控制与决策
Control and Decision

针对信息物理系统的自生成 ε -隐性最优欺骗攻击策略设计

单华晟, 李一刚

引用本文:
单华晟, 李一刚. 针对信息物理系统的自生成 ε -隐性最优欺骗攻击策略设计[J]. 控制与决策, 2024, 39(12): 4191-4199.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2023.1030

针对信息物理系统的自生成ϵ-隐性最优欺骗攻击策略设计

单华晟, 李一刚†

(沈阳工业大学 人工智能学院，沈阳 110870)

摘　要: 近年来,信息物理系统网络安全问题成为一大研究热点. 以攻击者角度研究攻击设计问题可有效评估系统对网络攻击的脆弱性并为设计网络保护措施提供理论依据. 鉴于此,在ϵ-隐性下研究针对信息物理系统远程状态估计的最优欺骗攻击设计问题. 首先,与需要额外滤波器和历史数据在线计算真实新息的相关结果不同,提出一种利用离线生成的攻击信号篡改传感器测量值以降低系统性能的自生成攻击模型,使攻击更易实现. 随后,推导得出该攻击下远程估计误差以量化攻击效果,并将攻击设计问题转化为多变量受限二次优化问题. 不同于相关结果的恒定均值,模型采用更具一般性的时变均值,使优化问题包含更多决策变量且相关结果中的攻击优化方法无法直接求解. 因此,利用K-L (Kullback-Leibler)散度和互信息的相关统计学性质将问题等价转化. 再结合拉格朗日乘数法和所提出的参数特征关联覆盖法得到最优攻击策略,使其在ϵ-隐性下最大化远程估计误差. 最后,通过仿真实例验证结果的有效性.

关键词: 信息物理系统；自生成攻击；远程状态估计；ϵ-隐性；K-L散度；互信息

中图分类号: TP273　　**文献标志码:** A

DOI: 10.13195/j.kzyjc.2023.1030

引用格式: 单华晟, 李一刚. 针对信息物理系统的自生成ϵ-隐性最优欺骗攻击策略设计 [J]. 控制与决策, 2024, 39(12): 4191-4199.

Optimal off-line generated ϵ-stealthy deception attack strategy design in cyber-physical system

SHAN Hua-sheng, LI Yi-gang†

(School of Artificial Intelligence，Shenyang University of Technology，Shenyang 110870，China)

Abstract: In recent years, network security of cyber-physical systems has become a hot research topic. Investigating the problem of designing attacks from the attacker's perspective can effectively evaluate the vulnerability of the system to network attacks, and provide theoretical basis for designing network protection methods. For this reason, this paper investigates the problem of designing the optimal ϵ-stealthy deception attacks against remote state estimation in cyber-physical systems. Firstly, different from the related results which require extra filters and historical data to calculate the true innovation online, this paper proposes a self-generated attack model which uses off-line generated attack signals to tamper with the sensor measurements and deteriorate the estimation performance, such that the attacks are more easily to be implemented. Subsequently, the remote estimation error under the attack is derived to quantify the attack effect, based on which, the attack design problem is transformed into a variable optimization problem. Since the model uses the more general time-varying mean, the optimization problem contains more decision variables, which cannot be solved directly by the attack optimization methods in the related results. Therefore, the problem is equivalently transformed by using the relevant statistical properties of K-L divergence and mutual information. Furthermore, by combining the Lagrange multiplier method and the optimization method with the covering by the related parameter characteristics, the optimal attack strategy is obtained to maximize the remote estimation error under the ϵ-stealthiness. Finally, simulation examples are given to verify the validity of the results.

Keywords: cyber-physical systems；self-generated attack；remote state estimation；ϵ-stealthiness；Kullback-Leibler divergence；mutual information

收稿日期: 2023-07-24；**录用日期:** 2024-03-05.

基金项目: 辽宁省自然科学基金博士启动项目 (2022-BS-178)；辽宁省教育厅基本科研项目青年项目 (JYTQN2023292).

责任编委: 俞立.

†通讯作者. E-mail: liyigang920407@163.com.

0 引 言

信息物理系统(cyber-physical system, CPS)将通信、控制、计算机等学科的相关技术深度融合[1],目前已渗透到智能电网[2-3]、智能交通[4]、远程医疗[5]、多线性感应牵引系统[6]等诸多实际领域. 此类系统将物理系统的状态信息通过无线网络传输到远程端[7]. 然而,无线网络传输过程中存在的恶意攻击、网络时延、随机丢包、有限带宽等因素会对系统性能造成严重破坏[8],因此针对CPS网络安全问题开展研究已迫在眉睫[9-11].

近年来,关于CPS网络安全问题的相关研究结果主要集中于拒绝服务攻击(denial-of-service, DoS)和欺骗攻击[12]. DoS攻击者通过阻断或干扰通信信道传输过程以降低信道传输质量,从而恶化系统性能[13-16]. 而欺骗攻击则通过拦截并篡改经由无线网络传输的数据以破坏系统性能,此类攻击可避免被检测器检测[17-20]. Liu等[21]首次将智能电网描述为静态系统并在此框架下研究欺骗攻击策略. 而后,Mo等[22]研究了针对线性动态系统传感器测量值的隐性攻击. 不仅如此,Mo等[23]还提出了重放攻击(replay data attack),此类攻击利用历史测量值替代真实测量值,使攻击更加难以检测. 尽管重放攻击隐性更高,但实施需要时间用于记录历史测量值且只对稳定系统具备隐性. 而基于新息的攻击策略无需此类假设,更易实施,因此针对此类攻击的研究结果更为详实. 具体而言,Guo等[19]研究了基于新息的最优严格隐性攻击,远程估计性能恶化仅达到较低的水平. 基于此,Guo等[20]将K-L散度引入以刻画在检测机制未知场景下攻击的隐性,并设计了最优非严格隐性攻击. 与文献[19-20]中零均值的攻击不同的是,Li等[24]设计了基于新息的具有恒定均值的最优非严格隐性攻击,使得远程估计性能恶化更为严重且仍能保证攻击隐性. 不仅如此,Guo等[25]和Li等[26]还考虑了一种攻击者通过额外滤波器获取系统边界信息的特殊情况. 与上述文献的隐性描述不同,Bai等[27]在ϵ-隐性下提出针对标量系统执行器的非严格隐性攻击. 随后,Bai等[28]和Kung等[29]将ϵ-隐性下针对执行器的攻击研究结果扩展至高阶系统.

文献[19-20, 24]中零均值攻击与恒定均值攻击是时变均值攻击的两个特例,并不具备一般性. 文献[27-29]研究了针对系统执行器的ϵ-隐性攻击,而在攻击实际CPS系统时针对其传感器测量值更为可行. 文献[20, 24, 30]基于新息的攻击需要额外滤波器和历史数据在线计算真实新息,而实际情况下系统历史传感器测量值难以获得,且对攻击者在线计算能力要求较为严苛. 文献[31]提出了一种可以完全消除攻击对检测残差影响的自生成攻击策略,该攻击策略并未优化攻击效果且仅对不稳定系统有效. 为解决上述问题,本文研究了仅需在可接受程度下牺牲一定的隐性便可进一步提升攻击效果的针对传感器测量值的时变均值自生成ϵ-隐性攻击策略设计问题.

综上所述,本文主要贡献如下:

1) 与文献[20, 24, 30]中需要额外滤波器和历史数据在线计算真实新息不同,本文提出一种自生成ϵ-隐性攻击模型. 该模型通过离线生成的攻击信号篡改传感器测量值以最大化远程估计误差,降低对攻击者在线计算能力要求,使攻击更易实现.

2) 与文献[19-20, 24]中零均值或恒定均值不同,本文攻击信号为更具一般性的时变均值,尽管其致使推导攻击下远程估计误差协方差等表达式因包含更多耦合项而变得更加困难,但有助于进一步提升攻击效果.

3) 与文献[19-20, 24]中最多含有两个决策变量的优化问题不同,本文优化问题包含更多决策变量,以致已有攻击优化方法无法对其直接求解. 因此,本文首先利用K-L散度和互信息的相关统计学性质将优化问题等价转化,然后结合拉格朗日乘数法和所提出的参数特征关联覆盖法得到最优攻击策略.

符号说明:$\mathbb{N}$和$\mathbb{R}$分别代表自然数集与实数集;X^{-1}、$\mathrm{tr}(X)$、X^{T}和$\det(X)$分别代表矩阵X的逆、迹、转置和行列式;$\mathrm{diag}[\cdot]$代表对角矩阵;$X>0$ ($X\geqslant 0$)代表矩阵X是(半)正定矩阵.

1 系统模型与问题描述

本文所研究的CPS如图1所示. 传感器将系统状态测量值通过无线网络传输到远程端后经由卡尔曼滤波器处理并以此估计系统状态. 攻击者试图在无线网络数据传输过程中拦截并篡改测量值以恶化系统估计性能.

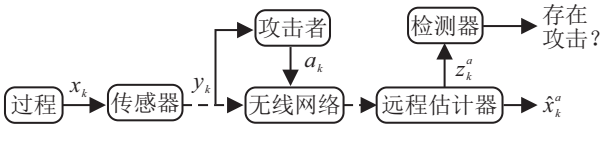

图1 CPS系统框图

1.1 系统模型

本文采用如下离散线性时不变过程刻画CPS:

$$\begin{cases} x_{k+1}=Ax_k+w_k, \\ y_k=Cx_k+v_k. \end{cases} \tag{1}$$

其中: $x_k \in \mathbb{R}^n$ 和 $y_k \in \mathbb{R}^m$ 分别为系统状态与传感器测量值, $w_k \in \mathbb{R}^n$ 和 $v_k \in \mathbb{R}^m$ 分别对应同为零均值i.i.d.高斯分布的过程噪声与传感器噪声, 协方差矩阵分别为 $Q > 0$ 和 $R > 0$. 在不失一般性的情况下, 假设 (A, C) 可测, $(A, \sqrt{Q})$ 可镇定. 初始状态 $x_0 \sim \mathcal{N}(0, \Sigma_0)$ 且与噪声 w_k 和 v_k 无关.

注1　本文中, CPS被建模为一个离散线性时不变过程, 这种建模框架具有很高的通用性, 涵盖许多实际CPS, 具备实际意义[32]. 例如, 四重水箱系统[33]、遥控直升机[34]、电力网络[32]和无人机[35]均被建模为离散线性时不变过程.

1.2　卡尔曼滤波器

为估计系统状态, 远程端采用卡尔曼滤波器处理所接收的测量值, 具体如下:

$$\begin{cases} \hat{x}_k^- = A\hat{x}_{k-1}, \\ P_k^- = AP_{k-1}A^{\mathrm{T}} + Q, \\ K_k = P_k^- C^{\mathrm{T}}(CP_k^- C^{\mathrm{T}} + R)^{-1}, \\ \hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-), \\ P_k = (I - K_k C)P_k^-. \end{cases} \tag{2}$$

其中: $\hat{x}_k^-$ 和 $\hat{x}_k$ 分别表示系统状态 x_k 的先验状态估计和后验状态估计; P_k^- 和 P_k 为相应误差协方差矩阵, 且有 $P_k^- \triangleq \mathrm{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^{\mathrm{T}}]$, $P_k \triangleq \mathrm{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^{\mathrm{T}}]$; K_k 为卡尔曼增益; 真实新息 $z_k = y_k - C\hat{x}_k^-$ 且 $z_k \sim \mathcal{N}(0, \Sigma)$.

由于 (A, C) 可测, 在任意初始条件下, 卡尔曼滤波都将呈指数级快速收敛[19-20]. 稳态误差协方差矩阵 $\bar{P}$ 和稳态卡尔曼增益 K 定义如下:

$$\bar{P} \triangleq \lim_{x \to \infty} P_k^-, \tag{3}$$

$$K \triangleq \bar{P}C^{\mathrm{T}}(C\bar{P}C^{\mathrm{T}} + R)^{-1}, \tag{4}$$

其中 $\bar{P}$ 是方程 $X = AXA^{\mathrm{T}} + Q - AXC^{\mathrm{T}}(CXC^{\mathrm{T}} + R)^{-1}CXA^{\mathrm{T}}$ 的唯一半正定解.

当系统受攻击时, 卡尔曼滤波器递归如下所示:

$$\begin{cases} y_k^a = y_k + a_k, \\ \hat{x}_k^{a-} = A\hat{x}_{k-1}^a, \\ \hat{x}_k^a = \hat{x}_k^{a-} + Kz_k^a, \\ z_k^a = y_k^a - C\hat{x}_k^{a-}, \end{cases} \tag{5}$$

其中 x_k^a、y_k^a 和 z_k^a 分别为攻击下的状态估计值、传感器测量值和新息.

1.3　攻击隐性条件

检测器根据所接收的新息序列的统计特性判断系统是否受到攻击. 为表征攻击下新息序列与真实新息序列间分布差异, 引入K-L散度以刻画在检测机制未知时攻击的隐性. K-L散度定义如下所示.

定义1　序列 x 与 y 间K-L散度定义为

$$D(x\|y) = \int_{-\infty}^{+\infty} \log \frac{f_x(\xi)}{f_y(\xi)} f_x(\xi)\mathrm{d}\xi. \tag{6}$$

其中: x 和 y 是两个随机序列; f_x 和 f_y 是对应的概率密度函数, 当且仅当 $f_x = f_y$ 时, $D(x\|y) = 0$.

在 $[0, N-1]$ 内执行 N 步攻击, N 由攻击者在攻击开始前决定. 基于定义1, 有限攻击步数为 N 的 ϵ-隐性[36]定义如下:

$$\frac{1}{N} D(z_{0:N-1}^a \| z_{0:N-1}) \leqslant \epsilon, \tag{7}$$

其中阈值 ϵ 为常数且 $\epsilon > 0$.

1.4　问题设置

在 ϵ-隐性下研究针对CPS的最优欺骗攻击设计问题. 具体而言, 攻击者试图在 $\frac{1}{N} D(z_{0:N-1}^a \| z_{0:N-1})$ 不超过阈值 ϵ 的前提下最大化 N 步内的远程状态估计误差. 问题可作如下描述:

$$\begin{aligned} &\max_{a_{0:N-1}} \ J = \sum_{k=0}^{N-1} \mathrm{tr}(P_k^a); \\ &\text{s.t. } \frac{1}{N} D(z_{0:N-1}^a \| z_{0:N-1}) \leqslant \epsilon. \end{aligned} \tag{8}$$

其中 $\epsilon > 0$ 且 $P_k^a \triangleq \mathrm{E}[(x_k - \hat{x}_k^a)(x_k - \hat{x}_k^a)^{\mathrm{T}}]$.

2　最优攻击策略设计

2.1　攻击模型设计

在不失一般性的前提下对攻击者提出如下假设.

假设1　攻击者能够获取系统信息, 即已知 A、C、Q、R 和 K[20].

假设2　攻击从卡尔曼滤波稳态开始, 即 $K_k = K, P_k = \bar{P}$[20].

与文献[20, 24, 30]中需在线计算真实新息不同, 本文攻击模型利用离线生成的攻击信号篡改传感器测量值以破坏系统性能, 对攻击者在线计算能力要求大幅降低, 使攻击更易实现. 具体内容如下.

定理1　针对系统(1)所提出的自生成攻击模型如下所示:

$$\begin{aligned} &a_0 = \theta_0, \\ &a_k = \sum_{i=0}^{k-1} CA[(I - KC)A]^{k-i-1} Ka_i + \theta_k, \ k \geqslant 1, \end{aligned} \tag{9}$$

其中 $\theta_k \sim \mathcal{N}(\mu_k, \Gamma)$ 且与真实新息相互独立.

证明　由式(2)和(5)可得

$$\begin{aligned} &\Delta\hat{x}_k = \hat{x}_k^a - \hat{x}_k = \\ &A\hat{x}_{k-1}^a + Kz_k^a - A\hat{x}_{k-1} - Kz_k = \end{aligned}$$

$$A\Delta\hat{x}_{k-1}+K[(y_k^a-CA\hat{x}_{k-1}^a)-$$
$$(y_k-CA\hat{x}_{k-1})]=$$
$$A\Delta\hat{x}_{k-1}-KCA\Delta\hat{x}_{k-1}+K(y_k^a-y_k)=$$
$$(I-KC)A\Delta\hat{x}_{k-1}+Ka_k. \tag{10}$$

攻击前后新息的差异为

$$\Delta z_k=z_k^a-z_k=$$
$$y_k^a-CA\hat{x}_{k-1}^a-(y_k-CA\hat{x}_{k-1})=$$
$$-CA\Delta\hat{x}_{k-1}+a_k. \tag{11}$$

由式(11)可得攻击下新息z_k^a为

$$z_k^a=z_k+\Delta z_k=z_k-CA\Delta\hat{x}_{k-1}+a_k. \tag{12}$$

设$a_k=CA\Delta\hat{x}_{k-1}+\theta_k$,其中$\theta_k\sim\mathcal{N}(\mu_k,\Gamma)$是与$z_k$相互独立的i.i.d高斯变量. 将$a_k$代入式(12)得

$$z_k^a=z_k+\theta_k, \tag{13}$$

其中$z_k^a\sim\mathcal{N}(\mu_k,\Sigma_a)$且有$\Sigma_a=\Sigma+\Gamma$.

结合式(10)和(11)可得攻击信号a_k为

$$a_k=CA\Delta\hat{x}_{k-1}+\theta_k=$$
$$CA[(I-KC)A\Delta\hat{x}_{k-2}+Ka_{k-1}]+\theta_k. \tag{14}$$

因攻击开始前存在等式$\Delta\hat{x}_{-1}=\hat{x}_{-1}^a-\hat{x}_{-1}=\hat{x}_{-1}-\hat{x}_{-1}=0$,故$k=0$时攻击信号应表示为

$$a_0=CA\Delta\hat{x}_{-1}+\theta_0=\theta_0. \tag{15}$$

最后,结合式(14)和(15)可得$k\geqslant 1$时攻击信号为

$$a_k=\sum_{i=0}^{k-1}CA[(I-KC)A]^{k-i-1}Ka_i+\theta_k. \tag{16}$$

定理1得证. □

注2 在文献[20, 24, 30]中,由于新息$z_k=y_k-C\hat{x}_k^-$中$\hat{x}_k^-$是基于$[y_0:y_{k-1}]$的x_k的先验MMSE估计,攻击者需利用额外滤波器和历史测量值计算真实新息. 然而,实际情况中系统的历史测量值难以获得,且该过程对攻击者在线计算能力要求较高. 因此,定理1提出了一种自生成攻击模型,由式(9)可知,攻击信号a_k仅与系统信息有关. 本文攻击模型被离线生成的攻击信号篡改的y_k^a替代真实测量值y_k,无需在线计算真实新息,降低对攻击者在线计算能力的要求,使攻击更易实现.

注3 下文使用$\bar{\mu}$表示序列$\mu_{0:N-1}=[\mu_0^{\mathrm{T}},\mu_1^{\mathrm{T}},\dots,\mu_{N-1}^{\mathrm{T}}]^{\mathrm{T}}$.

2.2 最优攻击设计

本节研究定理1自生成攻击模型的最优情况. 首先,推导该攻击模型下远程估计误差协方差以量化攻击效果并将攻击设计问题转化为多变量受限二次优化问题;随后利用K-L散度和互信息的相关统计学性质将现有攻击优化方法无法直接求解的含有较多决策变量的优化问题等价转化,并结合拉格朗日乘数法和所提出的参数特征关联覆盖法得到最优攻击策略.

为量化攻击效果,通过下述引理得到远程估计误差协方差P_k^a.

引理1 定理1攻击下的远程误差协方差P_k^a为

$$P_k^a=AP_{k-1}^aA^{\mathrm{T}}+Q+K\Sigma_aK^{\mathrm{T}}+K\mu_k\mu_k^{\mathrm{T}}K^{\mathrm{T}}-$$
$$\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}+\sum_{i=0}^{k-1}A^{i+1}K\mu_{k-1-i}\mu_k^{\mathrm{T}}K^{\mathrm{T}}-$$
$$KC\bar{P}+\sum_{i=0}^{k-1}K\mu_k\mu_{k-1-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(i+1)\mathrm{T}}. \tag{17}$$

证明 由文献[20]给出

$$P_k^a=$$
$$AP_{k-1}^aA^{\mathrm{T}}+Q+\mathrm{E}[Kz_k^az_k^{a\mathrm{T}}K^{\mathrm{T}}]-$$
$$\mathrm{E}[(x_k-\hat{x}_k^{a-})z_k^{a\mathrm{T}}K^{\mathrm{T}}]-\mathrm{E}[Kz_k^a(x_k-\hat{x}_k^{a-})^{\mathrm{T}}]. \tag{18}$$

注意到,由于攻击下的新息z_k^a不是零均值分布的,$\mathrm{E}[z_k^az_k^{a\mathrm{T}}]$并不是$z_k^a$的协方差,故$\mathrm{E}[Kz_k^az_k^{a\mathrm{T}}K^{\mathrm{T}}]$推导如下:

$$\mathrm{E}[Kz_k^az_k^{a\mathrm{T}}K^{\mathrm{T}}]=$$
$$\mathrm{E}[K(z_k^a-\mu_k+\mu_k)(z_k^a-\mu_k+\mu_k)^{\mathrm{T}}K^{\mathrm{T}}]\overset{\text{(a)}}{=}$$
$$K\Sigma_aK^{\mathrm{T}}+K\mu_k\mu_k^{\mathrm{T}}K^{\mathrm{T}}, \tag{19}$$

其中(a)由$\mathrm{E}[(z_k^a-\mu_k)\mu_k^{\mathrm{T}}]=\mathrm{E}[\mu_k(z_k^a-\mu_k)^{\mathrm{T}}]=0$得到.

在计算式(18)的最后两项前,给出以下方程[19]:

$$x_k-\hat{x}_k^{a-}=A^k(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}A^iw_{k-1-i}-$$
$$\sum_{i=0}^{k-1}A^{i+1}Kz_{k-1-i}^a, \tag{20}$$

$$z_k^a=z_k+\theta_k=$$
$$C[A(I-KC)]^k(x_0-\hat{x}_0^-)+v_k+$$
$$\sum_{i=0}^{k-1}C[A(I-KC)]^iw_{k-1-i}+\theta_k-$$
$$\sum_{i=0}^{k-1}C[A(I-KC)]^iAKv_{k-1-i}. \tag{21}$$

结合式(20)和(21),式(18)倒数第2项计算如下:

$$\mathrm{E}[(x_k-\hat{x}_k^{a-})z_k^{a\mathrm{T}}K^{\mathrm{T}}]=$$
$$\mathrm{E}\Big\{\Big[A^k(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}A^iw_{k-1-i}-$$
$$\sum_{i=0}^{k-1}A^{i+1}Kz_{k-1-i}^a\Big]z_k^{a\mathrm{T}}K^{\mathrm{T}}\Big\}=$$

$$\mathrm{E}\Big[\Big(A^k(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}A^i w_{k-1-i}\Big)z_k^{a\mathrm{T}}K^{\mathrm{T}}\Big]-$$
$$\mathrm{E}\Big[\sum_{i=0}^{k-1}A^{i+1}Kz_{k-1-i}^a z_k^{a\mathrm{T}}K^{\mathrm{T}}\Big]\overset{\text{(a)}}{=}$$
$$\mathrm{E}\Big[\Big(A^k(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}A^i w_{k-1-i}\Big)\Big(C[A(I-KC)]^k\times$$
$$(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}C[A(I-KC)]^i w_{k-1-i}\Big)^{\mathrm{T}}K^{\mathrm{T}}\Big]-$$
$$\mathrm{E}\Big[\sum_{i=0}^{k-1}A^{i+1}Kz_{k-1-i}^a z_k^{a\mathrm{T}}K^{\mathrm{T}}\Big]\overset{\text{(b)}}{=}$$
$$\mathrm{E}\Big[\Big(A^k(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}A^i w_{k-1-i}\Big)\Big(C[A(I-KC)]^k\times$$
$$(x_0-\hat{x}_0^-)+\sum_{i=0}^{k-1}C[A(I-KC)]^i w_{k-1-i}\Big)^{\mathrm{T}}K^{\mathrm{T}}\Big]-$$
$$\sum_{i=0}^{k-1}A^{i+1}K\mu_{k-1-i}\mu^{\mathrm{T}}K^{\mathrm{T}}\overset{\text{(c)}}{=}$$
$$\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}-\sum_{i=0}^{k-1}A^{i+1}K\mu_{k-1-i}\mu_k^{\mathrm{T}}K^{\mathrm{T}}. \tag{22}$$

其中: (a)是由于v_k为零均值高斯分布且与$(x_0-\hat{x}_0^-)$、w_k相互独立; (b)是由于不同时刻的z_k^a相互独立, 即$\mathrm{E}[z_i^a z_j^{a\mathrm{T}}]=\mu_i\mu_j^{\mathrm{T}}\ (i\neq j)$; (c)的推导类似文献[19, 24]中式(24).

与式(22)类似, 式(18)最后一项如下所示:

$$\mathrm{E}[Kz_k^a(x_k-\hat{x}_k^{a-})^{\mathrm{T}}]=$$
$$K\bar{C}P-\sum_{i=0}^{k-1}K\mu_k\mu_{k-1-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(i+1)\mathrm{T}}. \tag{23}$$

最后, 将式(19)、(22)和(23)代入(18)可得到定理1攻击下的远程误差协方差P_k^a. □

注4　与文献[19-20, 24]攻击下新息z_k^a为零均值或恒定均值不同, 本文时变均值的z_k^a致使P_k^a的推导因包含更多耦合项而更为困难. 因此, 本文利用攻击下新息的相关统计学特性与卡尔曼滤波器稳态误差协方差矩阵的性质推导得到P_k^a的表达式.

随后, 基于引理1, 优化目标J由如下定理给出.

定理2　定理1攻击下的优化目标J表示如下:

$$J=\sum_{k=0}^{N-1}\mathrm{tr}(P_k^a)=$$
$$\mathrm{tr}\Big(\sum_{i=0}^{N-1}A^i\bar{P}A^{i\mathrm{T}}+\sum_{j=0}^{N-2}\sum_{i=0}^{j}A^iQA^{i\mathrm{T}}-$$
$$\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^i\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}-\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iK\bar{C}PA^{i\mathrm{T}}+$$
$$\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iK\Sigma_aK^{\mathrm{T}}A^{i\mathrm{T}}+$$
$$\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iK\mu_{j-i}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}+$$
$$\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}A^{t-j+i+1}K\mu_{j-i}\mu_{j+1}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j)\mathrm{T}}+$$
$$\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}A^{t-j}K\mu_{j+1}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j+i+1)\mathrm{T}}\Big), \tag{24}$$

其中J表示0到k时刻P_k^a的迹之和.

证明　假设攻击从$k=0$时刻开始. 类似于文献[23], 在计算性能指标J前, 首先推导$k=0$时的误差协方差P_0^a. 类似式(18)、(19)和(22), P_0^a如下所示:

$$P_0^a=$$
$$AP_{-1}^aA^{\mathrm{T}}+Q+\mathrm{E}[Kz_0^az_0^{a\mathrm{T}}K^{\mathrm{T}}]-$$
$$\mathrm{E}[(x_0-\hat{x}_0^-)z_0^{a\mathrm{T}}K^{\mathrm{T}})]-\mathrm{E}[Kz_0^a(x_0-\hat{x}_0^-)^{\mathrm{T}}]=$$
$$\bar{P}+K\Sigma_aK^{\mathrm{T}}+K\mu_0\mu_0^{\mathrm{T}}K^{\mathrm{T}}-$$
$$\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}-KC\bar{P}. \tag{25}$$

将引理1中P_k^a递归至如下形式:

$$P_k^a=$$
$$A^k\bar{P}A^{k\mathrm{T}}+\sum_{i=0}^{k-1}A^iQA^{i\mathrm{T}}+\sum_{i=0}^{k}A^iK\Sigma_aK^{\mathrm{T}}A^{i\mathrm{T}}-$$
$$\sum_{i=0}^{k}A^i\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}-\sum_{i=0}^{k}A^iKC\bar{P}A^{i\mathrm{T}}+$$
$$\sum_{i=0}^{k}A^iK\mu_{k-i}\mu_{k-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}+$$
$$\sum_{j=0}^{k-1}\sum_{i=0}^{j}A^{k-j+i}K\mu_{j-i}\mu_{j+1}^{\mathrm{T}}K^{\mathrm{T}}A^{(k-j-1)\mathrm{T}}+$$
$$\sum_{j=0}^{k-1}\sum_{i=0}^{j}A^{k-j-1}K\mu_{j+1}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(k-j+i)\mathrm{T}}. \tag{26}$$

根据式(25)和(26)可得优化目标J. □

基于定理2, 将优化问题(8)重写为

$$\max_{\Sigma_a,\mu_{0:N-1}}\mathrm{tr}\Big(\Phi+\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iK\Sigma_aK^{\mathrm{T}}A^{i\mathrm{T}}+$$
$$\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iK\mu_{j-i}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}+$$
$$\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}A^{t-j+i+1}K\mu_{j-i}\mu_{j+1}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j)\mathrm{T}}+$$
$$\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}A^{t-j}K\mu_{j+1}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j+i+1)\mathrm{T}}\Big);$$

$$\text{s.t. } \frac{1}{N}D(z^a_{0:N-1}\|z_{0:N-1})\leqslant\epsilon. \tag{27}$$

其中

$$\Phi=\sum_{i=0}^{N-1}A^i\bar{P}A^{i\mathrm{T}}+\sum_{j=0}^{N-2}\sum_{i=0}^{j}A^iQA^{i\mathrm{T}}-\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^i\bar{P}C^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}-\sum_{j=0}^{N-1}\sum_{i=0}^{j}A^iKC\bar{P}A^{i\mathrm{T}}. \tag{28}$$

注5 由于式(24)中参数$\bar{P}$、A、C、Q、R、K、N均恒定,可知优化目标J仅与变量μ_k和Σ_a有关.

最优攻击策略可由下述定理得到.

定理3 z^a_k的最优协方差矩阵Σ_a和最优均值序列$\bar{\mu}$如下可得:

$$\Sigma_a=\Big(-\frac{2}{l}\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK+\Sigma^{-1}\Big)^{-1}, \tag{29}$$

$$\bar{\mu}=\alpha\xi. \tag{30}$$

其中l、ξ、α满足

$$(l,\alpha,\xi)=\arg\max_{l_i,\alpha_i,\xi_i}J\{[\Sigma_a(l_1),\bar{\mu}(\alpha_1,\xi_1)],[\Sigma_a(l_2),\bar{\mu}(\alpha_2,\xi_2)],\ldots\}, \tag{31}$$

$$\alpha_i^2\xi_i^{\mathrm{T}}\bar{\Sigma}\xi_i=N\Big(2\epsilon-\log\frac{|\Sigma|}{|\Sigma_a|}-\mathrm{tr}(\Sigma^{-1}\Sigma_a)+m\Big), \tag{32}$$

$$\bar{\Sigma}=\mathrm{diag}[\Sigma^{-1},\Sigma^{-1},\ldots,\Sigma^{-1}]\in\mathbb{R}^{Nm\times Nm}, \tag{33}$$

$$H=\begin{bmatrix}H_{11}&*&\ldots&*\\H_{21}&H_{22}&\ldots&*\\\vdots&\vdots&\ddots&\vdots\\H_{N1}&H_{N2}&\ldots&H_{NN}\end{bmatrix}, \tag{34}$$

H为对称矩阵且下三角元素满足$H_{ab}=\sum_{i=0}^{N-a}K^{\mathrm{T}}A^{i\mathrm{T}}A^{i+a-b}K$,$l_i$为矩阵$2N\bar{\Sigma}^{-1}H$的正特征值,$\xi_i$为对应的特征向量.

证明 在求解最优攻击策略前,首先利用K-L散度和互信息的相关统计学性质将优化问题(27)等价转化. 文献[28]证明,当z_k是i.i.d高斯分布且$z_k\sim\mathcal{N}(0,\Sigma)$时,存在

$$D(z^a_{0:N-1}\|z_{0:N-1})=\sum_{k=0}^{N-1}(I(z^a_{0:k-1};z^a_k)+D(z^a_k\|z_k)), \tag{35}$$

其中$I(z^a_{0:k-1};z^a_k)$表示$z^a_{0:k-1}$与z^a_k之间的互信息. 由于X与Y相互独立时,两者间互信息$I(X;Y)=0$,故优化问题(27)中约束条件等价转化为

$$\frac{1}{N}\sum_{k=0}^{N-1}D(z^a_k\|z_k)\leqslant\epsilon. \tag{36}$$

由于两高斯变量$x\sim\mathcal{N}(\mu_1,\Sigma_1)$、$y\sim\mathcal{N}(\mu_2,\Sigma_2)$间K-L散度为

$$D(x\|y)=\frac{1}{2}\Big(\log\frac{\det\Sigma_2}{\det\Sigma_1}+\mathrm{tr}(\Sigma_2^{-1}\Sigma_1)+(\mu_1-\mu_2)^{\mathrm{T}}\Sigma_2^{-1}(\mu_1-\mu_2)-m\Big),$$

其中m代表x、y维数. 可将式(36)推导为

$$\frac{1}{2}\Big(\log\frac{|\Sigma|}{|\Sigma_a|}+\mathrm{tr}(\Sigma^{-1}\Sigma_a)-m+\frac{1}{N}(\bar{\mu}^{\mathrm{T}}\bar{\Sigma}\bar{\mu})\Big)\leqslant\epsilon. \tag{37}$$

同时,基于$\mathrm{tr}(XY)=\mathrm{tr}(YX)$,将$J$等价转化为

$$\begin{aligned}J=&\ \mathrm{tr}\Big(\Phi+\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma_a\Big)+\sum_{j=0}^{N-1}\sum_{i=0}^{j}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\mu_{j-i}+\\&\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}\mu_{j+1}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j)\mathrm{T}}A^{t-j+i+1}K\mu_{j-i}+\sum_{t=0}^{N-2}\sum_{j=0}^{t}\sum_{i=0}^{j}\mu_{j-i}^{\mathrm{T}}K^{\mathrm{T}}A^{(t-j+i+1)\mathrm{T}}A^{t-j}K\mu_{j+1}=\\&\mathrm{tr}\Big(\Phi+\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma_a\Big)+\bar{\mu}^{\mathrm{T}}H\bar{\mu}.\end{aligned} \tag{38}$$

基于式(37)和(38),将优化问题(27)等价转化并改写为如下标准凸优化问题形式:

$$\begin{aligned}&\min_{\Sigma_a,\bar{\mu}}-\mathrm{tr}\Big(\Phi+\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma_a\Big)-\bar{\mu}^{\mathrm{T}}H\bar{\mu};\\&\text{s.t. }\frac{1}{2}\Big(\log\frac{|\Sigma|}{|\Sigma_a|}+\mathrm{tr}(\Sigma^{-1}\Sigma_a)-m+\frac{1}{N}(\bar{\mu}^{\mathrm{T}}\bar{\Sigma}\bar{\mu})\Big)\leqslant\epsilon.\end{aligned} \tag{39}$$

为解决上述优化问题,引入拉格朗日算子l并构造拉格朗日函数$\mathcal{L}$,有

$$\begin{aligned}\mathcal{L}(\Sigma_a,\bar{\mu},l)=&-\mathrm{tr}\Big(\Phi+\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma_a\Big)-\\&\bar{\mu}^{\mathrm{T}}H\bar{\mu}+l\cdot\frac{1}{2}\Big(\log\frac{|\Sigma|}{|\Sigma_a|}+\mathrm{tr}(\Sigma^{-1}\Sigma_a)-\\&m+\frac{1}{N}(\bar{\mu}^{\mathrm{T}}\bar{\Sigma}\bar{\mu})-2\epsilon\Big).\end{aligned} \tag{40}$$

通过Karush-Kuhn-Tucker(KKT)条件可得,优化问题的解满足如下方程:

$$\frac{\partial\mathcal{L}(\Sigma_a,\bar{\mu},l)}{\partial\Sigma_a}=-\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK+l\cdot\frac{1}{2}(\Sigma^{-1}-\Sigma_a^{-1})=0, \tag{41}$$

$$\frac{\partial \mathcal{L}(\Sigma_a,\bar{\mu},l)}{\partial \bar{\mu}} = -2H\bar{\mu} + l\cdot\frac{1}{N}\bar{\Sigma}\bar{\mu} = 0, \tag{42}$$

$$l\cdot\frac{l}{2}\Big(\log\frac{|\Sigma|}{|\Sigma_a|} + \mathrm{tr}(\Sigma^{-1}\Sigma_a) - \tag{43}$$

$$m + \frac{1}{N}(\bar{\mu}^{\mathrm{T}}\bar{\Sigma}\bar{\mu}) - 2\epsilon\Big) = 0.$$

由于$\sum\limits_{j=0}^{N-1}\sum\limits_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK = l\cdot\frac{1}{2}(\Sigma^{-1}-\Sigma_a^{-1})$不恒等于零,故$l>0$.

由式(41)和(42)推导可得

$$\Sigma_a = \Big(-\frac{2}{l}\sum_{j=0}^{N-1}\sum_{i=0}^{j}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK + \Sigma^{-1}\Big)^{-1}, \tag{44}$$

$$l\bar{\mu} = 2N\bar{\Sigma}^{-1}H\bar{\mu}. \tag{45}$$

式(45)表明,l为$2N\bar{\Sigma}^{-1}H$正特征值,$\bar{\mu}$为对应特征向量. 因此最优Σ_a和最优$\bar{\mu}$可由式(44)和等式$\bar{\mu}=\alpha\xi$解得. 其中l、ξ、α满足

$$(l,\alpha,\xi) =$$

$$\arg\max_{l_i,\alpha_i,\xi_i} J\{[\Sigma_a(l_1),\bar{\mu}(\alpha_1,\xi_1)],[\Sigma_a(l_2),\bar{\mu}(\alpha_2,\xi_2)],\ldots\},$$

$$\alpha_i^2\xi_i^{\mathrm{T}}\bar{\Sigma}\xi_i = N\Big(2\epsilon - \log\frac{|\Sigma|}{|\Sigma_a|} - \mathrm{tr}(\Sigma^{-1}\Sigma_a) + m\Big).$$

无论i取何值,总有Σ为对称正定矩阵,$K^{\mathrm{T}}A^{i\mathrm{T}}A^iK$为对称半正定矩阵,故

$$\Sigma K^{\mathrm{T}}A^{i\mathrm{T}}A^iK \sim \Sigma^{-\frac{1}{2}}\Sigma K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Sigma^{\frac{1}{2}}. \tag{46}$$

式(46)表明,$\Sigma K^{\mathrm{T}}A^{i\mathrm{T}}A^iK$的特征值总是非负的且至少有一个大于零. 基于此,由于$\bar{\Sigma}^{-1}H$第a行主对角线元素为$\sum\limits_{i=0}^{N-a}\Sigma K^{\mathrm{T}}A^{i\mathrm{T}}A^iK$,故$\bar{\Sigma}^{-1}H$的迹存在如下关系式:

$$\mathrm{tr}(\bar{\Sigma}^{-1}H) = \mathrm{tr}\Big(\sum_{j=0}^{N-1}\sum_{i=0}^{j}\Sigma K^{\mathrm{T}}A^{i\mathrm{T}}A^iK\Big) > 0. \tag{47}$$

式(47)表明$2N\bar{\Sigma}^{-1}H$必有正特征值. □

注6　互信息(mutual information)是信息论中一种重要的信息度量,用于衡量两个随机变量之间的依赖程度. 它量化了一个随机变量中所包含的另一随机变量的信息量,或在已知一个随机变量的情况下另一随机变量减小的不确定性[37].

本文利用两随机变量X与Y相互独立时两者间互信息$I(X;Y)=0$这一性质,将优化问题(8)中原始约束条件等价简化为式(36),进而实现对所转化的优化问题进行求解.

注7　由于已有攻击优化方法无法对带有多决策变量$\bar{\mu}$的优化问题(27)直接求解,首先利用K-L散度和互信息的相关统计学性质将优化问题等价转化,再结合拉格朗日乘数法和所提出的参数特征关联覆盖法得到最优攻击策略.

算法1　攻击信号a_k的生成.

step 1: 计算矩阵$\bar{\Sigma}$和H;

step 2: 根据定理3得到z_k^a的最优Σ_a和最优$\bar{\mu}$;

step 3: 计算θ_k的协方差矩阵$\Gamma=\Sigma_a-\Sigma$;

step 4: $k\in[0,N-1]$区间内的攻击信号a_k由定理1给出,其中$\theta_k\sim\mathcal{N}(\bar{\mu}(km+1:(k+1)m),\Gamma)$.

3　仿真实例

通过仿真实例验证本文结果的有效性. 首先考虑一个稳定的线性化最小相位四重水箱系统[33],系统矩阵参数如下:

$$A_1 = \begin{bmatrix} 0.9683 & 0 & 0.0819 & 0 \\ 0 & 0.9780 & 0 & 0.06377 \\ 0 & 0 & 0.9167 & 0 \\ 0 & 0 & 0 & 0.9355 \end{bmatrix},$$

$$C_1 = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \end{bmatrix},$$

$$Q_1 = \mathrm{diag}[0.25\ \ 0.25\ \ 0.25\ \ 0.25],$$

$$R_1 = \mathrm{diag}[0.5\ \ 0.5]. \tag{48}$$

图2给出了10 000次蒙特卡洛模拟时针对稳定的最小相位四重水箱系统的攻击效果.

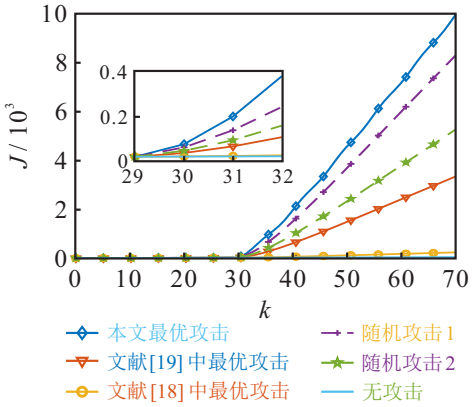

图2　针对稳定的最小相位四重水箱系统的攻击效果

对$k\in[30,69]$区间施加攻击,即攻击步数$N=40$. 由图2可知,针对稳定的最小相位四重水箱系统,与文献[19]中的攻击策略相比,当牺牲一定程度($\epsilon=1.5$)的隐性时,本文最优攻击策略使系统估计误差显著增大,验证了自生成ϵ-隐性攻击的有效性. 与文献[20]非严格隐性零均值攻击策略相比,本文最优攻击策略能够取得更大的估计误差,其主要原因是所提出的攻击模型引入了更具一般性的时变均值,使得所转化的优化问题具有更高的自由度. 与两种所提出攻击模型下的随机攻击策略相比,本文最优攻击策略能够最大化远程估计误差,充分说明了结果的最优性.

其次,考虑一个不稳定飞行器系统[38],系统矩阵

参数如下:

$$A_2=\begin{bmatrix}0.991\,1 & -0.120\,3 & -0.430\,2\\ 0.001\,7 & 0.990\,2 & -0.074\,7\\ 0 & 0.818\,7 & 0\end{bmatrix},$$

$$C_2=\mathrm{diag}[1\ \ 1\ \ 1],$$

$$Q_2=R_2=\mathrm{diag}[0.1\ \ \ 0.1\ \ \ 0.1]. \tag{49}$$

不同攻击策略经10 000次蒙特卡洛模拟针对不稳定的飞行器系统的攻击效果如图3所示.

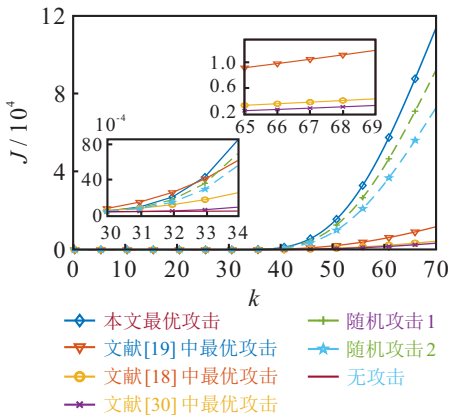

图 3 针对不稳定的飞行器系统的攻击效果

由图3可见,针对不稳定飞行器系统,与文献[19, 31]中的攻击策略相比,本文最优攻击策略在牺牲一定程度($\epsilon=1.5$)的隐性下,攻击效果获得显著提升,从而再次验证了自生成ϵ-隐性攻击的有效性. 与文献[20]非严格隐性零均值攻击策略相比,本文时变均值的最优攻击策略仍能取得更强的攻击效果. 类似地,与两种随机攻击策略相比,本文最优攻击策略的攻击效果最强,再次表明了结果的最优性.

最后,为验证攻击效果与隐性间存在权衡,图4给出不同阈值下针对稳定的四重水箱系统攻击效果. 图4结果表明,系统性能恶化程度随着阈值ϵ的增大而增大,表明攻击效果与隐性间存在权衡关系.

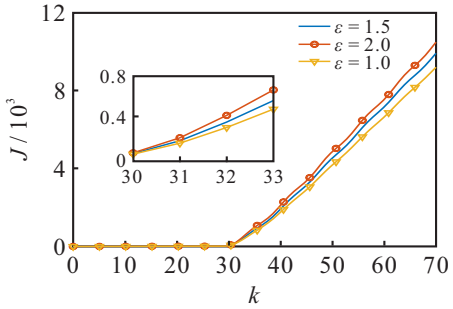

图 4 不同阈值ϵ下的攻击效果

4 结 论

本文研究了针对CPS的自生成ϵ-隐性最优欺骗攻击设计问题. 首先,提出了一种仅需系统信息即可离线生成攻击信号的自生成攻击模型,使攻击更易实现;随后,推导得出该攻击模型下的估计误差协方差以量化攻击效果,并将攻击设计问题转化为多变量受限的二次优化问题;进一步,利用K-L散度和互信息的相关统计学性质将其等价转化,再结合拉格朗日乘数法和所提出的参数特征关联覆盖法得到最优攻击策略;最后,通过仿真实例充分验证了本文结果的有效性.

参考文献(References)

[1] Derler P, Lee E A, Vincentelli A S. Modeling cyber-physical systems[J]. Proceedings of the IEEE, 2012, 100(1): 13-28.

[2] Yu W W, Wen G H, Yu X H, et al. Bridging the gap between complex networks and smart grids[J]. Journal of Control and Decision, 2014, 1(1): 102-114.

[3] Lu A Y, Yang G H. False data injection attacks against state estimation in the presence of sensor failures[J]. Information Sciences, 2020, 508: 92-104.

[4] Liu Y G, Xu B G, Ding Y H. Convergence analysis of cooperative braking control for interconnected vehicle systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(7): 1894-1906.

[5] Lee I, Sokolsky O, Chen S J, et al. Challenges and research directions in medical cyber-physical systems[J]. Proceedings of the IEEE, 2012, 100(1): 75-90.

[6] Xu D, Zhang W, Shi P, et al. Model-free cooperative adaptive sliding-mode-constrained-control for multiple linear induction traction systems[J]. Transactions on Cybernetics, 2020, 50(9): 4076-4086.

[7] An L W, Yang G H. Data-driven coordinated attack policy design based on adaptive L_2-gain optimal theory[J]. IEEE Transactions on Automatic Control, 2018, 63(6): 1850-1857.

[8] Amin S, Schwartz G A, Sastry S S. Security of interdependent and identical networked control systems[J]. Automatica: Journal of IFAC, 2013, 49(1): 186-192.

[9] 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[J]. 控制与决策, 2023, 38(8): 2093-2105.
(Yang G H, Lu A Y, An L W. A survey on secure state estimation of cyber-physical systems under cyber attacks[J]. Control and Decision, 2023, 38(8): 2093-2105.)

[10] 叶丹, 靳凯净, 张天予. 网络攻击下的信息物理系统安全性研究综述[J]. 控制与决策, 2023, 38(8): 2243-2252.
(Ye D, Jin K J, Zhang T Y. A survey on security of cyber-physical systems under network attacks[J]. Control and Decision, 2023, 38(8): 2243-2252.)

[11] An L W, Yang G H. Distributed secure state estimation for cyber-physical systems under sensor attacks[J]. Automatica, 2019, 107: 526-538.

[12] Cardenas A A, Amin S, Sastry S. Secure control:

Towards survivable cyber-physical systems[C]. The 28th International Conference on Distributed Computing Systems Workshops. Beijing, 2008: 495-500.

[13] 汪慕峰, 胥布工. DoS干扰攻击下的信息物理系统状态反馈稳定[J]. 控制与决策, 2019, 34(8): 1681-1687.
(Wang M F, Xu B G. State feedback stabilization of cyber-physical system under DoS jamming attacks[J]. Control and Decision, 2019, 34(8): 1681-1687.)

[14] 金丹, 吴麒, 陈博, 等. DoS攻击下信息物理系统的无模型H_∞控制[J]. 控制与决策, 2022, 37(10): 2565-2574.
(Jin D, Wu Q, Chen B, et al. Model-free H_∞ control for cyber-physical systems under DoS attacks[J]. Control and Decision, 2022, 37(10): 2565-2574.)

[15] Tang Y, Zhang D D, Ho D W C, et al. Event-based tracking control of mobile robot with denial-of-service attacks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(9): 3300-3310.

[16] Sun Y C, Yang G H. Periodic event-triggered resilient control for cyber-physical systems under denial-of-service attacks[J]. Journal of the Franklin Institute, 2018, 355(13): 5613-5631.

[17] 叶丹, 王吉言. 多传感器系统的最优线性欺骗攻击设计[J]. 控制与决策, 2019, 34(11): 2297-2302.
(Ye D, Wang J Y. Design of optimal linear deception attack for multi-sensor system[J]. Control and Decision, 2019, 34(11): 2297-2302.)

[18] 孙子文, 洪涛. 基于多信道博弈的ICPS虚假注入攻击防御策略[J]. 控制与决策, 2022, 37(5): 1357-1366.
(Sun Z W, Hong T. ICPS false injection attack defense strategy based on multi-channel game[J]. Control and Decision, 2022, 37(5): 1357-1366.)

[19] Guo Z Y, Shi D W, Johansson K H, et al. Optimal linear cyber-attack on remote state estimation[J]. IEEE Transactions on Control of Network Systems, 2017, 4(1): 4-13.

[20] Guo Z Y, Shi D W, Johansson K H, et al. Worst-case stealthy innovation-based linear attack on remote state estimation[J]. Automatica, 2018, 89: 117-124.

[21] Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids[J]. ACM Transactions on Information and System Security, 2011, 14(1): 1-33.

[22] Mo Y L, Sinopoli B. False data injection attacks in control systems[C]. Preprints of the 1st Workshop on Secure Control Systems. Piscataway: IEEE, 2010: 1.

[23] Mo Y L, Sinopoli B. Secure control against replay attacks[C]. The 47th Annual Allerton Conference on Communication, Control, and Computing. Monticello, 2009: 911-918.

[24] Li Y G, Yang G H. Optimal stealthy false data injection attacks in cyber-physical systems[J]. Information Sciences: An International Journal, 2019, 481(C): 474-490.

[25] Guo Z Y, Shi D W, Johansson K H, et al. Worst-case innovation-based integrity attacks with side information on remote state estimation[J]. IEEE Transactions on Control of Network Systems, 2019, 6(1): 48-59.

[26] Li Y G, Yang G H. Optimal innovation-based deception attacks with side information against remote state estimation in cyber-physical systems[J]. Neurocomputing, 2022, 500(C): 461-470.

[27] Bai C Z, Pasqualetti F, Gupta V. Security in stochastic control systems: Fundamental limitations and performance bounds[C]. American Control Conference. Chicago, 2015: 195-200.

[28] Bai C Z, Pasqualetti F, Gupta V. Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs[J]. Automatica, 2017, 82: 251-260.

[29] Kung E, Dey S, Shi L. The performance and limitations of ϵ- stealthy attacks on higher order systems[J]. IEEE Transactions on Automatic Control, 2017, 62(2): 941-947.

[30] Li Y G, Yang G H. Optimal stealthy innovation-based attacks with historical data in cyber-physical systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(6): 3401-3411.

[31] Zhang T Y, Ye D. False data injection attacks with complete stealthiness in cyber-physical systems: A self-generated approach[J]. Automatica, 2020, 120: 109117.

[32] Pasqualetti F, Dorfler F, Bullo F. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems[J]. IEEE Control Systems, 2015, 35(1): 110-127.

[33] Naha A, Teixeira A, Ahlén A, et al. Sequential detection of replay attacks[J]. IEEE Transactions on Automatic Control, 2023, 68(3): 1941-1948.

[34] Chen Y, Kar S, Moura J M F. Cyber-physical attacks with control objectives[J]. IEEE Transactions on Automatic Control, 2018, 63(5): 1418-1425.

[35] Chang Y H, Hu Q, Tomlin C J. Secure estimation based Kalman Filter for cyber-physical systems against sensor attacks[J]. Automatica, 2018, 95: 399-412.

[36] Zhang Q, Liu K, Xia Y, et al. Optimal stealthy deception attack against cyber-physical systems[J]. IEEE transactions on cybernetics, 2020, 50(9): 3963-3972.

[37] Latham P, Roudi Y. Mutual information[J]. Scholarpedia, 2009, 4(1): 1658.

[38] Hu L, Wang Z D, Han Q L, et al. State estimation under false data injection attacks: Security analysis and system protection[J]. Automatica, 2018, 87: 176-183.

作者简介

单华晟 (2000–), 男, 硕士生, 主要研究方向为信息物理系统网络安全性、隐性欺骗攻击策略优化设计, E-mail: shs3507@163.com;

李一刚 (1992–), 男, 讲师, 博士, 硕士生导师, 主要研究方向为信息物理系统网络安全性、隐性欺骗攻击策略优化设计, E-mail: liyigang920407@163.com.