基于鲁棒性能的信息物理融合系统乘性攻击检测

赵振根[1], 李渝哲[2†]

(1. 南京航空航天大学 自动化学院, 江苏 南京 211106; 2. 东北大学 流程工业综合自动化国家重点实验室, 辽宁 沈阳 110004)

摘要: 针对信息物理融合系统乘性攻击检测的难题, 本文提出两种基于鲁棒性能的乘性攻击检测与数据驱动实现策略. 首先, 利用互质分解和间隙度量理论, 建立信息物理融合系统和乘性攻击模型. 其次, 利用稳定裕度性能指标, 评估乘性攻击对闭环系统稳定性能的影响, 给出了稳定裕度的下界. 再次, 提出基于稳定裕度和基于残差鲁棒性能的乘性攻击检测策略, 并设计相应的检测阈值和检测逻辑. 进一步, 利用子空间辨识方法, 在线辨识稳定裕度和残差鲁棒性能, 提出基于鲁棒性能的乘性攻击检测的数据驱动实现方法. 最后, 利用飞行器系统的仿真, 验证了所提出的基于鲁棒性能的乘性攻击检测方法的有效性.

关键词: 信息物理融合系统; 攻击检测; 稳定裕度; 数据驱动

引用格式: 赵振根, 李渝哲. 基于鲁棒性能的信息物理融合系统乘性攻击检测. 控制理论与应用, 2022, 39(10): 1952 – 1960

DOI: 10.7641/CTA.2021.10541

Robust performance based multiplicative attack detection for cyber-physical systems

ZHAO Zhen-gen[1], LI Yu-zhe[2†]

(1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 211106, China;
2. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang Liaoning 110004, China)

Abstract: To address the problem of multiplicative attack detection in closed-loop cyber-physical systems, this paper proposes two methods of robust performance based multiplicative attack detection and their data-driven realization. First, apply the coprime factorization and gap metric to establish the models of cyber-physical systems and multiplicative attacks. Then, the stability margin is used to evaluate the effects of multiplicative attacks on the stability performance of closed-loop systems, and the low bound of stability margin is presented. Furthermore, the stability margin based and residual robust performance based attack detection schemes are proposed, and the corresponding detection thresholds and detection logics are designed. Moreover, with the aid of subspace identification, the data-driven realizations of the robust performance based attack detection are presented, via online identifying the stability margin and residual robust performance. Finally, the simulations on flight vehicle system are applied to verify the effectiveness of the robust performance based multiplicative attack detection methods.

Key words: cyber-physical systems; attack detection; stability margin; data-driven

Citation: ZHAO Zhengen, LI Yuzhe. Robust performance based multiplicative attack detection for cyber-physical systems. *Control Theory & Applications*, 2022, 39(10): 1952 – 1960

1 引言

随着计算机、通信和控制技术的发展, 信息物理融合系统广泛应用于现今的复杂工业系统. 与此同时, 在信息物理融合系统中, 大规模的子系统通过工业互联网连接, 使得系统的安全问题变得更为棘手. 如何确保信息物理融合系统平稳与高性能运行, 成为当下自动化领域的热点问题[1–2].

关于信息物理融合系统安全的研究, 主要分为两

收稿日期: 2021−06−24; 录用日期: 2021−09−30.

†通信作者. E-mail: yuzheli@mail.neu.edu.cn; Tel.: +86 13269108925.

本文责任编委: 朱善迎.

国家自然科学基金项目(62003161, 61890924, 61991404), 江苏省自然科学基金项目(BK20190399), 中央高校基本科研业务费项目(NS2020020, N180805002)资助, 飞行器自主控制技术教育部工程研究中心开放课题(NJ2020004)资助.

Supported by the National Natural Science Foundation of China (62003161, 61890924, 61991404), the Natural Science Foundation of Jiangsu Province (BK20190399), the Fundamental Research Funds for the Central Universities (NS2020020, N180805002) and the Research Funds for the Engineering Research Center of Aircraft Autonomous Control Technology, Ministry of Education (NJ2020004).

方面: 攻击设计和攻击防御. 从攻击者的角度, 主要研究攻击模型和安全性分析. 基于信息安全理论, 信息物理融合系统安全主要涉及可用性、完整性和保密性的问题. 拒绝服务攻击破坏信息的可用性[3], 而欺骗攻击破坏信息的完整性[4]. 相较于拒绝服务攻击, 欺骗攻击是由攻击者设计的更为精细的一种攻击, 对工业系统造成的影响和危害更大, 包括当前研究较多的虚假数据入攻击、零动态攻击和稀疏攻击等[5–6].

从防御者的角度, 信息物理融合系统安全主要研究攻击检测和弹性控制[7], 是防御者应对攻击设计的防御策略. 攻击检测的主要目的在于确定系统是否受到攻击. 攻击检测是攻击防御的关键环节, 为弹性控制提供决策信息[8]. 目前, 攻击检测研究主要存在两种思路: 一种思路是借鉴传统的故障检测方法用于攻击检测: 如基于未知输入观测器的方法[9]、基于故障检测滤波器的方法[10]、基于卡尔曼滤波器[11]和非线性观测器的检测方法[12]; 另一种思路是针对特定的攻击设计相应的检测策略: 如基于噪声设计的重放攻击检测[13]、基于博弈论的传感器攻击检测[14]、基于左零空间的隐蔽攻击检测[15]和基于序贯数据验证的线性欺骗攻击检测[16].

然而, 现有的攻击检测策略研究的攻击大多是加性的虚假数据注入攻击, 无法检测乘性篡改信息物理融合系统数据的攻击, 且很少评估攻击对系统性能的影响[17–18]. 近年来, 乘性攻击检测也得到一定的关注, 但主要是利用残差阈值对比的方法[19–20], 很少研究评估乘性攻击对系统性能的影响和利用性能变化进行攻击检测. 本文致力于研究基于鲁棒性能的乘性攻击检测策略, 相较于现有的攻击检测方法, 主要有以下几个方面的创新和贡献:

1) 利用互质分解和间隙度量理论, 提出一种信息物理融合系统和乘性攻击的建模方法, 并评估乘性攻击对信息物理融合系统的稳定性能的最坏影响.

2) 提出基于系统稳定裕度的乘性攻击检测策略, 设计相应的检测阈值和决策逻辑, 并给出基于稳定裕度的乘性攻击检测方法的数据驱动实现策略.

3) 提出基于残差鲁棒性能的乘性攻击检测策略, 设计相应的检测阈值和决策逻辑, 并给出基于残差鲁棒性能的乘性攻击检测方法的数据驱动实现策略.

本文组织结构如下: 第2部分介绍信息物理融合系统与乘性攻击模型; 第3部分提出乘性攻击对系统稳定性能的评估方法; 第4部分提出两种基于鲁棒性能的乘性攻击检测方法; 第5部分提出相应的数据驱动实现策略; 第6部分将提出的方法在飞行器系统上进行仿真验证; 第7部分总结本文的研究工作.

2　系统建模与问题描述

2.1　物理系统模型

实际的物理系统是一个非线性、时变的系统[12], 但为了突出乘性攻击对于系统性能影响的定量描述, 着重于对系统的显式分析, 获取更多结构性的结果, 用以评估攻击下的系统性能, 本文借鉴信息物理融合系统安全领域的建模思路[10,15], 采用线性离散时不变系统建立物理系统模型如下:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k), \\ y(k) = Cx(k) + Du(k), \end{cases} \tag{1}$$

其中: $x(k)\in\mathbb{R}^n, u(k)\in\mathbb{R}^{k_{\rm u}}, y(k)\in\mathbb{R}^{k_{\rm y}}$分别表示系统状态、控制输入和测量输出. 模型(1)对应的传递函数为 $y(z) = P_0(z)u(z)$, 其中$P_0(z) = (A, B, C, D)$. 为了简化阐述, 在传递函数的表示中, 一些情况省略z.

定义 1　系统$P_0(z)$的右互质分解为

$$P_0(z) = N(z)M^{-1}(z), \tag{2}$$

其中$M\in\mathcal{RH}_\infty, N\in\mathcal{RH}_\infty$是右互质的.

如果$M^*M + N^*N = I$, 则称右互质分解(2)是$P_0(z)$的归一化右互质分解.

定义 2　系统$P_0(z)$的左互质分解为

$$P_0(z) = \hat{M}^{-1}(z)\hat{N}(z), \tag{3}$$

其中$\hat{N}\in\mathcal{RH}_\infty, \hat{M}\in\mathcal{RH}_\infty$是左互质的.

如果$\hat{N}\hat{N}^* + \hat{M}\hat{M}^* = I$, 则称左互质分解(3)是$P_0(z)$的归一化左互质分解.

2.2　不确定性与乘性攻击模型

假设物理系统中存在参数摄动$\Delta_{\rm A}, \Delta_{\rm B}, \Delta_{\rm C}, \Delta_{\rm D}$, 其状态空间为$(A+\Delta_{\rm A}, B+\Delta_{\rm B}, C+\Delta_{\rm C}, D+\Delta_{\rm D})$, 不确定系统的传递函数的互质分解为

$$P_\Delta = (N + \Delta_{\rm N})(M + \Delta_{\rm M})^{-1}, \tag{4}$$

其中$\{\Delta_{\rm M}, \Delta_{\rm N}\}$为右互质分解因子不确定性. 假设模型不确定性满足如下约束:

$$\|\Delta\|_\infty = \| \begin{bmatrix} \Delta_{\rm M} \\ \Delta_{\rm N} \end{bmatrix} \|_\infty \leqslant \delta_\Delta, \tag{5}$$

其中: $\Delta = [\Delta_{\rm M}^{\rm T}\ \ \Delta_{\rm N}^{\rm T}]^{\rm T}$, δ_Δ表示右互质分解因子不确定性的范数上界.

在鲁棒控制理论中, 除了互质分解, 也可以利用间隙度量进行不确定性建模[21].

定义 3　令$P_i(z) = N_i(z)M_i^{-1}(z), i = 1, 2$为归一化右互质分解, 则$P_1(z)$和$P_2(z)$的有向间隙定义为

$$\vec{\delta}(P_1, P_2) = \inf_{\Phi\in{\rm H}_\infty} \| \begin{bmatrix} M_1 \\ N_1 \end{bmatrix} - \begin{bmatrix} M_2 \\ N_2 \end{bmatrix} \Phi \|_\infty, \tag{6}$$

其中$\Phi\in{\rm H}_\infty$表示任意稳定的传递函数.

系统$P_1(z)$和$P_2(z)$的间隙度量可以计算为

$$\delta(P_1,P_2)=\max\{\vec{\delta}(P_1,P_2),\vec{\delta}(P_2,P_1)\}. \quad (7)$$

任意两个系统的间隙度量: $\delta(P_1,P_2)\in[0,1]$. 如果$\delta(P_1,P_2)<1$, 则有

$$\delta(P_1,P_2)=\vec{\delta}(P_1,P_2)=\vec{\delta}(P_2,P_1).$$

当摄动半径δ相同, 由归一化互质分解和间隙度量定义的不确定系统集合相同[22]

$$P_\delta=\{P_\Delta:\|\Delta\|_\infty\leqslant\delta\}, \quad (8a)$$

$$\mathcal{B}(P_0,\delta)=\{P_\Delta:\delta(P_\Delta,P_0)\leqslant\delta\}, \quad (8b)$$

其中: P_δ表示与$P_0(z)$的归一化互质分解因子摄动范数界在δ内的系统集合, $\mathcal{B}(P_0,\delta)$表示以$P_0(z)$为球心, 以间隙度量δ为半径的系统集合.

典型信息物理融合系统如图1所示, 主要由物理系统、控制系统和通信网络组成. 乘性攻击是指攻击者通过窃听、虚假数据注入等手段乘性地篡改信息物理融合系统的数据, 在攻击实现形式上属于欺骗攻击, 包括乘性执行器攻击和传感器攻击[19–20]. 乘性的执行器攻击模型为

$$u_{\rm a}(k)=A^{\rm u}u(k),$$

其中: $A^{\rm u}\in\mathcal{R}^{k_{\rm u}\times k_{\rm u}}$表示执行器攻击矩阵, $u_{\rm a}(k)$表示攻击后的控制输入. 乘性的传感器攻击模型为

$$y_{\rm a}(k)=A^{\rm y}y(k),$$

其中: $A^{\rm y}\in\mathcal{R}^{k_{\rm y}\times k_{\rm y}}$表示传感器攻击矩阵, $y_{\rm a}(k)$表示攻击后的测量输出.

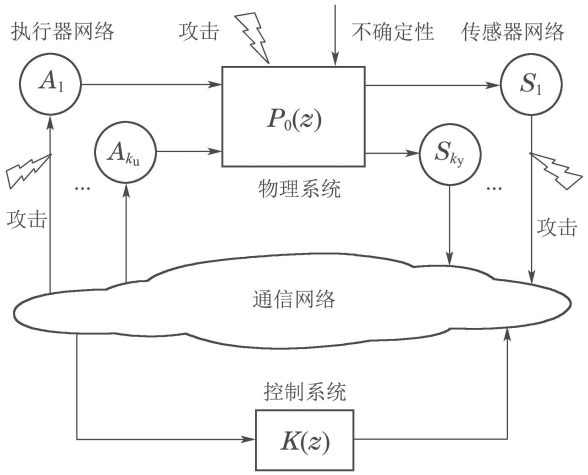

图 1 典型的信息物理融合系统

Fig. 1 Standard cyber-physical systems

定义矩阵表示$A_{\rm a}=A+\Delta_{\rm A}$, $B_{\rm a}=(B+\Delta_{\rm B})A^{\rm u}$, $C_{\rm a}=A^{\rm y}(C+\Delta_{\rm C})$, $D_{\rm a}=A^{\rm y}(D+\Delta_{\rm D})A^{\rm u}$, 则乘性攻击下系统的状态空间为$(A_{\rm a},B_{\rm a},C_{\rm a},D_{\rm a})$. 攻击系统归一化右互质分解为$P_{\rm a}=N_1M_1^{-1}$, 类似于式(4), 其可以改写为

$$P_{\rm a}=(N+N_{\rm a})(M+M_{\rm a})^{-1}, \quad (9)$$

其中$\{M_{\rm a},N_{\rm a}\}$表示由乘性攻击和模型不确定性引起的归一化右互质分解因子摄动. 因此,

$$\begin{bmatrix}M_{\rm a}\\N_{\rm a}\end{bmatrix}=\begin{bmatrix}M_1\\N_1\end{bmatrix}-\begin{bmatrix}M\\N\end{bmatrix}. \quad (10)$$

假设乘性攻击满足

$$\|\begin{bmatrix}M_{\rm a}\\N_{\rm a}\end{bmatrix}\|_\infty\leqslant\delta_{\rm a}, \quad (11)$$

其中$\delta_{\rm a}$表示乘性攻击和模型不确定性引起的归一化右互质分解因子摄动的范数上界. 在实际的信息物理融合系统中, 乘性攻击的信息通常很难获得, 所以$\delta_{\rm a}$是未知的.

2.3 问题描述

对于图1中的反馈互联系统$[P_0,K]$, 控制器$K(z)$必须能镇定物理系统$P_0(z)$. 由鲁棒控制理论, 所有镇定的控制器都可以表示为

$$K=(\hat{V}+Q\hat{N})^{-1}(\hat{U}+Q\hat{M})=(U+MQ)(V+NQ)^{-1}, \quad (12)$$

其中$Q\in\mathcal{RH}_\infty$是一个稳定的传递函数矩阵, 也称作Youla参数. $\hat{V},\hat{U},U,V$满足下述的Bezout等式:

$$\begin{bmatrix}\hat{V}&-\hat{U}\\-\hat{N}&\hat{M}\end{bmatrix}\begin{bmatrix}M&U\\N&V\end{bmatrix}=\begin{bmatrix}I&0\\0&I\end{bmatrix}, \quad (13)$$

其中传递函数的状态空间实现可以表示为

$$M=(A_{\rm F},B,F,I),\ N=(A_{\rm F},B,C+DF,D),$$
$$U=(A_{\rm F},-L,F,0),\ V=(A_{\rm F},-L,C+DF,I),$$
$$\hat{N}=(A_{\rm L},B+LD,C,D),\ \hat{M}=(A_{\rm L},L,C,I),$$
$$\hat{V}=(A_{\rm L},-B-LD,F,I),\ \hat{U}=(A_{\rm L},-L,F,0),$$

其中L和F为观测器增益和状态反馈增益, 且使得$A_{\rm L}=A+LC$和$A_{\rm F}=A+BF$稳定.

对于反馈互联系统$[P_0,K]$, 乘性攻击$\{M_{\rm a},N_{\rm a}\}$会引起系统闭环性能的变化. 如何定义合适的闭环性能指标、利用系统可测量的输入输出数据来评估性能的变化以及实现乘性攻击的检测, 是这篇文章致力研究和解决的重要课题.

3 信息物理融合系统稳定性能评估

针对乘性攻击下的信息物理融合系统, 利用稳定裕度来定义信息物理融合系统的稳定性能. 对于图1的典型信息物理融合系统, 将$P_0(z)$、不确定性和乘性攻击的组合用$P_{\rm a}(z)$替代, 在系统中引入参考信号v_1和噪声信号v_2, 则可得到图2所示的基于稳定裕度的乘性攻击检测原理图.

计算从参考信号v_1, 噪声信号v_2到系统的输入输出u,y的传递函数

$$H(P_{\rm a},K)=\begin{bmatrix}I\\P_{\rm a}\end{bmatrix}(I-KP_{\rm a})^{-1}[I\ \ K], \quad (14)$$

$H(P_{\rm a},K)$称为鲁棒性能矩阵. 如果$H(P_{\rm a},K)$是适定的, 且属于$\mathcal{RH}_\infty$, 则信息物理融合系统$[P_{\rm a},K]$是内稳定的. 定义其稳定裕度为

$$b(P_{\rm a},K)=\|H(P_{\rm a},K)\|_\infty^{-1}. \tag{15}$$

假设标称的闭环系统$[P_0,K]$的稳定裕度为$b(P_0,K)$, 且乘性攻击满足式(11), 则受攻击的闭环系统是稳定的, 当且仅当[23]

$$\arcsin\delta_{\rm a}+\arccos b(P_0,K)\leqslant\frac{\pi}{2}. \tag{16}$$

为了确定乘性攻击对闭环系统的最坏影响, 先引入系统图和逆图的概念.

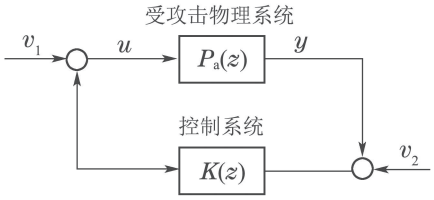

图 2 基于稳定裕度的乘性攻击检测

Fig. 2 Stability margin based multiplicative attack detection

定义 4　系统$P(z)$的图为有界的输入输出对所张成的空间

$$G_P=\{\begin{bmatrix}u\\y\end{bmatrix}:\begin{bmatrix}u\\y\end{bmatrix}=\begin{bmatrix}M\\N\end{bmatrix}\alpha\}, \tag{17}$$

其中$\alpha\in\mathcal{H}_2$为能量有界信号.

类似的, 定义控制器$K(z)$的逆图为

$$G_K'=\begin{bmatrix}0&I\\I&0\end{bmatrix},G_K=\begin{bmatrix}U\\V\end{bmatrix}\beta, \tag{18}$$

其中$\beta\in\mathcal{H}_2$为能量有界信号.

当$b(P_{\rm a},K)$取最小值时, 说明乘性攻击对系统的稳定性能影响最大. 下述定理给出受攻击闭环系统的稳定裕度的下界.

定理 1　假设$[P_0,K]$和受攻击的闭环系统$[P_{\rm a},K]$是稳定的, 则$[P_{\rm a},K]$的稳定裕度的下界为

$$\min b(P_{\rm a},K)=\cos[\arcsin\delta_{\rm a}+\arccos b(P_0,K)]. \tag{19}$$

证　令$G_K'^{\perp}$表示G_K'的垂直空间. 根据空间角度和稳定裕度的关系[23], G_{P_0}与$G_K'^{\perp}$的夹角为

$$\theta_0=\theta(G_{P_0},G_K'^{\perp})=\arccos b(P_0,K). \tag{20}$$

利用间隙度量的角度形式, $G_{P_{\rm a}}$与G_{P_0}的夹角满足

$$\theta_\Delta=\theta(G_{P_{\rm a}},G_{P_0})=\arcsin\delta_{\rm a}. \tag{21}$$

综合式(20)–(21), 则$G_{P_{\rm a}}$与G_K'的夹角

$$\theta_{\rm a}=\theta(G_{P_{\rm a}},G_K'^{\perp})\leqslant\theta(G_{P_{\rm a}},G_{P_0})+\theta(G_{P_0},G_K'^{\perp})\leqslant\theta_\Delta+\theta_0.$$

进而, 对于受攻击系统$P_{\rm a}\in\mathcal{B}(P_0,\delta_{\rm a})$, 其对应的稳定裕度$b(P_{\rm a},K)=\cos\theta_{\rm a}\geqslant\cos(\theta_\Delta+\theta_0)$. 因此, 受攻击的信息物理融合系统的稳定裕度的下界为

$$\cos[\arcsin\delta_{\rm a}+\arccos b(P_0,K)].$$

证毕.

注 1　根据式(16)和定理1, 当乘性攻击引起物理系统的摄动距离超过其标称闭环系统稳定裕度的容忍范围时, 即$\arcsin\delta_{\rm a}>\dfrac{\pi}{2}-\arccos b(P_0,K)$, 闭环系统的稳定性遭到破坏, 此类攻击为不稳定的乘性攻击. 不稳定乘性攻击使得系统的信号发散, 这类攻击可以通过分析信号的能量范数检测出来.

4　基于鲁棒性能的乘性攻击检测

4.1　基于稳定裕度的乘性攻击检测

根据闭环系统稳定裕度的变化, 可以检测系统是否受到攻击. 如果$b(P_{\rm a},K)$相对$b(P_\Delta,K)$下降, 说明攻击破坏系统稳定性能. 令性能评估函数为

$$J=b(P_{\rm a},K). \tag{22}$$

根据定理1, 不受攻击的信息物理融合系统的稳定裕度满足

$$\min b(P_\Delta,K)=\cos[\arcsin\delta_\Delta+\arccos b(P_\Delta,K)].$$

因此, 在无攻击情况下, 性能评估函数

$$J\geqslant\cos[\arcsin\delta_\Delta+\arccos b(P_\Delta,K)].$$

如果设计检测阈值为

$$J_{\rm th}=\cos[\arcsin\delta_\Delta+\arccos b(P_\Delta,K)]. \tag{23}$$

则乘性攻击的检测逻辑为

$$\begin{cases}J\geqslant J_{\rm th},\ \text{无攻击},\\J<J_{\rm th},\ \text{攻击}.\end{cases} \tag{24}$$

4.2　基于残差鲁棒性能的乘性攻击检测

基于残差鲁棒性能的乘性攻击检测原理如图3, 反馈控制器采用Youla参数化控制器结构.

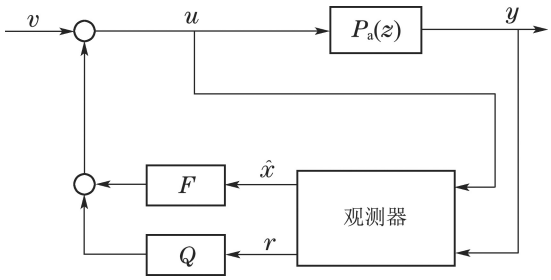

图 3 基于残差性能的乘性攻击检测

Fig. 3 Residual performance based multiplicative attack detection

系统的参考输入信号为v, 其输入输出信号可以表示为

$$\begin{bmatrix}u(z)\\y(z)\end{bmatrix}=\begin{bmatrix}I\\P_{\rm a}\end{bmatrix}(I-KP_{\rm a})^{-1}(\hat{V}+Q\hat{N})v(z). \tag{25}$$

残差生成器采用观测器实现, 其可以利用系统的左互质分解表示为

$$r(z)=[-\hat{N}\ \ \hat{M}]\begin{bmatrix}u(z)\\y(z)\end{bmatrix}. \tag{26}$$

将式(25)代入式(26), 可得

$$r(z)=T(z)v(z), \tag{27}$$

其中传递函数$T(z)$具有如下形式:

$$T(z)=[-\hat{N}\ \ \hat{M}]\begin{bmatrix}I\\P_{\rm a}\end{bmatrix}(I-KP_{\rm a})^{-1}(\hat{V}+Q\hat{N}).$$

定义攻击检测的性能评估函数为从v到r的传递函数的$\rm H_\infty$范数

$$J=\|T(z)\|_\infty. \tag{28}$$

针对不受攻击的信息物理融合系统, $T(z)$具有以下形式

$$\begin{aligned}T(z)=&[-\hat{N}\ \ \hat{M}]\begin{bmatrix}I\\P_\Delta\end{bmatrix}(I-KP_\Delta)^{-1}(\hat{V}+Q\hat{N})=\\&[-\hat{N}\ \ \hat{M}]\begin{bmatrix}M+\Delta_{\rm M}\\N+\Delta_{\rm N}\end{bmatrix}\cdot\\&[(\hat{V}+Q\hat{N})(M+\Delta_{\rm M})-\\&(\hat{U}+Q\hat{M})(N+\Delta_{\rm N})]^{-1}.\end{aligned}$$

根据Bezout等式(13), 上式可以转换为

$$\begin{aligned}T(z)=&[-\hat{N}\ \ \hat{M}]\begin{bmatrix}\Delta_{\rm M}\\\Delta_{\rm N}\end{bmatrix}\cdot\\&(I+[\hat{V}+Q\hat{N}\ \ -\hat{U}-Q\hat{M}]\begin{bmatrix}\Delta_{\rm M}\\\Delta_{\rm N}\end{bmatrix})^{-1}.\end{aligned}$$

设阈值$J_{\rm th}$为无攻击情况下$T(z)$的$\rm H_\infty$范数的最大值. 由于不确定性$\|\Delta\|_\infty\leqslant\delta_\Delta$, 因此, 阈值$J_{\rm th}$ 可以计算为

$$\begin{aligned}&J_{\rm th}=\\&\sup_{\|\Delta\|_\infty\leqslant\delta_\Delta}\|T(z)\|_\infty\leqslant\|[-\hat{N}\ \ \hat{M}]\|_\infty\cdot\\&\sup_{|\Delta\|_\infty\leqslant\delta_\Delta}\|\Delta(I+[\hat{V}+Q\hat{N}\ \ -\hat{U}-Q\hat{M}]\Delta)^{-1}\|_\infty\leqslant\\&\frac{\|[-\hat{N}\ \ \hat{M}]\|_\infty\|\Delta\|_\infty}{1-\|[\hat{V}+Q\hat{N}\ \ -\hat{U}-Q\hat{M}]\|_\infty\|\Delta\|_\infty}.\end{aligned}$$

上述不等式的右边是$\|\Delta\|_\infty$的单调递增函数, 当$\|\Delta\|_\infty$时函数取最大值. 因此, 攻击检测的阈值可以设计为

$$J_{\rm th}=\frac{\|[-\hat{N}\ \ \hat{M}]\|_\infty\delta_\Delta}{1-\|[\hat{V}+Q\hat{N}\ \ -\hat{U}-Q\hat{M}]\|_\infty\delta_\Delta}. \tag{29}$$

因此, 基于残差鲁棒性能的乘性攻击检测逻辑为

$$\begin{cases}J\leqslant J_{\rm th},\ \text{无攻击},\\J>J_{\rm th},\ \text{攻击}.\end{cases} \tag{30}$$

注 2 上述两种乘性攻击检测方法适用于不同结构的系统: 基于稳定裕度的方法适用于图2所示的信息物理融合系统, 而基于残差鲁棒性能的方法适用于图3所示的信息物理融合系统. 前一种方法对控制器的结构不作特定要求, 但要求系统的输出端存在噪声激励; 后一种方法不要求输出端存在噪声激励, 但要求控制器的结构是Youla参数化形式, 且残差信号内嵌在控制器中.

5 基于鲁棒性能的乘性攻击检测的数据驱动实现

利用子空间辨识[24], 本节提出数据驱动的基于鲁棒性能的乘性攻击检测策略.

5.1 数据驱动的$\rm H_\infty$范数估计

针对离散系统(1), 利用系统的输入输出数据估计传递函数矩阵$P_0(z)$的$\rm H_\infty$范数. 首先, 定义长度为q的序列向量

$$\psi_q(k)=[\psi(k)\ \ \cdots\ \ \psi(k+q-1)]^{\rm T}. \tag{31}$$

根据式 (1), 依次得到 $y(k),\cdots,y(k+q-1)$ 和 $x(k),u(k),\cdots,u(k+q-1)$的关系为

$$\begin{aligned}&y(k)=Cx(k)+Du(k),\\&y(k+1)=Cx(k+1)+Du(k+1)=\\&CAx(k)+CBu(k)+Du(k+1),\\&\vdots\\&y(k+q-1)=CA^{q-1}x(k)+\cdots+CBu(k+\\&\qquad q-2)+Du(k+q-1).\end{aligned}$$

则对应于离散系统模型(1)的序列模型为

$$y_q(k)=\Gamma_q x(k)+H_{u,q}u_q(k), \tag{32}$$

其中: $\Gamma_q=[C^{\rm T}\ \ (CA)^{\rm T}\ \ \cdots\ \ (CA^q-1)^{\rm T}]^{\rm T}$表示扩展可观性矩阵,

$$H_{u,q}=\begin{bmatrix}D&0&\cdots&0\\CB&D&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\CA^{q-2}B&\cdots&CB&D\end{bmatrix}, \tag{33}$$

表示输入–输出Toeplitz矩阵. 根据乘性算子理论[25], $P_0(z)$的$\rm H_\infty$范数可以通过下式进行估计:

$$\|P_0(z)\|_\infty=\lim_{q\to\infty}\sigma_{\max}(H_{u,q}), \tag{34}$$

其中$\sigma_{\max}(\cdot)$表示最大奇异值函数.

在数据驱动实现中, 考虑有限的序列长度q, 利用子空间辨识方法估计$H_{u,q}$. 进一步, 计算奇异值得出

$P_0(z)$的范数估计. 为了辨识$H_{u,q}$, 令N足够大, 定义数据矩阵

$$\Psi_{k,q} = [\psi_q(k) \ \cdots \ \psi_q(k+N-1)]. \tag{35}$$

则输入输出数据方程具有如下形式:

$$Y_{k,q} = \Gamma_q X_k + H_{u,q} U_{k,q}, \tag{36}$$

其中$X_k = [x(k) \ \cdots \ x(k+N-1)]$. 定义过去和将来的输入输出数据集如下:

$$Z_p = \begin{bmatrix} U_{k-p,p} \\ Y_{k-p,p} \end{bmatrix}, \ Z_q = \begin{bmatrix} U_{k,q} \\ Y_{k,q} \end{bmatrix}. \tag{37}$$

下述定理阐述$H_{u,q}$的数据驱动实现.

定理 2　假设下述条件成立, 1) $\mathrm{rank}(X_k) = n$. 2) $\mathrm{rank}(U_{k,q}) = qk_{\mathrm{u}}$. 3) $\mathrm{row}(X_k) \bigcap \mathrm{row}(U_{k,q}) = 0$.

对输入输出数据集做LQ分解如下:

$$\begin{bmatrix} Z_p \\ U_{k,q} \\ Y_{k,q} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix}, \tag{38}$$

则$H_{u,q}$可以估计为

$$\hat{H}_{u,q} = L_{32} L_{22}^{-1}. \tag{39}$$

证　利用子空间辨识的经典算法MOESP可以推导出[24].　　证毕.

5.2　基于稳定裕度的乘性攻击检测的数据驱动实现

$H(P_{\mathrm{a}}, K)$可以表示为如下状态空间:

$$\begin{aligned} x_h(k+1) &= A_h x_h(k) + B_h u_h(k), \\ y_h(k) &= C_h x_h(k) + D_h u_h(k), \end{aligned} \tag{40}$$

其中$u_h = [v_1^{\mathrm{T}} \ v_2^{\mathrm{T}}]^{\mathrm{T}}, y_h = [u^{\mathrm{T}} \ y^{\mathrm{T}}]^{\mathrm{T}}$.

实现数据驱动的基于稳定裕度的乘性攻击检测, 核心在于实时地估计闭环系统稳定裕度. 算法1给出稳定裕度的在线估计. 在线得到稳定裕度估计$\hat{b}(P_{\mathrm{a}}, K)$后, 令其为性能评估函数. 检测阈值可以采用式(23)或者利用无攻击的输入输出数据在线辨识, 攻击检测逻辑采用式(24).

算法 1　闭环稳定裕度的在线估计.

Step 1　初始化: 选择初始时刻k, N取足够大;

Step 2　for $i = 1 : M$;

Step 3　q从系统状态维数, 逐渐增大;

Step 4　采集时间段为k到$k+q+N-2$的v_1, v_2, u, y数据, 令$u_h = [v_1^{\mathrm{T}} \ v_2^{\mathrm{T}}]^{\mathrm{T}}, y_h = [u^{\mathrm{T}} \ y^{\mathrm{T}}]^{\mathrm{T}}$, 根据式(35)–(37), 构建相应的数据矩阵$Z_p, U_{k,q}, Y_{k,q}$;

Step 5　根据式(38)做LQ分解, 得到$\hat{H}_{u,q}$;

Step 6　计算$\hat{H}_{u,q}$的最大奇异值$\sigma_{\max}(\hat{H}_{u,q})$;

Step 7　如果相邻两次奇异值的差值小于0.001, 则终止循环;

Step 8　根据式 (15) 和式 (34), 稳定裕度估计为$\hat{b}(P_{\mathrm{a}}, K) = \dfrac{1}{\sigma_{\max}(\hat{H}_{u,q})}$;

Step 9　end.

5.3　基于残差鲁棒性能的乘性攻击检测的数据驱动实现

$T(z)$可以表示为如下状态空间:

$$\begin{aligned} x_l(k+1) &= A_l x_l(k) + B_l u_l(k), \\ y_l(k) &= C_l x_l(k) + D_l u_l(k), \end{aligned} \tag{41}$$

其中$u_l = v, y_l = r$. 类似地, 算法2给出残差鲁棒性能的在线估计.

同理, 在线得到$\|\hat{T}(z)\|_\infty$后, 令其为性能评估函数. 检测阈值可以采用式(29)或者利用无攻击的输入输出数据在线辨识, 攻击检测逻辑采用式(30).

算法 2　残差鲁棒性能的在线估计.

Step 1　初始化: 选择初始时刻k, N取足够大;

Step 2　for $i = 1 : M$;

Step 3　q从系统状态维数n, 逐渐增大;

Step 4　采集时间段为k到$k+q+N-2$的v, r的数据, 令$u_l = v, y_l = r$, 根据式(35)和式(37), 构建相应的数据矩阵$Z_p, U_{k,q}, Y_{k,q}$;

Step 5　根据式(38)做LQ分解, 得到$\hat{H}_{u,q}$;

Step 6　计算$\hat{H}_{u,q}$的最大奇异值$\sigma_{\max}(\hat{H}_{u,q})$;

Step 7　如果相邻两次奇异值的差值小于0.001, 则终止循环;

Step 8　$\|T(z)\|_\infty$的估计为$\sigma_{\max}(\hat{H}_{u,q})$;

Step 9　end.

注 3　本文针对两种典型的闭环系统(图2和图3), 分别提出基于稳定裕度和基于残差鲁棒性能的乘性攻击检测方法, 给出两种检测方法阈值设计形式(式(23)和式(29)), 并设计相应的数据驱动算法. 传统的乘性故障检测方法[26]主要是基于残差进行设计的. 与这些方法相比: 提出的基于稳定裕度的检测方法不依赖于残差生成, 可以实时检测系统的输入输出数据进行攻击检测; 提出的基于残差性能的检测方法, 给出新的阈值设计形式, 这种阈值只依赖于系统和控制器的互质分解; 评估乘性攻击对系统稳定性能的最坏影响, 给出不稳定攻击的实现条件.

6　仿真与分析

面对日益复杂对抗的飞行环境, 飞行器的安全性分析和攻击检测逐渐引起关注[27]. 为了验证所提出的基于鲁棒性能的乘性攻击检测算法的有效性, 应用一个通用的飞行器纵向模型[28]:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k), \\ y(k) = Cx(k), \end{cases} \tag{42}$$

其中状态量$x = [x_1 \ x_2 \ x_3]^{\mathrm{T}}$分别表示俯仰角、俯仰角速率和飞行速度. 控制器输入u为升降舵控制指令. 系统矩阵如下:

$$A=\begin{bmatrix}0.9944 & -0.1203 & -0.4302\\ 0.0017 & 0.9902 & -0.0747\\ 0 & 0.8187 & 0\end{bmatrix},$$

$$B=\begin{bmatrix}0.4252\\ -0.0082\\ 0.1813\end{bmatrix}^{\mathrm{T}},\ C=I_3.$$

飞行器由于质量变化或者空间环境变化引起动力学的变化, 假设这种变化引起系统矩阵A发生摄动. 令摄动矩阵Δ_{A}为

$$\Delta_{\mathrm{A}}=\mathrm{diag}\{0.01,0,0\},\tag{43}$$

则不确定系统$P_\Delta=(A+\Delta_{\mathrm{A}},B,C,0)$的归一化互质分解因子不确定性和间隙度量不确定性满足$\delta_\Delta=\delta(P,P_\Delta)=0.0255$.

6.1 基于稳定裕度的乘性攻击检测

系统传递函数矩阵$P_0=(A,B,C,0)$. 做右互质分解$P_0=NM^{-1}$, 其状态空间实现为

$$M=(A_{\mathrm{F}},B,F,I),\ N=(A_{\mathrm{F}},B,C,0).$$

选择控制器增益为

$$F=[-0.7981\ \ 0.8586\ \ 0.2925].\tag{44}$$

做左互质分解$P_0=\hat{M}^{-1}\hat{N}$, 其状态空间实现$\hat{N}=(A_{\mathrm{L}},B,C,0)$, $\hat{M}=(A_{\mathrm{L}},L,C,I)$, 选择观测器增益

$$L=\begin{bmatrix}-0.2928 & 0.0113 & -0.0323\\ 0.0141 & -0.0070 & 0.0023\\ 0.0093 & -0.0006 & 0.0001\end{bmatrix}.\tag{45}$$

控制器采用Youla参数化的控制器结构. 令$Q=0$, 则$K=\hat{V}^{-1}\hat{U}=UV^{-1}$为基于观测器的镇定控制器, 其具有如下形式:

$$K=(A+BF+LC,-L,F,0).$$

此时, 信息物理融合系统$[P_0,K]$的稳定裕度为$b(P_0,K)=0.3888$.

基于稳定裕度的攻击检测阈值可以设计为

$$J_{\mathrm{th}}=\cos[\arcsin\delta_\Delta+\arccos b(P,K)]=0.3652.$$

在实际系统中, 不确定系统实际对应的稳定裕度的边界在一些情况下很难获得. 当系统的不确定性特性完全未知时, 不确定系统的稳定裕度可以利用数据驱动方法进行在线估计. 取$N=5000$, 令过去数据序列长度$q=p$, 图4描述了估计的稳定裕度随着p,q的变化曲线. 当$q=p\geqslant 50$时, 稳定裕度的估计误差较小. 在数据驱动的阈值设计中, 选取p,q足够大, 则估计的阈值为$\hat{J}_{\mathrm{th}}=0.3678$.

假设系统受到乘性攻击, 攻击者按比例地篡改传感器的测量值, 将传感器第一个通道的测量值变成原来的2倍. 受攻击系统的状态空间表示为$P_{\mathrm{a}}=(A+\Delta_{\mathrm{A}},B,A^{\mathrm{y}}C,0)$, 攻击矩阵

$$A^{\mathrm{y}}=\mathrm{diag}\{2,1,1\},\tag{46}$$

受攻击系统稳定裕度的理论值$b(P_{\mathrm{a}},K)=0.3018$. 图5描述了数据驱动的基于稳定裕度的攻击检测过程. J_{th}和$\hat{J}_{\mathrm{th}}$分别表示基于模型设计的阈值和数据驱动估计的阈值, J表示基于稳定裕度的性能评估函数. 从曲线可以看出, 所设计的基于稳定裕度的攻击检测算法能较快地检测取系统中的乘性攻击. 相较于基于模型设计的阈值, 利用数据驱动估计的阈值检测时延更短, 需要消耗更多的存储和计算资源.

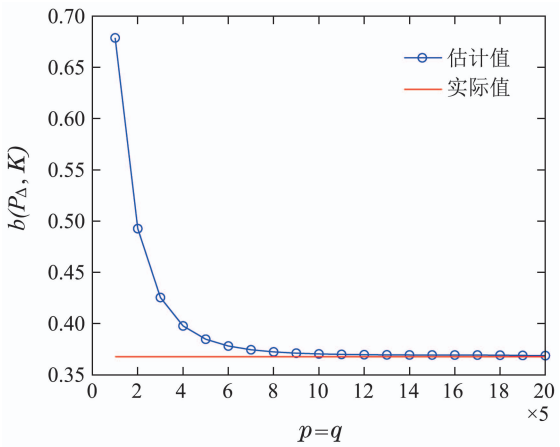

图 4 不确定系统稳定裕度的在线估计

Fig. 4 On-line estimation of stability margin of uncertain system

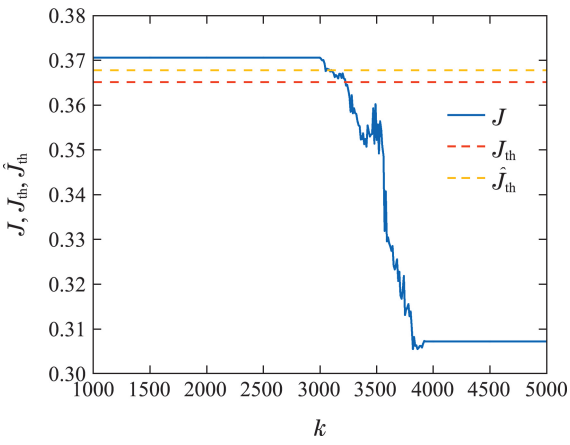

图 5 基于稳定裕度的乘性攻击检测

Fig. 5 Stability margin based multiplicative attack detection

6.2 基于残差鲁棒性能的乘性攻击检测

在基于残差鲁棒性能的攻击检的仿真中, 根据式(29), 计算出$J_{\mathrm{th}}=0.0675$. 数据驱动的阈值估计采取与基于稳定裕度类似的策略. 图6描述了残差鲁棒性能随着p,q的变化关系. 当p,q足够大时, 估计阈值收敛为$\hat{J}_{\mathrm{th}}=0.0385$.

同样, 假设乘性攻击具有式(46)的形式, 图7描述数据驱动的基于残差鲁棒性能的乘性攻击检测. J_{th}

和$\hat{J}_{\rm th}$分别表示基于模型设计的阈值和数据驱动估计的阈值, J表示残差性能评估函数. 从曲线可以看出, 基于残差鲁棒性能的乘性攻击检测算法也能较快地检测出系统中的乘性攻击. 类似地, 利用数据驱动估计阈值的乘性攻击检测算法的检测时延更短, 但需要消耗更多的计算和存储资源.

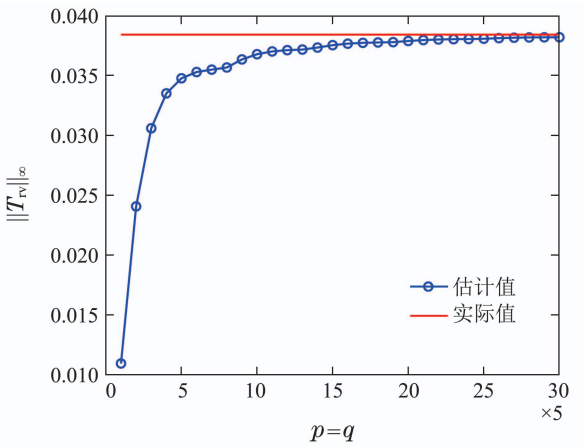

图 6　不确定系统残差鲁棒性能的在线估计

Fig. 6　On-line estimation of residual robust performance of uncertain system

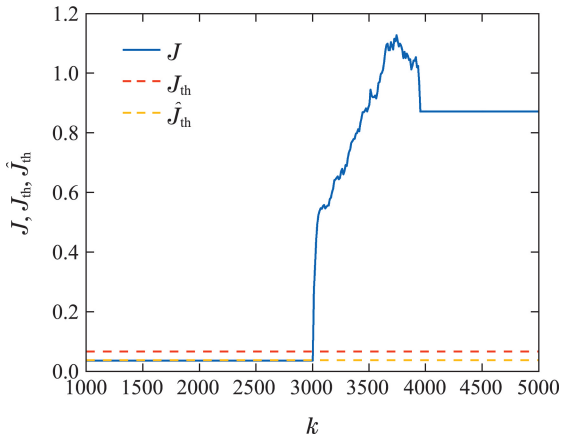

图 7　基于残差鲁棒性能的乘性攻击检测

Fig. 7　Residual robust performance based multiplicative attack detection

在上述实例中, 基于稳定裕度和基于残差鲁棒性能的检测方法都能较快地检测出系统中的乘性攻击, 对比图5和图7的攻击检测效果可以看出, 基于残差鲁棒性能的检测方法的检测时延更短.

7　结论

本文主要研究基于鲁棒性能的乘性攻击检测方法与其数据驱动实现问题, 分别提出基于稳定裕度和基于残差鲁棒性能两种乘性攻击检测策略. 两种检测策略中, 前者适用于标准的反馈互联系统, 后者适用于参考跟踪控制系统. 进一步, 针对两种基于鲁棒性能的乘性攻击检测算法, 提出数据驱动的实现策略, 使得提出的基于鲁棒性能的乘性攻击检测方法能更好的应用于实际工业信息物理融合系统. 今后的研究将聚焦乘性攻击检测与弹性控制一体化设计, 以提高信息物理融合系统的安全性能.

参考文献:

[1] DING D R, HAN Q L, GE X H, et al. Secure state estimation and control of cyber-physical systems: A survey. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(1): 176 – 190.

[2] LIU Ting, TIAN Jue, WANG Jiazhou, et al. Integrated security threats and defense of cyber-physical systems. *Acta Automatica Sinica*, 2019, 45(1): 5 – 24.
(刘烃, 田决, 王稼舟, 等. 信息物理融合系统综合安全威胁与防御研究. 自动化学报, 2019, 45(1): 5 – 24.)

[3] LI Y Z, SHI L, CHENG P, et al. Jamming attacks on remote state estimation in cyber-physical systems: a game-theoretic approach. *IEEE Transactions on Automatic Control*, 2015, 60(10): 2831 – 2836.

[4] XIE L, MO Y L, SINOPOLI B. Integrity data attacks in power market operations. *IEEE Transactions on Smart Grid*, 2011, 2(4): 659 – 666.

[5] DENG R L, XIAO G, LU R, et al. False data injection on state estimation in power systems-attacks, impacts, and defense: A survey. *IEEE Transactions on Industrial Informatics*, 2017, 13(2): 411 – 423.

[6] YE Dan, Wang Jiyan. Design of optimal linear deception attack for multi-sensor system. *Control and Decesion*, 2019, 31(11): 2297 – 2302.
(叶丹, 王吉言. 多传感器系统的最优线性欺骗攻击设计. 控制与决策, 2019, 31(11): 2297 – 2302.)

[7] DIBAJI S M, PIRANI M, FLAMHOLZ D B, et al. A systems and control perspective of CPS security. *Annual Reviews in Control*, 2019, 47: 394 – 411.

[8] DING D R, HAN Q L, XIANG Y, et al. A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing*, 2018, 275: 1674 – 1683.

[9] AMIN S, LITRICO X, SASTRY S S, et al. Cyber security of water SCADA systems-Part II: attack detection using enhanced hydrodynamic models. *IEEE Transactions on Control Systems*, 2013, 21(5): 1679 – 1693.

[10] PASQUALETTI F, DORFLER F, BULLO F. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 2013, 58(11): 2715 – 2729.

[11] ZHUANG Kangxi, SUN Ziwen. Establishing a detection model for denial of service attacks in industrial cyber physical systems. *Control Theory & Applications*, 2020, 37(3): 629 – 638.
(庄康熙, 孙子文. 针对工业信息物理系统中的拒绝服务攻击建立检测模型. 控制理论与应用, 2020, 37(3): 629 – 638.)

[12] KIM J, LEE C, SHIM H, et al. Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems having redundant sensors. *IEEE Transactions on Automatic Control*, 2019, 64(3): 1162 – 1169.

[13] MO Y L, CHABUKSWAR R, SINOPOLI B. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 2014, 22(4): 1396 – 1407.

[14] VAMVOUDAKIS K G, HESPANHA J P, SINOPOLI B, et al. Detection in adversarial environments. *IEEE Transactions on Smart Grid*, 2014, 59: 3209 – 3223.

[15] CHEN Y, KAR S, MOURA J M F. Dynamic attack detection in cyber-physical systems with side initial state information. *IEEE Transactions on Automatic Control*, 2017, 62(9): 4618 – 4624.

[16] LI Y Z, SHI L, CHEN T W. Detection against linear deception attacks on multi-sensor remote state estimation. *IEEE Transactions on Control of Network Systems*, 2018, 5(3): 846 – 856.

[17] 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.

[18] KUNG E, DEY S, SHI L. The performance and limitations of stealthy attacks on higher order systems. *IEEE Transactions on Automatic Control*, 2017, 62(2): 941 – 947.

[19] NA G, EUN Y. A multiplicative coordinated stealthy attack and its detection for cyber physical systems. *IEEE Conference on Control Technology and Applications*, Copenhagen, Denmark: IEEE, 2018, 1698 – 1703.

[20] KANELLOPOULOS A, VAMVOUDAKIS K G. A moving target defense control framework for cyber-physical systems. *IEEE Transactions on Automatic Control*, 2020, 65(3): 1029 – 1043.

[21] VINNICOMBE G. *Uncertainty and Feedback: Loop-Shaping and the v-Gap Metric*. London, UK: Imperial College Press, 2001.

[22] GEORGIOU T T, SMITH M C. Optimal robustness in the gap metric. *IEEE Transactions on Automatic Control*, 1990, 35(6): 673 – 686.

[23] QIU L, DAVISON E J. Feedback stability under simultaneous gap metric uncertainties in plant and controller. *Systems & Control Letters*, 1992, 18: 9 – 22.

[24] KATAYAMA T. *Subspace Methods for System Identification*. New York, NY, USA: Springer, 2005.

[25] ZHOU K M, DOYLE J C, GLOVER K. *Robust and Optimal Control. Upper Saddle River*, NJ, USA: Prentice-Hall, 1996.

[26] LI L L, LUO H, DING S X, et al. Performance-based fault detection and fault-tolerant control for automatic control systems. *Automatica*, 2019, 99: 308 – 316.

[27] MUNIRAJ D, FARHOOD M. Detection and mitigation of actuator attacks on small unmanned aircraft systems. *Control Engineering Practice*, 2019, 83: 188 – 202.

[28] HU L, WANG Z, HAN Q L, LIU X. State estimation under false data injection attacks: security analysis and system protection. *Automatica*, 2018, 87: 176 – 183.

作者简介:

赵振根 讲师, 硕士生导师, 目前研究方向为信息物理系统安全与无人机智能控制等, E-mail: zhaozhengen@nuaa.edu.cn;

李渝哲 教授, 博士生导师, 目前研究方向为网络化系统的状态估计、控制与优化、信息物理系统的安全与隐私及其在工业系统中的应用等, E-mail: yuzheli@mail.neu.edu.cn.