的切换次数.切换次数  $N_\vartheta(t_1, t_2)$  的上界为

$$N_\vartheta(t_1, t_2) \leq N_0 + \frac{t_2 - t_1}{\tau_D},$$

其中  $N_0$  称为抖动界,  $\tau_D$  表示平均驻留时间.

接下来,我们给出以下定理,证明当平均驻留时间足够大时,整体系统是稳定的.

**定理1** 对于无攻击闭环系统 (4) 和 (7),其切换由定义 1 规定,当平均驻留时间  $\tau_D > \tau_D^*$  且任意抖动界  $N_0 > 0$  时,该闭环系统是一致指数稳定的.平均驻留时间的下界  $\tau_D^*$  为

$$\tau_D^* = \frac{\ln \max_{l, o \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}} \frac{\lambda_{\max}(P_l)}{\lambda_{\min}(P_o)}}{\min_{l \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}} \frac{\lambda_{\min}(Q_l + P_l B_l R_l^{-1} B_l^T P_l)}{\lambda_{\max}(P_l)}}. \quad (8)$$

**证明** 对于每个激活的可控对  $l \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}$ , 考虑候选李雅普诺夫 (Lyapunov) 函数为

$$V_l(x(t)) = x^T(t)P_l x(t), \quad \forall x \in R^n, \quad (9)$$

其中  $P_l$  表示最优控制器 (7) 中黎卡提 (Riccati) 方程的解, 这表示  $V_l(x(t)) \geq 0$  且  $V_l(x(t)) = 0$  当且仅当  $x = 0_n$ . 李雅普诺夫函数  $V_l$  是正定的且径向无界的.

对  $V_l(x)$  求时间导数得到

$$\begin{aligned} \dot{V}_l(x) &= x^T(A - B_l R_l^{-1} B_l^T P_l)^T P_l x + x^T P_l (A - B_l R_l^{-1} B_l^T P_l) x \\ &= x^T (A^T P_l + P_l A - 2P_l B_l R_l^{-1} B_l^T P_l) x \\ &= x^T (-Q_l - P_l B_l R_l^{-1} B_l^T P_l) x = -x^T \hat{Q}_l x, \end{aligned} \quad (10)$$

其中  $\hat{Q}_l = Q_l + P_l B_l R_l^{-1} B_l^T P_l$ .

根据所选取的最优控制器可得  $Q_l \geq 0$ ,  $R_l > 0$ , 这意味着  $\dot{V}_l(x) \leq -x^T \hat{Q}_l x \leq 0$ . 此外, 根据对称矩阵瑞利-里茨不等式, 有  $\dot{V}_l(x) \leq -\lambda_{\min}(\hat{Q}_l) x^T x$ ,  $V_l(x) \leq \lambda_{\max}(P_l) x^T x$ . 因此, 可以得到

$$\dot{V}_l(x(t)) \leq -\lambda_{\min}(\hat{Q}_l) \frac{V_l(x)}{\lambda_{\max}(P_l)}. \quad (11)$$

由  $\lambda_{\min}(P_l) x^T x \leq V_l(x) \leq \lambda_{\max}(P_l) x^T x$  可得

$$V_l(x(t)) \leq \lambda_{\max}(P_l) x^T x \leq \lambda_{\max}(P_l) \frac{V_l(x)}{\lambda_{\min}(P_l)}. \quad (12)$$

此外, 对于任意控制器, 不等式 (11) 和 (12) 可以重写为

$$\dot{V}_l(x) \leq - \min_{l \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}} \frac{\lambda_{\min}(\hat{Q}_l) V_l(x)}{\lambda_{\max}(P_l)}, \quad (13)$$

$$V_l(x) \leq \max_{l, o \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}} \frac{\lambda_{\max}(P_l)}{\lambda_{\min}(P_o)} V_o(x). \quad (14)$$

为方便表示, 我们定义如下符号:

$$\theta = \min_{l \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}} \lambda_{\min}(\hat{Q}_l) / \lambda_{\max}(P_l), \quad (15)$$

$$\varpi = \max_{l, o \in \{1, 2, \dots, \text{card}(\tilde{B}_r)\}} \lambda_{\max}(P_l) / \lambda_{\min}(P_o). \quad (16)$$

不失一般性, 考虑时间区间  $[t_0, t_1]$ ,  $t_0 = 0$ . 根据定义 1, 切换信号  $\vartheta(t)$  等于执行模式的标号  $l$ . 假设第一次切换发生在  $t_0 + \Delta t$ , 根据式 (13), 我们有

$$V_{\vartheta(t_0 + \Delta t^-)}(t_0 + \Delta t^-) \leq e^{-\theta(t_0 + \Delta t - t_0)} V_{\vartheta(t_0)}(t_0). \quad (17)$$

由式 (14) 和 (17) 可得,

$$\begin{aligned} V_{\vartheta(t_0 + \Delta t)}(t_0 + \Delta t) &\leq \varpi V_{\vartheta(t_0 + \Delta t^-)}(t_0 + \Delta t^-) \\ &\leq \varpi e^{-\theta \Delta t} V_{\vartheta(t_0)}(t_0). \end{aligned} \quad (18)$$

在整个时间区间  $[t_0, t_1]$  范围内, 根据不等式 (17) 和 (18) 可得

$$\begin{aligned} V_{\vartheta(t)}(t) &\leq \varpi V_{\vartheta(t^-)}(t^-) \leq \varpi^{N_{\vartheta}(t_0, t) + 1} e^{-\theta t} V_{\vartheta(0)}(0) \\ &\leq \varpi^{N_0 + \frac{t-t_0}{\tau_D} + 1} e^{-\theta t} V_{\vartheta(0)}(0) \\ &\leq \varpi^{N_0 + 1} e^{\left(\frac{\ln \varpi}{\tau_D} - \theta\right)t} V_{\vartheta(0)}(0). \end{aligned} \quad (19)$$

因此, 选择适当的  $\tau_D$ , 可得  $\frac{\ln \varpi}{\tau_D} - \theta < 0$ , 即

$$\tau_D > \tau_D^* = \frac{\ln \max_{\forall l, o \in \{1, 2, \dots, \text{card}(\bar{B}_\Upsilon)\}} \frac{\lambda_{\max}(P_l)}{\lambda_{\min}(P_o)}}{\min_{\forall l \in \{1, 2, \dots, \text{card}(\bar{B}_\Upsilon)\}} \frac{\lambda_{\min}(\hat{Q}_l)}{\lambda_{\max}(P_l)}}. \quad (20)$$

式 (20) 定量揭示了系统切换时间与系统稳定性之间的关系, 明确给出了确保系统稳定的最小平均驻留时间. 由式 (19) 可得, 如果切换驻留时间  $\tau_D$  过小, 式 (19) 中的指数项  $\varpi^{N_{o+1}}$  将会主导系统响应, 使系统李雅普诺夫函数不再收敛, 破坏稳定性. 因此, 必须保持平均驻留时间  $\tau_D > \tau_D^*$ .

由定义 1 可知, 当  $t \rightarrow \infty$  时, 对于每个切换信号  $\vartheta \in N_\vartheta(0, t)$ , 整体系统是指数稳定的.

接下来, 进一步分析执行器的切换频率 (切换次数) 对系统稳定性的影响. 不失一般性, 假设在时刻  $t_i$ , 系统从可控模式  $l$  切换至可控模式  $o$ , 此时可得

$$V(t_i^+) = x^T(t_i) P_o x(t_i) \leq \lambda_{\max}(P_o) \|x(t_i)\|^2, \quad (21)$$

$$V(t_i) = x^T(t_i) P_l x(t_i) \geq \lambda_{\min}(P_l) \|x(t_i)\|^2. \quad (22)$$

据此, 切换过程中李雅普诺夫函数的增长比例可形式化定义为

$$\frac{V(t_i^+)}{V(t_i)} \leq \frac{\lambda_{\max}(P_o)}{\lambda_{\min}(P_l)} \triangleq \mu_{l,o}. \quad (23)$$

在整个时间区间  $[t_o, t_1]$  内, 有  $V(x(t)) \leq \mu^{N_\sigma(0,t)} e^{-\lambda_0 t} V(x(0))$ .

定义  $\bar{\mu} = \max_i \mu_i$ , 则  $\bar{\mu} = \max_{l,o} \frac{\lambda_{\max}(P_l)}{\lambda_{\min}(P_o)} = \varpi$ , 对  $V(t)$  进行对数变换可得

$$\ln V(t) \leq N_\sigma \ln \bar{\mu} - \theta(t - t_0) + \ln V(t_0).$$

为确保系统指数稳定, 需要当  $t \rightarrow \infty$  时,  $\ln V(t) \rightarrow -\infty$ , 因此需满足  $N_\sigma \ln \bar{\mu} < \theta(t - t_0)$ , 进一步可推出  $\frac{N_\sigma}{(t-t_0)} < \frac{\theta}{\ln \bar{\mu}} \triangleq \eta$ , 即单位时间内的切换次数必须小于阈值  $\eta$ , 否则李雅普诺夫函数不能保持下降趋势, 系统稳定性无法保证. 证明完毕.

### 3.3 虚假数据注入攻击检测机制

本小节将提出一个攻击检测滤波器来识别虚假数据注入攻击. 对于具有虚假数据注入攻击的信息物理系统 (5), 构造了攻击检测滤波器, 其具体表述如下:

$$\begin{cases} \dot{I}_f(t) = A I_f(t) + B_l u_l(t) + g(I_f(t) - x(t)), \\ f_{ua}(t) = I_f(t) - x(t), \end{cases} \quad (24)$$

其中  $I_f(t)$  表示攻击检测滤波器的状态值,  $f_{ua}$  表示残差误差,  $g$  是待设计的常数矩阵.

**定理 2** 对于所考虑 CPS 系统 (5), 残差误差  $f_{ua}$  与虚假数据注入攻击发生之间的关系可描述如下: 当 CPS 系统未受到攻击时,  $f_{ua}(t) = 0_n, \forall t \geq 0$ ; 当 CPS 系统受到攻击时,  $f_{ua}(t) \neq 0_n, \forall t \geq 0$ .

**证明** 考虑攻击检测滤波器 (24) 和 CPS 系统 (5), 该攻击检测滤波器重写为

$$\dot{f}_{ua}(t) = (A - B_l R_l^{-1} B_l^T P_l + g) f_{ua}(t) - B_l u_{l,a}(t), \quad (25)$$

其中  $f_{ua}$  的初始状态设为  $0_n$ , 即  $I_f(t) = x(t)$ .

首先, 我们提出以下分析. 在 CPS 未受到攻击的情况下, 即  $u_{l,a}(t) = 0_m$ . 由于攻击检测滤波器的初始状态设置为  $I_f(t) = x(t)$ , 根据攻击检测滤波器公式 (25) 可得  $\dot{f}_{ua}(t) = 0_n$ . 此外, 容易得到  $f_{ua}(t) \equiv 0_n$ . 当 CPS 受到攻击时, 即  $u_{l,a}(t) \neq 0_m$ , 根据式 (25) 可得  $\dot{f}_{ua}(t) \neq 0_n$ , 这进一步导致残差误差累积. 因此, 我们有  $f_{ua}(t) \neq 0_n$ .

此外, 如果 CPS 的初始状态  $x(0)$  未知, 则攻击检测滤波器 (24) 的初始状态  $I_f(0)$  可以任意设定. 从式 (25) 可以看出, 通过选择一个合适的常数矩阵  $g$ , 可以使常数矩阵  $A - B_l R_l^{-1} B_l^T P_l + g$  成为赫尔维茨 (Hurwitz) 稳定矩阵, 这表明残差误差系统 (25) 是渐近稳定的. 因此, 即使 CPS 的初始状态未知, 也仍然可以通过判断残差误差  $f_{ua}$  是否为 0 来识别 CPS 是否受到攻击. 如果可能, 尽量将  $I_f(0)$  设置在 CPS 的正常状态范围内, 以便攻击检测滤波器能够尽快实现稳定.

我们可以得出结论: 当  $f_{ua}(t) = 0_n, \forall t \geq 0$  时, CPS 系统 (5) 未受到攻击; 否则, CPS 系统 (5) 受到攻击. 证明完毕.

在实际的 CPS 中, 由于噪声和延迟因素的影响, 残差误差  $f_{ua}$  可能会不准确. 为增强所设计攻击检测滤波器在实际信息物理系统中的适应性, 进一步分析其在测量噪声扰动与通信/处理延迟条件下的鲁棒性表现.

考虑信息物理系统输出测量存在有界噪声、时延, 则信息物理系统模型可重写为

$$\begin{cases} \dot{x}(t) = Ax(t) + B_l u_l(t) + B_l u_{l,a}(t), \\ y(t - \tau) = x(t - \tau) + w(t - \tau), \|w(t)\| \leq \bar{w}, 0 \leq \tau \leq \tau_{\max}, \end{cases} \quad (26)$$

其中,  $\tau$  为时延,  $w(t)$  为有界噪声, 满足  $\|w(t)\| \leq \bar{w}, 0 \leq \tau \leq \tau_{\max}$ , 测量噪声有界, 时延在一定的范围内是保证系统稳定和的前提, 否则, 会导致系统损坏, 甚至引发系统不稳定. 这个范围内的值称为可接受值, 而超出此范围的值则称为不可接受值. 因此, 在实际应用中, 时延与噪声需要被限定, 以防止系统状态在控制器作用之前逃逸到不可接受的值<sup>[46,47]</sup>.

因此, 考虑有界噪声、时延情况下, 攻击检测滤波器 (24) 改写为

$$\begin{cases} \dot{I}_f(t) = A I_f(t) + B_l u_l(t) + g(I_f(t) - x(t - \tau)) + g w(t - \tau), \\ f_{ua}(t) = I_f(t) - x(t). \end{cases} \quad (27)$$

假设系统状态导数有界, 即  $\|\dot{x}(t)\| \leq \delta$ , 进一步可得  $\|\dot{x}(t)\| \leq \delta \Rightarrow \|x(t) - x(t - \tau)\| \leq \delta\tau$ . 该假设是合理的, 这是因为信息物理系统的物理层动态通常受实际设备限制, 例如, 电机转速、温度、压力等状态变量在真实系统中不可能无限增长或变化, 其状态导数假设有界是合理的<sup>[48,49]</sup>.

选取李雅普诺夫函数为

$$V_f = f_{ua}^T(t) P f_{ua}(t), P > 0. \quad (28)$$

对  $V_f(t)$  求导可得

$$\begin{aligned} \dot{V}_f = & f_{ua}^T(t) \left[ (A - B_l R_l^{-1} B_l^T P_l + g)^T P + P (A - B_l R_l^{-1} B_l^T P_l + g) \right] f_{ua}(t) \\ & + 2f_{ua}^T(t) P g (x(t) - x(t - \tau)) + 2f_{ua}^T(t) P g w(t - \tau). \end{aligned} \quad (29)$$

选择合适的常数矩阵  $g$  使得矩阵  $A - B_l R_l^{-1} B_l^T P_l + g$  是赫尔维茨稳定矩阵, 则

$$\tilde{Q} = - \left[ (A - B_l R_l^{-1} B_l^T P_l + g)^T P + P (A - B_l R_l^{-1} B_l^T P_l + g) \right] > 0. \quad (30)$$

由  $\|x(t) - x(t - \tau)\| \leq \delta\tau, \|w(t - \tau)\| \leq \bar{w}$ , 可得扰动上界为

$$\dot{V}_f(t) \leq -\lambda_{\min}(\tilde{Q}) \|f_{ua}(t)\|^2 + 2 \|f_{ua}(t)\| \|P g\| (\delta\tau + \bar{w}). \quad (31)$$

由不等式 (31) 可得, 当扰动项有界时, 残差误差  $f_{ua}(t)$  有界, 满足输入 - 状态有界性条件, 进一步有

$$\limsup_{t \rightarrow \infty} \|f_{ua}(t)\| \leq \frac{2 \|P g\| (\delta\tau + \bar{w})}{\lambda_{\min}(\tilde{Q})}. \quad (32)$$

即攻击检测滤波器估计误差可由输入扰动的幅值和时延上界所定量控制. 由此, 残差误差  $f_{ua}(t)$  的幅值也始终保持有界, 上述分析表明, 在存在测量噪声扰动及传输时延的环境下, 攻击检测滤波器估计误差系统具有良好的鲁棒性, 可保证检测系统在噪声与延迟复杂环境中正常工作.

根据上述分析, 在实际的 CPS 中, 需要制定一个攻击检测规则来消除噪声和时延导致的估计误差,

$$\begin{cases} \|f_{ua}(t)\| \geq \bar{f}_{ua}, & \text{受到攻击,} \\ \|f_{ua}(t)\| < \bar{f}_{ua}, & \text{没受到攻击,} \end{cases} \quad (33)$$

其中  $\bar{f}_{ua}$  表示攻击阈值, 是一个正常数. 当 CPS 系统未受到攻击时,  $\bar{f}_{ua}$  通常选择一个相对较小的正常数作为阈值.

### 3.4 DHR 调度算法

本小节将提出一种 DHR 调度算法来改变系统动态特性. 在 DHR 架构中, 在线执行器之间的异构度越大, 在线执行器同时被成功攻击的概率就越低. 此外, 在线执行器的可信度是 CPS 可靠性的先决条件. 因此, 我们使用历史可信度作为重要指标来表征执行器的可靠性. 为了整合 DHR 框架中的方案, 提出了一种可基于异构度和历史可信度进行优化的切换算法. 这种方法将通过异构执行器和模式切换来实现所需的不可预测性和安全性.

历史可信度是一种先验信息, 反映了执行器在受到攻击时的历史行为. 对于执行器模式  $u_l(t)$ , 其历史可信度可以表示为

$$h_l(t) = \frac{\sum_{i=1}^{\zeta} \chi_l(i)}{\zeta(t)}, \quad (34)$$

其中  $\zeta(t)$  表示切换次数,  $\chi_l(i)$  是与执行器  $l$  相关的状态转移函数.

$$\chi_l(i) = \begin{cases} 1, & \text{执行器 } u_l \text{ 上线,} \\ 0, & \text{执行器 } u_l \text{ 下线.} \end{cases} \quad (35)$$

设  $\Phi_a$  为替代执行器组合的集合,  $\Phi \in \Phi_a$  表示在线选择的执行器组合. 那么,  $\Phi$  的历史可信度可表示为  $H(t) = \sum_{l \in \Phi} \sum_{i=1}^{\zeta} \chi_l(i) / \zeta(t)$ .

需要注意的是, 可控对  $(A, B_l)$ ,  $l \in \{1, 2, \dots, \text{card}(\tilde{B}_\gamma)\}$  是从式 (2) 中列向量  $\beta_i$  的所有组合中筛选得到的. 这意味着执行器组合  $\Phi$  可能包含相同的列向量和控制信号  $u_i(t)$ , 从而导致一定程度的相似性. 定义  $L(B_l)$  为输入矩阵  $B_l$  中列向量  $\beta_i$  的集合. 替代执行器组合的异构度  $G_\Phi(t)$  可以描述为

$$G_\Phi(t) = G(B_l, l \in \Phi) = 1 - \frac{\bigcap_{l \in \Phi} L(B_l)}{\bigcup_{l \in \Phi} L(B_l)}. \quad (36)$$

考虑历史可信度和异构度对 DHR 架构安全性的影响, 采用逼近理想解排序 (technique for order preference by similarity to ideal solution, TOPSIS) 算法<sup>[50]</sup> 计算综合权重  $\alpha$ . 组合  $\Phi$  被选中的系数  $S_\Phi$  由下式给出:

$$S_\Phi = \alpha H(t) + (1 - \alpha) G_\Phi(t). \quad (37)$$

DHR 调度面临如下双目标优化权衡问题, 若某执行器历史上攻击次数较少, 可信度高 (即  $H(t)$  大), 说明其通道“安全可信”, 应优先历史表现良好稳定的执行器使用, 如果历史置信度过大会降低系统安全性; 备用执行器与主执行器结构差异大时 (即  $G_\Phi(t)$  大), 攻击者更难复用已掌握的系统模型或注入参数, 从而提高攻击规避能力, 但可能会增加系统不确定性. 因此需选择合适的系数  $\alpha$ , 使得系统在保证安全性的同时维持稳定.

为指导权重系数  $\alpha$  的选择, 定义最大期望鲁棒指数 (expected robustness index, ERI) 作为参考指标, 设攻击者针对主执行器设计攻击的成功概率为  $p_A$ , 历史上备用执行器集合被攻击的频率为  $p_H^l$ , 其异构度  $G_\Phi$  与抗攻击能力相关, 则备用执行器的期望鲁棒性可建模为  $\text{ERI}_\Phi = (1 - p_H^l) G_\Phi$ . 该指数越大, 说明备用执行器越鲁棒可信, 越适合被调度.

**算法 1** 基于历史可信度和异构度的 DHR 执行器调度算法.

**Require:** 给定  $\zeta(0), \alpha$  和时间阈值  $t_a$ , 找出所有执行器 ( $B$  矩阵的列) 的排列组合, 并推导出可控对  $(A, B_l)$  的子集, 记为  $\tilde{B}_\Upsilon$ .

**Ensure:** 异构执行器的数量为  $\text{card}(\tilde{B}_\Upsilon)$ , 同时在线服务的执行器数量为 2 或 3.

- 1:  $\zeta(0) \leftarrow 1, \alpha, P_l, \delta;$
- 2: 执行器被随机选择以构成集合  $\Phi$ . 从  $\Phi$  中随机选择一个可控对  $(A, B_l)$  作为主执行器, 而其他可控对  $(A, B_{l1}) \in \Phi$  则作为备用执行器;
- 3: **while**  $t > 0$  **do**
- 4:   **if**  $\|f_{ua}(t)\| < \bar{f}_{ua}$  **then**
- 5:     可控对  $(A, B_l)$  用于控制该 CPS 系统;
- 6:     当  $\|f_{ua}(t)\| < \bar{f}_{ua}$  的持续时间超过给定的时间阈值  $t_a$  时; 将当前执行器组下线并重置, 从候选可控对  $\tilde{B}_\Upsilon$  中随机选择执行器上线;
- 7:   **else**  $\{\|f_{ua}(t)\| \geq \bar{f}_{ua}\}$
- 8:     当同时在线服务的执行器数量为 3 时, 使用备用执行器对  $(A, B_{l1}), (A, B_{l2})$  的平均值来控制 CPS 系统, 即  $(B_{l1}u_{l1} + B_{l2}u_{l2})/2$ . 或者, 使用备用执行器对  $(A, B_{l1})$  来控制 CPS 系统;
- 9:      $H(t) = \sum_{l \in \Phi} \sum_{i=1}^{\zeta} \chi_l(i)/\zeta(t);$
- 10:      $G_\Phi(t) = G(B_l, l \in \Phi) = 1 - \frac{\prod_{l \in \Phi} L(B_l)}{\prod_{l \in \Phi} L(B_l)};$
- 11:      $S_\Phi = \alpha H(t) + (1 - \alpha) G_\Phi(t);$
- 12:     将当前受攻击的执行器下线并恢复这些执行器, 切换至系数  $S_\Phi$  最高的执行器组合;
- 13:   **end if**
- 14: **end while**

历史置信度  $H(t)$  与  $(1 - p_H^l)$  成正比, 为最大化  $\text{ERI}_\Phi$ , 即最小化攻击风险, 可设置

$$\frac{\alpha}{1 - \alpha} = \left( \frac{\partial \text{ERI}_\Phi}{\partial H} \right) / \left( \frac{\partial \text{ERI}_\Phi}{\partial G_\Phi} \right) = \frac{G_\Phi}{1 - p_H^l}, \quad (38)$$

进一步有  $\alpha = \frac{G_\Phi}{1 + G_\Phi - p_H^l}$ .

考虑到 DHR 架构的整体攻击抵抗能力可以通过异构度来评估, 历史可信度仅反映了在特定时间段内对执行器攻击的特征. 通过历史统计数据与差异度量可对上述比值进行经验学习或设定, 通过理论分析和实验验证, 参数  $\alpha$  取值范围在  $(0, 0.4)$  中时, 系统能够在历史可信度和异构度之间取得较好的平衡, 确保执行器切换的稳定性和安全性<sup>[51, 52]</sup>.

考虑到 DHR 架构在实际应用中的成本效益, 通常选择不超过 3 个执行器同时在线运行<sup>[33]</sup>. 从可控对  $(A, B_l), l \in \{1, 2, \dots, \text{card}(\tilde{B}_\Upsilon)\}$  中, 总共有  $C_{\text{card}(\tilde{B}_\Upsilon)}^3 + C_{\text{card}(\tilde{B}_\Upsilon)}^2$  种可能的组合. 基于历史可信度和异构度的执行器切换规则在算法 1 中给出.

## 4 安全控制算法及分析

本节首先提出一种防御控制算法, 然后对攻击下的防御控制进行分析以确保系统稳定性.

基于所提出的 DHR 安全方案, 所考虑的系统在可用执行器组合之间进行切换, 以确保安全性和最大的不可预测性, 如算法 1 所述. 具体而言, 系统切换到具有最高系数  $S_\Phi$  的执行器组合, 同时将受攻击的执行器从队列中移除, 以促进被隔离通道的及时恢复. 算法 2 展示了主动和被动防御控制算法.

定义  $\tilde{B}_\Upsilon^*$  为  $\tilde{B}_\Upsilon$  的一个子集, 表示已被攻击者攻陷的执行器集合. DHR 安全方案可以利用调度模块修复被攻击的执行器. 定义  $\Delta T$  为执行器的修复时间. 显然, 在  $\Delta T$  时间内攻击所有异构执行器是一个极低概率的事件. 因此, 更多可用的可控对可以增强 DHR 方案的性能. 接下来, 将给出一个定理, 以说明系统 (5) 在攻击下的稳定性.

**定理 3** 假设攻击者在时间  $\Delta T$  内无法攻陷所有可用执行器, 即  $\tilde{B}_\Upsilon \setminus \tilde{B}_\Upsilon^* \neq \emptyset$ . 在主动和被动防御控制 (算法 2) 的作用下, 闭环系统在攻击下仍然保持指数稳定.

**算法 2** 针对虚假数据注入攻击的主动与被动防御控制算法.

**Require:** 已知初始时间  $t = 0$  和初始状态  $x(0)$ ; 求所有可控对  $(A, B_l)$ , 记为集合  $\tilde{B}_\Upsilon$ .

```

1: for  $l = 1 : \text{card}(\tilde{B}_\Upsilon)$  do
2:   计算最优控制器 (7) 中的黎卡提方程矩阵  $P_l$  和控制增益  $K_l$ ;
3:   计算函数  $V_l(0) = x^T(0)P_l x(0)$ ;
4: end for
5: 引入切换时间序列  $\{t_1, t_2, \dots, t_k, \dots\}$ ;
6: 根据算法 1 激活可控对  $(A, B_l)$  和  $(A, B_{l1})$ , 形成集合  $\Phi$  并从中随机选择一个作为主执行器;
7: while  $\vartheta(t) = l$  and  $t - t_k \leq \tau_D^*$  do
8:   使用方程 (4) 和控制器 (7) 运行该系统;
9:   使用式 (24) 计算残差误差识别信号;
10: end while
11: if  $\|f_{ua}(t)\| < \bar{f}_{ua}$  then
12:   返回到步骤 6;
13: else
14:   将集合  $\Phi$  下线;
15:   启用被动防御方案, 运行算法 1;
16:   运行系统 (4) 和控制器 (7) 至少  $\tau_D^*$  时间长度, 然后返回到步骤 6;
17: end if
    
```

**证明** 考虑当前时刻与上一次系统切换之间的时间区间  $[t_\vartheta, t)$ . 如果在线主执行器未受到攻击, 或者攻击与上一次切换之间的时间区间满足  $t - t_\vartheta > \tau_D^*$ , 则可以使用定理 2 完成证明. 如果攻击与上一次切换之间的时间满足  $t - t_\vartheta \leq \tau_D^*$ , 则证明将分为两个步骤: (1) 当两个备用执行器未受到攻击时, 系统的状态轨迹  $\|x\|$  下降; (2) 当两个备用执行器受到攻击时, 闭环系统保持有界.

**步骤 1.** 将候选李雅普诺夫函数视为

$$V_l(x(t)) = \frac{1}{2}x^T(P_{l1} + P_{l2})x, \quad (39)$$

其中  $P_{l1}, P_{l2}$  分别表示备用执行器的黎卡提方程的解. 对  $V_l(t)$  求导得到

$$\begin{aligned} \dot{V}_l(x) &= x^T(t) \left( \frac{A^T P_{l1} + P_{l1} A}{2} - P_{l1} B_{l1} R_{l1}^{-1} B_{l1}^T P_{l1} \right) x(t) \\ &\quad + x^T(t) \left( \frac{A^T P_{l2} + P_{l2} A}{2} - P_{l2} B_{l2} R_{l2}^{-1} B_{l2}^T P_{l2} \right) x(t) \\ &= \frac{1}{2}x^T(t) (-Q_{l1} - P_{l1} B_{l1} R_{l1}^{-1} B_{l1}^T P_{l1}) (x(t)) \\ &\quad + \frac{1}{2}x^T (-Q_{l2} - P_{l2} B_{l2} R_{l2}^{-1} B_{l2}^T P_{l2}) (x(t)) \\ &\leq - \frac{x^T(t) (\hat{Q}_{l1} + \hat{Q}_{l2}) x(t)}{2} \leq 0, \end{aligned} \quad (40)$$

其中  $\hat{Q}_{l1} = Q_{l1} + P_{l1} B_{l1} R_{l1}^{-1} B_{l1}^T P_{l1}$ ,  $\hat{Q}_{l2} = Q_{l2} + P_{l2} B_{l2} R_{l2}^{-1} B_{l2}^T P_{l2}$ . 根据对称矩阵瑞利 - 里茨 (Rayleigh-Ritz) 不等式, 有  $\|x\|^2 \leq \frac{2V_l(t)}{\lambda_{\min}(P_{l1} + P_{l2})}$ , 这表明系统的状态轨迹  $\|x\|$  是递减的.

**步骤 2.** 考虑候选李雅普诺夫函数 (39), 对  $V_l(t)$  进行求导得到

$$\begin{aligned} \dot{V}_l(x) &= \frac{1}{2}x^T(t) (-Q_{l1} - P_{l1} B_{l1} R_{l1}^{-1} B_{l1}^T P_{l1}) x(t) \\ &\quad + \frac{1}{2}x^T(t) (-Q_{l2} - P_{l2} B_{l2} R_{l2}^{-1} B_{l2}^T P_{l2}) x(t) \\ &\quad + \frac{1}{2}x^T(t) P_l B_l u_{l,a} + \frac{1}{2}(B_l u_{l,a})^T P_l x(t) \\ &\leq - \frac{x^T(t) (\hat{Q}_{l1} + \hat{Q}_{l2}) x(t)}{2} + \|B_l\| \|u_{l,a}\| \|P_l\| \|x\|. \end{aligned} \quad (41)$$

根据假设 1 可知,  $\dot{V}_l(t)$  是有界的. 此外, 在时间区间  $[t, t_\theta + \tau_D^*]$  内,  $V_l(t)$  和  $\|x\|$  也是有界的. 当  $t - t_\theta > \tau_D^*$  时, 当前执行器集将下线, 并根据算法 1 切换执行器.

接下来的证明与定理 1 的证明过程类似, 此处省略. 证明完毕.

## 5 仿真与评估

### 5.1 典型 CPS 场景验证

为验证所提方法的有效性, 本文选择内置冗余执行器的 CPS, 比如四容水箱系统<sup>[53]</sup>, 采用 ECIES 算法进行密钥协商, 并利用 AES-256 对消息进行加密, 加密操作采用 CBC 模式, 使用 SHA-256 执行哈希操作, ECIES 采用 secp256k1 椭圆曲线. 实验环境配置为 Intel(R) Core(TM) i7-4790 @ 3.60 GHz 处理器, 20 GB DDR3 内存, 操作系统为 Windows 10. 在实验中, 传感器和执行器的加密模块 CPU 频率设置为 1.2 GHz, 同时将控制台的 CPU 频率设置为 3.6 GHz. 设置  $x(0) = \{x_1, x_2, x_3, x_4, x_5\} = [-0.31, 0.51, 0.11, 0.52, -0.37]^T$  作为初始状态. 仿真时间间隔设为  $[0, 40]$ . 在给定的修剪条件下, 系统 (1) 的系数矩阵  $A$ ,  $B$  和  $C$  如下方程所示:

$$A = \begin{bmatrix} -1.0649 & 0.0034 & -0.0000 & 0.9728 & 0.0000 \\ 0.0000 & -0.2492 & 0.0656 & -0.0000 & -0.9879 \\ 0.0000 & -22.5462 & -2.0457 & -0.0000 & 0.5432 \\ 8.1633 & -0.0057 & -0.0000 & -1.0478 & 0.0000 \\ 0.0000 & 1.7970 & -0.1096 & 0.0000 & -0.4357 \end{bmatrix},$$

$$B = \begin{bmatrix} -0.0062 & -0.0062 & -0.0709 & -0.1172 & -0.1172 & -0.0709 & 0.0003 \\ -0.0072 & 0.0072 & 0.0039 & 0.0188 & -0.0188 & -0.0039 & 0.0627 \\ 1.2456 & -1.2456 & -10.6058 & -9.2345 & 9.2345 & 10.6058 & 5.3223 \\ 2.7172 & 2.7172 & -2.4724 & -4.0101 & -4.0101 & -2.4724 & 0.0108 \\ -0.7497 & 0.7497 & -0.4923 & -1.1415 & 1.1415 & 0.4923 & -3.7367 \end{bmatrix},$$

$$C = [1, 0, 0, 0, 0].$$

在系统运行过程中, 传感器加密感知数据并将其传输至控制台; 控制台解密接收到的感知数据, 并加密控制指令以发送至执行器, 执行器在接收到指令后进行解密并根据指令控制被控系统状态. 本文采用循环运行 1000 次的方法以计算结果, 结果如图 2 所示, 可以看出传感器、执行器和控制台的平均计算开销分别是 2.506, 1.867 和 1.896 ms. 通过对各设备性能的对标分析, 可以观察到传感器的计算开销略高于执行器和控制台, 这表明在机密交互协议中, 传感器在数据加密过程中的性能是系统整体延迟的主要影响因素. 然而, 整个系统的计算开销仍然保持在可接受的范围内, 从而确保了信息物理系统控制的实时性.

本文主要展示在对抗环境中控制系统的结果, 以验证所提弹性控制方法的有效性. 未知攻击发生的时间间隔设置为  $t \in [20, 25]$ , 攻击检测阈值设置为  $\bar{f}_{ua} = 0.002$ . 图 3 展示了仅使用检测机制时, 系统状态在未知攻击信号下的演变. 可以看出, 受到未知攻击时系统状态仍然出现了明显的波动, 这是因为检测机制虽然检测出未知攻击的存在, 但由于没有采用安全策略, 系统状态依然受到攻击的影响, 未能保持稳定.

图 4 描绘了针对执行器攻击的主动和被动防御控制方案的切换信号. 可以看到, 切换信号的变化表明系统在不断适应攻击环境. 通过动态切换执行器, 系统能够有效隔离受攻击的执行器, 确保被控系统的正常工作. 这种切换机制不仅提高了系统的安全性, 也增强了其对未知攻击的抵抗能力.

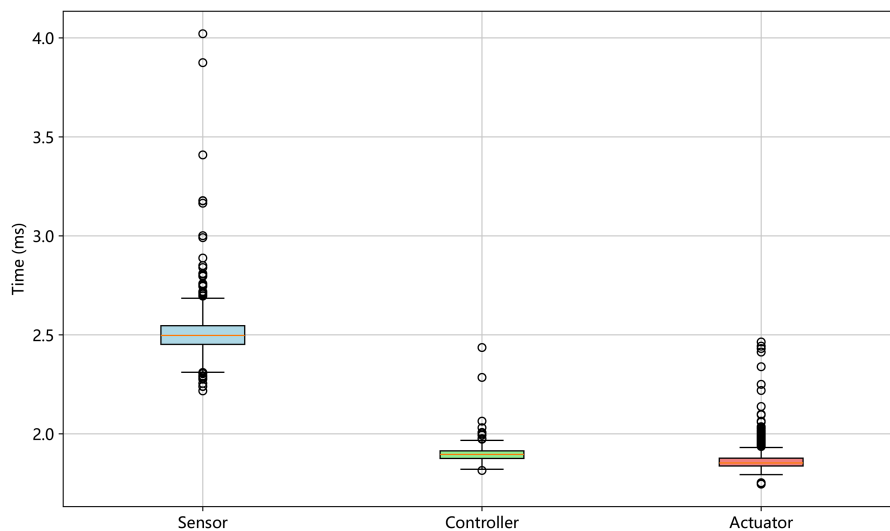


图 2 (网络版彩图) 机密交互协议计算性能.

Figure 2 (Color online) Performance of the confidential interaction protocol.

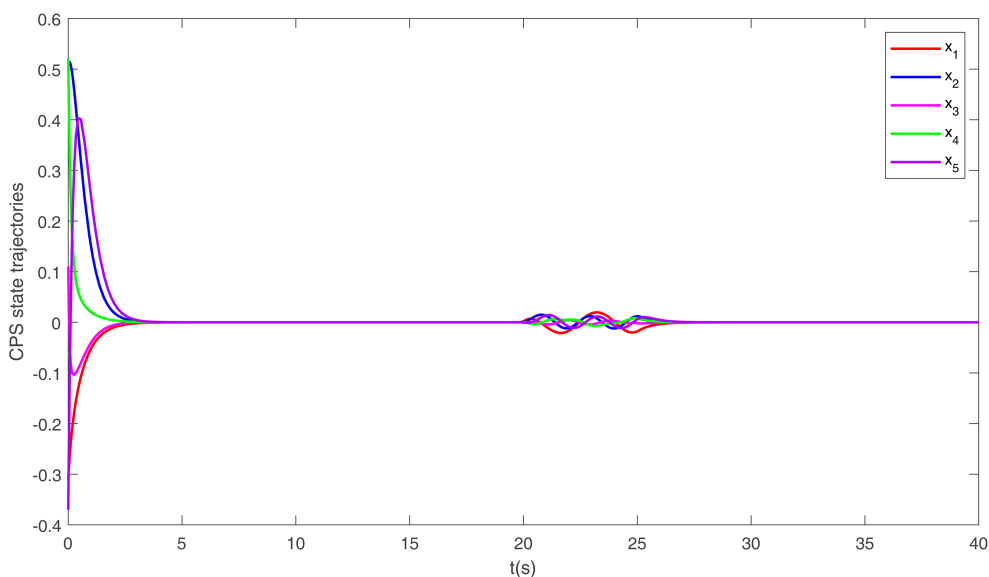


图 3 (网络版彩图) 在攻击信号下, 仅凭检测机制的收敛状态轨迹.

Figure 3 (Color online) Convergence state trajectory with only the detection mechanism under an attack signal.

图 5 描述了在使用主动和被动防御机制时攻击信号下的收敛状态. 结果显示, 系统状态在未知攻击发生时未受到明显扰动, 这表明所提出的防御机制有效地隔离了攻击对系统的影响, 确保了系统的稳定性和性能.

在图 6 中, 我们可以观察到攻击检测滤波器状态和残差误差的演变. 如图 6 所示, 检测滤波器成功检测到注入的信号. 此处呈现的仿真结果验证了本文所提检测方法的有效性.

## 5.2 智能电网案例验证

为全面验证所提方法的普适性及与现有防御策略的优势对比, 本文进一步针对智能电网场景的电压调节系统进行了扩展仿真实验. 在该场景中, 配电网电压需严格维持在额定值的  $\pm 5\%$  范围内 (如  $220 \pm 11$  V), 变电站通过自动调节分接开关、无功补偿设备等装置实现电压稳定控制. 攻击者可通过虚假数据注入攻击篡改执行器的调整电压数据, 进而错误控制分接开关与无功补偿设备, 导致电压越限 (过高或过低), 从而引发设备损坏或系统

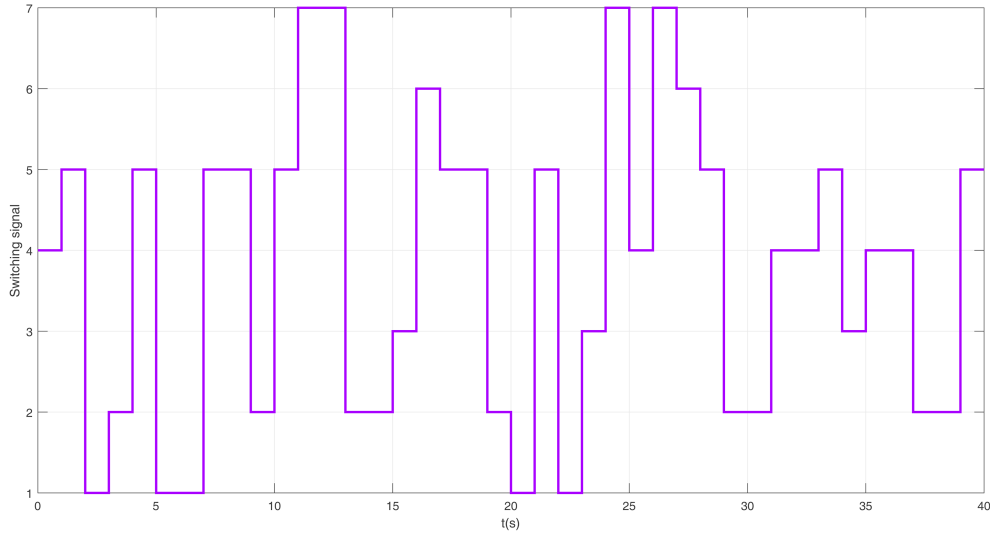


图 4 (网络版彩图) 主动和被动防御控制方案的切换信号。

Figure 4 (Color online) Switching signal for the proactive-reactive defense control scheme.

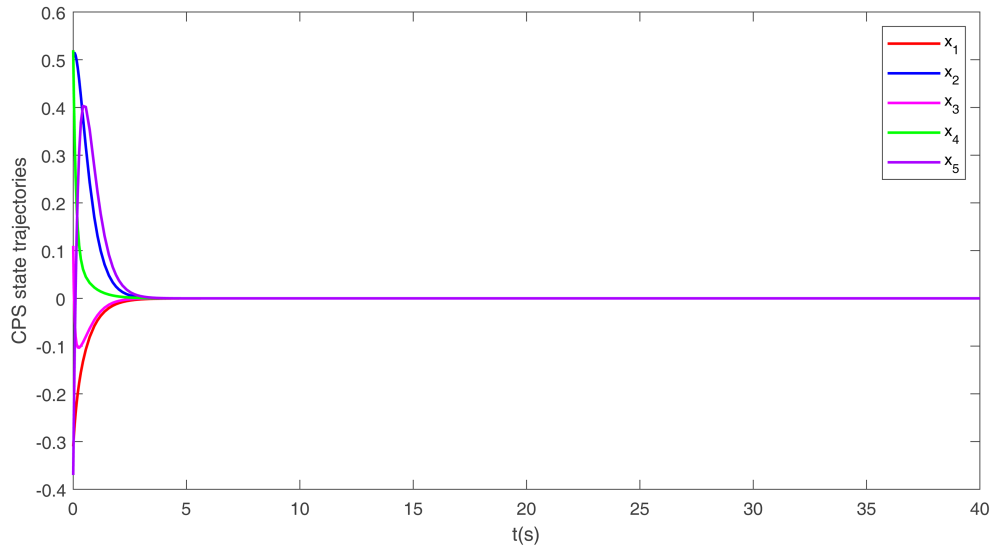


图 5 (网络版彩图) 采用主动和被动防御机制时的攻击信号收敛状态。

Figure 5 (Color online) Convergence state under an attack signal when utilizing the proactive-reactive defense mechanism.

崩溃等严重后果。

仿真设置方面, 实验基于 Python 3.12.8 环境构建, 总仿真时长设为 100 s, 采样周期为 1 s. 系统在第 15, 40, 60 和 75 s 这 4 个时间点分别遭受虚假数据注入攻击, 攻击者试图通过向执行器注入错误控制数据使系统电压偏离安全范围. 本实验将所提方法与静态冗余架构和移动目标防御两种典型防御策略进行了对比, 同时也与仅依赖加密的方法和仅依赖攻击检测机制的方法进行了性能比较.

图 7 展示了所提方法与静态冗余和移动目标防御在攻击场景下的任务执行时间对比. 静态冗余方法采用 3 个异构执行器, 通过多数表决机制确定最终执行器输出结果; 移动目标防御则以预设周期随机更换执行器. 实验假设系统在电压越限后会触发保护机制并能够进行人工恢复, 同时假设虚假数据注入攻击需持续输出一定时间的错误数据才能导致电压越限. 实验结果表明, 对于静态冗余方法, 随着攻击持续进行, 攻击者能够逐个攻破执行器, 导致系统在后续时间内频繁出现电压越限状况. 而移动目标防御策略虽不断切换执行器, 但当被攻击者

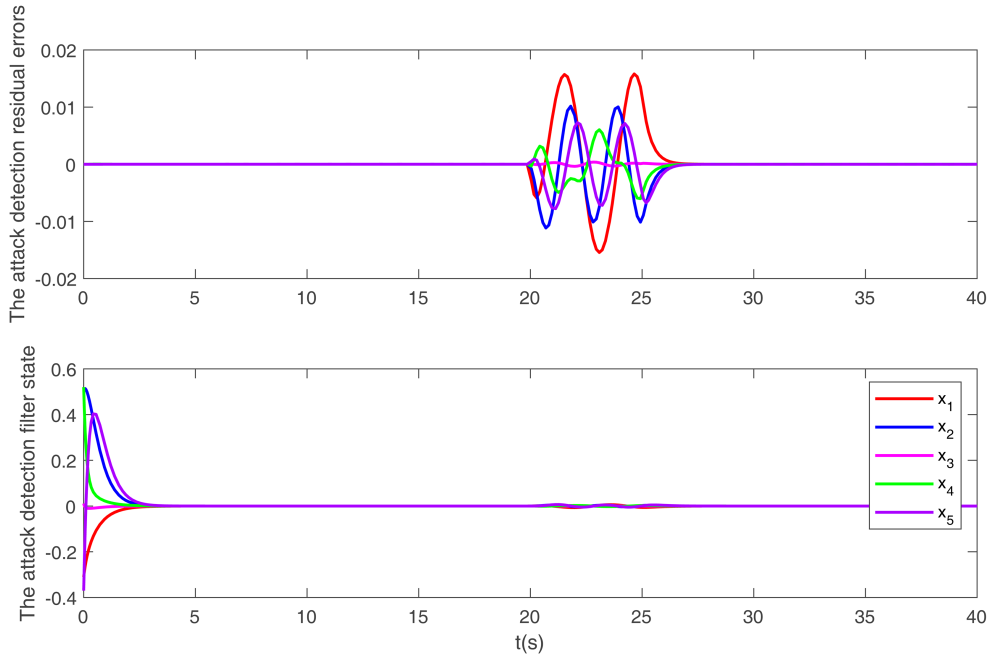


图 6 (网络版彩图) 攻击下的攻击检测滤波器状态和残差误差.  
 Figure 6 (Color online) Attack detection filter state and the residual error with attacks.

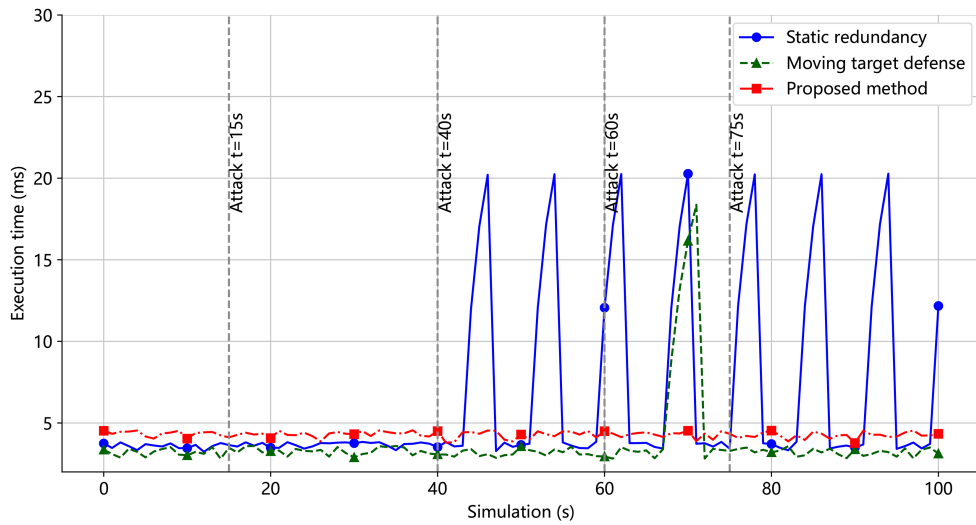


图 7 (网络版彩图) 不同防御策略在电压调节系统中的任务执行时间对比.  
 Figure 7 (Color online) Comparison of task execution time among different defense strategies in the voltage regulation system.

掌控的执行器处于激活状态且攻击持续时间足够长时,系统仍可能出现电压越限情况.相比之下,本文提出的防御方法能在检测到执行器被攻破输出错误数据时,在确保系统稳定所允许的切换频率范围内,触发执行器切换和清洗机制,有效阻断攻击者的虚假数据注入攻击路径.

图 8 呈现了所提方法与静态冗余和移动目标防御在攻击场景下的执行器输出结果对比.实验结果表明,当攻击者掌控的执行器处于在线状态时,移动目标防御方法仍可能执行错误的电压调控操作;静态冗余方法在攻击者成功攻破两个执行器后,持续输出错误的电压调控指令.而本文所提方法能够在执行器输出异常结果后,在确保系统稳定所允许的切换频率范围内,启动执行器下线与清洗操作,同时上线新的备用执行器,使攻击影响被及时隔离,且经过裁决机制后的系统输出结果始终保持正确的电压调控值,如图 9 所示.

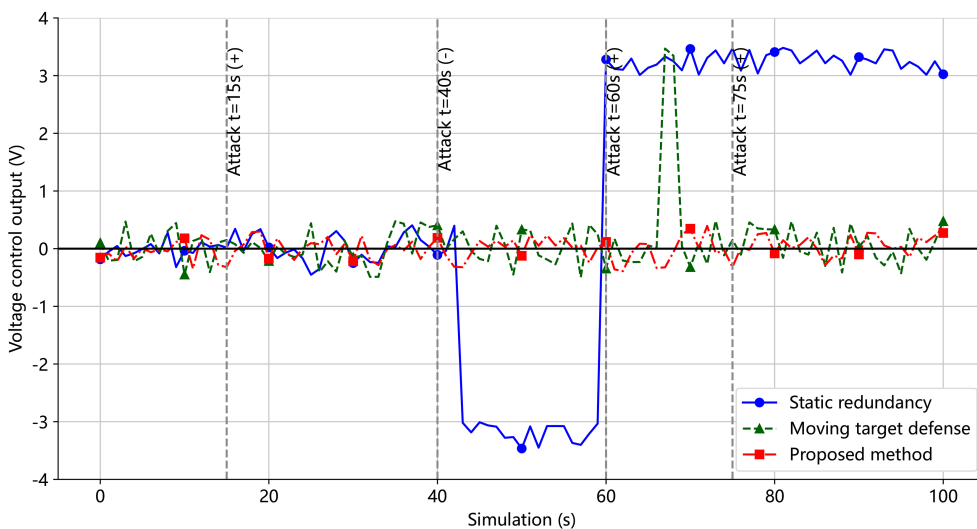


图 8 (网络版彩图) 不同防御策略在攻击场景下的电压调控输出对比.

Figure 8 (Color online) Comparison of voltage regulation outputs among different defense strategies in attack scenarios.

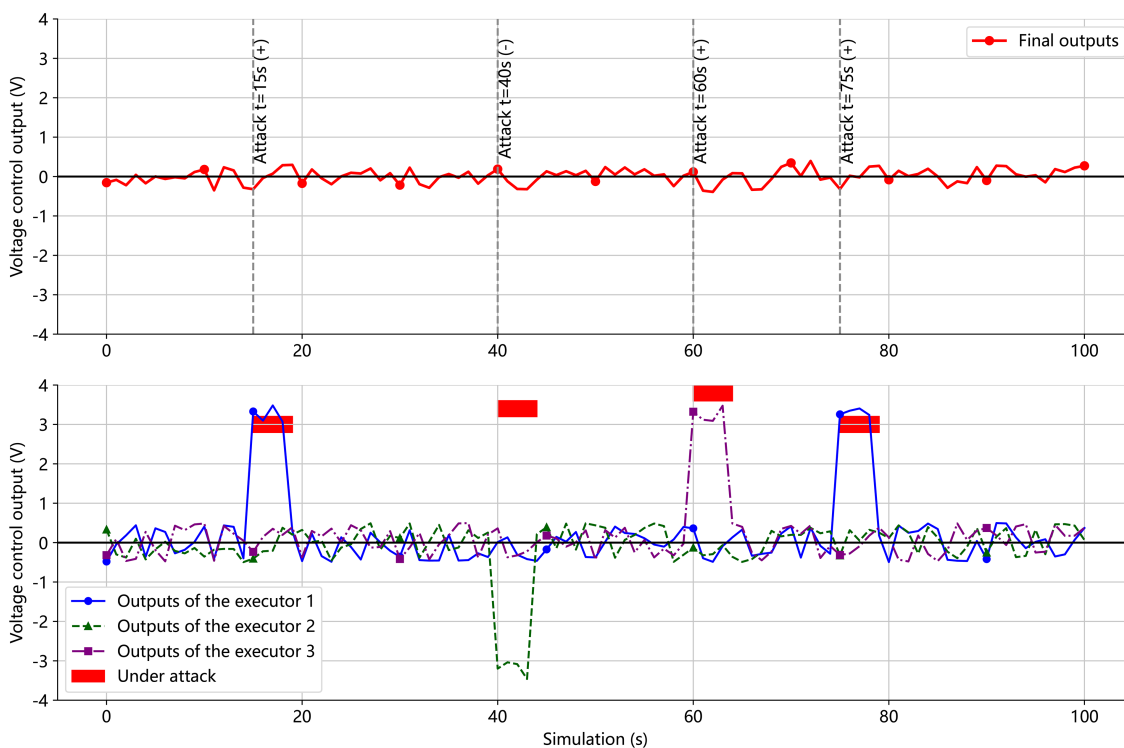


图 9 (网络版彩图) 所提方法在攻击场景下各执行器输出及最终系统输出.

Figure 9 (Color online) Outputs of individual actuators and final system outputs using the proposed method in attack scenarios.

进一步地, 图 10 展示了所提方法与仅依赖加密机制和仅依赖攻击检测机制的防御方法在攻击场景下的任务执行时间对比. 实验结果表明, 仅依赖加密方法虽能有效抵御窃听攻击, 防止攻击者获取整个系统参数, 但对于不依赖系统参数的虚假数据注入攻击则无能为力, 导致每次攻击均能成功突破防线. 基于检测机制的方法在开始的两次攻击中成功识别并隔离攻击, 使系统恢复正常运行, 但由于攻击者能够通过窃听攻击获取系统参数, 在后两次攻击中, 攻击者可调整虚假数据注入策略, 使注入的错误数据更具隐蔽性, 从而在持续一定时间后成功攻破系统. 相比之下, 本文所提方法通过集成主动防御、被动防御和机密交互机制, 不仅能够防止攻击者获取完

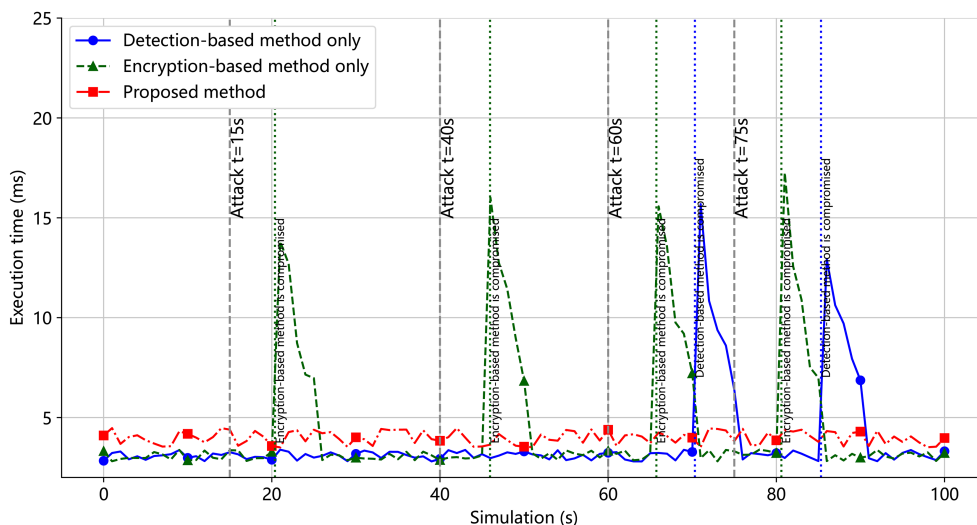


图 10 (网络版彩图) 所提方法与单一防御机制在攻击场景下的性能对比。

Figure 10 (Color online) Performance comparison between the proposed method and single defense mechanisms in attack scenarios.

整系统参数, 而且能够有效检测并隔离受攻击执行器, 同时动态调整系统构型, 从而在全部 4 次攻击中都成功维持系统稳定运行。

## 6 结论

本文提出了一种信息物理融合的主被动协同防御架构, 有效解决了 CPS 系统在虚假数据注入与窃听协同攻击下的安全控制挑战, 避免了现有研究中防御技术简单叠加导致的系统复杂度增加和资源效率低下等问题。该架构以 DHR 架构为基础, 将信息域的机密交互协议与物理域的攻击检测滤波器有机融合, 形成互补式防御体系。在信息域, 设计了保障数据传输安全且满足 CPS 实时性需求的机密交互协议, 并提出了能够平衡检测可靠性、异构度最大化和系统稳定性的 DHR 控制器调度算法; 在物理域, 提出的攻击检测滤波器弥补了 DHR 架构在异构资源受限、底层同源风险和切换频率约束等方面的不足。构建的主被动结合防御控制算法通过控制器动态切换实现主动防御, 增强系统不可预测性; 通过攻击检测滤波器及时识别并隔离受攻击控制器, 实现被动防御。基于李雅普诺夫稳定性理论和切换系统理论的严谨分析证明, 所提架构能够确保面临未知攻击时系统稳定性; 多种 CPS 应用场景的仿真实验进一步验证了该方法增强了系统抵御协同攻击能力的有效性。

## 参考文献

- 1 Duo W, Zhou M C, Abusorrah A. A survey of cyber attacks on cyber physical systems: recent advances and challenges. *IEEE CAA J Autom Sin*, 2022, 9: 784–800
- 2 Zhang Q R, Meng S Q, Wang L H, et al. Secure output-feedback control for cyber-physical systems under stealthy attacks. *Acta Autom Sin*, 2024, 50: 1363–1372 [张淇瑞, 孟思琪, 王兰豪, 等. 隐蔽攻击下信息物理系统的安全输出反馈控制. *自动化学报*, 2024, 50: 1363–1372]
- 3 Habib A A, Hasan M K, Alkhayyat A, et al. False data injection attack in smart grid cyber physical system: issues, challenges, and future direction. *Comput Electrical Eng*, 2023, 107: 108638
- 4 Reda H T, Anwar A, Mahmood A N, et al. A taxonomy of cyber defence strategies against false data attacks in smart grids. *ACM Comput Surv*, 2023, 55: 1–37
- 5 Han K, Zhang K, Wang Z P, et al. Resilient predictive load frequency control of multiarea interconnected power systems with privacy preserving and active detection against stealthy cyber attacks. *IEEE Int Things J*, 2025, 12: 9044–9057
- 6 Xin L, He G, Long Z. Stealthy false data injection attacks detection and classification in cyber-physical systems using deep reinforcement learning. *IEEE Trans Automat Sci Eng*, 2025, 22: 141–153

- 7 Cash M, Morales-Gonzalez C, Wang S, et al. On false data injection attack against building automation systems. In: Proceedings of International Conference on Computing, Networking and Communications (ICNC), 2023. 35–41
- 8 Wang J, Yang W, Chen G, et al. Security analysis and defense of multi-encoding mechanism against eavesdropping attacks. *IEEE Trans Netw Sci Eng*, 2025, 12: 3758–3769
- 9 Wang K, Gao H, Xu X, et al. An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks. *IEEE Sens J*, 2016, 16: 4051–4062
- 10 Chen Y J, Wang L C, Liao C H. Eavesdropping prevention for network coding encrypted cloud storage systems. *IEEE Trans Parallel Distrib Syst*, 2016, 27: 2261–2273
- 11 Min Z, Yang G, Sangaiah A K, et al. A privacy protection-oriented parallel fully homomorphic encryption algorithm in cyber physical systems. *J Wireless Com Netw*, 2019, 2019: 15
- 12 Anikin I, Alnajjar K. Secure data transmission in cyber-physical systems based on the new approach for stream cipher's gamma generation. In: *Cyber-Physical Systems*. Berlin: Springer, 2021. 333–346
- 13 Teranishi K, Sadamoto T, Chakraborty A, et al. Designing optimal key lengths and control laws for encrypted control systems based on sample identifying complexity and deciphering time. *IEEE Trans Automat Contr*, 2022, 68: 2183–2198
- 14 Nekouei E, Tanaka T, Skoglund M, et al. Information-theoretic approaches to privacy in estimation and control. *Annu Rev Control*, 2019, 47: 412–422
- 15 Cortés J, Dullerud G E, Han S, et al. Differential privacy in control and network systems. In: Proceedings of the 55th Conference on Decision and Control (CDC), 2016. 4252–4272
- 16 Ye H, Liu J, Wang W, et al. Secure and efficient outsourcing differential privacy data release scheme in cyber-physical system. *Future Generation Comput Syst*, 2020, 108: 1314–1323
- 17 Olowonni F O, Rawat D B, Liu C. Federated learning with differential privacy for resilient vehicular cyber physical systems. In: Proceedings of the 18th Annual Consumer Communications & Networking Conference (CCNC), 2021. 1–5
- 18 Yuan L, Wang K, Miyazaki T, et al. Optimal transmission strategy for sensors to defend against eavesdropping and jamming attacks. In: Proceedings of IEEE International Conference on Communications (ICC), 2017. 1–6
- 19 Jin Z, Ma M, Wang Z, et al. Optimal transmission schedule with privacy preservation for cyber-physical system against eavesdropping attack. *IEEE Signal Process Lett*, 2025, 32: 436–440
- 20 Feng Y, Wang F, Duan F, et al. Anonymous privacy-preserving consensus via mixed encryption communication. *IEEE Trans Circ Syst II*, 2022, 69: 3445–3449
- 21 Yin X, Li S. Verification of opacity in networked supervisory control systems with insecure control channels. In: Proceedings of IEEE Conference on Decision and Control (CDC), 2018. 4851–4856
- 22 Li Q X, Liu Y G, Li S B, et al. A survey of attack identification based on intrusion detection in cyber-physical systems. *Control Eng China*, 2022, 29: 1049–1057 [李沁雪, 刘永桂, 黎善斌, 等. 基于入侵检测的信息物理系统攻击识别综述. *控制工程*, 2022, 29: 1049–1057]
- 23 Harkat H, Camarinha-Matos L M, Goes J, et al. Cyber-physical systems security: a systematic review. *Comput Indust Eng*, 2024, 188: 109891
- 24 Lian Z, Shi P, Chen M. A survey on cyber-attacks for cyber-physical systems: modeling, defense, and design. *IEEE Int Things J*, 2025, 12: 1471–1483
- 25 Wu S, Luo H, Yin S, et al. A residual-driven secure transmission and detection approach against stealthy cyber-physical attacks for accident prevention. *IEEE Trans Inform Forensic Secur*, 2023, 18: 5762–5771
- 26 Liu Y, Cheng L, Ye D. Stealthy false data injection attacks against the summation detector in cyber-physical systems. *Trans Ind Cyb-Phy Sys*, 2024, 2: 391–403
- 27 Wu S, Luo H, Zhang J, et al. Coprime factorization-based encryption and attack detection for nonlinear cyber-physical systems using deep learning approach. *IEEE Trans Automat Sci Eng*, 2025, 22: 14020–14029
- 28 Wang J, Li X. Data-driven stealthy attacks detection in cyber-physical systems based on complex dynamical networks encoding strategy. *Intl J Robust Nonlinear*, 2025, 35: 3154–3165
- 29 Milinković S A, Lazić L R. Industrial PLC security issues. In: Proceedings of the 20th Telecommunications Forum (TELFOR), 2012. 1536–1539
- 30 Osen O L, Singh V, Hovden C, et al. Scalable, masterless, distributed, and redundant motor control for PLC using object orientation and recursivity. In: Proceedings of the 9th International Congress on Information and Communication Technology, 2024. 255–268
- 31 Tan J, Jin H, Zhang H, et al. A survey: when moving target defense meets game theory. *Comput Sci Rev*, 2023, 48: 100544
- 32 Yao Q, Xiong X L, Wang Y J, et al. Review of moving target defense: an analysis of vulnerability and applications in new scenarios. *Control Dec*, 2023, 38: 3025–3038 [姚倩, 熊鑫立, 王永杰, 等. 移动目标防御综述: 脆弱性分析及新场景应用. *控制与决策*, 2023, 38: 3025–3038]

- 33 Wu J. *Cyberspace Mimic Defense*. Berlin: Springer, 2020
- 34 Shi L, Miao Y, Ren J, et al. Game analysis and optimization for evolutionary dynamic heterogeneous redundancy. *IEEE Trans Netw Serv Manage*, 2023, 20: 4186–4197
- 35 Wu J X. Cyberspace's endogenous safety and security problem and the countermeasures. *Sci Sin Inform*, 2022, 52: 1929–1937 [郭江兴. 论网络空间内生安全问题及对策. *中国科学: 信息科学*, 2022, 52: 1929–1937]
- 36 Sun Q, Zhang K, Shi Y. Resilient model predictive control of cyber-physical systems under DoS attacks. *IEEE Trans Ind Inf*, 2020, 16: 4920–4927
- 37 Khojasteh M J, Khina A, Franceschetti M, et al. Learning-based attacks in cyber-physical systems. *IEEE Trans Control Netw Syst*, 2021, 8: 437–449
- 38 Das S, Ghosh A, Chatterjee D. Detection of false data injection attacks in cyber-physical systems. In: *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2024. 3249–3254
- 39 Cuan Z, Ding D W, Ren Y, et al. Adaptive fixed-time control for state-constrained high-order uncertain nonlinear cyber-physical systems under malicious attacks. *IEEE Trans Fuzzy Syst*, 2023, 31: 4285–4297
- 40 Wu C, Yao W, Pan W, et al. Secure control for cyber-physical systems under malicious attacks. *IEEE Trans Control Netw Syst*, 2021, 9: 775–788
- 41 Hankerson D, Menezes A, Vanstone S. *Guide to Elliptic Curve Cryptography*. Berlin: Springer, 2006
- 42 Li D, Li J, Di X, et al. Design of cross-plane colour image encryption based on a new 2D chaotic map and combination of ECIES framework. *Nonlinear Dyn*, 2023, 111: 2917–2942
- 43 Khasawneh S, Kadoch M. Hybrid cryptography algorithm with precomputation for advanced metering infrastructure networks. *Mobile Netw Appl*, 2018, 23: 982–993
- 44 Lewis F L, Vrabie D, Syrmos V L. *Optimal Control*. Hoboken: John Wiley & Sons, 2012
- 45 Hespanha J, Morse A. Stability of switched systems with average dwell-time. In: *Proceedings of the 38th IEEE Conference on Decision and Control*, 1999. 2655–2660
- 46 Celentano L, Basin M V. Optimal estimator design for LTI systems with bounded noises, disturbances, and nonlinearities. *Circ Syste Signal Process*, 2021, 40: 3266–3285
- 47 Xu L, Guo Q, Zeng H, et al. Communication topology optimization for time-delay cyber-physical microgrids under distributed control. *IEEE Trans Smart Grid*, 2024, 15: 5087–5101
- 48 Ghaffari V. Robust ultimately boundedness control of uncertain delayed systems under time-varying reference and external disturbance. *ISA Trans*, 2024, 146: 114–126
- 49 Ni J, Liu L, Liu C, et al. Fixed-time leader-following consensus for second-order multiagent systems with input delay. *IEEE Trans Ind Electron*, 2017, 64: 8635–8646
- 50 Forcina A, Silvestri L, de Felice F, et al. Exploring Industry 4.0 technologies to improve manufacturing enterprise safety management: a TOPSIS-based decision support system and real case study. *Saf Sci*, 2024, 169: 106351
- 51 Shao S S, Ji Y M, Zhang W L, et al. A DHR executor selection algorithm based on historical credibility and dissimilarity clustering. *Sci China Inf Sci*, 2023, 66: 212304
- 52 Yang K, Zhang F, Guo W. Fast construction algorithm for confidence targets based on historical information and heterogeneity of executive bodies. *J Inform Eng Univ*, 2021, 22: 694–698 [杨珂, 张帆, 郭威. 一种基于执行体的历史信息 and 异构性的置信目标快速构建算法. *信息工程大学学报*, 2021, 22: 694–698]
- 53 Giraldo J A, El Hariri M, Parvania M. Moving target defense for cyber-physical systems using IoT-enabled data replication. *IEEE Int Things J*, 2022, 9: 13223–13232

# Privacy-preserving resilience control for cyber-physical systems based on the dynamic heterogeneous redundancy architecture

Yukun NIU<sup>1</sup>, Lei HE<sup>1</sup>, Chuan HE<sup>2,3</sup>, Xiaopeng HAN<sup>1</sup>, Zhigang CAO<sup>1</sup> & Ding ZHOU<sup>1\*</sup>

1. *Purple Mountain Laboratories, Nanjing 211111, China*

2. *China Electric Power Research Institute, Nanjing 210003, China*

3. *School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China*

\* Corresponding author. E-mail: zhoudinghawk@163.com, zhouding@pmlabs.com.cn

**Abstract** This paper addresses the security control challenges of cyber-physical system (CPS) under coordinated false data injection and eavesdropping attacks. Existing research often relies on stacking multiple defense techniques to counter complex and diverse coordinated attacks, resulting in increased system complexity, interference among defense mechanisms, and reduced resource efficiency. To overcome these limitations, we propose a cyber-physical fusion proactive-reactive defense architecture. This architecture is built upon dynamic heterogeneous redundancy (DHR) and integrates confidential interaction protocols to defend against coordinated false data injection and eavesdropping attacks in the cyber domain. We design a lightweight confidential interaction protocol based on elliptic curve cryptography that ensures data transmission security while meeting CPS real-time requirements. Additionally, we propose a DHR scheduling algorithm based on historical trustworthiness and heterogeneity that balances detection reliability, heterogeneity maximization, and system stability. To address challenges faced by DHR architecture in practical applications—including limited heterogeneous resources, underlying homogeneity risks, and switching frequency constraints—we design an attack detection filter in the physical domain as a complementary measure. The developed proactive-reactive defense control algorithm enhances system unpredictability through dynamic controller switching (proactive defense) while timely identifying compromised controllers using attack detection filters (reactive defense). Theoretical analysis demonstrates that the closed-loop system maintains stability even under unknown attacks. Simulation experiments across various CPS scenarios confirm that our proposed resilient control method enhances the system's ability to resist coordinated false data injection and eavesdropping attacks, validating the effectiveness of our approach.

**Keywords** cyber-physical systems, dynamic heterogeneous redundancy, proactive-reactive defense, false data injection attack, eavesdropping attack