控制与决策
Control and Decision

DoS攻击下信息物理系统的无模型H_∞控制

金丹, 吴麒, 陈博, 俞立

引用本文:
金丹,吴麒,陈博,俞立. 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.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.0278

DoS攻击下信息物理系统的无模型H_∞控制

金 丹, 吴 麒, 陈 博†, 俞 立

(1. 浙江工业大学 信息工程学院，杭州 310023；2. 浙江工业大学 网络空间安全研究院，杭州 310023)

摘 要: 针对动态模型未知的信息物理系统在拒绝服务(DoS)攻击下的安全控制问题,提出无模型的H_∞控制方法,其中DoS攻击具有代价约束且连续攻击次数有界. 首先,利用量测数据设计丢包情形下的Smith预估器对当前状态进行预测,并给出了量测反馈H_∞控制器的结构形式;其次,利用博弈论将H_∞控制问题转化为二人零和博弈问题,从而给出控制器增益的设计方法;进一步,基于Q-learning方法设计模型未知下的控制器增益在线求解算法,实现系统的安全H_∞控制;最后,通过雕刻机平台的仿真和实验验证所提出方法的有效性.

关键词: 信息物理系统；DoS攻击；无模型；H_∞控制；博弈论；数据驱动

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.0278

引用格式: 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.

Model-free H_∞ control for cyber-physical systems under DoS attacks

JIN Dan, WU Qi, CHEN Bo†, YU Li

(1. College of Information Engineering，Zhejiang University of Technology，Hangzhou 310023，China；2. Institute of Cyberspace Security，Zhejiang University of Technology，Hangzhou 310023，China)

Abstract: This paper is concerned with the model-free H_∞ control for cyber-physical systems with unknown dynamic models in the presence of denial-of-service (DoS) attacks, where DoS attacks are considered to be cost-constrained, and the number of successive attacks is bounded. Firstly, a Smith predictor with dropouts is designed to predict the current state using the measurement data, and the structure of the measurement feedback H_∞ controller is given. Secondly, the H_∞ control problem is transformed into a two-player zero-sum game, and then the designing method of the controller gain is presented using the game theory. Furthermore, an online Q-learning algorithm is designed to solve the controller gain with unknown model parameters, and the secure H_∞ control of the system is realized. Finally, an experiment on the plat of Carving is given to demonstrate the effectiveness of the proposed methods.

Keywords: cyber-physical systems；DoS attacks；model-free；H_∞ control；game theory；data-driven

0 引 言

信息物理系统(CPS)作为计算、网络和物理对象之间紧密集成的系统[1],由于其完美的集成设计与优越性能,在基础设施和工业应用的发展中发挥着越来越重要的作用. 基于CPS的结构特点以及网络层的开放性,CPS在为人们提供便利的同时也易受到网络攻击. 比如,2020年上半年,委内瑞拉的电网遭到袭击造成全国大面积停电,以色列的自来水公司受到网络攻击,台湾的两家大型炼油厂受到勒索软件攻击等[2]. 频繁的网络攻击造成了巨大的经济损失,甚至威胁到社会安全. 因此,提高安全意识,加强网络层的安全防护势在必行.

拒绝服务(denial-of-service, DoS)攻击作为一种常见的网络攻击,它通过占用通信资源或设备资源,以禁止量测或控制信号的传输,导致通信信道中正常传输的数据包丢失[3]. 与网络化控制系统中常见的丢包现象不同,攻击诱导的丢包现象有其自身的特点. 由于攻击者往往具有蓄意性和破坏性,致使攻击诱导的丢包现象一般无规律可循,而由网络诱导的丢包现象通常会假设遵循一定的概率分布,这一假设在破坏者蓄意攻击情况下很难被满足. 此外,随着网络技术和硬件设施的发展,网络诱导的丢包现象越来越少,甚至可以忽略不计;与此相反,随着网络攻击日益频繁,攻击瞬间诱导的丢包现象不可避免,从而系统

收稿日期: 2021-02-11；**录用日期:** 2021-06-18.

基金项目: 国家自然科学基金项目(61973277, 62073292)；浙江省自然科学基金项目(LR20F030004).

责任编委: 林志赟.

†通讯作者. E-mail: bchen@zjut.edu.cn.

性能可能会出现大的瞬间波动. 如果不能很好地处理DoS攻击下的丢包问题,可能会导致网络崩溃,甚至破坏物理系统.

针对DoS攻击下的CPS安全控制问题,目前存在如下几种处理方法:随机系统的方法[4-5],博弈论方法[6-7],弹性控制方法[8-10],预测控制方法[11-12]等. 特别地,文献[4-5]基于随机系统理论,将攻击行为描述为马尔可夫分布,设计了一种DoS攻击下的最优控制策略. 文献[6]基于博弈论,给出了DoS攻击下的最优攻击策略和防御策略;文献[7]基于零和博弈理论刻画了DoS攻击的频率与持续时间,进而设计了安全弹性控制方法. 同样地,文献[8-9]也采用了弹性控制方法,不同之处在于文献[8]设计了基于传输次数的弹性控制机制,而文献[9]则设计了基于事件触发的弹性控制机制. 注意到,无论DoS攻击是发生在单通道[7,9,13],还是发生在多通道[8,11-12,14-15],上述工作都是基于模型所设计的安全控制方法. 事实上,大多数实际复杂系统的精准动力学模型很难获得.

为了克服控制方法对系统模型的依赖,近年来在控制领域中无模型的控制方法得到了快速发展. 文献[16-17]利用神经网络能够有效地避免系统模型,但是,在基于神经网络的辨识系统过程中所导致的近似误差为计算控制器增加了不确定性,从而降低了系统性能. 为了克服神经网络方法中近似误差增加的不足,强化学习作为一种解决无模型问题的有效方法得到了应用[18-21]. 文献[20]利用强化学习解决了系统模型未知情况下的H_∞控制问题,且不需要可允许的初始控制策略,但没有考虑网络攻击下的无模型控制问题. 文献[22]建立了对抗拥塞攻击的无模型自适应控制框架,设计了基于输入/输出数据的n-步预测补偿机制,但是在设计过程中利用伯努利分布来描述干扰攻击行为. 文献[23]针对线性系统中执行器攻击下系统模型未知的安全控制问题,利用异策略的强化学习方法,提出了一种基于数据的自适应积分滑模控制策略. 到目前为止,基于量测反馈设计CPS系统模型参数未知与DoS攻击下的安全控制方法鲜有报道.

综上分析,本文聚焦于传感器-控制器(S-C)通道遭受DoS攻击的CPS安全控制问题,其中DoS攻击无需满足某种概率分布. 当系统参数未知以及存在外部干扰时,本文利用可获得的历史数据和参考信息,通过求解二人零和博弈的鞍点设计安全的H_∞控制器. 本文的贡献概括如下:1) 在系统状态信息丢失和具有外部扰动的情况下,利用学习方法设计了基于数据驱动的控制器;2) 设计了基于量测反馈求解带有丢包的零和博弈问题的鞍点算法,实现了量测反馈控制策略的在线计算和调整;3) 在雕刻机实验平台验证了所提出算法的有效性.

1 问题描述

考虑如下的离散系统:

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

其中:$x(k)\in\boldsymbol{R}^n$是系统状态,$u(k)\in\boldsymbol{R}^m$是控制输入,$w(k)\in\boldsymbol{R}^l$是外部干扰输入,$y(k)\in\boldsymbol{R}^p$是系统输出,且$A\in\boldsymbol{R}^{n\times n}$、$B\in\boldsymbol{R}^{n\times m}$、$C\in\boldsymbol{R}^{p\times n}$、$D\in\boldsymbol{R}^{n\times l}$是常数矩阵. 假设$(A,B)$可控,$(A,C)$可观,且$w(k)\in L_2[0,\infty)$.

考虑如下的动力学系统,其对系统(1)提供了跟踪轨迹$v(k)$:

$$r(k+1)=Er(k), \tag{2}$$

$$v(k)=Fr(k). \tag{3}$$

其中:$r(k)\in\boldsymbol{R}^q$,$v(k)\in\boldsymbol{R}^p$,矩阵$E\in\boldsymbol{R}^{q\times q}$,$F\in\boldsymbol{R}^{p\times q}$.

如图1所示,当系统状态$x(k)$通过网络传输给控制器时,若S-C通道遭受到DoS攻击则可能会造成丢包现象. 此时,控制中心可获得的系统状态可表示为

$$x_\alpha(k)=\alpha(k)x(k)+(1-\alpha(k))x_\alpha(k-1).$$

其中:$\alpha(k)\in\{0,1\}$,$\alpha(k)=0$代表S-C通道发生DoS攻击,$\alpha(k)=1$代表S-C通道数据传输正常. 一般来说,入侵者的攻击能量是有限的,可假设入侵者的连续攻击次数是有界的,而由此造成的最大连续丢包数也是有界数,记为$\bar{n}_\alpha$.

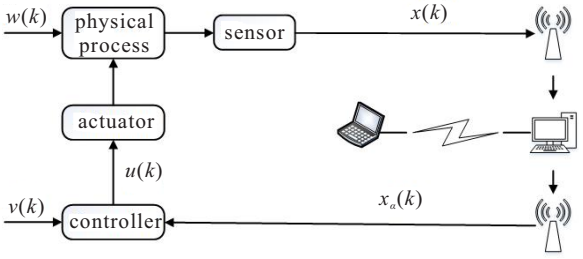

图1 DoS攻击下信息物理系统的框架

定义$X(k)=[x^{\mathrm{T}}(k),r^{\mathrm{T}}(k)]^{\mathrm{T}}$,于是系统(1)可增广为

$$\begin{cases} X(k+1)=A_1X(k)+B_1u(k)+D_1w(k),\\ y(k)=C_1X(k). \end{cases} \tag{4}$$

其中:$A_1=\begin{bmatrix}A & 0\\ 0 & E\end{bmatrix}$,$B_1=\begin{bmatrix}B\\ 0_{n\times m}\end{bmatrix}$,$D_1=\begin{bmatrix}D\\ 0_{n\times l}\end{bmatrix}$,

$C_1=[C,\ 0_{p\times q}]$.

本文将从主动防御的角度出发,设计最优的H_∞控制器使得系统在遭受DoS攻击下,系统输出$y(k)$依然能够跟踪参考信号$v(k)$. 因此,可考虑系统的跟踪误差

$$e(k)=Cx(k)-Fr(k). \tag{5}$$

注意到,控制器的最优性是针对某一性能指标而言,为了避免参考轨迹对控制器设计的影响,本文选取如下折扣型的性能函数(值函数):

$$J(k)=\sum_{i=k}^{\infty}\lambda^{i-k}\big[e^{\mathrm{T}}(i)Qe(i)+u^{\mathrm{T}}(i)Ru(i)-\gamma^2w^{\mathrm{T}}(i)w(i)\big]. \tag{6}$$

其中: $Q^{\mathrm{T}}=Q\geqslant 0, R=R^{\mathrm{T}}>0, 0<\lambda\leqslant 1$是折扣因子,$\gamma\geqslant 0$是衰减水平. 如果系统满足如下方程,则称系统的$L_2$-增益小于或等于$\gamma$:

$$\sum_{k=0}^{\infty}\lambda^k\big[e^{\mathrm{T}}(k)Qe(k)+u^{\mathrm{T}}(k)Ru(k)\big]\leqslant\gamma^2\sum_{k=0}^{\infty}\lambda^kw^{\mathrm{T}}(k)w(k).$$

根据文献[24], H_∞控制问题可转化为二人零和博弈问题,控制策略玩家试图最小化值函数,而干扰策略玩家试图最大化值函数. 一旦找到博弈问题的鞍点,也就意味着H_∞控制器的获得. 在$u(k)$可容许的情况下,选取博弈问题的值函数

$$V(k)=J(k),$$

于是最优的值函数

$$V^*(k)=\min_{u(k)}\max_{w(k)}\sum_{i=k}^{\infty}\lambda^{i-k}\big[e^{\mathrm{T}}(i)Qe(i)+u^{\mathrm{T}}(i)Ru(i)-\gamma^2w^{\mathrm{T}}(i)w(i)\big]. \tag{7}$$

为保证解的存在性,假设

$$\min_{u(k)}\max_{w(k)}V(k)=\max_{w(k)}\min_{u(k)}V(k),$$

这是鞍点$(u^*(k),w^*(k))$存在的充分性条件. 基于以上分析,本文中的H_∞控制问题可描述为如下的二人零和博弈问题:

$$\min_{u(k)}\max_{w(k)}\sum_{i=k}^{\infty}\lambda^{i-k}\big[e^{\mathrm{T}}(i)Qe(i)+u^{\mathrm{T}}(i)Ru(i)-\gamma^2w^{\mathrm{T}}(i)w(i)\big];$$

s.t. 式(4). \tag{8}

接下来,将通过求解带有丢包的二人零和博弈问题设计无模型Q-learning 算法.

注1　针对最优跟踪控制问题,根据文献[25],控制律可以表示成$u(k)=k_xx(k)+k_rr(k)$的形式,易见控制输入与参考轨迹的性质有关. 为解决最优控制问题,参考轨迹必须渐近收敛到零,否则当$k\to\infty$时,$u(k)$不会趋于零,进而导致性能指标趋向于无穷,这就失去了最优控制的意义. 为了放松这个限制条件,本文引入折扣型的性能指标. 如果矩阵E是Hurwitz矩阵,则λ可取$(0,1]$上的任意值;如果矩阵E不是Hurwitz矩阵,则λ需在$(0,1)$上取值[22].

2　无模型的控制器设计

本节针对系统模型参数未知的情况,利用量测数据求解二人零和博弈问题的鞍点$(u^*(k),w^*(k))$.

定义在k时刻之前发生的连续丢包数为$n_\alpha(k)$,若在k时刻,$x(k)$传输成功,则$n_\alpha(k)=0$. 显然,$0\leqslant n_\alpha(k)\leqslant\bar{n}_\alpha$. 于是,在$k$时刻,控制器通过网络得到的最新状态量为

$$x_\alpha(k)=x(k-n_\alpha(k)). \tag{9}$$

若在k时刻S-C通道发生DoS攻击,则可利用历史状态数据、控制输入和干扰输入来预测当前状态$x(k)$,即

$$x(k)=A^{n_\alpha(k)}x(k-n_\alpha(k))+\sum_{i=1}^{n_\alpha(k)}A^{i-1}Bu(k-i)+\sum_{i=1}^{n_\alpha(k)}A^{i-1}Dw(k-i).$$

根据文献[26],本文将利用量测数据构造一个带有丢包的Smith预估器,定义k时刻可获得的量测向量为

$$z(k)=\big[\underbrace{x_\alpha^{\mathrm{T}}(k)\ \dots\ x_\alpha^{\mathrm{T}}(k)}_{\bar{n}_\alpha+1}\rightarrow\leftarrow\underbrace{u^{\mathrm{T}}(k-1)\ \dots\ u^{\mathrm{T}}(k-n_\alpha(k))}_{\bar{n}_\alpha}\rightarrow\leftarrow\underbrace{w^{\mathrm{T}}(k-1)\ \dots\ w^{\mathrm{T}}(k-n_\alpha(k))}_{\bar{n}_\alpha}\ r^{\mathrm{T}}(k)\big]^{\mathrm{T}},$$

则增广状态向量

$$X(k)=MN(n_\alpha(k))z(k). \tag{10}$$

其中

$$M=\begin{bmatrix}I & A & \dots & A^{\bar{n}_\alpha} & B & AB & \dots & A^{\bar{n}_\alpha-1}B\\ 0 & 0 & \dots & 0 & 0 & 0 & \cdots & 0\end{bmatrix}\rightarrow\leftarrow\begin{bmatrix}D & AD & \dots & A^{\bar{n}_\alpha-1}D & 0\\ 0 & 0 & \dots & 0 & I\end{bmatrix},$$

$N(n_\alpha(k))$是与n_α有关的矩阵,且它的维数为$[n(\bar{n}_\alpha+1)+m\bar{n}_\alpha+l\bar{n}_\alpha+q][n(\bar{n}_\alpha+1)+m\bar{n}_\alpha+l\bar{n}_\alpha+q]$. 定义$z_\alpha(k)=N(n_\alpha(k))z(k), n_\alpha(k)=0,1,\dots,\bar{n}_\alpha$,则

$z_\alpha(k)$计算如下:

当$n_\alpha(k)=0$时,有

$$z_\alpha(k)=\big[\underbrace{x_\alpha^{\mathrm{T}}(k)\,0\,\dots\,0}_{\bar{n}_\alpha+1}\ \underbrace{0\,\dots\,0}_{\bar{n}_\alpha}\ \underbrace{0\,\dots\,0}_{\bar{n}_\alpha}\ r^{\mathrm{T}}(k)\big]^{\mathrm{T}};$$

当$n_\alpha(k)=1$时,有

$$z_\alpha(k)=\big[\underbrace{0\,x_\alpha^{\mathrm{T}}(k)\,\dots\,0}_{\bar{n}_\alpha+1}\ \underbrace{u^{\mathrm{T}}(k-1)\,\dots\,0}_{\bar{n}_\alpha}\rightarrow$$
$$\leftarrow\underbrace{w^{\mathrm{T}}(k-1)\,\dots\,0}_{\bar{n}_\alpha}\ r^{\mathrm{T}}(k)\big]^{\mathrm{T}};$$

当$n_\alpha(k)=\bar{n}_\alpha$时,有

$$z_\alpha(k)=$$
$$\big[\underbrace{0\,0\,\dots\,x_\alpha^{\mathrm{T}}(k)}_{\bar{n}_\alpha+1}\ \underbrace{u^{\mathrm{T}}(k-1)\,\dots\,u^{\mathrm{T}}(k-\bar{n}_\alpha)}_{\bar{n}_\alpha}\rightarrow$$
$$\leftarrow\underbrace{w^{\mathrm{T}}(k-1)\,\dots\,w^{\mathrm{T}}(k-\bar{n}_\alpha)}_{\bar{n}_\alpha}\ r^{\mathrm{T}}(k)\big]^{\mathrm{T}}.$$

由$z_\alpha(k)$的定义可知,$z_\alpha(k)$均由已知的信息构成,于是增广状态向量可表示为已知量测向量$z_\alpha(k)$的形式,即

$$X(k)=Mz_\alpha(k). \tag{11}$$

进而,最优的控制输入和最差的干扰输入可表示为

$$u^*(k)=-K_1^*X(k)=-\bar{K}_1^*z_\alpha(k),$$
$$w^*(k)=-K_2^*X(k)=-\bar{K}_2^*z_\alpha(k).$$

其中:$\bar{K}_1^*=K_1^*M,\bar{K}_2^*=K_2^*M$. 显然,最优量测反馈控制器增益$\bar{K}_1^*$和量测反馈干扰增益$\bar{K}_2^*$的维数与最大丢包数$\bar{n}_\alpha$有关.

在导出主要结果前,给出如下引理.

引理1 关于H_∞控制问题,给定任意稳定的控制输入和干扰输入,即

$$u(k)=-K_1X(k)=-\bar{K}_1z_\alpha(k), \tag{12}$$
$$w(k)=-K_2X(k)=-\bar{K}_2z_\alpha(k). \tag{13}$$

选取折扣因子λ使得$\lambda^{0.5}E$为稳定阵,则值函数可表示为

$$V(k)=X^{\mathrm{T}}(k)PX(k)=z_\alpha^{\mathrm{T}}(k)\bar{P}z_\alpha(k), \tag{14}$$

其中矩阵$P>0$,且$\bar{P}=\bar{P}^{\mathrm{T}}=M^{\mathrm{T}}PM>0$.

证明 由文献[21]中引理1中的证明可知,对于任意稳定的控制策略和干扰策略,控制输入和干扰输入分别满足式(12)和(13),则值函数具有如下的形式:

$$J(k)=X^{\mathrm{T}}(k)PX(k).$$

其中:$P=\begin{bmatrix}P_{11}&P_{12}\\P_{21}&P_{22}\end{bmatrix}$,且$P_{11},P_{12},P_{21}$和$P_{22}$满足

$$P_{11}=$$
$$\sum_{i=0}^{\infty}\lambda^iA_c^{i\mathrm{T}}(C^{\mathrm{T}}QC+K_{11}^{\mathrm{T}}RK_{11}-\gamma^2K_{21}^{\mathrm{T}}K_{21})A_c^i,$$
$$P_{12}=$$
$$\sum_{i=0}^{\infty}\lambda^i[A_c^{i\mathrm{T}}(C^{\mathrm{T}}QC+K_{11}^{\mathrm{T}}RK_{11}-\gamma^2K_{21}^{\mathrm{T}}K_{21})G+$$
$$A_c^{i\mathrm{T}}(-C^{\mathrm{T}}QF+K_{11}^{\mathrm{T}}RK_{12}-\gamma^2K_{21}^{\mathrm{T}}K_{22})E^i],$$
$$P_{21}=P_{12}^{\mathrm{T}},$$
$$P_{22}=$$
$$\sum_{i=0}^{\infty}\lambda^i[G^{\mathrm{T}}(C^{\mathrm{T}}QC+K_{11}^{\mathrm{T}}RK_{11}-\gamma^2K_{21}^{\mathrm{T}}K_{21})G+$$
$$G^{\mathrm{T}}(-C^{\mathrm{T}}QF+K_{11}^{\mathrm{T}}RK_{12}-\gamma^2K_{21}^{\mathrm{T}}K_{22})E^i+$$
$$E^{i\mathrm{T}}(-F^{\mathrm{T}}QC+K_{12}^{\mathrm{T}}RK_{11}-\gamma^2K_{22}^{\mathrm{T}}K_{21})G+$$
$$E^{i\mathrm{T}}(F^{\mathrm{T}}QF+K_{12}^{\mathrm{T}}RK_{12}-\gamma^2K_{22}^{\mathrm{T}}K_{22})E^i].$$

$A_c=A_1-B_1K_1-D_1K_2$, $G=\sum_{j=0}^{i-1}A_c^{i-j-1}(-BK_{12}-DK_{22})E^j$均是稳定的. 因为$\lambda^{0.5}E$是稳定的,故$P_{22}$是有界的. 将$X(k)$用$z_\alpha(k)$代替,可得

$$J(k)=X^{\mathrm{T}}(k)PX(k)=z_\alpha^{\mathrm{T}}(k)\bar{P}z_\alpha(k),$$

其中$\bar{P}=M^{\mathrm{T}}PM>0$. □

结合性能指标函数(6)和引理1,贝尔曼方程可表示成量测向量$z_\alpha(k)$的形式,即

$$z_\alpha^{\mathrm{T}}(k)\bar{P}z_\alpha(k)=$$
$$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)+u^{\mathrm{T}}(k)Ru(k)-$$
$$\gamma^2w^{\mathrm{T}}(k)w(k)+\lambda z_\alpha^{\mathrm{T}}(k+1)\bar{P}z_\alpha(k+1), \tag{15}$$

其中$Q_z=M^T(C\ -F)^{\mathrm{T}}Q(C\ -F)M$.

定义关于量测向量的哈密尔顿方程为

$$H(z_\alpha(k),u(k),w(k))=$$
$$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)+u^{\mathrm{T}}(k)Ru(k)-\gamma^2w^{\mathrm{T}}(k)w(k)+$$
$$\lambda z_\alpha^{\mathrm{T}}(k+1)\bar{P}z_\alpha(k+1)-z_\alpha^{\mathrm{T}}(k)\bar{P}z_\alpha(k).$$

针对上述方程求偏导,$\partial H(z_\alpha(k),u(k),w(k))/\partial u(k)=0$和$\partial H(z_\alpha(k),u(k),w(k))/\partial w(k)=0$,从而获得如下最优控制输入$u^*(k)$和最差干扰输入$w^*(k)$,即

$$u^*(k)=-\bar{K}_1^*z_\alpha(k), \tag{16}$$
$$w^*(k)=-\bar{K}_2^*z_\alpha(k). \tag{17}$$

其中

$$\bar{K}_1^*=(\Lambda_{11}-\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21})^{-1}(\bar{\eta}_1-\Lambda_{12}\Lambda_{22}^{-1}\bar{\eta}_2), \tag{18}$$
$$\bar{K}_2^*=(\Lambda_{22}-\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12})^{-1}(\bar{\eta}_2-\Lambda_{21}\Lambda_{11}^{-1}\bar{\eta}_1), \tag{19}$$

且$\bar{P}^*$满足带有丢包的博弈黎卡提方程(GARE)

$$\bar{P} = \lambda A_2^{\mathrm{T}}\bar{P}A_2 + Q_z - \bar{\eta}^{\mathrm{T}}\Lambda^{-1}\bar{\eta}, \tag{20}$$

$$\Lambda = \begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix} = \begin{bmatrix}R+\lambda B_2^{\mathrm{T}}\bar{P}B_2 & \lambda B_2^{\mathrm{T}}\bar{P}D_2\\ \lambda D_2^{\mathrm{T}}\bar{P}B_2 & \lambda D_2^{\mathrm{T}}\bar{P}D_2-\gamma^2 I\end{bmatrix} = \begin{bmatrix}R+\lambda B_1^{\mathrm{T}}PB_1 & \lambda B_1^{\mathrm{T}}P\\ \lambda D_1^{\mathrm{T}}PB_1 & \lambda D_1^{\mathrm{T}}P-\gamma^2 I\end{bmatrix}, \tag{21}$$

$$\bar{\eta} = \begin{bmatrix}\bar{\eta}_1\\ \bar{\eta}_2\end{bmatrix} = \begin{bmatrix}\lambda B_2^{\mathrm{T}}\bar{P}A_2\\ \lambda D_2^{\mathrm{T}}\bar{P}A_2\end{bmatrix} = \begin{bmatrix}\lambda B_1^{\mathrm{T}}PA_1\\ \lambda D_1^{\mathrm{T}}PA_1\end{bmatrix} M := \eta M, \tag{22}$$

$$A_2 = M^+A_1M,\ B_2 = M^+B_1,$$
$$D_2 = M^+D_1,\ M^+ = M^{\mathrm{T}}(MM^{\mathrm{T}})^{-1}.$$

由式(21)和(22)可知,最优量测反馈控制器增益和最差量测反馈干扰增益可重写为

$$\bar{K}_1^* = K_1^*M,\ \bar{K}_2^* = K_2^*M,$$

其中K_1^*和K_2^*可以通过求解无攻击下的二人零和博弈问题的方法得到. 此时,关于系统(4)的鞍点$(u^*(k),\ w^*(k))$表示为

$$u^*(k) = -K_1^*X(k),\ w^*(k) = -K_2^*X(k).$$

其中控制器增益K_1^*和干扰增益K_2^*满足

$$K_1^* = (\Lambda_{11}-\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21})^{-1}(\eta_1-\Lambda_{12}\Lambda_{22}^{-1}\eta_2),$$
$$K_2^* = (\Lambda_{22}-\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12})^{-1}(\eta_2-\Lambda_{21}\Lambda_{11}^{-1}\eta_1),$$

且常数矩阵$P^* = P^{*\mathrm{T}} > 0$满足普通的GARE,即

$$P = \lambda A_1^{\mathrm{T}}PA_1 + Q_1 - \eta^{\mathrm{T}}\Lambda^{-1}\eta,$$
$$Q_1 = (C\ -F)^{\mathrm{T}}Q(C\ -F).$$

根据文献[24],为保证在严格反馈稳定策略中得到唯一的鞍点,增广系统(4)还需满足如下不等式:

$$I-\gamma^{-2}\lambda D_1^{\mathrm{T}}PD_1 > 0,$$
$$R+\lambda B_1^{\mathrm{T}}PB_1 > 0.$$

下一步,将要讨论带有丢包的GARE的解的最优性.

引理2　令$\tilde{X}(k) = \lambda^{k/2}X(k)$,其中折扣因子$\lambda\in(0,1]$,如果$X(k)$是渐近稳定或有界的,则当$k\to\infty$时,$\tilde{z}_\alpha(k) = \lambda^{k/2}z_\alpha(k)\to 0$.

证明　令$\tilde{X}(k) = \lambda^{k/2}X(k)$,类似于文献[21]中引理2的证明,若$\lambda = 1$,$E$是Hurwitz矩阵,则$X(k)\to 0$是渐近稳定的,进而有$k\to\infty$时$\tilde{X}(k)\to 0$;若$\lambda\in(0,1)$,$X(k)$是渐近稳定或者有界的,则均有$k\to\infty$时$\tilde{X}(k)\to 0$. 这意味着$\tilde{u}(k) = -K_1\tilde{X}(k)\to 0$,$\tilde{w}(k) = -K_2\tilde{X}(k)\to 0$. 此外,$\tilde{X}(k)\to 0$意味着$\tilde{r}(k) = \lambda^{k/2}r(k)\to 0$和$\tilde{x}(k) = \lambda^{k/2}x(k)\to 0$,同时$\tilde{x}_\alpha(k) = \lambda^{k/2}x_\alpha(k) = \lambda^{k/2}x(k-\bar{n}_\alpha)\to 0$. 从而$u(k) = \lambda^{-k/2}\tilde{u}(k), w(k) = \lambda^{-k/2}\tilde{w}(k)$和$r(k) = \lambda^{-k/2}\tilde{r}(k)$. 再由方程(14)和$\tilde{z}(k) = \lambda^{k/2}z(k)$可知,$\tilde{z}(k)$可重写为

$$\begin{aligned}\tilde{z}(k) = [&\underbrace{\tilde{x}_\alpha^{\mathrm{T}}(k)\ \dots\ \tilde{x}_\alpha^{\mathrm{T}}(k)}_{\bar{n}_\alpha+1}\to\\ &\leftarrow\underbrace{\tilde{u}^{\mathrm{T}}(k-1)\ \dots\ \tilde{u}^{\mathrm{T}}(k-n_\alpha(k))}_{\bar{n}_\alpha}\to\\ &\leftarrow\underbrace{\tilde{w}^{\mathrm{T}}(k-1)\ \dots\ \tilde{w}^{\mathrm{T}}(k-n_\alpha(k))}_{\bar{n}_\alpha}\ \tilde{r}^{\mathrm{T}}(k)]^{\mathrm{T}}.\end{aligned}$$

显然,若$X(k)$渐近稳定或有界,则当$k\to\infty$时,$\tilde{z}(k)\to 0$,即$\tilde{z}_\alpha(k) = N(n_\alpha(k))\tilde{z}(k) = N(n_\alpha(k))\lambda^{k/2}z(k) = \lambda^{k/2}z_\alpha(k)\to 0$. □

定理1　考虑增广系统(4),选取值函数(7),并且选择折扣因子λ使得$\lambda^{0.5}E$是稳定的,则当$k\to\infty, \tilde{z}_\alpha(k)\to 0$时,通过方程(16)和(17)获取的鞍点$(u^*(k), w^*(k))$能够最小化值函数(7).

证明　将方程(16)和(17)代入下式:

$$\begin{aligned}&\tilde{z}_\alpha^{\mathrm{T}}(\infty)\bar{P}^*\tilde{z}_\alpha(\infty) - \tilde{z}_\alpha^{\mathrm{T}}(k)\bar{P}^*\tilde{z}_\alpha(k) =\\ &\sum_{i=k}^{\infty}[\lambda^{i+1}z_\alpha^{\mathrm{T}}(i+1)\bar{P}^*z_\alpha(i+1) - \lambda^i z_\alpha^{\mathrm{T}}(i)\bar{P}^*z_\alpha(i)] =\\ &\sum_{i=k}^{\infty}[\tilde{z}_\alpha^{\mathrm{T}}(i)(\lambda A_2^{\mathrm{T}}\bar{P}^*A_2 - \bar{P}^*)\tilde{z}_\alpha(i) + \Xi],\end{aligned} \tag{23}$$

其中$\bar{P}^*$是博弈黎卡提方程(20)的解,并且

$$\begin{aligned}\Xi = &\tilde{z}_\alpha^{\mathrm{T}}(i)\eta_1^{*\mathrm{T}}\tilde{u}^*(i) + \tilde{z}_\alpha^{\mathrm{T}}(i)\eta_2^{*\mathrm{T}}\tilde{w}^*(i) + \tilde{u}^{*\mathrm{T}}(i)\eta_1^*\tilde{z}_\alpha(i) +\\ &\tilde{u}^{*\mathrm{T}}(k)\Lambda_{12}^*\tilde{w}^*(i) + \tilde{w}^{*\mathrm{T}}(i)\eta_2^*\tilde{z}_\alpha(i) + \tilde{w}^{*\mathrm{T}}(i)\Lambda_{21}^*\tilde{u}^*(i) +\\ &\lambda\tilde{u}^{*\mathrm{T}}(i)B_2^{\mathrm{T}}\bar{P}^*B_2\tilde{u}^*(i) + \lambda\tilde{w}^{*\mathrm{T}}(i)D_2^{\mathrm{T}}\bar{P}^*D_2\tilde{w}^*(i).\end{aligned}$$

由引理2可得$\tilde{z}_\alpha(\infty)\to 0$,且方程(23)可简化为

$$\begin{aligned}0 = &\sum_{i=k}^{\infty}\left[\tilde{z}_\alpha^{\mathrm{T}}(i)(\lambda A_2^{\mathrm{T}}\bar{P}^*A_2 - \bar{P}^*)\tilde{z}_\alpha(i) + \Xi\right] +\\ &\tilde{z}_\alpha^{\mathrm{T}}(k)\bar{P}^*\tilde{z}_\alpha(k).\end{aligned} \tag{24}$$

对贝尔曼方程(15)进行简单的计算,可以得到关于量测向量的Lyapunov方程

$$\begin{aligned}\bar{P}^* = &Q_z + \bar{K}_1^{*\mathrm{T}}R\bar{K}_1^* - \gamma^2\bar{K}_2^{*\mathrm{T}}\bar{K}_2^* +\\ &\lambda(A_2 - B_2\bar{K}_1^* - D_2\bar{K}_2^*)^{\mathrm{T}}\bar{P}(A_2 - B_2\bar{K}_1^* - D_2\bar{K}_2^*).\end{aligned}$$

将$\bar{P}^*$代入方程(24),使得

$$\begin{aligned}&\sum_{i=k}^{\infty}[\tilde{z}_\alpha^{\mathrm{T}}(i)(-Q_z - \bar{K}_1^{*\mathrm{T}}R\bar{K}_1^* + \gamma^2\bar{K}_2^{*\mathrm{T}}\bar{K}_2 +\\ &\eta_1^{\mathrm{T}}\bar{K}_1^* + \eta_1^{\mathrm{T}}\bar{K}_2^* - \bar{K}_1^{*\mathrm{T}}\lambda B_1^{\mathrm{T}}M^{+\mathrm{T}}\bar{P}^*M^+B_1\bar{K}_1^* +\end{aligned}$$

$$\bar{K}_1^{*\mathrm{T}}\eta_1+\bar{K}_1^{*\mathrm{T}}\Lambda_{12}\bar{K}_2^*+\bar{K}_2^{*\mathrm{T}}\eta_2-\bar{K}_2^{*\mathrm{T}}\Lambda_{21}\bar{K}_1^*-$$
$$\bar{K}_2^{*\mathrm{T}}\lambda D_1^{\mathrm{T}}M^{+\mathrm{T}}\bar{P}^*M^+D_1\bar{K}_2^*)\tilde{z}_\alpha(i)+\Xi\big]+$$
$$\tilde{z}_\alpha^{\mathrm{T}}(k)\bar{P}^*\tilde{z}_\alpha(k)=0. \tag{25}$$

此外,将$\tilde{z}_\alpha(k),\tilde{u}^*(k)$和$\tilde{w}^*(k)$代入值函数(7),使得

$$V(k)=\lambda^{-k}\sum_{i=k}^{\infty}(\tilde{z}_\alpha^{\mathrm{T}}(i)M^{\mathrm{T}}Q_1M\tilde{z}_\alpha(i)+$$
$$\tilde{u}^{*\mathrm{T}}(i)R\tilde{u}^*(i)-\gamma^2\tilde{w}^{*\mathrm{T}}(i)\tilde{w}^*(i)). \tag{26}$$

接下来,将方程(25)的两边乘上λ^{-k}再加到方程(26)中,则值函数为

$$V(k)=$$
$$z_\alpha^{\mathrm{T}}(k)\bar{P}^*z_\alpha(k)+$$
$$\lambda^{-k}\sum_{i=k}^{\infty}\begin{bmatrix}\tilde{u}^*(i)+\bar{K}_1^*\tilde{z}_\alpha(i)\\ \tilde{w}^*(i)+\bar{K}_2^*\tilde{z}_\alpha(i)\end{bmatrix}^{\mathrm{T}}\Lambda\begin{bmatrix}\tilde{u}^*(i)+\bar{K}_1^*\tilde{z}_\alpha(i)\\ \tilde{w}^*(i)+\bar{K}_2^*\tilde{z}_\alpha(i)\end{bmatrix}.$$

显然,当$\tilde{u}^*(i)=-\bar{K}_1^*\tilde{z}_\alpha(i)$和$\tilde{w}^*(i)=-\bar{K}_2^*\tilde{z}_\alpha(i)$,即$u^*(i)=-\bar{K}_1^*z_\alpha(i),w^*(i)=-\bar{K}_2^*z_\alpha(i)$时,$V(k)$取到最优值. □

定理1说明了系统鞍点$(u^*(k),w^*(k))$可以基于模型信息获得. 然而,在实际中系统的精准模型很难获得,为了克服对系统模型的依赖,接下来将利用无模型Q-learning方法求解系统的鞍点.

令$Z(k)=[z_\alpha^{\mathrm{T}}(k),u^{\mathrm{T}}(k),w^{\mathrm{T}}(k)]^{\mathrm{T}}$,定义$Q$-函数为

$$Q(k)=$$
$$e^{\mathrm{T}}(k)Qe(k)+u^{\mathrm{T}}(k)Ru(k)-\gamma^2w^{\mathrm{T}}(k)w(k)+$$
$$\lambda z_\alpha^{\mathrm{T}}(k+1)\bar{P}z_\alpha(k+1)=$$
$$Z^{\mathrm{T}}(k)GZ(k)+\lambda Z^{\mathrm{T}}(k)\begin{bmatrix}A_2\\ B_2\\ D_2\end{bmatrix}^{\mathrm{T}}\bar{P}\begin{bmatrix}A_2\\ B_2\\ D_2\end{bmatrix}Z(k)=$$
$$Z^{\mathrm{T}}(k)HZ(k). \tag{27}$$

其中

$$G=\begin{bmatrix}Q_z & 0 & 0\\ 0 & I & 0\\ 0 & 0 & -\gamma^2I\end{bmatrix}, \tag{28}$$

$$H^j=\begin{bmatrix}H_{zz} & H_{zu} & H_{zw}\\ H_{uz} & H_{uu} & H_{uw}\\ H_{wz} & H_{wu} & H_{ww}\end{bmatrix}=$$
$$\begin{bmatrix}\lambda A_2^{\mathrm{T}}\bar{P}A_2+Q_z & \eta_1^{\mathrm{T}} & \eta_2^{\mathrm{T}}\\ \eta_1 & \Lambda_{11} & \Lambda_{12}\\ \eta_2 & \Lambda_{21} & \Lambda_{22}\end{bmatrix}. \tag{29}$$

针对Q-函数分别关于$u(k)$和$w(k)$求偏导,可得

$$u^*(k)=-\bar{K}_1^*z_\alpha(k),\ w^*(k)=-\bar{K}_2^*z_\alpha(k),$$

其中

$$\bar{K}_1^*=(H_{uu}-H_{uw}H_{ww}^{-1}H_{wu})^{-1}\times$$
$$(H_{uz}-H_{uw}H_{ww}^{-1}H_{wz}),$$
$$\bar{K}_2^*=(H_{ww}-H_{wu}H_{uu}^{-1}H_{uw})^{-1}\times$$
$$(H_{wz}-H_{wu}H_{uu}^{-1}H_{uz}).$$

于是,关于Q-函数的贝尔曼方程为

$$Q(k)=z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)+u^{\mathrm{T}}(k)Ru(k)-$$
$$\gamma^2w^{\mathrm{T}}(k)w(k)+\lambda Q(k+1). \tag{30}$$

任意给定$Q_0(k)\geqslant 0,Q_j(k)=Z^{\mathrm{T}}(k)H^jZ(k)$,则由$H^j$计算可得$\bar{K}_1^j$和$\bar{K}_2^j$,于是贝尔曼方程(30)可进行迭代更新$Q$-函数,有

$$Z^{\mathrm{T}}(k)H^{j+1}Z(k)=$$
$$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)+u^{j\mathrm{T}}(k)Ru^j(k)-\gamma^2w^{j\mathrm{T}}(k)w^j(k)+$$
$$\lambda Z^{\mathrm{T}}(k+1)H^{j+1}Z(k+1). \tag{31}$$

由于矩阵Q_z中含有系统参数信息,不能直接用于计算$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)$. 下一步,将对$z_\alpha(k)$进行分割,剥离出$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)$中的已知部分和未知部分. 令

$$z_\alpha(k)=\begin{bmatrix}z_{\alpha1}(k)\\ z_{\alpha2}(k)\\ r(k)\end{bmatrix},\ M=\begin{bmatrix}I & \bar{M} & 0\\ 0 & 0 & I\end{bmatrix}.$$

其中:$z_{\alpha1}(k)$是$z_\alpha(k)$中从第1行到第n行元素构成的向量,$z_{\alpha2}(k)$为$z_\alpha(k)$除去$z_{\alpha1}(k)$和$r(k)$部分余下的向量. 显然,当$\alpha(k)=0$时,$z_{\alpha1}(k)=x(k)$,当$\alpha(k)=1$时,$z_{\alpha1}(k)=0$. 按照这种分割方法,方程(31)右边的第1项可扩展为

$$z_\alpha^{\mathrm{T}}(k)Q_zz_\alpha(k)=$$
$$z_{\alpha1}^{\mathrm{T}}(k)C^{\mathrm{T}}QCz_{\alpha1}(k)+z_{\alpha2}^{\mathrm{T}}(k)\bar{M}^{\mathrm{T}}C^{\mathrm{T}}QC\bar{M}z_{\alpha2}(k)-$$
$$-2r^{\mathrm{T}}(k)F^{\mathrm{T}}QCz_{\alpha1}(k)-2r^{\mathrm{T}}(k)F^{\mathrm{T}}QC\bar{M}z_{\alpha2}(k)+$$
$$r^{\mathrm{T}}(k)F^{\mathrm{T}}QFr(k).$$

令

$$\theta_1(k)=Z^{\mathrm{T}}(k)\otimes Z^{\mathrm{T}}(k)-\lambda Z^{\mathrm{T}}(k+1)\otimes Z^{\mathrm{T}}(k+1),$$
$$\theta_2(k)=z_{\alpha2}^{\mathrm{T}}(k)\otimes z_{\alpha2}^{\mathrm{T}}(k),$$
$$\theta_3(k)=z_{\alpha2}^{\mathrm{T}}(k)\otimes r^{\mathrm{T}}(k),$$
$$\theta(k)=[\theta_1(k),-\theta_2(k),2\theta_3(k)].$$

经过简单的计算,方程(31)可转化为如下形式:

$$\theta(k)\begin{bmatrix}\mathrm{vec}(H^{j+1})\\ \mathrm{vec}(\bar{M}^{\mathrm{T}}C^{\mathrm{T}}QC\bar{M})\\ \mathrm{vec}(F^{\mathrm{T}}QC\bar{M})\end{bmatrix}=$$

$$u^{j\mathrm{T}}(k)Ru^j(k)-\gamma^2 w^{j\mathrm{T}}(k)w^j(k)+r^{\mathrm{T}}(k)F^{\mathrm{T}}QFr(k)+ z_{\alpha 1}^{\mathrm{T}}(k)C^{\mathrm{T}}QCz_{\alpha 1}(k)-2r^{\mathrm{T}}(k)F^{\mathrm{T}}QCz_{\alpha 1}(k):=\phi^j(k), \tag{32}$$

进而方程(32)可简写为

$$\theta(k)\mathrm{vec}(\bar{H}^{j+1})=\phi^j(k), \tag{33}$$

其中

$$\mathrm{vec}(\bar{H}^{j+1})=\begin{bmatrix}\mathrm{vec}(H^{j+1})\\ \mathrm{vec}(\bar{M}^{\mathrm{T}}C^{\mathrm{T}}QC\bar{M})\\ \mathrm{vec}(F^{\mathrm{T}}QC\bar{M})\end{bmatrix}.$$

取初始时刻 $k_0\geqslant n+q$,并搜集 s 组数据,则

$$\Theta\mathrm{vec}(\bar{H}^{j+1})=\Phi^j. \tag{34}$$

其中

$$\Theta=[\theta^{\mathrm{T}}(k_0),\dots,\theta^{\mathrm{T}}(k_0+s)]^{\mathrm{T}},$$
$$\Phi=[\phi^{j\mathrm{T}}(k_0),\dots,\phi^{j\mathrm{T}}(k_0+s)]^{\mathrm{T}}.$$

注意到矩阵 Θ 里有许多零列,可以经过一系列初等列变换,将 Θ 所有零列换到最后几列,即

$$\Theta L=[\bar{\Theta},\ 0],$$

其中 L 是初等列变换矩阵. 此时,方程(34)可转化为

$$\bar{\Theta}\mathrm{vec}(\bar{H}^{j+1})_{L1}=\Phi^j, \tag{35}$$

其中

$$\begin{bmatrix}\mathrm{vec}(\bar{H}^{j+1})_{L1}\\ \mathrm{vec}(\bar{H}^{j+1})_{L2}\end{bmatrix}=L^{-1}\mathrm{vec}(\bar{H}^{j+1}).$$

基于方程(35),在 $\bar{\Theta}^{\mathrm{T}}\bar{\Theta}$ 满秩的条件下,可得

$$\mathrm{vec}(\bar{H}^{j+1})_{L1}=(\bar{\Theta}^{\mathrm{T}}\bar{\Theta})^{-1}\bar{\Theta}^{\mathrm{T}}\Phi^j.$$

注意到,计算量测反馈控制器增益和量测反馈干扰增益所需要的信息都包含在 $\mathrm{vec}(\bar{H}^{j+1})_{L1}$ 中,因此,可以利用 $\mathrm{vec}(\bar{H}^{j+1})_{L1}$ 重构矩阵 H^{j+1},而 H^{j+1} 中未知的元素可选择任意值. 在此情形下,量测反馈控制器增益和量测反馈干扰增益可按如下方式更新:

$$\bar{K}_1^{j+1}=(H_{uu}^{j+1}-H_{uw}^{j+1}(H_{ww}^{j+1})^{-1}H_{wu}^{j+1})^{-1}\times (H_{uz}^{j+1}-H_{uw}^{j+1}(H_{ww}^{j+1})^{-1}H_{wz}^{j+1}), \tag{36}$$

$$\bar{K}_2^{j+1}=(H_{ww}^{j+1}-H_{wu}^{j+1}(H_{uu}^{j+1})^{-1}H_{uw}^{j+1})^{-1}\times (H_{wz}^{j+1}-H_{wu}^{j+1}(H_{uu}^{j+1})^{-1}H_{uz}^{j+1}). \tag{37}$$

由以上推导可知,利用历史数据在线求解鞍点的Q-learning算法可归纳为算法1.

算法1　基于Q-learning在线求解鞍点的算法.

step 1: 在时间区间 $[k_0,k_0+s]$ 内,选取初始量测反馈控制策略 $\bar{K}_1^0$ 和量测反馈干扰策略 $\bar{K}_2^0$ 为稳定策略,选取 $H^0=0$,计算 Θ 和 $z_\alpha(k)$;

step 2: 选取常数 $\epsilon>0$, 令迭代指标数 $j=0$;

step 3: 重复;

step 4: 求解方程(35)得到 $\mathrm{vec}(\bar{H}^{j+1})_{L1}$,并重构 H^{j+1};

step 5: 求解方程(36)得到 $\bar{K}_1^{j+1}$;

step 6: 求解方程(37)得到 $\bar{K}_2^{j+1}$;

step 7: $j\leftarrow j+1$;

step 8: 直到 $\|H^j-H^{j-1}\|<\epsilon$;

step 9: $j^*\leftarrow j$;

step 10: 利用 $u(k)=-\bar{K}_1^{j^*}z_\alpha(k)$ 作为近似最优的 H_∞ 控制输入.

关于求解二人零和博弈问题的Q-learning算法收敛性可参见文献[20]. 通过Q-learning算法求解所求得的二人零和博弈问题解,可以以任意精度逼近鞍点,并且无需系统模型的参数信息. 换言之,H_∞ 控制器能够通过在线的无模型Q-learning方法求解.

注2　由方程(27)可知矩阵 H 的维数为 $[(n+m+l)(\bar{n}_\alpha+1)+q][(n+m+l)(\bar{n}_\alpha+1)+q]$. 为了保证最小二乘法的应用,在学习的过程中需要构造一组数据集,其长度至少为 $[(n+m+l)(\bar{n}_\alpha+1)+q]^2/2$,以保证算法中相关矩阵满足秩的条件,才能开始迭代过程.

3　数值仿真和实验验证

本节利用雕刻机平台(如图2)对算法1的有效性进行验证.

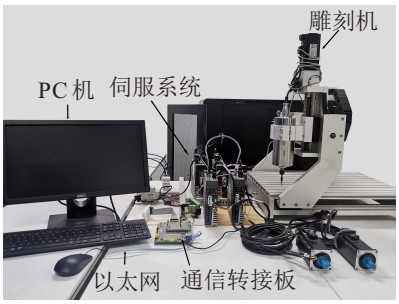

图 2　雕刻机平台

根据文献[27]中的雕刻机模型,并将其简化为单轴形式,取采样时间为 $T=0.01\,\mathrm{s}$,则离散后带有扰动的雕刻机模型为

$$x(k+1)=\begin{bmatrix}1 & 0.010\,0\\ 0 & 0.484\,2\end{bmatrix}x(k)+\begin{bmatrix}0\\ 0.044\,1\end{bmatrix}u(k)+\begin{bmatrix}0\\ 0.01\end{bmatrix}w(k),$$
$$y(k)=[1\ 0]x(k),$$

其中 $x(k)$ 的两个分量分别代表雕刻机刀头的位置和速度. 系统的输出是刀头的位置. 当信息在S-C通道进行传输时,可能会遭受DoS攻击,假设攻击诱导的最大连续丢包数为 $\bar{n}_\alpha=2$.

3.1 数值仿真

选取性能指标参数$Q=100, R=1$,选取外部系统参数$E=1, F=1$. 给定期望位置$r(k)=10$. 为便于比较,首先根据方程(18)和(19)计算最优的量测反馈控制器增益和最差的量测反馈干扰增益分别为

$$\begin{aligned}&\bar{K}_1^*=\\&[-5.676\,7\ \ -0.109\,6\ \ -5.676\,7\ \ -0.109\,8\rightarrow\\&\leftarrow -5.676\,7\ \ -0.109\,9\ \ -0.004\,8\ \ -0.004\,8\rightarrow\\&\leftarrow -0.001\,1\ \ -0.001\,1\ \ -5.676\,7\,],\\&\bar{K}_2^*=[1.287\,2\quad 0.024\,8\quad 1.287\,2\quad 0.024\,9\rightarrow\\&\qquad\leftarrow 1.287\,2\quad 0.024\,9\quad 0.001\,1\ \ 0.001\,1\rightarrow\\&\qquad\leftarrow -0.000\,2\quad -0.000\,2\quad -1.287\,2].\end{aligned}$$

然后,选择稳定的初始量测反馈控制器增益$\bar{K}_1^0$和初始量测反馈干扰增益$\bar{K}_2^0$分别为

$$\begin{aligned}&\bar{K}_1^0=\\&[-2.000\,0\ \ -0.010\,0\ \ -2.000\,0\ \ -0.024\,8\rightarrow\\&\leftarrow -2.000\,0\ \ -0.032\,0\ \ -0.000\,4\ \ -0.001\,1\rightarrow\\&\leftarrow -0.000\,1\ \ -0.000\,2\ \ 4.000\,0]\\&\bar{K}_2^0=[1.500\,0\quad 0.010\,0\quad 1.500\,0\quad 0.019\,8\rightarrow\\&\qquad\leftarrow 1.500\,0\quad 0.024\,6\quad 0.000\,4\quad 0.000\,8\rightarrow\\&\qquad\leftarrow 0.000\,1\quad -0.000\,2\quad -2.000\,0].\end{aligned}$$

算法1经过3次迭代后,分别发生在第404、第952和第1 391个采样时刻点,H^j中与量测反馈控制器增益有关的矩阵块分别收敛到最优的矩阵块,如图3所示.

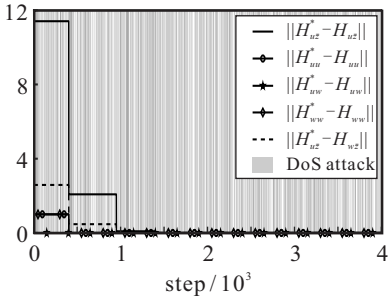

图3 H^j和H^*中相对应的分块矩阵的误差

同时,增益$\bar{K}_1^j$和$\bar{K}_2^j$分别收敛到

$$\begin{aligned}&\bar{K}_1^3=\\&[-5.676\,8\ \ -0.109\,4\ \ -5.676\,8\ \ -0.109\,8\rightarrow\\&\leftarrow -5.676\,8\ \ -0.109\,9\ \ -0.004\,8\ \ -0.004\,8\rightarrow\\&\leftarrow -0.001\,1\ \ -0.001\,1\ \ -5.676\,7],\\&\bar{K}_2^3=[1.287\,2\quad 0.024\,6\quad 1.287\,2\quad 0.024\,8\rightarrow\\&\qquad\leftarrow 1.287\,2\quad 0.024\,8\quad 0.001\,1\quad 0.001\,1\rightarrow\\&\qquad\leftarrow -0.000\,2\quad -0.000\,2\quad -1.287\,2].\end{aligned}$$

关于增益$\|\bar{K}_1^j-\bar{K}_1^*\|$和$\|\bar{K}_2^j-\bar{K}_2^*\|$的变化趋势如图4所示.

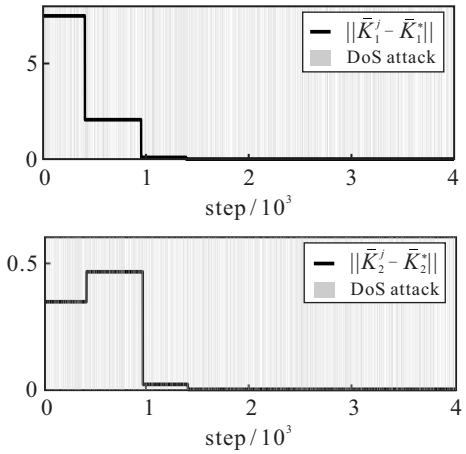

图4 $\|\bar{K}_1^j-\bar{K}_1^*\|$和$\|\bar{K}_2^j-\bar{K}_2^*\|$的变化趋势

注意到,在执行算法的过程中,系统存在随机丢包现象(如图5),且不同时刻连续丢包数不同. 此外,在系统学习的过程中,随机探测信号被加到控制输入$u(k)$中,以使系统受到充分的激励. 由图6可知,在第

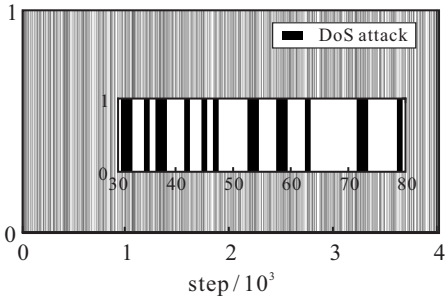

图5 系统丢包情况

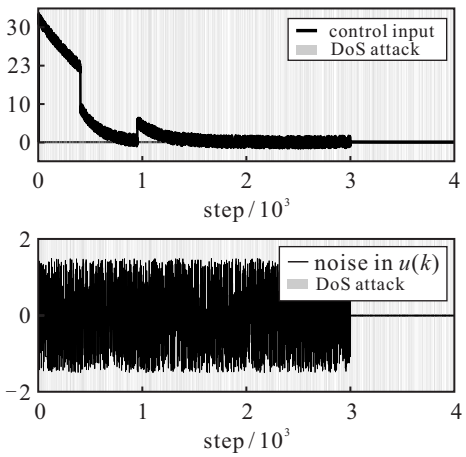

图6 控制输入和探测信号

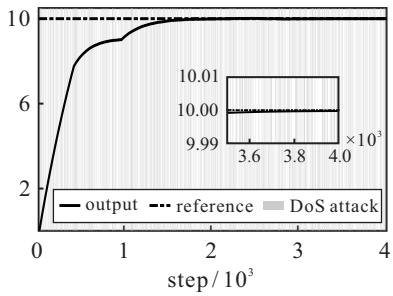

图7 系统输出轨迹和参考信号轨迹

1 391个采样时刻之后,尽管DoS攻击依然随机发生,但控制器增益不再进行更新;在第3 000个采样时刻后,探测信号不再加入到控制输入中,控制输入就渐近收敛到零. 在这样的环境下,由图7可知,系统的输出依然能够渐近跟踪参考信号,验证了所提出算法的有效性.

3.2　实验验证

雕刻机的机械部分主要由伺服电机、伺服驱动器以及相关支持组件组成. 为了实时收到电机的信息,基于嵌入式系统的接口设备被用来作为信号的中继站. 电机的物理信息(包括位置和速度等)将通过内置在伺服驱动器中的传感器以固定频率采样获得,然后通过CANopen协议被发送到接口板. 接口板与PC机之间通过以太网进行数据传输,即接口板收到伺服驱动器的数据后,立即将数据转发给PC机.

在实验平台上,设置与仿真环境同样的算法参数,给平台赋予同样初始量测反馈控制器增益和量测反馈干扰增益. 经过2次迭代,增益分别收敛到

$$\begin{aligned}&\bar{K}_1^2=\\&[-5.677\,6\quad -0.109\,7\quad -5.677\,6\quad -0.109\,9\rightarrow\\&\leftarrow -5.677\,6\quad -0.110\,3\quad -0.004\,8\quad -0.004\,8\rightarrow\\&\leftarrow -0.001\,1\quad -0.001\,1\quad -5.677\,6],\\&\bar{K}_2^2=[1.287\,4\quad 0.024\,9\quad 1.287\,4\quad 0.025\,0\rightarrow\\&\qquad\leftarrow 1.287\,4\quad 0.025\,1\quad 0.001\,1\quad 0.001\,1\rightarrow\\&\qquad\leftarrow -0.000\,2\quad -0.000\,2\quad -1.287\,5].\end{aligned}$$

图8表明了增益$\|\bar{K}_1^j-\bar{K}_1^*\|$和$\|\bar{K}_2^j-\bar{K}_2^*\|$的变化趋势.

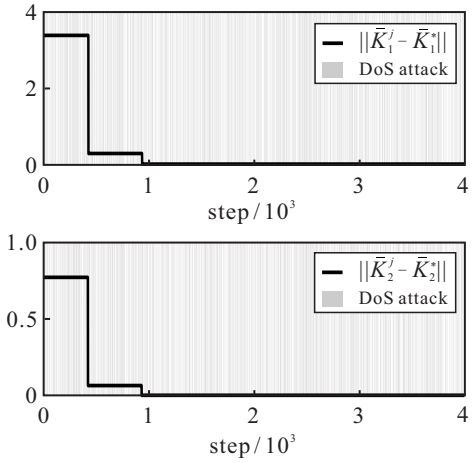

图8　$\|\bar{K}_1^j-\bar{K}_1^*\|$和$\|\bar{K}_2^j-\bar{K}_2^*\|$的变化趋势

实验数据表明,平台在采样时刻423和933处进行了增益的更新,2次更新后,量测反馈控制器增益误差$\|\bar{K}_1^j-\bar{K}_1^*\|$为0.001 9,这与仿真时量测反馈控制器增益的最终误差为0.000 2不同. 在同样的参数设置和环境设置下,造成增益更新次数不一样和收敛误差不一致的主要原因在于实验会受到外部环境因素的影响. 另外,图9刻画了平台的控制输入,图10表明了基于数据驱动的量测反馈控制方法能够稳定实际实验平台,并使得系统平台输出跟踪上参考信号,从而验证了算法的有效性.

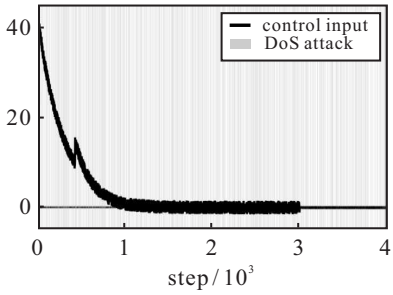

图9　控制输入

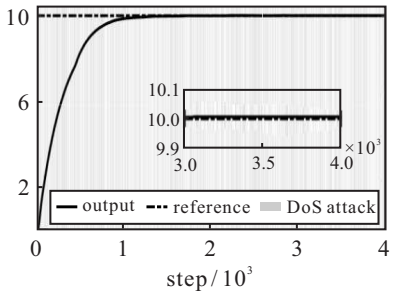

图10　平台输出轨迹和参考信号轨迹

4　结　论

针对DoS攻击下参数未知的CPS安全控制问题,本文提出了无模型的安全H_∞控制方法. DoS攻击导致某些状态量丢失,从而无法实现闭环系统的实时状态反馈. 本文利用历史量测数据构造了带有丢包的Smith预估器,用量测变量替代状态变量,实现了闭环系统的实时反馈. 系统参数的未知导致无法基于模型设计控制器增益,本文利用历史数据设计了量测反馈控制器,并通过无模型的Q-learning算法在线学习和迭代求解GARE获得量测反馈控制器增益,最终实现了系统的无模型H_∞控制. 最后,通过雕刻机的数值仿真和平台实验验证了算法的有效性. 下一步,可将本文的方法延伸至CPS双通道遭受网络攻击的情况或CPS协调一致控制等问题.

参考文献(References)

[1] Ali S, Balushi T A, Nadir Z, et al. Cyber security for cyber physical systems[M]. Cham: Springer, 2018: 1-10.

[2] 艾科网信. 2020年上半年十大网络安全时间[EB/OL]. [2021-02-01]. http://www.acknetworks.com/news/ shownews.php?id=221.

[3] Mahmoud M S, Hamdan M M, Baroudi U A. Modeling and control of cyber-physical systems subject to cyber attacks: A survey of recent advances and challenges[J].

Neurocomputing, 2019, 338: 101-115.

[4] Befekadu G K, Gupta V, Antsaklis P J. Risk-sensitive control under Markov modulated denial-of-service (DoS) attack strategies[J]. IEEE Transactions on Automatic Control, 2015, 60(12): 3299-3304.

[5] Befekadu G K, Gupta V, Antsaklis P J. Risk-sensitive control under a class of denial-of-service attack models[C]. Proceedings of the 2011 American Control Conference. San Francisco, 2011: 643-648.

[6] Orojloo H, Azgomi M A. A game-theoretic approach to model and quantify the security of cyber-physical systems[J]. Computers in Industry, 2017, 88: 44-57.

[7] Wu C W, Wu L G, Liu J X, et al. Active defense-based resilient sliding mode control under denial-of-service attacks[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 237-249.

[8] de Persis C, Tesi P. Input-to-state stabilizing control under denial-of-service[J]. IEEE Transactions on Automatic Control, 2015, 60(11): 2930-2944.

[9] Dolk V S, Tesi P, de Persis C, et al. Event-triggered control systems under denial-of-service attacks[J]. IEEE Transactions on Control of Network Systems, 2017, 4(1): 93-105.

[10] 杨飞生, 汪璟, 潘泉, 等. 网络攻击下信息物理融合电力系统的弹性事件触发控制[J]. 自动化学报, 2019, 45(1): 110-119.
(Yang F S, Wang J, Pan Q, et al. Resilient event-triggered control of grid cyber-physical systems against cyber attack[J]. Acta Automatica Sinica, 2019, 45(1): 110-119.)

[11] Li T X, Chen B, Yu L, et al. Active security control approach against DoS attacks in cyber-physical systems[J]. IEEE Transactions on Automatic Control, 2021, 66(9): 4303-4310.

[12] 孙洪涛, 彭晨, 王志文. DoS攻击下的信息物理系统事件触发预测控制设计[J]. 控制与决策, 2019, 34(11): 2303-2309.
(Sun H T, Peng C, Wang Z W. Event-triggered predictive control of cyber-physical systems under DoS attacks[J]. Control and Decision, 2019, 34(11): 2303-2309.)

[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] Lai S Y, Chen B, Li T X, et al. Packet-based state feedback control under DoS attacks in cyber-physical systems[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2019, 66(8): 1421-1425.

[15] Lu A Y, Yang G H. Input-to-state stabilizing control for cyber-physical systems with multiple transmission channels under denial of service[J]. IEEE Transactions on Automatic Control, 2018, 63(6): 1813-1820.

[16] Sun J, Qi G Q, Zhu Z Q. A sparse neural network based control structure optimization game under DoS attacks for DES frequency regulation of power grid[J]. Applied Sciences, 2019, 9(11): 2217.

[17] Wei Q L, Liu D R, Lin H Q. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems[J]. IEEE Transactions on Cybernetics, 2016, 46(3): 840-853.

[18] Al-Tamimi A, Lewis F L, Abu-Khalaf M. Model-free Q-learning designs for linear discrete-time zero-sum games with application to H_∞ control[J]. Automatica, 2007, 43(3): 473-481.

[19] Valadbeigi A P, Sedigh A K, Lewis F L. H_∞ static output-feedback control design for discrete-time systems using reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(2): 396-406.

[20] Liu Y Y, Wang Z S, Shi Z. H_∞ tracking control for linear discrete-time systems via reinforcement learning[J]. International Journal of Robust and Nonlinear Control, 2020, 30(1): 282-301.

[21] Kiumarsi B, Lewis F L, Modares H, et al. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics[J]. Automatica, 2014, 50(4): 1167-1175.

[22] Qiu X J, Wang Y C, Xie X P, et al. Resilient model-free adaptive control for cyber-physical systems against jamming attack[J]. Neurocomputing, 2020, 413: 422-430.

[23] Huang X, Zhai D, Dong J X. Adaptive integral sliding-mode control strategy of data-driven cyber-physical systems against a class of actuator attacks[J]. IET Control Theory & Applications, 2018, 12(10): 1440-1447.

[24] Basar T, Bernhard P. H_∞ optimal control and related minimax design problems: A dynamic game approach[M]. The 2nd editor. New York: Springer Science & Business Media, 1995: 177-212.

[25] Lewis F L, Vrabie D L, Syrmos V L. Optimal control[M]. Hoboken: John Wiley & Sons, Inc., 2012: 49-106.

[26] Astrom K J, Hang C C, Lim B C. A new Smith predictor for controlling a process with an integrator and long dead-time[J]. IEEE Transactions on Automatic Control, 1994, 39(2): 343-345.

[27] Wu X, Dong H, She J H, et al. High-precision contour-tracking control of ethernet-based networked motion control systems[J]. IEEJ Journal of Industry Applications, 2020, 9(1): 1-10.

作者简介

金丹 (1982−), 女, 博士生, 从事强化学习、最优控制的研究, E-mail: djin@zjut.edu.cn;

吴麒 (1993−), 男, 博士生, 从事系统辨识、网络化运动控制的研究, E-mail: wq93science@163.com;

陈博 (1984−), 男, 教授, 博士生导师, 从事信息融合、安全估计与控制等研究, E-mail: bchen@zjut.edu.cn;

俞立 (1961−), 男, 教授, 博士生导师, 从事网络化控制系统、信息融合等研究, E-mail: lyu@zjut.edu.cn.

(责任编辑：孙艺红)