假数据注入攻击下信息物理融合系统的稳定性研究

彭大天 [1,2] 董建敏 [1,2] 蔡忠闽 [1,2] 张长青 [3] 彭勤科 [1,2]

摘 要 假数据注入 (False data injection, FDI) 攻击由于其隐蔽性特点, 严重威胁着信息物理融合系统 (Cyber-physical systems, CPS) 的安全. 从攻击者角度, 本文主要研究了 FDI 攻击对 CPS 稳定性的影响. 首先, 给出了 FDI 攻击模型, 从前向通道和反馈通道分别注入控制假数据和测量假数据. 接着, 提出了 FDI 攻击效力模型来量化 FDI 攻击对 CPS 状态估计值和测量残差的影响. 在此基础上, 设计了一个攻击向量协同策略, 并从理论上分析出操纵 CPS 稳定性的攻击条件: 攻击矩阵 H 和系统矩阵 A 的稳定性及时间参数 k_a 的选取时机. 数值仿真结果表明 FDI 攻击协同策略能够有效操纵两类 (含有稳定和不稳定受控对象) 系统的稳定性. 该研究进一步揭示了 FDI 攻击的协同性, 对保护 CPS 安全和防御网络攻击提供了重要参考.

关键词 信息物理融合系统, 网络空间安全, 假数据注入攻击, 卡尔曼滤波

引用格式 彭大天, 董建敏, 蔡忠闽, 张长青, 彭勤科. 假数据注入攻击下信息物理融合系统的稳定性研究. 自动化学报, 2019, **45**(1): 196−205

DOI 10.16383/j.aas.2018.c180331

On the Stability of Cyber-physical Systems Under False Data Injection Attacks

PENG Da-Tian[1,2] DONG Jian-Min[1,2] CAI Zhong-Min[1,2] ZHANG Chang-Qing[3] PENG Qin-Ke[1,2]

Abstract Due to the stealthiness behavior, false data injection (FDI) attacks severely threaten the security of cyber-physical systems (CPS). From the attackers′ perspective, this paper mainly studies how FDI attacks impact the stability of CPS. First, we give the FDI attack model where the false control and measurement data are injected into the forward and feedback channels, respectively. Then, we propose an FDI effectiveness model to quantify the attack impact on the state estimation and measurement residue of CPS. On this basis, we design a coordination strategy associated with attack vector and further derive the theoretical attack conditions to manipulate the stability of CPS, which are related to the stability of attack matrix H and system matrix A and the selected moment of time parameter k_a. Finally, numerical simulations indicate that FDI attacks can effectively manipulate the stability of CPS including two classes of controlled plants: stable and unstable. This study further reveals the coordination behavior of FDI attacks, which provides important reference for securing the CPS and defending cyber attacks.

Key words Cyber-physical systems (CPS), cyberspace security, false data injection (FDI) attacks, Kalman filter

Citation Peng Da-Tian, Dong Jian-Min, Cai Zhong-Min, Zhang Chang-Qing, Peng Qin-Ke. On the stability of cyber-physical systems under false data injection attacks. *Acta Automatica Sinica*, 2019, **45**(1): 196−205

信息物理融合系统 (Cyber-physical systems, CPS) 是物理动态过程 (Physical dynamics) 在网络空间 (Cyber) 中高度集成的新型化网络控制系统[1], 利用计算、通信和控制等先进技术分析信息, 并通过反馈机制实现对物理过程的实时控制, 是实现工业 4.0 和中国制造 2025 的最关键技术[2]. 典型的 CPS 有智能能源系统[3]、智联网[4]、无人系统[5]、智能核电工业系统[6] 等. CPS 基本架构由物理层 (传感器、执行器和物理对象), 有线和无线通信传输设备组成的网络层和监控层 (控制器、估计器和检测器) 组成. 信息传输网络是 CPS 最基本的网络单元, 实现信息互联互通. 然而在 CPS 安全领域, 信息传输网络的引入可能增加了物理层动态过程的安全风险[7], 充斥于网络空间的各类网络攻击易引发控制性能的下降, 进而可能造成巨大经济损失, 甚至危及人员生命安全.

2015 年 IEEE 专题[8] 研究表明 Cyber networks 和 Physical dynamics 高度融合是 CPS 智能化关键之所在, 同时强调人因 (运营人员或攻击者) 决策是 CPS 安全风险主要问题源之一. 例如,

收稿日期 2018-05-25 录用日期 2018-08-14
Manuscript received May 25, 2018; accepted August 14, 2018
国家自然科学基金 (60774086, 61772415), 国家留学基金 (201706280191, 201706280220) 资助
Supported by National Natural Science Foundation of China (60774086, 61772415) and China Scholarship Council (201706280191, 201706280220)
本文责任编委 陈积明
Recommended by Associate Editor CHEN Ji-Ming
1. 西安交通大学智能网络与网络安全教育部重点实验室 西安 710049 2. 西安交通大学电子与信息工程学院系统工程研究所 西安 710049 3. 西安交通大学电子与信息工程学院计算机科学与技术系 西安 710049
1. Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Xi′an Jiaotong University, Xi′an 710049 2. System Engineering Institute, School of Electronic and Information Engineering, Xi′an Jiaotong University, Xi′an 710049 3. Department of Computer Science and Technology, School of Electronic and Information Engineering, Xi′an Jiaotong University, Xi′an 710049

常见的分布式拒绝服务 (Distributed denial of service, DDoS) 攻击[9] 大致分为两个阶段: 1) 在不同时刻和网络拓扑节点上试图控制多个终端机; 2) 协同地发送大量数据包形成洪流至多个目标 IP 服务器, 使其频繁响应该访问请求, 用以过度消耗大量的带宽, 路由及计算资源甚至造成超负荷瘫痪, 从而使合法用户无法请求访问, 影响正常的网络服务. 为了检测和防御这类攻击[10], 综合性策略应该从以下几个方面考虑: 1) 更新增强型防火墙、安全补丁和反病毒软件; 2) 提升身份验证密码保护机制和入侵检测系统的性能; 3) 开发有效的系统访问请求响应 (DNS) 协议用以识别恶意非法用户; 4) 提出新型的路由资源调度, 将访问请求分流管理, 降低恶意攻击的风险; 5) 构建安全可靠的网络架构用以故障恢复和数据备份. 在工业网络攻击中存在一类数据完整性攻击: 重放攻击 (Replay attacks), 攻击者通过网络非法接入, 侦听到系统处于稳定运行阶段的传感测量数据, 然后重复发送该数据包形成虚拟的测量输出至控制中心, 致使检测系统误认为该测量输出是合理满意的, 同时攻击者可以任意注入恶意控制命令去影响受控对象的控制性能. 为了检测该类攻击, 利用 Neyman Pearson 检测器和半正定规划设计了一种物理水印检测数据的完整性[11]. 从攻防博弈的角度, 试图找到控制性能和检测能力的纳什平衡, 在可接受的控制性能前提下, 尽可能缓解重放攻击对系统的不利影响[12]. 2017 年 5 月爆发了一种席卷全球的新型勒索病毒[13]: WannaCry, 它利用针对 Windows 操作系统的网络层协议的漏洞或垃圾邮件、恶意广告链接等方式感染主机, 并自动执行 RSA-2048 加密程序迅速锁定存储器上的文件系统, 这类加密机制一般用户无法 (暴力) 破解, 必须付费才能获取解密口令. 随后 Microsoft 发布了 Critical 补丁用以修补易为该病毒所利用的漏洞. 更新杀毒软件、备份数据、关闭闲置的网络端口和删除垃圾邮件等常用措施能有效预防此类病毒[14].

近年来, 欺骗攻击由于隐蔽性和极具破坏性特点成为网络攻击研究中的热点. 例如著名网络战标志性事件 StuxNet[15], 攻击者利用监督控制和数据采集系统 (Supervisory control and data acquisition, SCADA) 的网络漏洞注入蠕虫病毒, 目的是感染核浓缩工作的离心机控制系统, 使得监测结果显示离心机正常工作, 而实际已严重失控. 作为欺骗攻击的典型代表, 假数据注入 (False data injection, FDI) 攻击利用信息传输网络漏洞, 在传感器或执行器上注入攻击者精心设计的假数据, 改变传感器测量值或控制器控制指令, 确保绕过检测器的坏值检测同时影响物理动态过程的控制性能. 因此, FDI 攻击对 CPS 的安全威胁几乎难以避免[16−17]. 其最早由 Liu 等学者提出[18], 在电力系统中通过篡改传感数据改变状态估计, 同时能避免被基于最小二乘的坏值检测器发现, 文献 [18] 给出 FDI 攻击的定义并阐述了隐蔽性和难以检测性等特点. 然而该工作仅考虑了传感器端的测量数据篡改, 并未考虑在执行器端注入控制假数据. 本文给出的攻击模型同时考虑了执行器端和传感器端的假数据注入.

在此基础上, 人们开始研究 FDI 攻击在网络控制系统的应用. Kwon 等基于线性时不变 (Linear time invariant, LTI) 系统提出 Deception attack model[19] 并制定了三类混合攻击策略研究对系统状态估计值和测量残差的影响. Covert misappropriation attacks[20] 主要构建基于反馈控制的 Covert agent, 与原系统控制器对抗, 实现对控制系统影响并保持 FDI 攻击的隐蔽性. Stealthy integrity attacks[21] 利用最大扰动状态可达集来衡量对控制系统的影响程度. 基于输出跟踪网络控制系统, Pang 等[22] 提出了 Two-channel FDI attacks 影响系统输出跟踪误差. Coding scheme[23] 用于检测 FDI 攻击, 使用状态估计误差和残差两个指标来量化 FDI 攻击对系统性能的影响. 这类研究给出了基于控制系统的 FDI 攻击研究框架, 但是往往假设受控对象为不稳定系统 (即系统矩阵至少存在一个不稳定特征值), 攻击者利用执行器端注入的控制假数据抵消掉原系统控制器的稳定调节功能, 相当于受控对象实际上没有收到任何有效的控制指令, 致使系统失稳, 同时在传感器端对测量值进行篡改达到攻击隐蔽性的目的. 当受控对象为稳定系统时, 这类攻击策略将失去效力. 本文将主要针对稳定的受控对象设计 FDI 攻击协同策略. 文献 [22] 尽管考虑了受控对象的稳定性, 但研究的输出跟踪闭环控制不具有一般性. 本文以最基本控制单元为研究对象, 所提出的 FDI 攻击协同策略能够拓展到其他控制系统.

最优攻击策略也是许多学者专注研究的热点问题. 例如 Stealthy control signal attacks[24] 提出了两目标优化模型: 降低 FDI 攻击可检测性和增大控制代价. FDI 攻击通过篡改电表数据, 触发安全约束经济性调度子系统的负载再分布机制, 从而引起发电功率再分配[25], 使攻击者获取非法经济利益[26]. 在此基础上, 我们已有的工作研究了 FDI 攻击能够操纵区域边际价格以帮助电力生产方利益联盟实现非法收入[27]. 这类研究往往利用 FDI 攻击协同性特点增大系统控制成本或获取经济利益, 并未考虑 FDI 攻击协同策略如何影响系统的稳定性. 本文将研究 FDI 攻击操纵系统稳定性的协同攻击条件.

FDI 攻击的实现往往假设攻击者完全掌握控制系统模型参数、网络拓扑结构 (及 Jacobian 矩阵)、最优控制器和检测器检测方法等信息. 这类假设看

似勉强, 但是在网络攻击领域具有一定的合理性, 因为攻击者可以利用网络漏洞侦听足够的系统运行数据, 进行信息综合, 从而获取与系统模型相关的先验知识. Liang 等回顾了 FDI 攻击面向电力系统的研究现状[28], 并以 2015 年乌克兰停电事件[29] 为例揭示了 FDI 攻击假设条件的合理性和攻击协同性特点.

本文着眼于最基本控制单元并给出最具一般性的 FDI 攻击模型, 直观地揭示 FDI 攻击如何影响系统内部状态和外部测量. 提出的 FDI 攻击协同策略进一步拓展了已有工作的研究内容, 适用于所有稳定和不稳定的受控对象, 能够有效分析其对 CPS 稳定性影响. 本文主要贡献包括:

1) 从攻击者角度, 构建 FDI 攻击效力模型来量化 FDI 攻击对 CPS 性能的影响程度, 并提出一个攻击向量协同策略, 理论上分析了控制假数据和测量假数据对系统状态估计误差偏差量和残差偏差量的影响.

2) 基于给定的攻击向量协同策略, 理论上分析控制假数据和测量假数据对系统测量输出和实际输出的影响, 并给出操纵 CPS 稳定性的攻击条件.

3) 通过对稳定和不稳定 LTI 系统的数值仿真, 验证 FDI 攻击效力模型和协同策略的有效性.

本文组织架构如下: 第 1 节介绍正常 (无攻击) 情况下, CPS 控制系统的基本组成单元; 第 2 节介绍 FDI 攻击效力模型; 第 3 节提出 FDI 攻击协同策略并分析了对 CPS 稳定性的影响; 第 4 节进行数值仿真及结果分析; 第 5 节给出本文工作的结论.

1 CPS 控制系统

CPS 控制系统最基本的控制单元包括受控对象、状态估计器、坏值检测器和控制器. 考虑受控对象为随机离散 LTI 系统, 其动力学模型如下:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + \omega_k \\ y_k = Cx_k + \xi_k \end{cases} \tag{1}$$

其中, $x_k \in \mathbf{R}^n$, $u_k \in \mathbf{R}^p$ 和 $y_k \in \mathbf{R}^q$ 分别是系统状态变量、控制输入和测量输出; A, B 和 C 是系统矩阵、输入矩阵和输出矩阵; ω_k 和 ξ_k 分别表示过程噪声和测量噪声. 通常假设系统噪声是相互独立且均值为零的高斯白噪声, 即 $\omega_k \sim \mathrm{N}(0, \Omega)$ 和 $\xi_k \sim \mathrm{N}(0, \Xi)$, Ω 和 Ξ 分别是其协方差矩阵. (A, B) 能控和 (A, C) 能观. 假设系统初始状态 $x_0 = 0$.

卡尔曼滤波常用作 CPS 状态估计器和坏值检测器, 其基本方程如下:

$$\begin{cases} P_{0|0} = 0,\ \hat{x}_{0|0} = 0 \\ \hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \\ P_{k|k-1} = AP_{k-1|k-1}A^{\mathrm{T}} + \Omega \\ z_k = y_k - C\hat{x}_{k|k-1} \\ K_k = P_{k|k-1}C^{\mathrm{T}}(CP_{k|k-1}C^{\mathrm{T}} + \Xi)^{-1} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k z_k \\ P_{k|k} = P_{k|k-1} - K_k C P_{k|k-1} \end{cases} \tag{2}$$

其中, 第 k 时刻 $\hat{x}_{k|k}$ 表示状态估计, $P_{k|k}$ 是估计误差协方差矩阵, K_k 是卡尔曼增益矩阵, z_k 是测量残差.

当卡尔曼滤波达到稳态时, $P = \lim_{k\to\infty} P_{k|k}$ 和 $K = \lim_{k\to\infty} K_k$. z_k 服从高斯分布, 其均值为 0, 协方差矩阵 $V = CPC^{\mathrm{T}} + \Xi$, 即 $z_k \sim \mathrm{N}(0, V)$.

定义检测指数 $g_k = z_k^{\mathrm{T}} P^{-1} z_k$, 其满足自由度为 q 的 χ^2 分布. 坏值检测要求当 $g_k \geq \tau$ 时触发故障报警, 该时刻对应的系统状态值和测量值被认定为坏值, 将弃用或删除. 否则, 认定其为正常值, 将通过坏值检测器, 其中常数 τ 表示检测阈值.

定义状态估计误差 $e_k = x_k - \hat{x}_{k|k}$, 其中初始条件 $e_0 = 0$, 其与测量残差满足动力学模型, 如下:

$$\begin{cases} e_{k+1} = (A - KCA)e_k + (I - KC)\omega_k - K\xi_k \\ z_{k+1} = C(Ae_k + \omega_k) + \xi_k \end{cases} \tag{3}$$

正常情况下 (无恶意攻击), $\mathrm{E}\{\omega_k\} = 0$ 和 $\mathrm{E}\{\xi_k\} = 0$, 且卡尔曼滤波最优增益 K 保证矩阵 $(A - KCA)$ 是稳定矩阵. 当系统处于稳态时, 状态估计误差的期望值趋于 0, 即 $\lim_{k\to\infty} \mathrm{E}\{e_k\} = 0$, 同时, z_k 均能通过坏值检测器的检测.

基于卡尔曼滤波的最优状态估计, 线性二次型调节器 (Linear-quadratic regulator, LQR) 常作为 CPS 的最优状态反馈控制器. 在无限时域内最小化性能指标 $\min J = \min \sum_{k=0}^{\infty} (\hat{x}_k^{\mathrm{T}} Q \hat{x}_k + u_k^{\mathrm{T}} R u_k)$, 其中 Q 和 R 是正定加权矩阵, 最优控制序列如下:

$$u_k = r_c - L\hat{x}_k \tag{4}$$

其中, 常数 r_c 代表参考输入, L 表示控制增益. $L = (R + B^{\mathrm{T}}SB)^{-1}B^{\mathrm{T}}SA$, S 是离散代数 Riccati 方程: $S = A^{\mathrm{T}}SA - A^{\mathrm{T}}SB(R + B^{\mathrm{T}}SB)^{-1}B^{\mathrm{T}}SA + Q$ 的唯一正定解.

综上, 在卡尔曼滤波和 LQR 共同作用下, 无论 CPS 控制对象是否稳定 (矩阵 A 的所有特征值都处于单位圆内, 则认为控制对象稳定, 否则, 认为不稳定), 其闭环控制系统总能达到稳定, 即矩阵 $(A - KCA)$ 和 $(A - BL)$ 都是稳定矩阵.

2　FDI 攻击效力模型

假定 CPS 信息传输网络存在安全漏洞, 允许攻击者侦听和篡改系统运行数据并掌握 CPS 的网络拓扑和模型参数. 例如系统内部人员在闭环系统的前向和反馈通道恶意注入特定的假数据到执行器和传感器. 基于模型 (1), FDI 攻击模型有如下形式:

$$\begin{cases} x'_{k+1} = Ax'_k + B(u'_k + u^a_k) + \omega_k \\ y'_k = Cx'_k + y^a_k + \xi_k \end{cases} \tag{5}$$

其中, $u^a_k \in \mathbf{R}^p$ 和 $y^a_k \in \mathbf{R}^q$ 分别表示攻击者在 k 时刻注入到执行器和传感器的控制假数据和测量假数据. 由于假数据的注入, 使 x'_k, u'_k 和 y'_k 不同于原系统 (1) 的状态变量、控制输入和测量输出, 并且卡尔曼滤波基本方程 (2) 中系统状态估计方程和残差方程也发生变化, 如下:

$$\begin{cases} \hat{x}'_{k|k-1} = A\hat{x}'_{k-1|k-1} + Bu'_{k-1} \\ z'_k = y'_k - C\hat{x}'_{k|k-1} \\ \hat{x}'_{k|k} = \hat{x}'_{k|k-1} + K_k z'_k \end{cases} \tag{6}$$

同时, 最优控制序列变成 $u'_k = r_c - L\hat{x}'_k$, 而系统状态估计误差变为 $e'_k = x'_k - \hat{x}'_{k|k}$, 其中初始条件 $e'_0 = 0$, 相应地, 动力学模型 (3) 有如下变化形式:

$$\begin{cases} e'_{k+1} = (A - KCA)e'_k + (B - KCB)u^a_k - \\ \qquad Ky^a_{k+1} + (I - KC)\omega_k - K\xi_k \\ z'_{k+1} = C(Ae'_k + Bu^a_k + \omega_k) + y^a_{k+1} + \xi_k \end{cases} \tag{7}$$

为了有效量化 FDI 攻击对 CPS 影响, 常用系统状态估计误差偏差量 ($\Delta e_k = e'_k - e_k$) 和残差偏差量 ($\Delta z_k = z'_k - z_k$) 来表示. 因此, 给定式 (3) 和式 (7), 得到如下动力学模型:

$$\begin{cases} \Delta e_{k+1} = (A - KCA)\Delta e_k - Ky^a_{k+1} + \\ \qquad (B - KCB)u^a_k \\ \Delta z_{k+1} = CA\Delta e_k + y^a_{k+1} + CBu^a_k \end{cases} \tag{8}$$

本文称该动力学模型为 FDI 攻击效力模型. 显然, Δe_k 和 Δz_k 是关于攻击向量 $[\boldsymbol{u}_a^{\mathrm{T}}, \boldsymbol{y}_a^{\mathrm{T}}]^{\mathrm{T}}$ 的函数, $\boldsymbol{u}_a = [u^a_0, \cdots, u^a_k]^{\mathrm{T}}$, $\boldsymbol{y}_a = [y^a_0, \cdots, y^a_k]^{\mathrm{T}}$. 通过设计不同攻击向量, 可以得到多种协同策略, 对系统造成不同程度的破坏.

3　FDI 攻击协同策略

考虑 CPS 的执行器和传感器网络拓扑结构, 攻击者能够协同设计控制假数据和测量假数据, 注入到 CPS 引发系统内部状态偏离原工作点, 甚至可能驱使稳定的闭环系统失稳, 同时利用假数据的欺骗性和隐蔽性, 避免触发坏值检测器报警. 具体地, 对于 FDI 攻击效力模型 (8) 来说, 一旦 FDI 攻击协同策略成功实现, 随着时间 k 递增, Δe_k 将达到攻击者期望的任一工作点甚至无界而 Δz_k 有界. 前者考虑了 FDI 攻击对系统状态的稳定性影响, 后者确保注入的假数据能够通过坏值检测器的检测.

引理 1. 给定 FDI 攻击模型 (5), 若攻击向量的协同策略满足

$$u^a_{k+1} = \begin{cases} u^{\max}_a, & k > k_a \text{ 且 } u^a_{k_a} \geq 0 \\ Hu^a_k, & 0 \leq k \leq k_a \\ -u^{\max}_a, & k > k_a \text{ 且 } u^a_{k_a} < 0 \end{cases} \tag{9}$$

$$y^a_{k+1} = -C(A\Delta e_k + Bu^a_k) + z_k \tag{10}$$

其中, k_a 表示 u^a_k 处于上界 $u^{\max}_a$ 或下界 $-u^{\max}_a$ 的时刻值, 向量 $u^{\max}_a \in \mathbf{R}^p$ 中任意元素都是正实数 ($u^a_{k_a} = u^{\max}_a$ 或 $-u^{\max}_a$), 矩阵 $H \in \mathbf{R}^{p\times p}$ 是不稳定矩阵, 即至少存在一个特征值处在单位圆之外. 对于 FDI 攻击效力模型 (8), 可得

$$\lim_{k\to\infty} \|\Delta e_k\|_2 = \begin{cases} e^{\max}_a, & \text{若 } A \text{ 稳定} \\ \infty, & \text{否则} \end{cases} \tag{11}$$

$$\lim_{k\to\infty} \|\Delta z_k\|_2 < M \tag{12}$$

其中, $e^{\max}_a$ 和 M 是正实数, $\|\cdot\|_2$ 表示 Euclid 范数.

证明. 将式 (10) 代入式 (8) 中的 Δz_{k+1}, 可知 $\Delta z_{k+1} = z_k$. 已知 $z_k \sim \mathrm{N}(0, V)$, 则

$$\lim_{k\to\infty} \|\Delta z_k\|_2 = [\mathrm{tr}(V)]^{\frac{1}{2}}$$

易知式 (12) 成立.

将式 (10) 代入式 (8) 中的 Δe_{k+1}, 可得

$$y^a_k = -C\Delta e_k + z_k \tag{13}$$

$$\Delta e_{k+1} = A\Delta e_k + Bu^a_k. \tag{14}$$

联立式 (9), (13) 和 (14), 写出如下增广系统:

$$\begin{cases} f_{k+1} = \Lambda f_k \\ y^a_k = \Psi f_k + z_k \end{cases} \tag{15}$$

其中, $f_k = \begin{bmatrix} \Delta e_k \\ u^a_k \end{bmatrix}$, $\Lambda = \begin{bmatrix} A & B \\ 0_{p\times n} & H \end{bmatrix}$, $\Psi = \begin{bmatrix} -C & 0_{q\times p} \end{bmatrix}$.

由于矩阵 H 是不稳定的, 则系统矩阵 Λ 至少存在一个特征值处在单位圆之外, 该增广系统的内部状态 Δe_k, u^a_k 和外部输出 y^a_k 都会随时间 k 递增而

发散, 趋于不稳定状态; 直到 $k=k_a$, u_k^a 将收敛于其边界值. 这时, 随时间 k 递增, 当原受控对象 (1) 的系统矩阵 A 是稳定矩阵时, Δe_k 和 y_k^a 将趋于稳定, 分别收敛于 $e_a^{\max}$ 和 $-y_a^{\max}$ ($y_a^{\max}$ 是正实数); 当 A 不稳定时, Δe_k 和 y_k^a 将继续发散, 趋于无穷大. 加之, (A,B) 能控和 (A,C) 能观, 易知式 (11) 成立. □

注 1. 上述引理, 除得到式 (11) 和式 (12) 外, 还可得到

$$\lim_{k\to\infty} y_k^a=\begin{cases}-y_a^{\max}, & \text{若 } A \text{ 稳定}\\ -\infty, & \text{否则}\end{cases} \tag{16}$$

注 2. 对于式 (13), 由于初始条件 $\Delta e_0=0$, 可进一步得到 y_k^a 的形式如下:

$$\begin{aligned} y_k^a &= z_k - C(A\Delta e_{k-1}+Bu_{k-1}^a)=\cdots=\\ &\quad z_k - C(A\Delta e_0+\sum_{i=0}^{k-1}A^{k-1-i}Bu_i^a)=\\ &\quad z_k-\sum_{i=0}^{k-1}CA^{k-1-i}Bu_i^a \end{aligned} \tag{17}$$

由式 (16) 和式 (17) 易知, y_k^a 的收敛性由矩阵 A 的稳定性和 u_k^a 的收敛性决定.

注 3. 对于增广系统 (15), 有如下讨论:

1) 系统矩阵 A 直接反映了原受控对象 (1) 的稳定特性. 当 A 不稳定时, 随着时间 k 递增, 只要有界 $u_k^a\neq 0$, FDI 攻击效力模型总能得到 Δe_k 无界且 Δz_k 有界. 攻击者利用原受控对象不稳定特性达到 FDI 攻击的目的, 从一定程度上, 此类攻击策略相对容易实现. 许多工作[19, 21−23] 已涉及类似结论. 本文给出的引理不仅归纳了已有工作的特定情形 (即矩阵 A 不稳定), 还主分析了矩阵 A 稳定的情形下, 攻击者如何设计协同攻击策略以影响 CPS 的稳定性.

2) 当控制假数据序列 u_k^a, $k\in[0,k_a]$ 呈发散状态时, 由于矩阵 H 不稳定, 无论 A 是否稳定, FDI 攻击都将有效作用于 CPS 使之无法收敛, 其中矩阵 H 的特征值大小代表 FDI 攻击的效力, 决定 Δe_k 变化快慢及剧烈程度; 当 u_k^a, $k\in(k_a,\infty)$ 处于边界上, 不稳定矩阵 H 将不起作用, 这时, 测量假数据序列 y_k^a 收敛的充分条件是矩阵 A 稳定, 否则, y_k^a 继续呈发散状态 (16). 可见, 时间参数 k_a 选取的时机直接决定了发散序列 u_k^a 的长度及边界值 $u_a^{\max}$ 的大小, 进而影响 $e_a^{\max}$ 和 $y_a^{\max}$ 的取值. 这里, $e_a^{\max}$ 的大小直接反映了原系统内部状态估计在 FDI 攻击前后的偏差程度 (见式 (8)). 因此, 攻击者可通过任意选取时间参数 k_a 改变原系统内部状态达到攻击者期望的任一工作点, 并最终决定了 FDI 攻击协同策略对 CPS 的攻击效力. 该结论拓展了已有研究工作.

引理 1 给出的攻击向量协同策略实际上分析了 FDI 攻击对 CPS 内部状态和用于坏值检测的残差的影响. 下面进一步分析由于 FDI 攻击引起的系统内部状态变量的变化对 CPS 稳定性的影响.

首先我们定义: FDI 攻击模型 (5) 中 y_k' 是 FDI 攻击下系统的测量输出, $y_k^r=Cx_k'+\xi_k$ 表示系统的实际输出. 也就是说, $y_k^r=y_k'-y_k^a$ 是测量假数据注入之前的输出值, y_k' 是测量假数据注入之后的输出值. 对于闭环控制系统, 假设常数 r_o 表示系统测量输出参考跟踪值, 则跟踪误差为 $\epsilon_k=r_o-y_k'$.

定理 1. 给定 FDI 攻击模型 (5), 若攻击向量的协同策略满足式 (9) 和式 (10), 则系统的测量输出和实际输出形式分别为

$$\lim_{k\to\infty}\mathrm{E}\{y_k'\}\le M_1 \tag{18}$$

$$\lim_{k\to\infty}\mathrm{E}\{y_k^r\}=\begin{cases}\bar{y}_a^{\max}, & \text{若 } A \text{ 稳定}\\ \infty, & \text{否则}\end{cases} \tag{19}$$

其中, M_1 表示正实数.

证明. 根据式 (6), 可得

$$\begin{cases}\hat{x}_{k+1}'=A\hat{x}_k'+Bu_k'+Kz_{k+1}'\\ y_{k+1}'=CA\hat{x}_k'+CBu_k'+z_{k+1}'\end{cases}$$

定义 δ 表示增量计算符号, 如 $\hat{x}_k^\delta=\hat{x}_k'-\hat{x}_{k-1}'$, $y_k^\delta=y_k'-y_{k-1}'$, $z_k^\delta=z_k'-z_{k-1}'$ 和 $u_k^\delta=u_k'-u_{k-1}'=-L\hat{x}_k^\delta$. 上式可写成

$$\begin{cases}\hat{x}_{k+1}^\delta=(A-BL)\hat{x}_k^\delta+Kz_{k+1}^\delta\\ y_{k+1}^\delta=C(A-BL)\hat{x}_k^\delta+z_{k+1}^\delta\end{cases}$$

由于 $z_k\sim\mathrm{N}(0,V)$, 且 $\lim_{k\to\infty}\|\Delta z_k\|_2=0$, 易知 $\mathrm{E}\{z_k^\delta\}=0$. 对于上面的增量模型, z_k^δ 相当于系统噪声. 它的期望形式可写成

$$\begin{cases}\mathrm{E}\{\hat{x}_{k+1}^\delta\}=(A-BL)\mathrm{E}\{\hat{x}_k^\delta\}\\ \mathrm{E}\{y_{k+1}^\delta\}=C(A-BL)\mathrm{E}\{\hat{x}_k^\delta\}\end{cases} \tag{20}$$

测量输出的跟踪误差 $\epsilon_k=r_o-y_k'$ 可重写为 $\epsilon_{k+1}=r_o-y_k'-(y_{k+1}'-y_k')=\epsilon_k-y_{k+1}^\delta$, 其期望形式为

$$\mathrm{E}\{\epsilon_{k+1}\}=\mathrm{E}\{\epsilon_k\}-C(A-BL)\mathrm{E}\{\hat{x}_k^\delta\} \tag{21}$$

联立式 (20) 和式 (21) 得到一个新的增广系统. 由于矩阵 $A-BL$ 是稳定矩阵, 该增广系统是稳定系统, 随着时间递增, 最终必然能达到新的平衡态. 因此, 可得

$$\lim_{k\to\infty}\mathrm{E}\{\hat{x}_k^{\delta}\}=0$$
$$\lim_{k\to\infty}\mathrm{E}\{y_k^{\delta}\}=0$$
$$\lim_{k\to\infty}\mathrm{E}\{\epsilon_k\}=0$$

进一步, 可得 $\lim_{k\to\infty}\mathrm{E}\{y_k'\}=r_o\le M_2$, 即式 (18) 成立. 同时可得 $\lim_{k\to\infty}\mathrm{E}\{y_k^r\}=r_o-\lim_{k\to\infty}\mathrm{E}\{y_k^a\}$. 由引理 1 可知 y_k^a 满足式 (16), 所以式 (19) 成立, 且 $\bar{y}_a^{\max}\approx r_o+y_a^{\max}$. □

注 4. FDI 攻击的隐蔽性体现在: 随着时间 k 递增, 测量输出 y_k' 可达渐近稳定, 以欺骗坏值检测器, 而系统实际输出 y_k^r 是否收敛依赖于原受控对象的系统矩阵 A 的稳定性. 当 A 不稳定时, y_k^r 将无法收敛[22]; 当 A 稳定时, y_k^r, $k\in[0,k_a]$ 呈发散状态, 直到 $k=k_a$, y_k^r 收敛于攻击者期望的任一工作点. FDI 攻击协同策略对 CPS 攻击效力的结果是影响系统实际输出的收敛性而不改变系统测量输出的稳定性.

注 5. FDI 攻击的协同性表现在: 从执行器端注入的控制假数据用来扰动系统状态, 进而影响受控对象实际输出的收敛性, 而从传感器端注入的测量假数据用来消除系统状态改变所引发的不利影响, 避免被坏值检测器发现. 这大致分为三个步骤, 如图 1 所示.

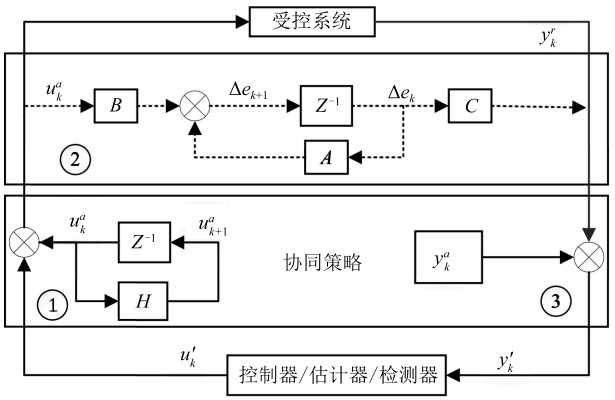


图 1　FDI 攻击协同策略架构

Fig. 1　Framework of coordination strategy under FDI attacks

步骤 1. 设计 u_k^a. 根据式 (9), 攻击者主要设计攻击矩阵 H 和时间参数 k_a. 其中, H 的维度由执行器网络拓扑结构决定, H 和 k_a 的数值应依据攻击者的攻击意图而定. 特别地, 初始值 $u_0^a\neq 0$.

步骤 2. 评估攻击效力. 根据 FDI 攻击效力模型 (8), 量化系统状态估计误差偏差量和残差偏差量, 使之满足式 (11) 和式 (12).

步骤 3. 设计 y_k^a. 根据式 (10), 其维度由传感器网络拓扑结构决定, 其数值满足式 (17). 这体现了攻击者具有对防御者 (坏值检测器) 的欺骗能力和隐藏特性, 是 FDI 攻击成功与否关键所在.

4　数值仿真及结果分析

本节通过数值仿真验证本文提出的 FDI 攻击协同策略的有效性. 考虑两输入两输出的随机离散 LTI 系统作为 CPS 控制系统的受控对象.

4.1　稳定矩阵 A

给定稳定矩阵 A^1, 参考输入 r_c^1 和其他系统参数 B, C, Ω, Ξ

$$A^1=\begin{bmatrix}0.21 & 0.37 & 0.04\\ 0.61 & 0.58 & 0.03\\ 0.63 & 0.45 & 0.31\end{bmatrix},\ r_c^1=\begin{bmatrix}0.1\\ 0.1\end{bmatrix}$$
$$B=\begin{bmatrix}0.71 & 0.25\\ 0.98 & 0.88\\ 0.27 & 0.74\end{bmatrix},\ \Omega=\begin{bmatrix}10^{-6} & 0 & 0\\ 0 & 2\times10^{-6} & 0\\ 0 & 0 & 10^{-6}\end{bmatrix}$$
$$A^2=\begin{bmatrix}0.23 & 0.67 & 0.56\\ 0.42 & 0.94 & 0.12\\ 0.31 & 0.34 & 0.17\end{bmatrix},\ r_c^2=\begin{bmatrix}1\\ 1\end{bmatrix}$$
$$C=\begin{bmatrix}0.14 & 0.89 & 0.30\\ 0.01 & 0.20 & 0.66\end{bmatrix},\ \Xi=\begin{bmatrix}10^{-4} & 0\\ 0 & 2\times10^{-4}\end{bmatrix}\tag{22}$$

计算出系统信噪比约为 $[0.69,\ 0.28]^{\mathrm{T}}$, 同时得到卡尔曼增益 K 和控制增益 L

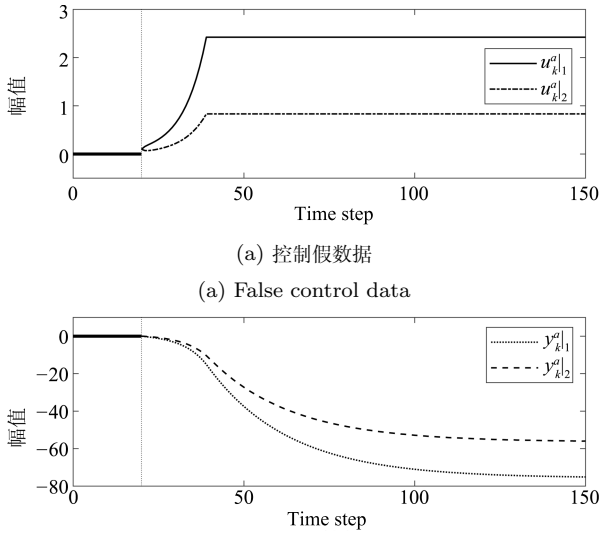
$$K=\begin{bmatrix}0.59 & -0.30\\ 1.19 & -0.59\\ 0.59 & -0.30\end{bmatrix},\ L=\begin{bmatrix}0.27 & 0.35\\ 0.26 & 0.31\\ 0.04 & 0.08\end{bmatrix}^{\mathrm{T}}\tag{23}$$

在 $k=0,1,\cdots,19$ 时, 闭环 LTI 系统在卡尔曼滤波和 LQR 共同作用下处于稳定运行状态; 在 $k=20$ 时, 攻击者开始发动 FDI 攻击. 此时, 根据协同攻击策略 (9), 设计控制假数据的初始值、攻击矩阵和时间参数如下:

$$u_{20}^a=\begin{bmatrix}0.1\\ 0.1\end{bmatrix},\ H=\begin{bmatrix}0.95 & 0.61\\ 0.23 & 0.49\end{bmatrix},\ k_a=40\tag{24}$$

图 2 是 FDI 攻击对稳定 LTI 系统的协同策略的控制假数据和测量假数据的变化情况. 由于 $k=0,1,\cdots,19$ 时刻没有发动任何攻击, 其值均为 0. 随着时间 $k\ge 20$ 递增, u_k^a 不断增加直至 $k_a=40$ 时, 可确定 u_k^a 的上界, 收敛于 $u_a^{\max}=[2.42,\ 0.83]^{\mathrm{T}}$. 同时, y_k^a 不断减小直至收敛于 $-y_a^{\max}=[-75.13,\ -55.99]^{\mathrm{T}}$, 结果符合式 (16), 其中下标 1 和下标 2 分

别表示系统的两路输入或输出. 从能量平衡角度, 在执行器端注入的控制假数使系统能量增加, 在传感器端注入的测量假数据抵消注入的能量, 当 u_k^a 收敛于 $u_a^{\max}$, y_k^a 收敛于 $-y_a^{\max}$, 最终使系统总能量保持一定平衡, 体现了 FDI 攻击的隐蔽性.



(a) 控制假数据

(a) False control data

(b) 测量假数据

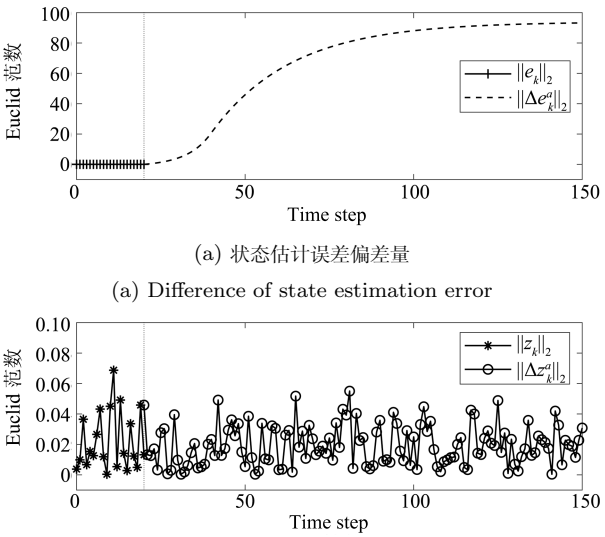
(b) False measurement data

图 2 FDI 攻击对稳定 LTI 系统的协同策略

Fig. 2 Coordination strategy under FDI attacks against stable LTI system

图 3 是 FDI 攻击协同策略对稳定 LTI 系统的攻击效力评估结果. 由于 $k = 0, 1, \cdots, 19$ 时刻没有发动任何攻击, 分别得到原闭环 LTI 系统的状态估计误差值的 Euclid 范数 $\|e_k\|_2$ 和残差值的 Euclid 范数 $\|z_k\|_2$. 随着时间 $k \geq 20$ 递增, $\|\Delta e_k^a\|_2$ 不断增加, 最终收敛于 $e_a^{\max} = 93.19$, 同时, $\|\Delta z_k^a\|_2$ 仍然保持原系统残差值的幅值波动水平, 低于检测阈值 0.08. 可见, FDI 攻击的真实意图体现在系统内部状态收敛性完全依赖于 u_k^a 的收敛性, 而 FDI 攻击隐蔽性体现在外部测量残差保持着与攻击前相同的收敛性, 这与式 (11) 和式 (12) 的结论一致, 验证了 FDI 攻击效力模型 (8) 的有效性.

图 4 是 FDI 攻击对稳定 LTI 系统测量输出和实际输出的影响. 由于 $k = 0, 1, \cdots, 19$ 时刻没有发动任何攻击, 分别得到原闭环 LTI 系统输出 y_k, 由于存在过程噪声和测量噪声, 测量输出值存在一定波动. 随着时间 $k \geq 20$ 递增, FDI 攻击真实意图体现在系统实际输出 y_k^r 不断增加直至达到上界, 收敛于 $\bar{y}_a^{\max} = [75.36, 56.15]^{\mathrm{T}}$, 而 FDI 攻击隐蔽性体现在系统测量输出 y_k' 保持攻击前系统输出 y_k 几乎相同波动水平, 收敛于 $[0.24, 0.17]^{\mathrm{T}}$, 这与命题的结果式 (18) 和式 (19) 一致. 这也揭示了 FDI 攻击能够操纵系统实际输出的稳定性, 而保持稳定的测量输出能欺骗 CPS 运营中心的坏值检测器.



(a) 状态估计误差偏差量

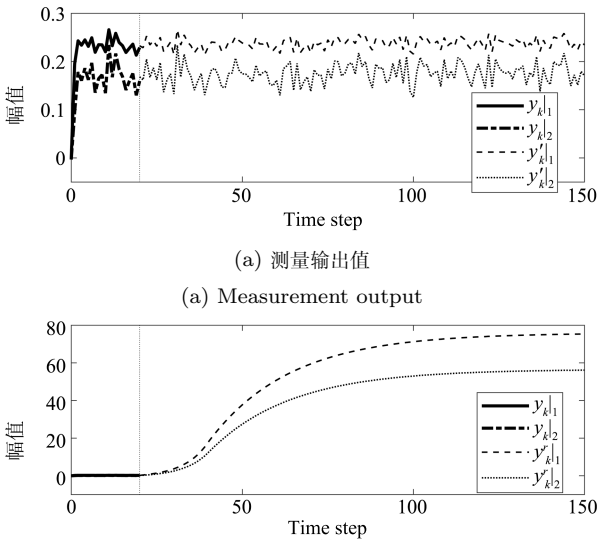
(a) Difference of state estimation error

(b) 残差偏差量

(b) Difference of the residue

图 3 FDI 攻击对稳定 LTI 系统的攻击效力

Fig. 3 FDI attack effectiveness on stable LTI system



(a) 测量输出值

(a) Measurement output

(b) 实际输出值

(b) Real output

图 4 FDI 攻击下稳定 LTI 系统输出

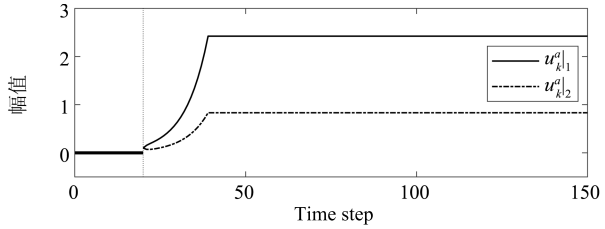
Fig. 4 Outputs of stable LTI system under FDI attacks

4.2 不稳定矩阵 A

给定不稳定矩阵 A^2, 参考输入 r_c^2 和其他系统参数不变, 如式 (22) 所示, 计算出系统信噪比约为 $[6.92, 2.75]^{\mathrm{T}}$. 针对不稳定的 LTI 系统在卡尔曼滤波和 LQR 共同作用下处于稳定运行状态, 其卡尔曼增益 K (同式 (23) 中 K) 和控制增益 L 为

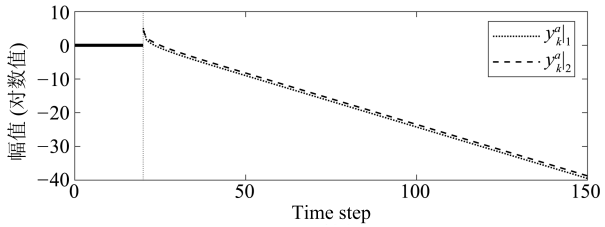
$$L=\begin{bmatrix}0.19 & 0.40 & 0.11\\ 0.22 & 0.42 & 0.12\end{bmatrix} \tag{25}$$

该案例中, 攻击者发动 FDI 攻击时间和协同策略与第 4.1 节的案例相同. 图 5 是 FDI 攻击对不稳定 LTI 系统的协同策略控制假数据和测量假数据的变化情况. 与图 2 比较, 随着时间 $k\geq 20$ 递增, 即使控制假数据 u_k^a 收敛于 $u_a^{\max}$, 测量假数据 y_k^a 也无法收敛, 这与注 1 和注 2 结论一致, 原因在于不稳定矩阵 A 影响着式 (17) 的计算结果. 这里取对数值 $-\lg(|y_k^a|)$ 表示趋于无穷的 y_k^a.



(a) 控制假数据

(a) False control data



(b) 测量假数据

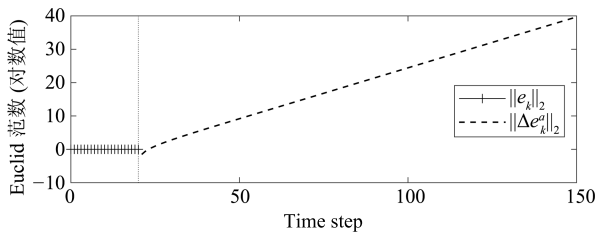
(b) False measurement data

图 5　FDI 攻击对不稳定 LTI 系统的协同策略

Fig. 5　Coordination strategy under FDI attacks against unstable LTI system

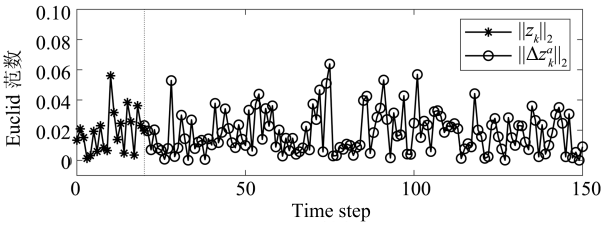
图 6 是 FDI 攻击对不稳定 LTI 系统的攻击效力评估结果. 与图 3 比较, 随着时间 $k\geq 20$ 递增, 即使控制假数据 u_k^a 收敛于 $u_a^{\max}$, FDI 攻击真实意图体现在状态估计误差偏差量的 Euclid 范数 $\|\Delta e_k^a\|_2$ 不断增加, 也无法收敛, 而 FDI 攻击的隐蔽性体现在残差偏差量的 Euclid 范数 $\|\Delta z_k^a\|_2$ 仍然保持攻击前残差值的幅值波动水平, 低于检测阈值 0.05. 这与注 3 结论一致. 这里取对数值 $\lg(\|\Delta e_k^a\|_2)$ 表示趋于无穷的 $\|\Delta e_k^a\|_2$.

图 7 是 FDI 攻击对不稳定 LTI 系统测量输出和实际输出的影响. 与图 4 比较, 随着时间 $k\geq 20$ 递增, 即使控制假数据 u_k^a 收敛于 $u_a^{\max}$, FDI 攻击真实意图体现在系统实际输出 y_k^r 不断增加, 也无法收敛, 而 FDI 攻击隐藏性体现在系统测量输出 y_k' 保持攻击前系统输出 y_k 几乎相同波动水平, 收敛于 $[2.73,\ 1.23]^{\mathrm{T}}$, 这符合式 (18) 和式 (19) 以及注 4 的结论. 这里取对数值 $\lg(|y_k^r|)$ 方便表示趋于无穷的 y_k^r.



(a) 状态估计误差偏差量

(a) Difference of state estimation error

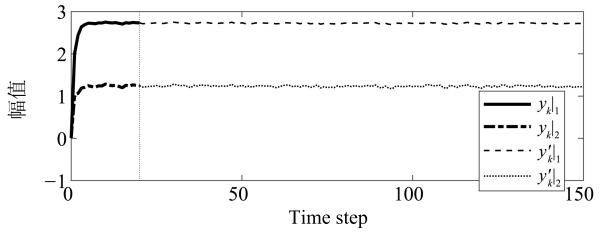


(b) 残差偏差量

(b) Difference of the residue

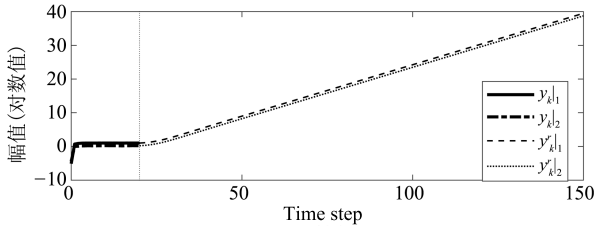
图 6　FDI 攻击对不稳定 LTI 系统的攻击效力

Fig. 6　FDI attack effectiveness on unstable LTI system



(a) 测量输出值

(a) Measurement output



(b) 实际输出值

(b) Real output

图 7　FDI 攻击下不稳定 LTI 系统输出

Fig. 7　Outputs of unstable LTI system under FDI attacks

考虑不同的信噪比和稳定性的受控对象 LTI 系统, 两个仿真案例的数值结果表明文中所提出的 FDI 攻击协同策略能够操纵 CPS 稳定性, 关键在于设计的攻击矩阵 H 和系统矩阵 A 的稳定性和时间参数 k_a 的选取时机.

5 结论

本文从攻击者的角度阐述了 FDI 攻击对 CPS 稳定性的影响. 基于 CPS 的基本控制单元, 给出了具有一般性 FDI 攻击模型, 允许攻击者同时从传感器和执行器网络注入假数据. 从系统状态估计误差和残差的角度提出了 FDI 攻击效力模型, 用来量化对 CPS 控制性能的影响, 进而设计了 FDI 攻击向量的协同策略. 从理论上分析了系统状态估计误差偏差量和残差偏差量变化情况, 证明了 FDI 攻击能影响系统实际输出的收敛性而不改变系统测量输出的稳定性. 数值仿真结果表明了 FDI 攻击可任意操纵 CPS 的稳定性, 同时能有效躲避坏值检测器.

本文揭示了 FDI 攻击协同性特点, 能够在网络拓扑的不同注入点 (执行器终端和传感器终端) 实现欺骗攻击. 总的来说, 与其他网络攻击的不同之处在于 FDI 攻击不是从物理意义上摧毁坏值检测器, 而是着眼于网络信息层, 误导坏值检测器做出错误检测结果而不自知; 同时面向不同模型参数的控制系统, FDI 攻击通过调控攻击矩阵 H 和时间参数 k_a, 灵活设计假数据注入值, 进而操纵该控制系统收敛于攻击者期望的任一工作点. 本文也揭示了 CPS 系统存在严重的网络安全漏洞, 威胁着各类控制管理调度系统的正常运行. 未来工作, 我们将提出攻击检测方案和防御策略, 保障 CPS 安全可靠运行.

References

1 Lee J, Bagheri B, Kao H A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 2015, **3**: 18−23

2 Mosterman P J, Zander J. Industry 4.0 as a cyber-physical system study. *Software and Systems Modeling*, 2016, **15**(1): 17−29

3 Deng Jian-Ling, Wang Fei-Yue, Chen Yao-Bin, Zhao Xiang-Yang. From Industries 4.0 to Energy 5.0: concept and framework of intelligent energy systems. *Acta Automatica Sinica*, 2015, **41**(12): 2003−2016
(邓建玲, 王飞跃, 陈耀斌, 赵向阳. 从工业 4.0 到能源 5.0: 智能能源系统的概念、内涵及体系框架. 自动化学报, 2015, **41**(12): 2003−2016)

4 Wang Fei-Yue, Zhang Jun. Internet of minds: the concept, issues and platforms. *Acta Automatica Sinica*, 2017, **43**(12): 2061−2070
(王飞跃, 张俊. 智联网: 概念、问题和平台. 自动化学报, 2017, **43**(12): 2061−2070)

5 Bai Tian-Xiang, Wang Shuai, Shen Zhen, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel robotics and parallel unmanned systems: framework, structure, process, platform and applications. *Acta Automatica Sinica*, 2017, **43**(2): 161−175
(白天翔, 王帅, 沈震, 曹东璞, 郑南宁, 王飞跃. 平行机器人与平行无人系统: 框架、结构、过程、平台及其应用. 自动化学报, 2017, **43**(2): 161−175)

6 Wang Fei-Yue, Sun Qi, Jiang Guo-Jin, Tan Ke, Zhang Jun, Hou Jia-Chen, et al. Nuclear energy 5.0: new formation and system architecture of nuclear power industry in the new IT era. *Acta Automatica Sinica*, 2018, **44**(5): 922−934
(王飞跃, 孙奇, 江国进, 谭珂, 张俊, 侯家琛, 等. 核能 5.0: 智能时代的核电工业新形态与体系架构. 自动化学报, 2018, **44**(5): 922−934)

7 Alguliyev R, Imamverdiyev Y, Sukhostat L. Cyber-physical systems and their security issues. *Computers in Industry*, 2018, **100**: 212−223

8 Sandberg H, Amin S, Johansson K H. Cyberphysical security in networked control systems: an introduction to the issue. *IEEE Control Systems*, 2015, **35**(1): 20−23

9 Zargar S T, Joshi J, Tipper D. A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys and Tutorials*, 2013, **15**(4): 2046−2069

10 Mölsä J. Mitigating denial of service attacks: a tutorial. *Journal of Computer Security*, 2005, **13**(6): 807−837

11 Weerakkody S, Mo Y L, Sinopoli B. Detecting integrity attacks on control systems using robust physical watermarking. In: Proceedings of the 53rd IEEE Annual Conference on Decision and Control. Los Angeles, USA: IEEE, 2014. 3757−3764

12 Miao F, Pajic M, Pappas G J. Stochastic game approach for replay attack detection. In: Proceedings of the 52nd IEEE Annual Conference on Decision and Control. Florence, Italy: IEEE, 2013. 1854−1859

13 Ehrenfeld J M. WannaCry, cybersecurity and health information technology: a time to act. *Journal of Medical Systems*, 2017, **41**(7): 104

14 Mohurle S, Patil M. A brief study of wannaCry threat: ransomware attack 2017. *International Journal of Advanced Research in Computer Science*, 2017, **8**(5): 1938−1940

15 Langner R. Stuxnet: dissecting a cyberwarfare weapon. *IEEE Security and Privacy*, 2011, **9**(3): 49−51

16 McLaughlin S, Konstantinou C, Wang X Y, Davi L, Sadeghi A R, Maniatakos M, et al. The cybersecurity landscape in industrial control systems. *Proceedings of the IEEE*, 2016, **104**(5): 1039−1057

17 Khorrami F, Krishnamurthy P, Karri R. Cybersecurity for control systems: a process-aware perspective. *IEEE Design and Test*, 2016, **33**(5): 75−83

18 Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 2011, **14**(1): Article No. 13

19 Kwon C, Liu W Y, Hwang I. Security analysis for cyber-physical systems against stealthy deception attacks. In: Proceedings of the 2013 American Control Conference. Washington DC, USA: IEEE, 2013, 3344−3349

20 Smith R S. Covert misappropriation of networked control systems: presenting a feedback structure. *IEEE Control Systems*, 2015, **35**(1): 82−92

21 Mo Y L, Sinopoli B. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 2016, **61**(9): 2618−2624

22 Pang Z H, Liu G P, Zhou D H, Hou F Y, Sun D H. Two-channel false data injection attacks against output tracking control of networked systems. *IEEE Transactions on Industrial Electronics*, 2016, **63**(5): 3242−3251

23 Miao F, Zhu Q Y, Pajic M, Pappas G J. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Transactions on Control of Network Systems*, 2017, **4**(1): 106−117

24 Zhang R, Venkitasubramaniam P. Stealthy control signal attacks in linear quadratic gaussian control systems: detectability reward tradeoff. *IEEE Transactions on Information Forensics and Security*, 2017, **12**(7): 1555−1570

25 Liu X, Li Z Y, Shuai Z K, Wen Y F. Cyber attacks against the economic operation of power systems: a fast solution. *IEEE Transactions on Smart Grid*, 2017, **8**(2): 1023−1025

26 Liu C S, Zhou M, Wu J, Long C N, Kundur D. Financially motivated FDI on SCED in real-time electricity markets: attacks and mitigation. *IEEE Transactions on Smart Grid*, DOI: 10.1109/TSG.2017.2784366, 2017.

27 Peng D T, Dong J M, Jian J N, Peng Q K, Zeng B, Mao Z H. Economic-Driven FDI Attack in Electricity Market. In: International Conference on Science of Cyber Security. Beijing, China: Springer, 2018. 216−224

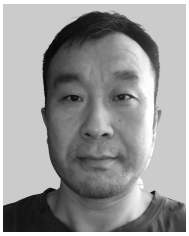
28 Liang G Q, Zhao J H, Luo F J, Weller S R, Dong Z Y. A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, 2017, **8**(4): 1630−1638

29 Liang G Q, Weller S R, Zhao J H, Luo F J, Dong Z Y. The 2015 Ukraine blackout: implications for false data injection attacks. *IEEE Transactions on Power Systems*, 2017, **32**(4): 3317−3318

彭大天 西安交通大学系统工程研究所博士研究生. 主要研究方向为机器学习和信息物理融合系统安全.
E-mail: pengdatian@stu.xjtu.edu.cn
(**PENG Da-Tian** Ph. D. candidate at the System Engineering Institute, Xi′an Jiaotong University. His research interest covers machine learning and cyber-physical systems security.)

董建敏 西安交通大学智能网络与网络安全教育部重点实验室博士研究生. 主要研究方向为人机交互, 机器学习和网络安全.
E-mail: jianmind23@stu.xjtu.edu.cn
(**DONG Jian-Min** Ph. D. candidate at the Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Xi′an Jiaotong University. Her research interest covers human-machine interaction, machine learning, and network security.)

蔡忠闽 西安交通大学智能网络与网络安全教育部重点实验室教授. 1998 年和 2004 年获得西安交通大学自动控制专业学士学位和系统工程专业博士学位. 主要研究方向为网络安全, 人机交互行为分析和机器学习.
E-mail: zmcai@sei.xjtu.edu.cn
(**CAI Zhong-Min** Professor at the Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Xi′an Jiaotong University. He received his bachelor degree in automatic control and Ph. D. degree in systems engineering from Xi′an Jiaotong University in 1998 and 2004, respectively. His research interest covers cyber security, human-machine interface behavior analysis, and machine learning.)



张长青 西安交通大学博士, 高级工程师. 主要研究方向为复杂工业过程的先进智能及模型控制技术.
E-mail: zhchqhing@sina.com
(**ZHANG Chang-Qing** Ph. D., senior engineer at Xi′an Jiaotong University. His research interest covers advanced intelligent algorithm and model-based control technology in the complex industrial process.)

彭勤科 西安交通大学系统工程研究所教授. 1983 年, 1986 年和 1990 年获得西安交通大学理学学士、系统工程专业硕士和博士学位. 主要研究方向为大数据挖掘和信息物理融合系统安全与优化. 本文通信作者.
E-mail: qkpeng@xjtu.edu.cn
(**PENG Qin-Ke** Professor at the System Engineering Institute, Xi′an Jiaotong University. He received his bachelor degree in applied mathematics, master and Ph. D. degrees in system engineering from Xi′an Jiaotong University in 1983, 1986 and 1990, respectively. His research interest covers big data mining and security and optimization of cyber-physical systems. Corresponding author of this paper.)