自动化学报
ACTA AUTOMATICA SINICA

针对信息物理系统远程状态估计的隐蔽虚假数据注入攻击

金增旺 刘茵 刁靖东 王震 孙长银 刘志强

Stealthy False Data Injection Attacks on Remote State Estimation of Cyber-physical Systems

JIN Zeng-Wang, LIU Yin, DIAO Jing-Dong, WANG Zhen, SUN Chang-Yin, LIU Zhi-Qiang

在线阅读 View online: https://doi.org/10.16383/j.aas.c240527

针对信息物理系统远程状态估计的隐蔽虚假数据注入攻击

金增旺[1, 2, 3] 刘 茵[1] 刁靖东[4] 王 震[1] 孙长银[5] 刘志强[1]

摘 要 从攻击者的角度探讨信息物理系统 (Cyber-physical system, CPS) 中隐蔽虚假数据注入 (False data injection, FDI) 攻击的最优策略. 选取 Kullback-Leibler (K-L) 散度作为攻击隐蔽性的评价指标, 设计攻击信号使得攻击保持隐蔽且最大程度地降低 CPS 远程状态估计的性能. 首先, 利用残差的统计特征计算远程状态估计误差协方差, 将 FDI 最优策略问题转化为二次约束优化问题. 其次, 在攻击隐蔽性的约束下, 运用拉格朗日乘子法及半正定规划推导出最优策略. 最后, 通过仿真实验验证所提方法与现有方法相比在隐蔽性方面具有显著优势.

关键词 信息物理系统, 虚假数据注入攻击, Kullback-Leibler 散度, 远程状态估计

引用格式 金增旺, 刘茵, 刁靖东, 王震, 孙长银, 刘志强. 针对信息物理系统远程状态估计的隐蔽虚假数据注入攻击. 自动化学报, 2025, **51**(2): 1−10

DOI 10.16383/j.aas.c240527 **CSTR** 32138.14.j.aas.c240527

Stealthy False Data Injection Attacks on Remote State Estimation of Cyber-physical Systems

JIN Zeng-Wang[1, 2, 3] LIU Yin[1] DIAO Jing-Dong[4] WANG Zhen[1] SUN Chang-Yin[5] LIU Zhi-Qiang[1]

Abstract The optimal strategy for stealthy false data injection (FDI) attacks in cyber-physical system (CPS) is explored from the attacker′s perspective. The Kullback-Leibler (K-L) divergence is selected as the evaluation index of attack stealthiness, and the attack signal is designed to keep the attack stealthy and minimize the performance of CPS remote state estimation. First, the statistical characteristics of the residuals are used to calculate the error covariance of remote state estimation, which transforms the FDI optimal strategy problem into a quadratically constrained optimization problem. Second, under the constraint of attack stealthiness, the optimal policy is derived using Lagrange multiplier method and semi-positive definite programming. Finally, simulation experiments are conducted to verify that the method proposed in this paper has significant advantages in terms of stealthiness compared with existing methods.

Key words Cyber-physical system (CPS), false data injection (FDI) attacks, Kullback-Leibler (K-L) divergence, remote state estimation

Citation Jin Zeng-Wang, Liu Yin, Diao Jing-Dong, Wang Zhen, Sun Chang-Yin, Liu Zhi-Qiang. Stealthy false data injection attacks on remote state estimation of cyber-physical systems. *Acta Automatica Sinica*, 2025, **51**(2): 1−10

近年来, 深度融合计算、通信和控制能力的可控可信可扩展的信息物理系统 (Cyber-physical system, CPS) 备受关注. 作为深度实践工业化和信息化“两化融合”的综合技术体系, CPS 正在引领新一轮的技术变革和产业革命, 赋能工业生产[1]、智能电网[2]、智能交通[3]、智能船舶[4]、辅助医疗[5] 以及国防军事[6] 等领域的智能化创新应用. 然而相比于传统控制系统, CPS 具有更广泛的攻击面, 攻击者可以针对不同网络层次发起恶意攻击从而造成巨大的破坏[7].

攻击者通过恶意软件感染、数据篡改等手段, 干扰或破坏 CPS 正常运行, 可能导致系统无法准确感知和响应物理环境变化, 进而造成物理设备故障, 甚至可能威胁生命财产安全. 近年来, 针对网络攻击下 CPS 的安全问题, 学术界进行了广泛研究.

收稿日期 2024-07-26 录用日期 2024-10-28
Manuscript received July 26, 2024; accepted October 28, 2024
国家重点研发计划 (2022YFB3104005), 国家自然科学基金 (U21B2008, U23B2039), 2022 年度太仓市基础研究计划 (TC2022JC17), 宁波市自然科学基金 (2021J046) 资助
Supported by National Key Research and Development Program of China (2022YFB3104005), National Natural Science Foundation of China (U21B2008, U23B2039), Basic Research Programs (2022) of Taicang of China (TC2022JC17), and Ningbo Natural Science Foundation of China (2021J046)
本文责任编委 李永明
Recommended by Associate Editor LI Yong-Ming
1. 西北工业大学网络空间安全学院 西安 710072 2. 西北工业大学长三角研究院 太仓 215400 3. 西北工业大学宁波研究院 宁波 315103 4. 中国空间技术研究院钱学森空间技术实验室 北京 100094 5. 安徽大学人工智能学院 合肥 230031
1. School of Cybersecurity, Northwestern Polytechnical University, Xi′an 710072 2. Yangtze River Delta Research Institute of Northwestern Polytechnical University, Taicang 215400 3. Ningbo Institute of Northwestern Polytechnical University, Ningbo 315103 4. Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing 100094 5. School of Artificial Intelligence, Anhui University, Hefei 230031

其中最常见的网络攻击为拒绝服务 (Denial of service, DoS) 攻击和虚假数据注入 (False data injection, FDI) 攻击. 在 DoS 攻击中, 攻击者通过恶意消耗通信或计算资源来破坏数据的可用性和可交换性, 令其无法成功被传送, 从而影响系统性能[8]. 现有关于 DoS 攻击的研究多聚焦在有限攻击资源下最大化降低受损系统性能. 文献 [9−11] 提出最优 DoS 攻击调度方案, 从而决定何时阻塞网络通信通道使得系统状态估计误差最大化. 与 DoS 攻击不同, FDI 攻击通过拦截和修改通信通道传输的系统数据来降低系统状态估计性能, 同时不被系统检测器发现. 尽管大多数 CPS 会配备异常状态检测器[12], 但攻击者通过精心设计的策略规避检测器, 使得 FDI 攻击对系统状态估计性能的影响更为严重. 从防御者的角度看, 抵御 FDI 攻击对于 CPS 的稳定性至关重要. 文献 [13] 设计动态输出反馈控制器, 使系统可达集始终保持在安全区域内. 文献 [14] 提出均值趋同控制方法从而保护节点初始状态信息的隐私. 文献 [15] 提出一种求和检测器来处理隐蔽 FDI 攻击, 这种检测方法不仅使用当前的入侵信息, 还整合所有历史信息以揭示潜在的威胁.

此外, 从攻击者的角度看, 如何设计 FDI 攻击也尤为重要, 因其表明 CPS 对于网络攻击的脆弱性. 文献 [16] 设计了一种隐蔽 FDI 攻击方案, 通过篡改传感器测量值并注入外部控制输入, 从而在控制回路中引起扰动, 同时绕过系统 χ^2 检测器的检测. 攻击者设计的 FDI 攻击隐蔽性也各不相同, 文献 [17] 研究了基于残差的严格隐蔽 FDI 攻击, 但系统状态估计性能的下降仅达到相对较低水平. 随后, 文献 [18] 研究了基于残差的非严格隐蔽攻击, 并通过引入 Kullback-Leibler (K-L) 散度来量化攻击的隐蔽性. 与上述文献的攻击位置不同, 文献 [19] 研究了在执行器上发动 FDI 攻击, 修改控制信号以降低系统性能. 上述攻击中针对传感器的攻击更为实际, 因为在大多数 CPS 中, 传感器读数是唯一通过通信通道传输并可能遭受攻击的数据. 此外, 以上文献主要研究能够持续无限时间的 FDI 攻击, 但在实际应用中上述攻击往往会受到能量限制而难以实现. 因此, 本文针对 CPS 传感器设计有限时间内的最优隐蔽 FDI 攻击策略.

为进一步刻画和分析 FDI 攻击对于 CPS 性能的影响, 本文主要研究针对 CPS 远程状态估计的隐蔽 FDI 攻击策略设计问题. 本文假设攻击者能够获取系统模型的完整信息, 并拦截、修改网络中传输的传感器数据. 本文以 K-L 散度作为衡量攻击策略隐蔽性的指标, 旨在设计最优攻击策略, 以最大程度降低 CPS 远程状态估计性能, 同时绕过系统异常状态检测器的检测. 本文的主要贡献总结如下:

1) 与文献 [19] 在传感器上进行无限时间的隐蔽 FDI 攻击不同, 本文的隐蔽 FDI 攻击是在有限时间范围内实施, 并选取 K-L 散度作为衡量攻击隐蔽性的指标. 这种类型的攻击更为实用, 因为在实际情况中攻击者通常会受到能量限制的影响.

2) 在本文提出的隐蔽 FDI 攻击下, 利用残差的统计特征推导受损系统估计误差协方差的递归公式, 选取有限时间内受损系统的估计误差协方差作为评价指标, 来评估在 FDI 攻击下 CPS 远程状态估计性能的退化情况. 随后, 将给定 K-L 散度阈值的最优隐蔽攻击策略设计问题转化为二次约束优化问题.

3) 本文采用拉格朗日乘子法和半正定规划法, 求解二次约束优化问题, 从而得到最优攻击策略使得系统状态估计误差最大化, 同时保证 FDI 攻击的隐蔽性.

符号说明: $\mathbf{R}_n$ 表示 n 维欧几里得空间, $\mathrm{tr}(A)$ 表示矩阵 A 的迹, $|A|$ 表示矩阵 A 的行列式, A^{T} 表示矩阵 A 的转置, I_n 表示 n 阶单位矩阵, $A>0$ ($A\geq 0$) 表示矩阵 A 为正定 (半正定) 矩阵, $x_t\sim \mathrm{N}(0,\ W)$ 表示向量 x_t 服从均值为 0、方差为 W 的高斯分布.

本文的结构安排如下: 第 1 节介绍系统模型和约束优化问题建模; 第 2 节根据攻击模型推导出最优攻击策略; 第 3 节进行实验验证; 第 4 节总结全文.

1 问题描述

本节将详细介绍所研究的 CPS 结构, 并深入阐述系统中远程状态估计器与异常状态检测器等关键组件的工作机制, 最后提出 FDI 攻击模型.

1.1 系统模型

根据文献 [20], CPS 结构如图 1 所示. 智能传感器负责采集和初步处理传感器测量数据, 随后将处理后的残差数据发送给远程状态估计器. 远程状态估计器接收到数据后, 利用卡尔曼滤波器作进一步处理, 以准确估计系统状态. 在此过程中, 攻击者在通信通道中注入恶意信号, 增加系统远程状态估

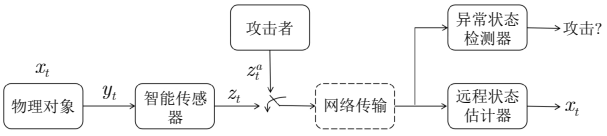


图 1 FDI 攻击下 CPS 结构图

Fig. 1 Diagram of CPS structure under FDI attack

计误差.

考虑如下离散线性时不变 (Linear and time-invariant, LTI) 系统, 其动态模型如下所示

$$\begin{cases} x_{t+1} = Ax_t + \omega_t \\ y_t = Cx_t + \nu_t \end{cases} \tag{1}$$

其中, $x_t \in \mathbf{R}_n$ 和 $y_t \in \mathbf{R}_m$ 分别是系统状态和传感器测量值, A、C 是合适维数的矩阵. 假设 ω_t 和 ν_t 是满足 $\omega_t \sim \mathrm{N}(0,\ W)$ 和 $\nu_t \sim \mathrm{N}(0,\ V)$ 的高斯白噪声, $W \geq 0$, $V > 0$. 假设 $(A,\ C)$ 是可检测的, $(A,\ \sqrt{W})$ 是可控的.

注 1. 在 CPS 中, 传感器观测过程中存在多种不同来源的热噪声、环境噪声, 由于这些噪声是独立的随机信号, 经过充分累加后其和的分布趋向于高斯分布[21], 故将其建模为高斯白噪声以简化系统.

远程状态估计器通过通信通道接收到传感器测量值, 采用卡尔曼滤波器对系统观测值进行处理, 以进行系统远程状态估计[22]

$$\begin{cases} \hat{x}_t^- = A\hat{x}_{t-1} \\ P_t^- = AP_{t-1}A^{\mathrm{T}} + W \\ K_t = P_t^- C^{\mathrm{T}} \left(CP_t^- C^{\mathrm{T}} + V\right)^{-1} \\ \hat{x}_t = \hat{x}_t^- + K_t(y_t - C\hat{x}_t^-) \\ P_t = (I_n - K_t C)P_t^- \end{cases} \tag{2}$$

其中, $\hat{x}_t^-$ 和 $\hat{x}_t$ 分别是系统状态的先验估计和后验估计, P_t^- 和 P_t 分别是先验估计协方差矩阵和后验估计协方差矩阵, K_t 是卡尔曼增益矩阵. 初始状态估计值 $\hat{x}_0 = 0$.

文献 [23] 指出在系统 $(A,\ C)$ 可检测的条件下, 卡尔曼滤波器在任何初始条件下都会以指数速度快速收敛. 因此随着系统的不断运行, 卡尔曼增益将收敛到矩阵 K. 假设卡尔曼滤波器已经处于稳定状态, 稳态下的卡尔曼增益 K 为

$$K = PC^{\mathrm{T}}(CPC^{\mathrm{T}} + V)^{-1} \tag{3}$$

其中 $P := \lim_{t\to\infty} \mathrm{E}[(x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)^{\mathrm{T}}]$ 是稳态下系统状态 x_t 的先验估计误差协方.

根据文献 [24], P 满足如下形式

$$P = APA^{\mathrm{T}} + W - APC^{\mathrm{T}}(CPC^{\mathrm{T}} + V)^{-1}CPA^{\mathrm{T}} \tag{4}$$

然后, 将系统残差 z_t 通过网络发送给远程状态估计器, 其定义为

$$z_t := y_t - C\hat{x}_t^- = C(x_t - \hat{x}_t^-) + \nu_t \tag{5}$$

系统处于稳态下的残差服从高斯分布 $z_t \sim \mathrm{N}(0,\ S)$, 即

$$S = CPC^{\mathrm{T}} + V \tag{6}$$

1.2　攻击检测模型

本节将介绍系统异常状态检测器所采用的 χ^2 检测模型. 为判断系统是否受到攻击, 在 CPS 的远程状态估计器上配备一个基于残差的 χ^2 检测器.

K-L 散度是描述两个概率分布之间距离的非负度量, 因此将其作为 FDI 攻击隐蔽性的衡量指标. χ^2 攻击检测构建的检测函数是攻击后的系统残差 z_t^a 和攻击前的系统残差 z_t 两者的 K-L 散度, 其定义如下

$$D(z_t^a \| z_t) = \int_{-\infty}^{+\infty} f_{z_t^a}(x) \lg \frac{f_{z_t^a}(x)}{f_{z_t}(x)} \mathrm{d}x \tag{7}$$

其中, $D(z_t^a \| z_t)$ 是 z_t^a 和 z_t 的 K-L 散度, f_{z_t} 是攻击前的系统残差 z_t 的概率分布函数, $f_{z_t^a}$ 是攻击后的系统残差 z_t^a 的概率分布函数. 根据 K-L 散度的定义可知 $D(z_t^a \| z_t) = 0$ 当且仅当 $f_{z_t} = f_{z_t^a}$. 此外, K-L 散度通常是不对称的, 即 $D(z_t^a \| z_t) \neq D(z_t \| z_t^a)$.

异常状态检测器将检测函数与设定的阈值 δ 进行比较, 警报触发机制如下

$$\begin{cases} D(z_t^a \| z_t) > \delta, \ \text{触发警报} \\ D(z_t^a \| z_t) \leq \delta, \ \text{不触发警报} \end{cases} \tag{8}$$

当 K-L 散度低于设定阈值时, 异常状态检测器未检测到任何攻击威胁, 系统被认为处于正常运行状态. 反之, 系统可能受到攻击, 异常状态检测器将发出警报. 特别地, 若攻击者发动的 FDI 攻击未触发异常状态检测器的警报, 则认为此次攻击是隐蔽的.

注 2. 本文选择 K-L 散度作为 FDI 的隐蔽性指标, 攻击者通过构造合理的攻击序列使得攻击前后的残差具有类似的统计特性 (即均服从零均值的高斯分布, 但方差不同), 以规避异常状态检测器的触发.

1.3　隐蔽攻击模型

本节将展示攻击者在不触发异常状态检测器的情况下, 为式 (1) 中描述的系统生成虚假数据, 使得系统状态估计值偏离实际数据.

我们首先提出如下假设:

假设 1. 攻击者对系统模型有充分了解, 即上述系统模型中的矩阵 A、C、W、V 和卡尔曼滤波增益 K 均为攻击者所知.

假设 2. 攻击者能够通过传感器与控制器之间的通信通道拦截原始数据并注入攻击者精心设计的虚假数据.

假设 3. 攻击从时刻 0 开始.

在时刻 t, 攻击策略被定义为

$$z_t^a = T_t z_t + b_t \tag{9}$$

其中, $T_t \in \mathbf{R}_{m\times m}$ 是攻击矩阵, $b_t \sim \mathrm{N}(0,\ \Phi)$ 是独立于 z_t 的高斯随机变量. 因为 $z_t \sim \mathrm{N}(0,\ S)$, 所以 z_t^a 服从高斯分布 $\mathrm{N}(0,\ S^a)$.

为实现隐蔽攻击, 攻击者需要在 K-L 散度不超过阈值的情况下, 最大限度地提高远程状态估计器误差, 将其建模为如下约束优化问题, 即:

$$\begin{aligned} &\max_{P_t^a} J = \sum_{i=0}^{N} \mathrm{tr}(P_t^a) \\ &\text{s.t.} \quad D(z_t^a||z_t) \le \delta \end{aligned} \tag{10}$$

其中 P_t^a 是攻击策略 z_t^a 下系统状态的估计误差协方差矩阵, N 是攻击的持续时间. 变量 J 被定义为 P_t^a 的迹, 从时刻 0 累加到时刻 N. 因此选用 J 的大小来表征 CPS 远程状态估计性能的退化情况.

注 3. 不同于文献 [18] 的问题 P1, 因为攻击者在攻击前知道攻击时间 N 的上限, 所以攻击的目标是最大化有限时间范围内估计误差协方差 P_t^a 的迹, 而不是在每个时刻最大化 P_t^a 的迹.

注 4. 与文献 [18], [20] 中的问题类似, 本文研究在攻击策略非严格隐蔽的情况下最优攻击的构造问题, 即 $\delta > 0$. 此外, 如果 $\delta = 0$, 这个问题已经在文献 [12], [17] 中研究.

2 系统性能分析

在本节中设计最优攻击策略以最大限度地降低系统状态估计性能, 同时保持攻击的隐蔽性.

具体地说, 为表征受损系统远程状态估计性能的下降情况, 首先推导出有限时间范围内估计误差协方差的迹. 然后, 将攻击策略的设计问题转化为 $D(z_t^a||z_t) \le \delta$ 约束下的约束优化问题. 最后, 将问题分解并分别利用拉格朗日乘子法和半正定规划方法, 从而求得最优攻击策略.

2.1 估计误差协方差演化

为分析受损系统远程状态估计性能下降情况, 本节推导受损系统远程状态估计误差协方差的递归.

在本文提出的 $z_t^a = T_t z_t + b_t$ 线性攻击下, 对系统 (1) 远程估计器的状态估计如下

$$\begin{cases} \hat{x}_t^{a-} = A\hat{x}_{t-1}^a \\ \hat{x}_t^a = \hat{x}_t^{a-} + K(y_t^a - C\hat{x}_t^{a-}) \end{cases} \tag{11}$$

其中 $\hat{x}_t^{a-}$ 和 $\hat{x}_t^a$ 为攻击后系统状态的先验估计和后验估计.

由于线性攻击策略 (9) 满足可行性约束 (10), 在攻击检测器无法检测到任何异常的情况下, 远程估计器产生的状态估计值 $\hat{x}_t^a$ 将偏离真实的系统状态, 以下定理总结在这种攻击下估计误差协方差的演化情况.

定理 1. 对于受损系统 (11), 在线性攻击 (9) 下, 远程状态估计器估计误差协方差 P_t^a 满足如下递归形式

$$\begin{aligned} P_t^a = {} & AP_{t-1}^a A^{\mathrm{T}} + W + KS^a K^{\mathrm{T}} - \\ & PC^{\mathrm{T}} T_t^{\mathrm{T}} K^{\mathrm{T}} - KT_t CP \end{aligned} \tag{12}$$

其中 S^a 是残差 z_t^a 的方差.

证明. 首先, 参考文献 [17] 中引理 4.1 的证明, 给出以下公式

$$\begin{aligned} P_t^a = {} & AP_{t-1}^a A^{\mathrm{T}} + W + \mathrm{E}[Kz_t^a z_t^{a\mathrm{T}} K^{\mathrm{T}}] - \\ & \mathrm{E}[(x_t - \hat{x}_t^{a-}) z_t^{a\mathrm{T}} K^{\mathrm{T}}] - \mathrm{E}[Kz_t^a (x_t - \hat{x}_t^{a-})^{\mathrm{T}}] \end{aligned} \tag{13}$$

其中第三项推导如下:

$$\mathrm{E}[Kz_t^a z_t^{a\mathrm{T}} K^{\mathrm{T}}] = K\mathrm{E}[z_t^a z_t^{a\mathrm{T}}] K^{\mathrm{T}} = KS^a K^{\mathrm{T}} \tag{14}$$

在计算式 (13) 的最后两项之前, 参照文献 [17] 中式 (21) 和式 (23) 的推导, 可以得到如下公式

$$\begin{aligned} x_t - \hat{x}_t^{a-} = {} & A^t (x_0 - \hat{x}_0^-) + \sum_{i=0}^{t-1} A^i \omega_{t-1-i} - \\ & \sum_{i=0}^{t-1} A^{i+1} K z_{t-1-i}^a \end{aligned} \tag{15}$$

$$\begin{aligned} z_t^a = {} & T_t C[A(I_n - KC)]^t (x_0 - \hat{x}_0^-) + \\ & \sum_{i=0}^{t-1} T_t C[A(I_n - KC)]^i \omega_{t-1-i} + b_t + T_t \nu_t - \\ & \sum_{i=0}^{t-1} T_t C[A(I_n - KC)]^i AK\nu_{t-1-i} \end{aligned} \tag{16}$$

然后, 根据式 (15) 和式 (16), 则式 (13) 的第四项如下

$$\begin{aligned} & \mathrm{E}[(x_t - \hat{x}_t^{a-}) z_t^{a\mathrm{T}} K^{\mathrm{T}}] = \\ & \quad \mathrm{E}\left[(A^t (x_0 - \hat{x}_0^-) + \sum_{i=0}^{t-1} A^i \omega_{t-1-i} - \right. \\ & \quad \left. \sum_{i=0}^{t-1} A^{i+1} K z_{t-1-i}^a) z_t^{a\mathrm{T}} K^{\mathrm{T}} \right] = \\ & \quad \mathrm{E}\left[(A^t (x_0 - \hat{x}_0^-) + \sum_{i=0}^{t-1} A^i \omega_{t-1-i}) z_t^{a\mathrm{T}} K^{\mathrm{T}} \right] - \\ & \quad \mathrm{E}\left[\sum_{i=0}^{t-1} A^{i+1} K z_{t-1-i}^a z_t^{a\mathrm{T}} K^{\mathrm{T}} \right] \overset{a}{=} \end{aligned}$$

$$\mathrm{E}\left[(A^t(x_0-\hat{x}_0^-)+\sum_{i=0}^{t-1}A^i\omega_{t-1-i})z_t^{a\mathrm{T}}K^{\mathrm{T}}\right]\overset{b}{=}$$
$$\mathrm{E}\Bigg[(A^t(x_0-\hat{x}_0^-)+\sum_{i=0}^{t-1}A^i\omega_{t-1-i})\cdot$$
$$(T_tC[A(I_n-KC)]^t(x_0-\hat{x}_0^-)+$$
$$\sum_{i=0}^{t-1}T_tC[A(I_n-KC)]^i\omega_{t-1-i})K^{\mathrm{T}}\Bigg]\overset{c}{=}$$
$$PC^{\mathrm{T}}T_t^{\mathrm{T}}K^{\mathrm{T}} \tag{17}$$

其中等式 (a) 是因为 z_t^a 是服从高斯分布的独立随机变量, 因此对所有的 $i\neq j$, 都有 $\mathrm{E}[z_i^az_j^{a\mathrm{T}}]=0$. 等式 (b) 是因为式 (16) 的后三项独立于式 (15), 而式 (16) 的后两项是零均值高斯分布变量. 由于假设攻击从稳态开始, 而且 b_t 服从零均值高斯分布, 可以得出 $\mathrm{E}[x_t-\hat{x}_t^-]=0$ 和 $\mathrm{E}[(x_t-\hat{x}_t^-)b_t]=0$. 此外, 还可以根据文献 [17] 中的式 (24) 推导出等式 (c).

同样地, 式 (13) 的最后一项被计算为

$$\mathrm{E}[Kz_t^a(x_t-\hat{x}_t^{a-})^{\mathrm{T}}]=KT_tCP \tag{18}$$

将式 (14)、式 (17)、式 (18) 替换到式 (13) 中, 则远程状态估计误差协方差矩阵的递归为

$$P_t^a=AP_{t-1}^aA^{\mathrm{T}}+W+KS^aK^{\mathrm{T}}-PC^{\mathrm{T}}T_t^{\mathrm{T}}K^{\mathrm{T}}-KT_tCP$$

□

定理 1 给出在该攻击策略下系统估计误差协方差矩阵 P_t^a 的演化情况. 由于攻击策略改变残差分布情况, 因此有必要研究估计误差协方差矩阵 P_t^a 与 z_t^a 分布之间的关系. 式 (13) 根据文献 [17] 给出, 其最后三项计算为式 (14)、式 (17) 和式 (18), 从而得到式 (19). 将 P_t^a 描述为 z_t^a 和 T_t 的函数, 这有助于找到最优的攻击策略.

然后, 基于定理 1, 可以将式 (10) 的系统状态估计性能退化变量 J 的演化情况总结为如下定理.

定理 2. 在本文所提出的攻击策略 (9) 下, 受损系统 (11) 的状态估计性能退化变量 J 为

$$J=\mathrm{tr}\Bigg(\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iWA^{i\mathrm{T}}+\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKS^aK^{\mathrm{T}}A^{i\mathrm{T}}-\sum_{j=0}^{N}\sum_{i=0}^{j}A^iPC^{\mathrm{T}}T_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}+\sum_{i=0}^{N}A^iPA^{i\mathrm{T}}-\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKT_{j-i}CPA^{i\mathrm{T}}\Bigg) \tag{19}$$

证明. 为计算受损系统状态估计性能退化变量 J, 首先根据式 (11) 及假设 3 推导出 P_0^a 为

$$\begin{aligned}P_0^a&=\mathrm{E}[(x_0-\hat{x}_0^a)(x_0-\hat{x}_0^a)^{\mathrm{T}}]=\\&\mathrm{E}[(x_0-\hat{x}_0^{a-}-Kz_0^a)(x_0-\hat{x}_0^{a-}-Kz_0^a)^{\mathrm{T}}]=\\&\mathrm{E}[(x_0-\hat{x}_0^{-}-Kz_0^a)(x_0-\hat{x}_0^{-}-Kz_0^a)^{\mathrm{T}}]=\\&P+KS^aK^{\mathrm{T}}-\mathrm{E}[(x_0-\hat{x}_0^-)z_0^{a\mathrm{T}}K^{\mathrm{T}}]-\\&\mathrm{E}[Kz_0^a(x_0-\hat{x}_0^-)^{\mathrm{T}}]\end{aligned} \tag{20}$$

其中式 (20) 的第三项为

$$\begin{aligned}&\mathrm{E}[(x_0-\hat{x}_0^-)z_0^{a\mathrm{T}}K^{\mathrm{T}}]=\\&\mathrm{E}[(x_0-\hat{x}_0^-)(T_0(C(x_0-\hat{x}_0^-)+\nu_0)+b_0)^{\mathrm{T}}K^{\mathrm{T}}]=\\&PC^{\mathrm{T}}T_0^{\mathrm{T}}K^{\mathrm{T}}\end{aligned} \tag{21}$$

同样地, 其第四项为

$$\mathrm{E}[Kz_0^a(x_0-x_0^-)^{\mathrm{T}}]=KT_0CP \tag{22}$$

结合式 (20) ~ 式 (22) 可得到

$$P_0^a=P+KS^aK^{\mathrm{T}}-PC^{\mathrm{T}}T_0^{\mathrm{T}}K^{\mathrm{T}}-KT_0CP \tag{23}$$

根据定理 1, 当 $t\geq 1$ 时, P_t^a 可以被改写成以下形式

$$\begin{aligned}P_t^a=&AP_{t-1}^aA^{\mathrm{T}}+W+KS^aK^{\mathrm{T}}-\\&PC^{\mathrm{T}}T_t^{\mathrm{T}}K^{\mathrm{T}}-KT_tCP=\\&A^tPA^{t\mathrm{T}}+\sum_{i=0}^{t-1}A^iWA^{i\mathrm{T}}+\sum_{i=0}^{t}A^iKS^aK^{\mathrm{T}}A^{i\mathrm{T}}-\\&\sum_{i=0}^{t}A^iPC^{\mathrm{T}}T_{t-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}-\sum_{i=0}^{t}A^iKT_{t-i}CPA^{i\mathrm{T}}\end{aligned} \tag{24}$$

然后基于式 (24), 受损系统状态估计性能退化变量 J 计算如下

$$\begin{aligned}J=&\mathrm{tr}(P_0^a+P_1^a+\cdots+P_N^a)=\\&\mathrm{tr}\Bigg(\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iWA^{i\mathrm{T}}+\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKS^aK^{\mathrm{T}}A^{i\mathrm{T}}-\\&\sum_{j=0}^{N}\sum_{i=0}^{j}A^iPC^{\mathrm{T}}T_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}+\sum_{i=0}^{N}A^iPA^{i\mathrm{T}}-\\&\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKT_{j-i}CPA^{i\mathrm{T}}\Bigg)\end{aligned}$$

□

为分析受损系统状态估计性能的退化情况, 首先利用式 (11) 及假设 3 得到初始估计误差协方差矩阵 P_0^a. 然后, 根据定理 1, 通过将式 (23) 代入式 (12) 得到 P_t^a, 最后通过递归式 (24) 得到系统状态估计性能退化变量 J.

2.2 最优攻击策略

基于定理 1 中得到的估计误差协方差和定理 2 中得到的状态估计性能退化变量 J, 在本节中推导约束优化问题 (10) 的最优解.

根据文献 [18], 当 $z_t \sim \mathrm{N}(0,\ S)$, $z_t^a \sim \mathrm{N}(0,\ S^a)$ 时, K-L 散度表示为

$$D(z_t^a||z_t) = \int_{-\infty}^{+\infty} f_{z_t^a}(x)\lg\frac{f_{z_t^a}(x)}{f_{z_t}(x)}\mathrm{d}x = \frac{1}{2}\left(\mathrm{tr}(S^{-1}S^a) - m + \lg\frac{|S|}{|S^a|}\right)$$

其中 m 是残差 z_t^a 的维数, 基于定理 2, 约束优化问题 (10) 等价于

$$\begin{aligned}\max_{S^a,T_t}\quad & \mathrm{tr}\left(\sum_{i=0}^{N}A^iPA^{i\mathrm{T}} + \sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iWA^{i\mathrm{T}} + \right.\\ & \sum_{j=0}^{N}\sum_{i=0}^{j}A^iKS^aK^{\mathrm{T}}A^{i\mathrm{T}} - \\ & \sum_{j=0}^{N}\sum_{i=0}^{j}A^iPC^{\mathrm{T}}T_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}} - \\ & \left.\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKT_{j-i}CPA^{i\mathrm{T}}\right)\\ \text{s.t.}\quad & \frac{1}{2}\left(\mathrm{tr}(S^{-1}S^a) - m + \lg\frac{|S|}{|S^a|}\right) \le \delta \end{aligned} \tag{25}$$

为求解约束优化问题 (25), 将其分解为任意攻击矩阵 T_t 下关于残差方差 S^a 的约束优化问题 (28) 和给定最优的残差方差 S^{a*} 下关于攻击矩阵 T_t 的约束优化问题 (34). 先利用拉格朗日乘子法求解问题 (28) 得到最优情况下残差的方差 S^{a*}, 再通过半正定规划法解出问题 (34) 的最优攻击矩阵 T_t^*.

定理 3. 对于任何给定的 $T_i(i \in [0,\ \cdots,\ N])$, 假设 z_t^{a*} 是任意给定 T_t 的约束优化问题 (25) 的最优解, 则 z_t^{a*} 服从零均值高斯分布, 并满足

$$S^{a*} = \left(S^{-1} - \frac{2}{\mu}\Sigma_S\right)^{-1} \tag{26}$$

其中

$$\Sigma_S = \sum_{j=0}^{N}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK$$

μ 是拉格朗日乘子, 满足

$$\mu > 2\min_{1\le i\le m}\lambda_i$$

以及

$$2\delta + m = \sum_{i=1}^{m}\left[\frac{1}{1-\frac{2}{\mu}\lambda_i} + \lg\left(1-\frac{2}{\mu}\lambda_i\right)\right] \tag{27}$$

其中, δ 是 K-L 散度的阈值, $\lambda_i(i \in [1,\ \cdots,\ m])$ 是 $\Sigma_S S$ 的特征值.

证明. 当系统处于稳态时, 先验估计误差协方差 P 趋于稳定. 因此, 对于任意给定的攻击矩阵 T_t, 求解约束优化问题 (25) 的最优解 S^{a*} 等价于求解以下问题

$$\begin{aligned}S^{a*} = \arg\max_{S^a}\quad & \mathrm{tr}\left(\sum_{j=0}^{N}\sum_{i=0}^{j}A^iKS^aK^{\mathrm{T}}A^{i\mathrm{T}}\right)\\ \text{s.t.}\quad & \frac{1}{2}\left(\mathrm{tr}(S^{-1}S^a) - m + \lg\frac{|S|}{|S^a|}\right) \le \delta\end{aligned} \tag{28}$$

则定义拉格朗日函数为

$$\mathcal{L}_p(S_t^a,\ \mu) = -\mathrm{tr}\left(\sum_{j=0}^{N}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iKS^a\right) + \frac{\mu}{2}\left(\mathrm{tr}(S^{-1}S^a) - m + \lg\frac{|S|}{|S^a|} - 2\delta\right) \tag{29}$$

其中, μ 是唯一标量且 $\mu \ge 0$.

描述约束优化问题解的 KKT (Karush-Kuhn-Tucker) 条件为

$$\frac{\partial\mathcal{L}_p(S^a,\ \mu)}{\partial S^a} = -\sum_{j=0}^{N}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK + \frac{\mu}{2}S^{-1} - \frac{\mu}{2}(S^a)^{-1} = 0 \tag{30}$$

$$\frac{\partial\mathcal{L}_p(S^a,\ \mu)}{\partial\mu} = \frac{1}{2}\left(\mathrm{tr}(S^{-1}S^a) - m + \lg\frac{|S|}{|S^a|} - 2\delta\right) = 0 \tag{31}$$

其中, 式 (30) 是因为 $\frac{\partial\ln|A|}{\partial A} = |A|(A^{-1})^{\mathrm{T}}/|A| = (A^{-1})^{\mathrm{T}}$, 因此 $\frac{\partial\ln|S^a|}{\partial S^a} = (S^a)^{-1}$.

根据式 (30), 可以得到

$$S^{a*} = \left(S^{-1} - \frac{2}{\mu}\Sigma_S\right)^{-1} \tag{32}$$

其中

$$\Sigma_S = \sum_{j=0}^{N}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK$$

同样可以注意到 $\Sigma_S = \mu S^{-1}/2 - \mu(S^a)^{-1}/2$, 则 μ 不能恒等于 0, 即拉格朗日乘子 $\mu > 0$. 根据式 (31) 和式 (32) 可以得到

$$2\delta + m = \mathrm{tr}(S^{-1}S^a) + \lg\frac{|S|}{|S^a|} = \sum_{i=1}^{m}\left[\frac{1}{1-\frac{2}{\mu}\lambda_i} + \lg\left(1-\frac{2}{\mu}\lambda_i\right)\right] \tag{33}$$

其中 λ_i ($i \in [1, \cdots, m]$) 为 $\Sigma_S S$ 的特征值. □

为求解约束优化问题 (28), 然后采用拉格朗日乘子法得到 KKT 条件为式 (30) 和式 (31), 从而得到最优解.

对于任意给定 $T_i (i \in [0, \cdots, N])$, 定理 3 得到 z_t^a 最优分布. 为进一步优化攻击效果, 下一步是确定能够最大化系统估计性能退化的攻击矩阵 T_t. 由于根据定理 3 已经得到残差最优分布的方差 S^{a*}, 因此约束优化问题 (25) 可分解为如下形式

$$\begin{aligned}\min_{T_t} \quad & \mathrm{tr}\Bigg(\sum_{j=0}^{N}\sum_{i=0}^{j} A^i P C^{\mathrm{T}} T_{j-i}^{\mathrm{T}} K^{\mathrm{T}} A^{i\mathrm{T}} + \\ & \sum_{j=0}^{N}\sum_{i=0}^{j} A^i K T_{j-i} C P A^{i\mathrm{T}}\Bigg) \\ \text{s.t.} \quad & \Phi = S^a - T_t S T_t^{\mathrm{T}} \geq 0 \end{aligned} \tag{34}$$

其中 Φ 是 b_t 的方差.

接着, 将第一个累加符号展开, 将优化问题 (34) 进一步改写为如下形式

$$\begin{aligned}\min_{T_t} \quad & \mathrm{tr}\Bigg(KT_N CP + \sum_{i=0}^{1} A^i K T_{N-1} C P A^{i\mathrm{T}} + \cdots + \\ & \sum_{i=0}^{N} A^i K T_0 C P A^{i\mathrm{T}}\Bigg) \\ \text{s.t.} \quad & \Phi = S^a - T_t S T_t^{\mathrm{T}} \geq 0 \end{aligned} \tag{35}$$

从上述公式中可以发现, 求解约束优化问题 (35) 等价于在每个时刻 $t \in [0, \cdots, N]$ 最小化 $\mathrm{tr}(\sum_{i=0}^{N-t} A^i \cdot K T_t C P A^{i\mathrm{T}})$, 其结果总结为如下定理.

定理 4. 利用 CVX 工具箱求解以下半正定规划问题, 得到最优攻击矩阵 T_t^* ($t \in [0, \cdots, N]$)

$$\begin{aligned} & T_t^* = \arg\min_{T_t} \ \mathrm{tr}(\Sigma_T T_t) \\ & \text{s.t.} \quad \begin{bmatrix} S^a & T_t \\ T_t^{\mathrm{T}} & S^{-1} \end{bmatrix} \geq 0, \ t = 0, \cdots, N \end{aligned} \tag{36}$$

其中

$$\Sigma_T = \sum_{i=0}^{N-t} C P A^{i\mathrm{T}} A^i P C^{\mathrm{T}} S^{-1}$$

证明. 根据线性攻击策略 (8) 和可行性约束 (9) 可得, 攻击者设计的攻击策略可行性约束必须满足以下条件

$$\Phi = S^a - T_t S T_t^{\mathrm{T}} \geq 0 \tag{37}$$

使用舒尔补将上述约束条件改写为一个线性矩阵不等式, 即

$$\begin{bmatrix} S^a & T_t \\ T_t^{\mathrm{T}} & S^{-1} \end{bmatrix} \geq 0$$

□

根据上述定理, 生成最优 FDI 攻击信号的步骤如下:

1) 根据定理 1 和定理 2, 推导出系统在有限时间内的系统状态估计性能退化变量 J.

2) 为确保每个时间步长内攻击的隐蔽性, 在 K-L 散度 $D(z_t^a || z_t) \leq \delta$ 的约束下, 将最优攻击策略设计问题转化为求解二次约束优化问题 (25), 并将其分解为约束优化问题 (28) 和约束优化问题 (34),

3) 定理 3 利用拉格朗日乘子法求解约束优化问题 (28), 得到残差 z_t^a 的最优分布. 在此基础上, 定理 4 利用半正定规划方法求解约束优化问题 (34), 得到最大程度降低 CPS 状态估计性能的最优攻击矩阵 T_t, 根据式 (9) 计算出攻击信号.

此外, 为生成最优攻击信号, 算法 1 实现步骤如下所示:

算法 1. 攻击信号生成

1) 根据定理 3 及定理 4 计算得到最优攻击矩阵 T_t

2) for each $t \in [0, \cdots, N]$

3) 利用攻击者已知的系统参数知识计算式 (9) 中 z_t^a;

4) 将攻击信号 z_t^a 注入到通信网络中;

5) end for

首先, 攻击者基于定理 3 和定理 4 求解得到最优攻击矩阵 T_t. 然后, 在每个时刻利用攻击者已知的系统参数的知识, 计算出最优攻击信号 z_t^a 并注入通信网络.

3　仿真实验

本节运用 MATLAB 仿真平台来验证本文提出的隐蔽 FDI 攻击的可行性和优越性.

为更好展示攻击效果, 假设 CPS 在 $t \in [30, 60]$ s 的时间范围内遭受 FDI 攻击, 设定攻击检测目标函数 K-L 散度的检测阈值 δ 为 1.5. 为将攻击结果与文献 [17] 和文献 [18] 进行比较, 我们考虑稳定系统的参数如下

$$A_1 = \begin{bmatrix} 0.7 & 0.2 \\ 0.05 & 0.64 \end{bmatrix}, \ C = \begin{bmatrix} 0.5 & -0.8 \\ 0 & 0.7 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix}, \ V = \begin{bmatrix} 1 & 0 \\ 0 & 0.8 \end{bmatrix}$$

以及不稳定系统的参数为 A_2、C、W、V, 其中 $A_2 = \begin{bmatrix} 1 & 0.2 \\ 0.05 & 1 \end{bmatrix}$.

首先, 考虑在线性攻击下稳定系统和不稳定系统的状态估计性能退化情况. 在 [0, 30] s 内, 系统

已进入稳态, FDI 攻击从 $t = 30$ s 开始, 到 $t = 60$ s 停止. 图 2 和图 3 分别展示在不同攻击策略下稳定系统和不稳定系统的状态估计性能退化情况. 红色虚线表示本文设计的最优攻击策略的结果 ($\delta = 1.5$), 绿色虚线和蓝色虚线分别代表文献 [17] 和文献 [18] 中设计的零均值高斯分布攻击的效果. 相较于文献 [17], 本文考虑非严格隐蔽攻击. 结果表明, 本文提出的最优攻击策略所造成的系统状态估计性能退化 J 大于文献 [17–18] 的攻击策略. 相较于文献 [18], 本文致力于在有限时间内求解最优攻击策略, 从而使得本文攻击更具有优越性.

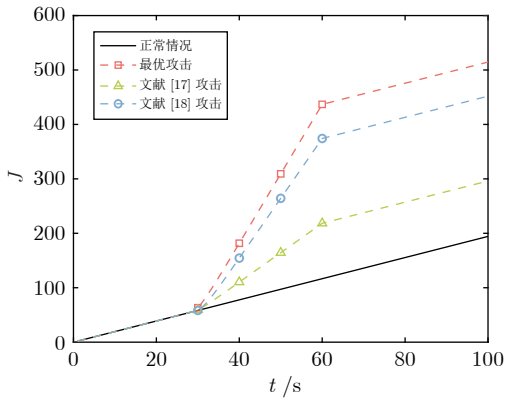


图 2 稳定系统状态估计性能的退化情况

Fig. 2 Degradation of state estimation performance of the stable system

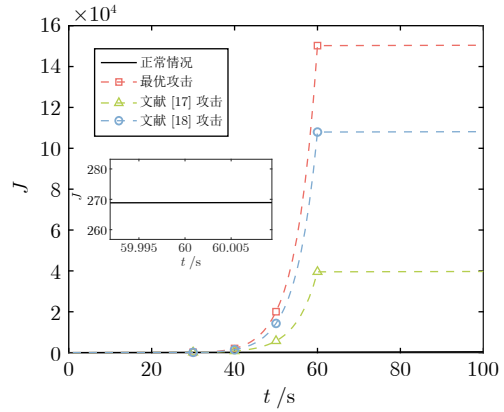


图 3 不稳定系统状态估计性能的退化情况

Fig. 3 Degradation of state estimation performance of the unstable system

然后, 考虑稳定系统的估计误差协方差迹和残差统计特征的演化情况. 在图 4 中实际值是通过 10 000 次蒙特卡洛实验计算得到, 即 $\mathrm{E}[(x_t - \hat{x}_t^-)\cdot(x_t - \hat{x}_t^-)^{\mathrm{T}}]$. 由图 4 可知, 不管系统是否受到攻击, 系统估计误差协方差迹的理论值和实际值都能很好地保持一致, 但遭受攻击后, 系统估计误差协方差大幅增加. 图 5 分析了 10 000 次蒙特卡洛模拟得到的数据, 展示在本文提出的最优攻击策略下系统残差统计特征 (S, $D(z_t^a||z_t)$) 的演化情况. 其中, 蓝线表示残差协方差 S 的迹, 绿线表示 z_t^a 和 z_t 的 K-L 散度, 红线表示设定的 K-L 散度阈值 δ. 由于蒙特卡洛模拟存在随机误差, 因此每条曲线都呈现微小波动. 从图 5 中可以观察到, 最优攻击策略使得系统估计状态偏离真实状态, 造成较大的估计误差, S 的大小在时间段 [30, 60] s 内都有所增加, 且结合图 4 可以知道, 受到 FDI 攻击后, 系统估计误差协方差也随之大幅增加, 即本文攻击能够对系统产生预定攻击效果. 值得注意的是, 在本文的攻击策略下, 受损系统实际 K-L 散度均小于等于设定的阈值 δ, 因此本文攻击策略具有良好的隐蔽性和显著的破坏性.

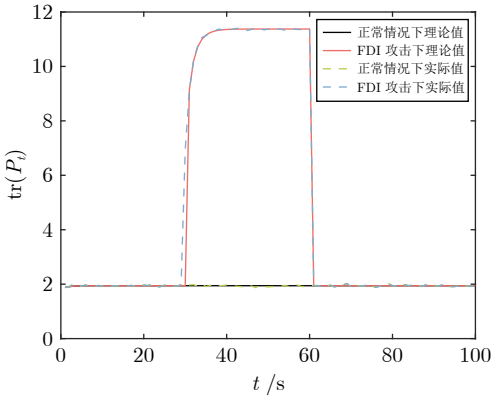


图 4 稳定系统估计误差协方差迹的演化情况

Fig. 4 Evolution of trace of estimation error covariance of the stable system

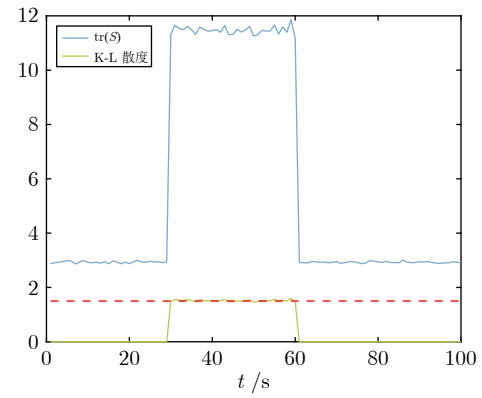


图 5 稳定系统残差统计特征的演化情况

Fig. 5 Evolution of statistical characteristics of the residuals of the stable system

最后, 考虑在不同阈值 δ 下系统状态估计性能的退化情况. 对于不同的 K-L 散度阈值, 在本文的最优线性攻击下, 状态估计性能的退化情况如图 6

所示. 结果表明, K-L 散度的阈值 δ 越大, 攻击造成的系统估计误差 J 越大, 这与攻击的隐蔽性和攻击结果之间存在基本权衡的理论是一致的.

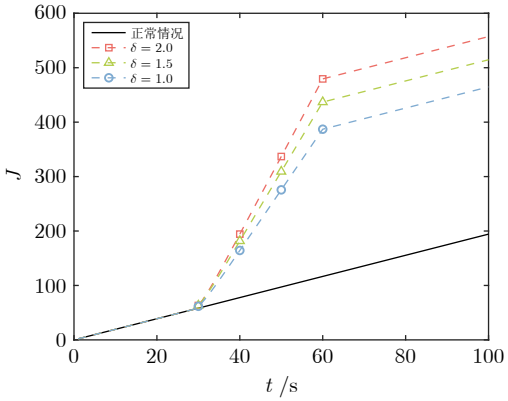


图 6　在不同阈值 δ 下稳定系统状态估计性能的退化情况

Fig. 6　Degradation of state estimation performance of the stable system at different thresholds δ

4　结论

本文研究了针对 CPS 远程状态估计的隐蔽 FDI 最优攻击策略. 首先, 基于 K-L 散度指标, 提出一种零均值高斯分布的线性攻击模型. 随后, 在此攻击框架下推导出受损系统远程状态估计误差协方差的演化, 并据此计算有限时间范围内估计误差协方差矩阵的迹, 以表征受损系统状态估计性能的下降情况. 接着, 采用拉格朗日乘子法和半正定规划法求解二次约束优化问题, 从而得到 FDI 最优攻击策略. 该策略既能最大程度地降低 CPS 远程状态估计性能, 又能确保 FDI 攻击的隐蔽性. 最后, 通过与现有工作的仿真对比, 验证本文算法具有更强大的攻击效果和更优越的攻击隐蔽性.

References

1　Wu W Q, Song C Y, Zhao J, Xu Z H. Physicsinformed gated recurrent graph attention unit network for anomaly detection in industrial cyber-physical systems. *Information Sciences*, 2023, **629**: 618−633

2　Muhammad N N, Neetesh S, Alvaro C, Santiago G, Pete B. Smart grid cyber-physical situational awareness of complex operational technology attacks: A review. *ACM Computing Surveys*, 2023, **55**(10): 1−36

3　LMarwa O, Fahd N A, Rana A, Majdi K, Mohammed A, Mohamed I A, et al. Artificial intelligence for traffic prediction and estimation in intelligent cyber-physical transportation systems. *IEEE Transactions on Consumer Electronics*, 2024, **70**(1): 1706−1715

4　Li Hong-Yang, Wei Mu-Heng, Huang Jie, Qiu Bo-Hua, Zhao Ye, Luo Wen-Cheng, et al. Survey on cyber-physical systems. *Acta Automatica Sinica*, 2019, **45**(1): 37−50
(李洪阳, 魏慕恒, 黄洁, 邱伯华, 赵晔, 骆文成, 等. 信息物理系统技术综述. 自动化学报, 2019, **45**(1): 37−50)

5　Liu W, Zhao F, Shankar A, Maple C, Peter J D, Kim B G, et al. Explainable AI for medical image analysis in medical cyber-physical systems: Enhancing transparency and trustworthiness of IoMT. *IEEE Journal of Biomedical and Health Informatics*, DOI: 10.1109/JBHI.2023.3336721

6　Mouhyemen K, Karel H, Amr M, Khaled A H, Mohammad M H. Mobile target coverage and tracking on drone-be-gone UAV cyber-physical testbed. *IEEE Systems Journal*, 2017, **12**(4): 3485−3496

7　Yang Guang-Hong, Lu An-Yang, An Li-Wei. A survey on secure state estimation of cyber-physical systems under cyber attacks. *Control and Decision*, 2023, **38**(8): 2093−2105
(杨光红, 芦安洋, 安立伟. 网络攻击下的信息物理系统安全状态估计研究综述. 控制与决策, 2023, **38**(8): 2093−2105)

8　Ding K M, Li Y Z, Daniel E Q, Subhrakanti D, Shi L. A multi-channel transmission schedule for remote state estimation under DoS attacks. *Automatica*, 2017, **78**: 194−201

9　Qin J H, Li M L, Shi L, Yu X H. Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks. *IEEE Transactions on Automatic Control*, 2017, **63**(6): 1648−1663

10　Li Y, Zhu S Y, Chen C L, Guan X P. Optimal denial-of-service attack strategy on state estimation over infinite-time horizon. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021, **68**(8): 2860−2864

11　Ye D, Song Y B. Optimal periodic DoS attack with energy harvester in cyber-physical systems. *Neurocomputing*, 2020, **390**: 69−77

12　Guo H B, Sun J, Pang Z H, Liu G P. Event-based optimal stealthy false data-injection attacks against remote state estimation systems. *IEEE Transactions on Cybernetics*, 2023, **53**(10): 6714−6724

13　Zhang Qi-Rui, Meng Si-Qi, Wang Lan-Hao, Liu Kun, Dai Wei. Secure output-feedback control for cyber-physical systems under stealthy attacks. *Acta Automatica Sinica*, 2024, **50**(7): 1001−1010
(张淇瑞, 孟思琪, 王兰豪, 刘坤, 代伟. 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**(7): 1001−1010)

14　Ying Chen-Duo, Wu Yi-Ming, Xu Ming, Zheng Ning, He Xiong-Xiong. Privacy-preserving average consensus control formultiagent systems under deception attacks. *Acta Automatica Sinica*, 2023, **49**(2): 425−436
(应晨铎, 伍益明, 徐明, 郑宁, 何熊熊. 欺骗攻击下具备隐私保护的多智能体系统均值趋同控制. 自动化学报, 2023, **49**(2): 425−436)

15　Ye D, Zhang T Y. Summation detector for false data injection attack in cyber-physical systems. *IEEE Transactions on Cybernetics*, 2020, **50**(6): 2338−2345

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

17　Guo Z Y, Shi D W, Karl H J, Shi L. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 2016, **4**(1): 4−13

18　Guo Z Y, Shi D W, Karl H J, Shi L. Worst-case stealthy innovation-based linear attack on remote state estimation. *Automatica*, 2018, **89**: 117−124

19　Rijha S, Syed A P, Syed T A. False data injection attacks on networked control systems. *Journal of Control and Decision*, 2023, **11**(4): 650−659

20　Li Y G, Yang G H. Worst-case ϵ-stealthy false data injection attacks in cyber-physical systems. *Information Sciences*, 2020, **515**: 352−364

21　Andrew C B. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 1941, **49**(1): 122−136

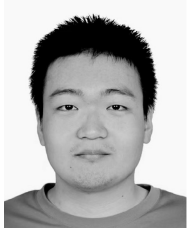
22　Shang J, Chen T W. Optimal stealthy integrity attacks on remote state estimation: The maximum utilization of historical data. *Automatica*, 2021, **128**: Artacle No. 109555

23 Anderson B D, Moore J B. Optimal filtering. *Courier Corporation*. New York: Dover Publications, 2005.

24 Pang Z H, Fu Y, Guo H B, Sun J. Analysis of stealthy false data injection attacks against networked control systems: Three case studies. *Journal of Systems Science and Complexity*, 2023, **36**(4): 1407−1422

金增旺 西北工业大学网络空间安全学院副教授. 主要研究方向为信息物理系统安全, 多智能体系统安全, 网络攻防下无人系统的安全估计与安全控制, 故障诊断与容错控制.
E-mail: jin_zengwang@nwpu.edu.cn
(**JIN Zeng-Wang** Associate professor at the School of Cybersecurity, Northwestern Polytechnical University. His research interest covers security of cyber-physical system, security of multi-agent system, secure estimation and secure control of unmanned system under cyber attack and defense, fault diagnosis and tolerant control.)

刘 茵 西北工业大学网络空间安全学院硕士研究生. 主要研究方向为信息物理系统安全.
E-mail: liuyin828@mail.nwpu.edu.cn
(**LIU Yin** Master student at the School of Cybersecurity, Northwestern Polytechnical University. Her main research interest is cyber-physical system security.)



刁靖东 中国空间技术研究院钱学森空间技术实验室博士. 主要研究方向为空间信息融合, 多源多目标跟踪和集值系统辨识. 本文通信作者.
E-mail: diaojingdong@spacechina.com
(**DIAO Jing-Dong** Ph.D. at the Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology. His research interest covers spatial information fusion, multi-source multi-target tracking, and set-valued system identification. Corresponding author of this paper.)

王 震 西北工业大学网络空间安全学院教授. 主要研究方向为人工智能, 网络空间智能对抗, 智能无人系统基础与应用.
E-mail: wzhen@nwpu.edu.cn
(**WANG Zhen** Professor at the School of Cybersecurity, Northwestern Polytechnical University. His research interest covers artificial intelligence, intelligent countermeasures in cyberspace, and foundation and application of intelligent unmanned system.)

孙长银 安徽大学人工智能学院教授. 主要研究方向为智能控制与优化, 强化学习, 神经网络.
E-mail: cysun@ahu.edu.cn
(**SUN Chang-Yin** Professor at the School of Artificial Intelligence, Anhui University. His research interest covers intelligent control and optimization, reinforcement learning, and neural networks.)

刘志强 西北工业大学网络空间安全学院教授. 主要研究方向为网络化系统, 故障诊断及应用.
E-mail: zqliu@nwpu.edu.cn
(**LIU Zhi-Qiang** Professor at the School of Cybersecurity, Northwestern Polytechnical University. His research interest covers networked system, fault diagnosis and application.)