

# DoS攻击下多线性系统多信道传输调度问题研究

徐鑫, 王慧敏

(东北大学信息科学与工程学院, 沈阳市 110819, 中国)

**摘要:** 近年来,信息物理系统(Cyber-Physical Systems, CPSs)的应用越来越广泛,但由于CPSs本身具备开放性,易遭受网络攻击,并且攻击者越来越智能,因此,有必要开展CPSs安全性的相关研究.本文主要考虑一个具有多个远程状态估计子系统的信息物理系统在DoS攻击下的交互过程.在每个系统中,每个传感器监控各自的系统,并由调度器为各个传感器的数据包分配通道,将其本地估计发送给远程状态估计器,其目标是最小化总估计误差协方差.为了更接近实际应用场景,考虑在多信道传输过程中,通道信号会受到不同环境影响,因此在不同环境的信道传输数据,需要消耗的能量有所不同.调度器和攻击者对于通道的选择,需要满足通道对最低能量的需求才能进行传输和攻击.对于攻击者而言,考虑其更加智能,因此如果对一条通道攻击后仍然有剩余能量并满足其余通道要求,便可同时选择攻击其他通道进行攻击,进而实现和调度器相反的目标.在此基础上,构造一个双人零和博弈,并采用纳什Q学习算法求解双方的最优策略,为研究信息物理系统安全状态估计提供研究思路.

**关键词:** 信息物理系统; DoS攻击; 远程状态估计; 传输调度; 零和博弈

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.0938

引用格式: 徐鑫,王慧敏. DoS攻击下多线性系统多信道传输调度问题研究[J]. 控制与决策.

## Research on Multi-Channel Transmission Scheduling Problem of Multi-Linear Systems under DoS Attacks

Xu Xin, Wang Huimin

(College of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

**Abstract:** In recent years, the application of Cyber-Physical Systems (CPSs) is becoming more and more widespread. Since the openness of CPSs and network attackers are becoming more and more intelligent, it is necessary to study the security of CPSs. This paper mainly considers the interaction process of a CPS with several remote state estimation subsystems under DoS attacks. In each system, each sensor monitors its respective system and a scheduler assigns channels to packets from each sensor and sends its local estimates to the remote state estimator with the goal of minimizing the total estimation error covariance. In order to be closer to the real application scenario, the channel signals are considered to be affected by different environments during multi-channel transmission, so that the energy consumed to transmit data in channels of different environments varies. The scheduler and the attacker need to meet the minimum energy requirements in order to transmit data and attack. For the attacker, it is considered to be more intelligent, so if there is still residual energy after attacking one channel and satisfying the remaining channel requirements, it can choose to attack other channels at the same time, thus achieving the opposite goal as the scheduler. A two-player zero-sum game is constructed and the Nash Q learning algorithm is used to solve the optimal policies of both sides, which provides research ideas for studying the secure state estimation of CPSs.

**Keywords:** Cyber-physical system; DoS attack; Remote state estimation; Transmission scheduling; Zero-sum game

## 0 引言

在未来的科技发展进程中,万物互联将作为其中重要的实现目标之一,信息物理系统(Cyber-Physical systems, CPSs)占据着其中的核心地位<sup>[1]</sup>.在这个前提

下,对于CPS需要有着更高的开放性,然而随着开放性的提高,伴随而来的则是一些恶意攻击者对于CPS系统安全的威胁<sup>[2]</sup>.由于信息物理系统的应用越来越广泛,并且在公共的基础设施建设(如储能电站<sup>[3]</sup>、无人

收稿日期: 2022-05-26; 录用日期: 2022-10-21.

基金项目: 国家自然科学基金面上项目(No.61973062),中央高校基本科研业务费(No.N2104021),广东省基础与应用基础研究基金(No.2022A1515011143).

†通讯作者. E-mail: wanghuimin@ise.neu.edu.cn.

化交通等)中占比越来越高,那么一旦遭到恶意攻击威胁,就很可能导致网络中断、传输受阻、控制失灵等一系列问题,进而会造成无法挽回的影响<sup>[4]</sup>.

CPS所面临的环境具有开放性和复杂性,并且包含着控制层、网络层和物理层,在每一层都容易受到安全威胁<sup>[5]</sup>.文献[6]分别从攻击者和防御者角度出发,系统地分析总结了现有研究结果针对不同类型攻击的攻击策略及防御策略设计.文献[7]考虑了DoS攻击过程会造成数据包丢包的情况,研究了CPS安全控制问题.

目前,针对CPS受到DoS攻击的安全问题,很多是从攻击者角度进行研究,其中多为研究DoS攻击者的最优策略调度问题.针对实际的大型CPS通常采用多信道技术进行数据传输,为此文献[8-10]进行了DoS攻击调度设计.文献[11]提出的DoS攻击,在节约能量消耗<sup>[12]</sup>、增强隐蔽性<sup>[13]</sup>、配置简单<sup>[14]</sup>情况下,设计了基于信道跳变的DoS攻击,并且该攻击具备周期性.在此基础上,Gan等人<sup>[11]</sup>提出了多个创新点,即在周期性攻击下,加入了能量限制.文献[15]针对DoS攻击者的能量分配问题,设计了相关算法进行求解.DoS攻击也会发生在反馈通道上阻止数据信息传递,文献[16]考虑了这种情况并提出了调度策略设计.上述研究成果都是单方面站在DoS攻击者角度对于攻击策略的设计.Li等人<sup>[2]</sup>首次考虑了一种博弈论方法,它提供了一种处理这些交互式决策问题的替代方法.由此开始,在基于博弈论的研究方法下,正式开展了在信息物理系统中关于攻击者和传感器双方在传输方面的调度博弈.文献[5,17]中,DoS攻击被建模为一个二进制过程,包括考虑传感器是否发送数据包以及攻击者是否要发动攻击.为了详细说明交互过程,Li等人<sup>[18]</sup>在2017年将模型扩展到基于SINR的框架中,在该框架中,首次提出传感器和攻击者在单信道交互过程中考虑能量限制的问题,并在此基础上提出了一种求解最优能量控制的方法.文献[4]指出,多信道网络通信可以作为一种有效的技术,通过多个信道传输信号来缓解突发通信流量的问题.2017年,Ding等人<sup>[19]</sup>首次对先前的研究进行改进,采用到了多通道传输模型,在多通道的基础上模拟了先前的互动决策过程.此外,Zhang等人<sup>[20]</sup>在2019年研究了基于博弈论方法设计了单个DoS攻击下的多线性动态系统的多通道传输调度问题.Yuan等人<sup>[21]</sup>在2020年采用了博弈论相关方法研究了DoS攻击下信息物理系统多信道传输的弹性控制问题.

在信息物理系统中,通常集成了大量的关键基

础设施,对于系统中众多传感器的安装通常是分散的<sup>[22,23]</sup>,即存在多个传感器分散放置来监控一个或者多个动态过程.文献[24]考虑了一类多通道多传感器网络系统的远程状态估计问题.文献[25]针对一类线性多子系统在DoS攻击下的安全状态估计问题,提出了基于强化学习的传感器调度策略.本文拟针对多线性时不变系统,以先前博弈交互的研究成果为基础,仍考虑在DoS攻击下,多传感器由调度器统一调度将数据包发送给远程状态估计器的状态估计过程.相较于先前的研究成果,本文考虑了更实际的应用场景,在多信道传输过程中,通道信号会受到不同环境影响,因此在不同环境的信道传输数据,需要消耗的能量有所不同.调度器和攻击者对于通道的选择,需要满足通道对最低能量的需求才能进行传输和攻击.对于攻击者而言,考虑其更加智能,因此如果对一条通道攻击后仍然有剩余能量并满足其余通道要求,便可同时选择其他通道进行攻击.在上述条件下,构建攻防双方的博弈过程,通过多智能体强化学习中的纳什Q学习算法求解出攻防双方的最优策略,为信息物理系统的安全防御提供依据.

## 1 问题描述

### 1.1 系统模型

考虑如下线性时不变系统:

$$\begin{cases} x_i(k+1) = A_i x_i(k) + w_i(k) \\ y_i(k) = C_i x_i(k) + v_i(k), \quad i \in G \end{cases} \quad (1)$$

其中,  $G = \{1, 2, \dots, M\}$ ;  $x_i(k) \in R^{n_i}$ 表示第*i*个子系统的状态向量;  $y_i(k) \in R^{m_i}$ 表示第*i*个子系统的测量输出;  $A_i, C_i$ 表示第*i*个子系统的系统矩阵;  $w_i(k)$ 和 $v_i(k)$ 是第*i*个子系统独立同分布高斯噪声,满足 $E[w_i(k)w_i(j)'] = \delta_{kj}Q_i$  ( $Q_i \geq 0$ ),  $E[v_i(k)v_i(j)'] = \delta_{kj}R_i$  ( $R_i \geq 0$ ),  $E[w_i(k)v_i(j)'] = 0, \forall j, k \in N, i = 1, 2, \dots, N$ ; 初始状态 $x_i(0)$ 是一个高斯零均值随机向量并且满足协方差矩阵 $\Pi_i(0) \geq 0$ ; 假设 $(A_i, C_i)$ 是可观的,  $(A_i, \sqrt{Q_i})$ 是可控的.

### 1.2 局部状态估计

在现代信息物理系统中,传感器通常被设计成“智能”的,以提高系统的估计和控制性能.其中,每个传感器在时间步长*k*处测量到对应的系统后,运行局部卡尔曼滤波,根据到时间*k*为止收集到的所有测量值来估计过程的状态,然后将其局部估计值传递给远程估计器.

根据信息物理系统当前状态下的本地估计值,可以计算出第*i*个子系统的本地状态的最小误差估计

值 $\hat{x}_{i,k}^s$ 和对应的误差协方差矩阵 $P_{i,k}^s$ 如式(2)和(3)所示:

$$\hat{x}_{i,k}^s = E[x_{i,k}|y_{i,1}, y_{i,2}, \dots, y_{i,k}]. \quad (2)$$

$$P_{i,k}^s = E[(x_{i,k} - \hat{x}_{i,k}^s)(x_{i,k} - \hat{x}_{i,k}^s)'|y_{i,1}, y_{i,2}, \dots, y_{i,k}]. \quad (3)$$

通过式(4)所示的卡尔曼滤波算法来求解信息物理系统本地状态的最小均方误差估计值 $\hat{x}_{i,k}^s$ 以及对应的协方差 $P_{i,k}^s$ :

$$\begin{aligned} \tilde{x}_{i,k} &= A_i \hat{x}_{i,k-1}^s, \\ \tilde{P}_{i,k}^s &= A_i P_{i,k-1}^s A_i' + Q_i, \\ K_{i,k} &= \tilde{P}_{i,k}^s C_i' [C_i \tilde{P}_{i,k}^s C_i' + R_i]^{-1}, \\ \hat{x}_{i,k} &= A_i \hat{x}_{i,k-1}^s + K_{i,k} (y_{i,k} - C_i \tilde{x}_{i,k}), \\ P_{i,k}^s &= (I_{n_{x,i}} - K_{i,k} C_i) \tilde{P}_{i,k}^s. \end{aligned} \quad (4)$$

经过卡尔曼滤波算法推导,并定义式(5)所示的算子可求得传感器一侧的稳态误差协方差 $\bar{P}_i$ :

$$\begin{aligned} h_i(X) &\triangleq A_i X A_i' + Q_i, \\ g_i(X) &\triangleq X - X C_i' (C_i X C_i' + R_i)^{-1} C_i X. \end{aligned} \quad (5)$$

稳定状态下的误差协方差可由 $g_i \circ h_i(X)$ 的唯一正半定解给出,其中 $g_i \circ h_i(X)$ 表示的是迭代的函数形式,即 $g_i(h_i(X))$ ;假设卡尔曼滤波器已经达到稳态,即 $P_{i,k}^s = \bar{P}_i$ .

### 1.3 多信道通信

在获得状态估计 $\hat{x}_{i,k}^s$ 之后,每个传感器通过 $N$ 信道通信网络将其作为数据包发送给远程估计器.在实际应用中,由于信号衰落、噪声等因素的存在,数据包可能无法成功到达远程估计器,因此多信道并不总是可靠的<sup>[26]</sup>.为了测量非理想分组传输,引入了包错误率(PER),对于任何调制方案,都随着信噪比(SNR)的降低而降低.对于第 $i$ 个信道,常规 $SNR_i$ 定义为式(6):

$$SNR_i = \frac{p_i^s}{\sigma_i}. \quad (6)$$

其中 $p_i^s$ 是第 $i$ 个信道传感器采用的传输能量, $\sigma_i$ 是第 $i$ 个信道的加性白噪声能量.

DoS攻击者可以阻塞信道以干扰数据包的传输,如果第 $i$ 个信道受到DoS攻击,则第 $i$ 个信道的信噪比 $SNR_i$ 可以重写为信干噪比 $SINR_i$ 如式(7):

$$SINR_i = \frac{p_i^s}{\sigma_i + p_i^a}. \quad (7)$$

其中, $p_i^a$ 是攻击者对第 $i$ 个信道施加的干扰能量.

在本文中,每个传感器需要从 $N$ 个信道中选择一个信道传输数据包.在这种情况下,当多个传感器选择同一信道时,传输的数据包可能会相互干扰.假

设 $k$ 时刻 $M$ 个传感器和DoS攻击者的通道选择分别表示为 $(l_{1,k}, l_{2,k}, \dots, l_{M,k})$ 和 $l_{a,k}$ . $M$ 个传感器采用的传输能量为 $(p_1^s, p_2^s, \dots, p_M^s)$ .

受文献[20]启发,当多个传感器选择了同一条通道进行传输数据包,彼此之间则会造成干扰,进而影响了传输成功率,表现形式如式(8):

$$SINR_{l_i} = \frac{p_i^s}{\sum_{j \neq i} \delta_{l_i l_j} \tau p_j^s + \sigma_{l_i} + p_{l_i}^a}. \quad (8)$$

上述模型是对信干噪比的扩展, $\sum_{j \neq i} \delta_{l_i l_j} \tau p_j^s$ 表示为多传感器在一条通道时候彼此之间产生的干扰,其中的系数 $\tau$ 表示同一信道上不同数据包之间同时传输的干扰程度.

本文中,传感器和远程估计器之间的通信基于正交幅度调制(QAM),从数字通信理论的角度描述PER和SINR之间的关系,对于第 $l_i$ 条通道,有如下式(9)和(10):

$$PER_{l_i} = 2Q(\sqrt{\alpha SINR_{l_i}}). \quad (9)$$

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\eta^2/2) d\eta. \quad (10)$$

假设远程估计器使用循环冗余校验(CRC)来检测符号错误.因此,每个传感器和遥感传感器之间 $\hat{x}_{i,k}^s$ 的传输可以描述如式(11)所示的二元随机过程 $\{\gamma_{i,k}\}$ , $k \in N, i \in G$ :

$$\gamma_{i,k} = \begin{cases} 1 & \text{数据包到达} \\ 0 & \text{其他情况} \end{cases} \quad (11)$$

对于每个传感器,定义 $\lambda_{i,k}$ 作为 $\hat{x}_{i,k}^s$ 数据包的到包率,因此有式(12):

$$\lambda_{i,k} \triangleq P[\gamma_{i,k} = 1] = 1 - PER_{l_i}. \quad (12)$$

假设 $M$ 个数据包的到达是独立的,即:

$$\begin{aligned} P(\gamma_{1,k} = 1, \gamma_{2,k} = 1, \dots, \gamma_{M,k} = 1) \\ = P(\gamma_{1,k} = 1)P(\gamma_{2,k} = 1) \dots P(\gamma_{M,k} = 1). \end{aligned} \quad (13)$$

对于每个传感器, $SINR_{l_i}$ 取决于传感器传输能量、DoS攻击者的干扰能量、多个数据包在一个信道上的拥塞造成的干扰、信道本身的附加噪声.数据包到达率 $\lambda_{i,k}$ 随着 $SINR_{l_i}$ 的增加而增加.然后,传感器和攻击者的不同选择将导致不同的数据包到达率,并进一步影响远程估计器的性能.

**注1** 本文考虑传感器将本地状态估计值 $\hat{x}_{i,k}^s$ 通过网络发送给远程估计器过程中可能遭遇DoS攻击的情况,通常,攻击者通过向无线网络中发送大量冗余数据包阻塞网络,此时,如果传感器传输通道于攻击者攻击通道为同一个信道 $i$ 时,将影响数据包到达率 $\lambda_{i,k}$ .受文献[18-20]启发,本文考虑DoS攻击对传感

器数据传输的影响,将第*i*个信道信噪比( $SNR_i$ )重写为式(7)所示的信号-干扰-噪声比率( $SINR_i$ ),其中加入了攻击者能量信息 $p_i^a$ ,进而影响到包率 $\lambda_{i,k}$ ,并最终影响远程状态估计器的估计性能。

#### 1.4 远程状态估计

当传感器一方将含有状态估计值的数据包发送到远程状态估计器时,远程状态估计器将自身的估计值更新为传感器传输过来的估计值,否则,远程状态估计器将会通过上一次交互得到的估计值进行重新预测,具体方式如式(14):

$$\gamma_{i,k} = \begin{cases} \hat{x}_{i,k}^s & \text{数据包到达} \\ A_i \hat{x}_{i,k-1}^s & \text{其他情况} \end{cases} \quad (14)$$

在上式中, $\hat{x}_{i,k}^s$ 为第*i*个传感器在第*k*个过程中远程状态估计器的估计值, $\hat{x}_{i,k-1}^s$ 为第*i*个传感器在第*k-1*个交互过程的远程状态估计器的估计值。

远程状态估计器的估计误差协方差满足式(15):

$$P_{i,k} \triangleq E[(x_{i,k} - \hat{x}_{i,k}^s)(x_{i,k} - \hat{x}_{i,k}^s)'] \quad (15)$$

因此远程状态估计器的误差协方差 $P_{i,k}$ 服从式(16)所示的递归:

$$P_{i,k} = \begin{cases} \bar{P}_i & \text{数据包到达} \\ h_i(P_{i,k-1}) & \text{其他情况} \end{cases} \quad (16)$$

其中, $P_{i,k-1}$ 表示第*i*个传感器上一个博弈交互过程对应的远程状态估计器的估计误差协方差, $h_i(\cdot)$ 表示卡尔曼滤波算法中的误差协方差预测函数,因此第*i*个传感器对应的取值集合为 $\{\bar{P}_i, h_i(\bar{P}_i), h_i^2(\bar{P}_i), \dots\}$ 。

远端状态估计器的估计值的误差协方差的期望值如式(17)所示:

$$E[P_{i,k}] = \lambda_{i,k} \bar{P}_i + (1 - \lambda_{i,k}) h_i(E[P_{i,k-1}]) \quad (17)$$

#### 1.5 问题总结

调度器的目标是帮助远程估计器在不浪费能源的情况下获得足够准确的状态估计.相反,DoS攻击者的目的是通过干扰数据包传输来降低远程估计器的性能.我们假设在线信息可用,也就是说,调度器和攻击者都将根据当前状态和以前的数据做出决策.在目标相反的情况下,我们将通过博弈方法研究能量受限的多通道通信网络中多个传感器和恶意攻击者的决策过程。

## 2 求解策略

### 2.1 提出求解方案

传感器可以通过反馈机制获得数据分组是否

成功发送到远程估计器的信息,以及估计误差协方差 $P_{i,k}$ 的完整知识.同时,攻击者还可以通过数据嗅探器推断知识 $P_{i,k}$ .传感器和攻击者根据实时信息在多通道通信网络中做出合理的决策.本节将开发一种博弈方法来模拟传感器和攻击者之间的交互决策过程,并使用纳什Q学习算法进行求解使传感器和攻击者获得最优策略。

传感器在调度器的调度下旨在帮助远程估计器获得准确的状态估计,而攻击者则试图降低远程估计器的性能.在目标相反的情况下,调度器和攻击者分别尝试最小化和最大化彼此的目标,除此之外也需要考虑能量消耗,因此,从攻击者角度得到目标函数表达式如式(18):

$$J_A \triangleq \sum_{k=1}^{+\infty} \beta^k \left[ \sum_{i=1}^M Tr\{E[P_{i,k}]\} + \delta_s \sum_{i=1}^M p_{i,k}^s - \delta_a \sum_{i=1}^N p_{i,k}^a \right] \quad (18)$$

式中 $\delta_s, \delta_a \geq 0$ 是加权参数,攻击者寻求最小化其能量消耗,而不仅仅是最大化误差协方差.此外,攻击者还需要消耗一定的传感器能量.另一方面,调度器具有完全相反的目标函数表达式,如式(19)所示:

$$J_S = -J_A \triangleq \sum_{k=1}^{+\infty} \beta^k \left[ - \sum_{i=1}^M Tr\{E[P_{i,k}]\} - \delta_s \sum_{i=1}^M p_{i,k}^s + \delta_a \sum_{i=1}^N p_{i,k}^a \right] \quad (19)$$

双方的目标都是最大限度地发挥各自的目标功能.接下来将构造一个双人零和博弈来模拟这个交互过程。

### 2.2 构建马尔可夫决策过程

定义一个游戏 $\mathcal{G} = \langle \mathcal{L}, \mathcal{S}, \mathcal{M}, \mathcal{P}, \mathcal{J} \rangle$ ,其中, $\mathcal{L}$ 表示参与博弈的两方; $\mathcal{S}$ 表示参与博弈双方各自的动作集合,即 $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2)$ ;博弈的状态空间定义为 $\mathcal{M}$ ;状态转移概率定义为 $P(m_{k+1}|m_k, s) \in \mathcal{P} : \mathcal{M}_k \times \mathcal{S} \rightarrow \mathcal{M}_{k+1}$ ,表示当前状态 $m_k$ 采取 $\mathcal{S}$ 集合中的*s*动作转移到下一个状态 $m_{k+1}$ 的概率;对于每个参与者,都能够从游戏中获得奖励,即 $\mathcal{J} : \mathcal{M} \times \mathcal{S} \rightarrow \mathcal{R}$ .五元组的详细定义如下:

玩家:鉴于多个传感器具有相同的目标函数,我们将多个传感器经由调度器统一调配,表示为 $\mathcal{L}_1$ .相反,攻击者具有相反的目标函数.因此,我们将攻击者视为另一个玩家 $\mathcal{L}_2$ .假定 $\mathcal{L}_1$ 和 $\mathcal{L}_2$ 为两个理性参与者。

动作:在每个时刻*k*,每个传感器都需要选择通过哪个通道发送其数据包,该过程由调度器控制.同时,攻击者需要根据其自身的能量分配来选择一个或者多个通道来发起DoS攻击.因此,在时刻*k*,传感

器的动作表示为:  $\vec{l}_k^s = (l_{1,k}^s, l_{2,k}^s, \dots, l_{M,k}^s)$ , 其中关于  $l_{i,k}^s, i \in \{1, 2, \dots, M\}, l_{i,k}^s \in \{1, 2, \dots, N\}$  表示第  $i$  个传感器在时间  $k$  选择第  $l_{i,k}^s$  信道; 假设DoS攻击者在  $k$  时刻选择  $z$  个通道进行攻击, 相应操作为  $\vec{l}_k^a = (l_{1,k}^a, l_{2,k}^a, \dots, l_{z,k}^a)$ , 其中  $l_{i,k}^a \in \{1, 2, \dots, N\}$  表示DoS攻击者在  $k$  时刻选择第  $l_{i,k}^a$  信道. 双方采用的动作在游戏中将统一表示  $(\vec{l}_k^s, \vec{l}_k^a) \in \mathcal{S}$ . 相较于已有的研究成果, 本文在动作选择方面考虑需要根据实际环境的能量限制来决定哪些通道可以选择, 而不再是任意通道均可供传感器和攻击者使用, 在后续的算法求解过程中进行了约束. 因此需定义  $M$  个传感器采用的传输能量为  $(p_1^s, p_2^s, \dots, p_M^s)$ ,  $p^a$  是DoS攻击者使用的总干扰能量,  $p_i^a$  是DoS攻击者对第  $i$  个通道使用的干扰能量,  $N$  个通道各自的能量限度为  $(p_1^c, p_2^c, \dots, p_N^c)$ . 对于第  $i$  个传感器选择第  $j$  个通道, 需要保证  $p_i^s \geq p_j^c$  才可以保证采用  $j$  通道进行传输. 对于攻击者同理, 但是如果攻击者选择  $j$  通道进行攻击, 除了保证  $p^a \geq p_j^c$  以外, 如果  $p^a - p_j^c > 0$  并且当前时刻剩余能量可满足对其他通道进行攻击, 那么则可以继续进行选择通道攻击, 相较于已有的研究结果, 攻击者采用了多通道攻击方式并且更符合实际.

**状态:** 博弈交互的每个过程中, 各个传感器都需要知道当前的估计误差协方差, 表示为  $P_{k_2} = h^{k_2-k_1}(\bar{P})$ , 其中,  $P_{k_2}$  表示第  $k_2$  博弈回合时的远程状态估计器的误差协方差,  $k_1$  是传感器一方发送的数据包最近一次成功到达远程状态估计器的博弈回合数; 因此每个回合远程状态估计器的误差协方差可以表示为  $P_k = h^{i-1}(\bar{P})$ , 其中,  $i$  为博弈交互过程中传感器一方发送数据包连续没有成功到达远程状态估计器的回合数, 因此取值集合有限, 在游戏中表示为  $m_k = [m_1(k), m_2(k), \dots, m_M(k)]$ , 为了描述状态更简便, 可以将状态定义为  $m_i(k) \triangleq k - \max_{0 \leq k' \leq k} \{k' : \gamma_{i,k'} = 1\}$ , 表示当前时刻  $k$  和传感器  $i$  的数据包成功到达远程估计器的最近时间之间的间隔.

**状态转移概率:** 在时刻  $k$ , 状态  $m_i(k)$  的状态可能为  $Z_{i,k} = \{0, 1, 2, \dots, k\}$ . 定义从状态集合  $Z_{i,k-1}$  到  $Z_{i,k}$  的转移概率如下:

$$T_{i,k}(a, b) \triangleq P[m_i(k+1) = b | m_i(k) = a] = \begin{cases} \lambda_{i,k}, & b = 0 \\ 1 - \lambda_{i,k}, & b = a + 1 \\ 0, & \text{其它} \end{cases} \quad (20)$$

其中  $T_{i,k}(a, b)$  表示  $a + 1$  行,  $b + 1$  列的矩阵, 并且  $a \in \{0, 1, 2, \dots, k-1\}$  和  $b \in \{0, 1, 2, \dots, k\}$  是状态  $Z_{i,k-1}$  和  $Z_{i,k}$  可能的索引值.  $k$  时刻游戏状态被描述

为  $m_k \in Z_{1,k} \times Z_{2,k} \times \dots \times Z_{M,k}$ . 根据上述矩阵, 当状态为  $m_{k-1}$  时, 下一个状态有  $2^M$  种取值, 为了更好的说明状态转移概率, 我们定义集合  $\xi_1 = \{j | m_j(k) = 0\}$  以及集合  $\xi_2 = \{G - \xi_1\}$ , 因此状态转移概率如式(21):

$$P(m_k | m_{k-1}, \vec{l}_k^s, \vec{l}_k^a) = \begin{cases} \prod_{i=1}^M \lambda_{i,k}, & \xi_1 = G \\ \prod_{i \in \xi_1} \lambda_{i,k} \prod_{j \in \xi_2} (1 - \lambda_{j,k}), & \xi_1 \subset G \\ \prod_{i=1}^M (1 - \lambda_{i,k}), & \xi_1 = \emptyset \end{cases} \quad (21)$$

**能量限制:** 由于传感器和攻击者的电池有限, 能量限制在信息物理系统中很常见. 本文的目标是在能源成本和系统性能之间实现理想的折衷. 为了简单起见, 我们固定了不同信道的传输能量水平  $p_j^c (j = 1, 2, \dots, N)$  和攻击能量水平  $p^a$ .

**奖励:** 对于调度器来说, 参与了一轮游戏之后, 便会获得相应的奖励, 因此, 对于玩家  $\mathcal{L}_1$  (调度器) 的单步奖励描述为:

$$r_k(m_{k-1}, \vec{l}_k^s, \vec{l}_k^a) = - \sum_{i=1}^M Tr\{E[P_{i,k}]\} - \delta_s \sum_{i=1}^M p_{i,k}^s + \delta_a \sum_{i=1}^N p_{i,k}^a \quad (22)$$

其中,  $m_{k-1}$  表示前一时刻的状态,  $\vec{l}_k^s$  表示调度器的动作选择,  $\vec{l}_k^a$  表示攻击者的动作选择,  $p_{i,k}^s$  表示传感器消耗的能量,  $p_{i,k}^a$  表示攻击者的能量消耗,  $-\sum_{i=1}^M Tr\{E[P_{i,k}]\}$  用来表征对于状态的影响. 对于参与方  $\mathcal{L}_2$  (攻击者) 的奖励正好是相反的.

**注2** 假设转移概率和奖励是静态的, 与时间指标无关, 这也为后面的求解提供了一些便捷.

在上述零和博弈过程构建好之后, 将采用博弈论的相关理论定理以及多智能体强化学习当中的纳什Q学习算法对该过程进行求解, 经过多次的蒙特卡罗迭代, 得到双方的最优传输策略.

### 2.3 求解博弈

由于目标函数的计算复杂, 求解在实践中具有挑战性. 有一些众所周知的方法可以解决这个问题, 例如, 值迭代和策略迭代. 这些方法很有效, 但需要了解过渡函数以及环境中所有状态的回报. 但是, 传感器和攻击者可能无法访问此类信息或执行计算. 因此, 为了解决信息的局限性或计算问题, 我们通过无模型强化学习的方法获得最优值. 为了解决随机博弈的策略设计问题, 首先需要引入随机博弈的均衡, 最终将策略设计问题转为求解博弈的纳什均衡问题. 下面给出纳什均衡定义以及相关定理:

**定义1** <sup>[27]</sup> 在给定的一个有  $n$  个玩家的游戏  $\mathcal{G}$  中,  $S_i$  表示为参与人  $i$  的策略集, 定义  $S =$

$S_1 \times S_2 \times \dots \times S_n$ 为总策略集合.定义  $J \triangleq (j_1(\pi), j_2(\pi), \dots, j_n(\pi))$ ,  $\pi \in S$  表示为所有参与者的收益,其中  $j_i(\pi)$  为参与人  $i$  的收益函数.对于每个参与人  $i \in \{1, 2, \dots, n\}$ , 在策略集合  $\pi = (\pi_1, \pi_2, \dots, \pi_n)$  中选择策略  $\pi_i$ , 得到收益  $j_i(\pi)$ . 如果  $\pi^* \in S$  作为一个纳什均衡策略<sup>[25]</sup>, 就应该满足式(23):

$$j_i(\pi_i^*, \pi_{-i}^*) \geq j_i(\pi_i, \pi_{-i}^*), \forall i, \pi_i \in S_i \quad (23)$$

其中  $\pi_{-i}$  是除参与人  $i$  外的所有参与人的策略集合. 当不等式(23)对所有参与人和所有可行策略都是严格成立的(用  $>$  替代  $\geq$ ) 时, 该均衡为严格纳什均衡.

**定理1** <sup>[27]</sup> 对于任何具有有限策略集的对策, 该对策中至少存在一个混合策略纳什均衡.

**定理2** 在博弈论框架下考虑CPS中调度器和攻击者的交互决策过程, 双方的最优策略构成了这个博弈的纳什均衡.

**证明** 针对所有智能传感器, 其纯策略最多有  $L = N^M$  种 ( $N^M$  为不考虑其能量限制时), 这些纯策略分别表示为  $\vec{l}_1^{s,pure}, \vec{l}_2^{s,pure} \dots \vec{l}_L^{s,pure}$ ; 智能传感器的混合策略为  $\vec{l}^{s,mixed}(\pi_1, \pi_2, \dots, \pi_L)$ , 表示  $\vec{l}_m^{s,mixed} = \vec{l}_m^{s,pure}$  的概率为  $\pi_m$ , 并且  $0 \leq \pi_m \leq 1, \sum_{m=1}^L \pi_m = 1$ . 同理, 对于智能攻击者, 在其总能量一定条件下, 纯策略最多有  $D = A_N^N + A_N^{N-1} + \dots + A_N^1$  种, 分别表示为  $\vec{l}_1^{a,pure}, \vec{l}_2^{a,pure} \dots \vec{l}_D^{a,pure}$ ; 智能攻击者的混合策略为  $\vec{l}^{a,mixed}(\mu_1, \mu_2, \dots, \mu_D)$ , 表示  $\vec{l}_m^{a,mixed} = \vec{l}_m^{a,pure}$  的概率为  $\mu_m$ , 并且  $0 \leq \mu_m \leq 1, \sum_{m=1}^D \mu_m = 1$ .

根据定理1, 对于存在有限策略集的任何博弈过程中至少存在一个混合策略纳什均衡解. 证明成立.  $\square$

根据以上分析, 我们可求出式(18)和(19)的最优解, 即双方的最优策略, 并总是作为双方博弈的纳什均衡存在. 求解该博弈过程的纳什均衡解将使用如下纳什Q学习算法, 具体算法的实现如算法1所示, 其中  $\varepsilon$  为一个充分小的正数, 表征收敛精度.

**算法1** step 1:  $k = 0$  时初始化Q表、双方动作、状态.

step 2: 循环迭代  $k = 1$  到 end.

step 3: 双方观察状态  $m$  和确定动作  $\vec{l}_k^s$  和  $\vec{l}_k^a$ .

step 4: 根据式(23)计算即时奖励.

step 5: 当  $\|Q_{k+1} - Q_k\| < \varepsilon$  时, 得到单步纳什值以及纳什均衡解:

$$\begin{aligned} NashQ_k(m') = & \max_{\pi_s} \min_{\pi_a} \sum_{\vec{l}_k^s, \vec{l}_k^a} Q_k(m_k, \vec{l}_k^s, \vec{l}_k^a) \\ & \times \pi_s(\vec{l}_k^s) \pi_a(\vec{l}_k^a) \end{aligned}$$

step 6: 更新Q表:

$$\begin{aligned} Q_{k+1}(m, \vec{l}_k^s, \vec{l}_k^a) = & (1 - \alpha_k) Q_k(m, \vec{l}_k^s, \vec{l}_k^a) \\ & + \alpha_k (r(m, \vec{l}_k^s, \vec{l}_k^a) \\ & + \beta NashQ_k(m')) \end{aligned}$$

step 7: 根据  $\epsilon - greedy$  算法选取下一步动作:

$$(\vec{l}_k^s, \vec{l}_k^a) = \begin{cases} \text{随机选择,} & ex \in [0, \epsilon] \\ \text{纳什均衡策略,} & ex \in (\epsilon, 1] \end{cases}$$

step 8: 更新状态和动作.

step 9:  $k$  是否为 end, 是则结束循环, 否则继续 step 3.

下面将给出纳什Q学习过程的收敛条件.

**定理3** 纳什Q学习过程在满足下面条件的情况下必然会收敛到最优值:

(1) 对于经历每个状态  $m \in \mathcal{M}$ , 都会有不同的联合动作  $(\vec{l}_k^s, \vec{l}_k^a) \in \mathcal{S}$ , 在学习过程, 状态和联合动作构成的实数对被无限次访问, 并且参与者仅更新与当前状态和动作相对应的Q值.

(2) 对于更新表达式中的学习率  $\alpha_k$  满足  $\alpha_k \in [0, 1), \sum_{k=0}^{+\infty} \alpha_k = +\infty, \sum_{k=0}^{+\infty} \alpha_k^2 < +\infty$ .

上述定理在文献[28]中进行过证明, 在仿真实例部分, 将进一步说明如何实现收敛条件. 对于每个  $i \in G$ , 将设置最终状态  $h^K(\bar{P}_i)$ , 即在学习过程中, 如果  $j \geq K$ , 则有  $h^j(\bar{P}_i) = h^K(\bar{P}_i)$ .

**注3** 本文通过构造一个双人零和博弈来模拟双方的交互过程, 同时在注2的假设条件下, 建立了纳什Q学习算法1来求解该博弈过程的纳什均衡解, 求解过程中, 博弈双方需要已知信道能量上限  $(p_1^c, p_2^c, \dots, p_N^c)$  以及传感器传输能量  $(p_1^s, p_2^s, \dots, p_M^s)$  和攻击者使用的总干扰能量  $p^a$ , 传感器不需要获取攻击者在每个信道的实时能量消耗. 算法1通过给定初始值, 进而在所有可能的动作集合中不断试错、迭代的方式寻找最优值. 定理3给出了所提算法通过大量迭代和合适地选择学习率可以收敛到最优值.

## 3 仿真实例

### 3.1 仿真参数

考虑如下系统参数:

$$A_1 = 1.2, C_1 = 0.7, Q_1 = R_1 = 0.8$$

$$A_2 = 1, C_2 = 0.8, Q_2 = R_2 = 0.75$$

通过在传感器上执行卡尔曼滤波过程, 可以获得稳态误差协方差:

$$\bar{P}_1 = 0.9245, \bar{P}_2 = 0.6347$$

假设无线通信网络中有三条信息传输通道,其中通道噪声 $\sigma_i = 0.5, i = 1, 2, 3$ ; 并且网络参数 $\alpha = 2$ ,系数 $\tau = 0.5$ ,用来描述一个信道同时传输干扰的程度.三条通道的能量限度分别为 $p_1^c = 1, p_2^c = 2, p_3^c = 3$ ,也就是需要通过每条通道,那么传感器或者攻击者至少要具备当前通道消耗的能量才能选择;假设传感器1和传感器2的传输能量分别为 $p_1^s = 3, p_2^s = 2$ ;攻击者具备的总干扰能量为 $p^a = 3$ .

我们将纳什Q学习算法应用在无限时间的马尔可夫调度器与攻击者的博弈过程中,其中,传感器和攻击者的加权参数分别为 $\delta_s = 0.5, \delta_a = 1$ ,折损系数 $\beta = 0.96$ ,学习率定义为 $\alpha_k = 10/[15 + \text{count}(m, \vec{l}_k^s, \vec{l}_k^a)]$ ,其中 $\text{count}(m, \vec{l}_k^s, \vec{l}_k^a)$ 表示为在第 $k$ 个时间步前实数对 $(m, \vec{l}_k^s, \vec{l}_k^a)$ 出现的次数,这里的学习效率满足Q学习收敛条件.在新的学习步骤中,较少访问的状态和动作对将受到更多关注.调度器和攻击者在每一步选择行动的随机性、非零丢弃率和足够大的迭代次数可以保证模拟中每个状态和行动的访问.设置最终状态,对于在 $k$ 步的时候,远程状态估计器的状态 $m_k = (h_1^{i_1(k)}(\bar{P}_1), h_2^{i_2(k)}(\bar{P}_2)), i_1(k), i_2(k) \in \{0, 1, 2\}$ .

### 3.2 仿真结果

在100000个学习时间步后,可以得到在不同状态下的双方采取不同动作的Q值均收敛.由于涉及到的状态维数较大,因此在下方以更容易实现的 $(\bar{P}_1, \bar{P}_2)$ 状态为例,其不同联合动作下Q值的收敛情况如图1所示.以 $(1,1,(1,2))$ 为例,其表示两个传感器均选择通道1进行数据传输,攻击者选择通道1和通道2进行DoS攻击.

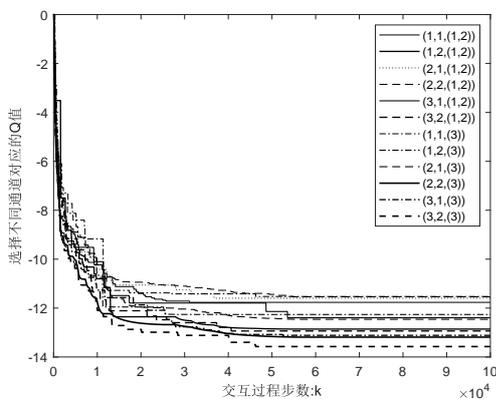


图1 状态 $(\bar{P}_1, \bar{P}_2)$ 下不同联合动作 $(\vec{l}_k^s, \vec{l}_k^a)$ 的Q值

除此之外,除了收敛的Q值之外,每个玩家在不同状态下的通道选择概率也是收敛的,收敛的概率值即为调度器和攻击者在不同状态下的纳什均衡策略,即为双方的最优策略,由于涉及的状态维数较大,

图2是以所处状态 $(\bar{P}_1, \bar{P}_2)$ 情况为例,在有攻击者的情况下,给出了两个传感器不同通道选择策略的收敛情况.从图中可以看出状态 $(\bar{P}_1, \bar{P}_2)$ 下两个传感器通道选择策略最终趋于 $(1,2)$ 的传输策略,即传感器1选择通道1进行数据传输,传感器2选择通道2进行数据传输.表1总结了在不同状态下攻防双方的最优策略.

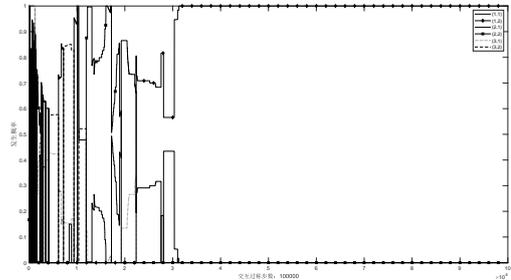


图2 状态 $(\bar{P}_1, \bar{P}_2)$ 下传感器的策略学习过程

表1 不同状态下调度器和DoS攻击者的最优策略

状态	调度器策略	攻击者策略
$\bar{P}_1, \bar{P}_2$	[0,1,0,0,0]	[0,1]
$h(\bar{P}_1), \bar{P}_2$	[0,0,0.7072,0.2928,0,0]	[0.661,0.339]
$h^2(\bar{P}_1), \bar{P}_2$	[0,0,0.377,0.623,0,0]	[0.3512,0.6488]
$\bar{P}_1, h(\bar{P}_2)$	[0,0,0.5088,0.4913,0,0]	[0,1]
$h(\bar{P}_1), h(\bar{P}_2)$	[0,0,0.6472,0.3528,0,0]	[0.3482,0.6518]
$h^2(\bar{P}_1), h(\bar{P}_2)$	[0.4562,0.5447,0,0,0,0]	[0.4451,0.5549]
$\bar{P}_1, h^2(\bar{P}_2)$	[0.1796,0.8204,0,0,0,0]	[0.5864,0.4136]
$h(\bar{P}_1), h^2(\bar{P}_2)$	[1,0,0,0,0,0]	[0.9724,0.0276]
$h^2(\bar{P}_1), h^2(\bar{P}_2)$	[0,0,0,0.5,0.5,0]	[0,1]

图3表示调度器在没有采取防御(即调度器调度传感器均匀分布选择通道,不具备修改策略功能)与在采取防御下不断修改自身策略两种情况下 $M$ 个子系统的状态估计误差协方差矩阵迹的累加和 $\sum_{i=1}^M Tr(P_{i,k})$ 值的对比.从图3容易得出结论,不断学习获取知识并采取了防御策略之后能够提供更准确的状态估计结果.

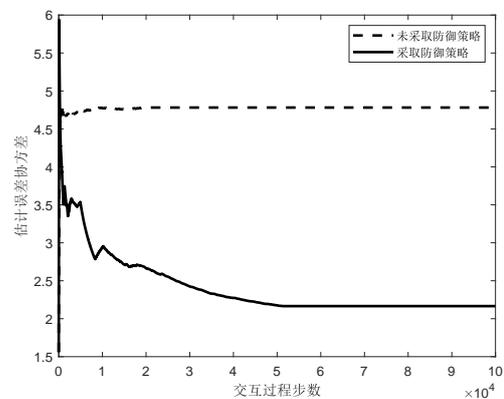


图3 调度器是否采取防御策略下性能比较

## 4 结论

考虑到实际的信息物理系统的传感器处于分

散部署,监控多个动态过程,因此本文主要考虑非线性时不变系统模型下,多个传感器通过调度器的决策,将数据包经过无线信道发送给远程状态估计器,在此过程会存在DoS攻击对多条无线信道进行阻塞以降低估计性能.针对该过程,考虑了传感器与攻击者均在各自的能量限制、环境对于通道信号的影响、攻击者可攻击多条通道等条件下分析双方的交互过程以及传输调度的策略求解问题,通过建立攻防双方的数学模型,构建了一个零和博弈过程,并采用多智能体强化学习中的纳什Q学习算法,求解出了双方在通道传输方面的纳什均衡策略,最后通过仿真算例,验证了所提方法的有效性.本文的工作仅限于博弈中只有一个攻击者的情况,未来将会基于多人一般和博弈的纳什Q学习算法,开展含有多方攻击者加入时CPS安全状态估计的研究工作.

#### 参考文献(References)

- [1] 李少远, 殷翔. 信息物理世界如何实现人机共融协同?[J]. 上海交通大学学报, 2021, 55(A1): 5-6.  
(Li S, Yin X. How to achieve Human-Machine cooperation in Cyber-Physical systems?[J]. Journal of Shanghai Jiaotong University, 2021, 55(A1): 5-6.)
- [2] Li Y, Shi L, Cheng P, et al. Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach [J]. IEEE Transactions on Automatic Control, 2015, 60(10): 2831-2836.
- [3] 高夏翔,李相俊,杨锡运. 集中式超大规模储能电站信息物理系统建模与可靠性评估[J]. 控制与决策, 2022, 37(5): 1309-1319.  
(Gao X, Li X J, Yang X Y. Modeling and reliability assessment of centralized ultra large scale energy storage power station cyber physical system[J]. Control and Decision, 2022, 37(5): 1309-1319.)
- [4] Incel O D. A survey on multi-channel communication in wireless sensor networks [J]. Computer Networks, 2011, 55(13): 3081-3099.
- [5] Zhang H, Cheng P, Shi L, et al. Optimal DoS attack scheduling in wireless networked control system [J]. IEEE Transactions on Control Systems Technology, 2015, 24(3): 843-852.
- [6] Duo W, Zhou M, and Abusorrah A, A survey of cyber attacks on cyber physical systems: recent advances and challenges[J]. IEEE/CAA Journal Of Automatica Sinica, 2022, 9(5): 784-800.
- [7] 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.)
- [8] Peng L, Shi L, Cao X, et al. Optimal attack energy allocation against remote state estimation [J]. IEEE Transactions on Automatic Control, 2017, 63(7): 2199-2205.
- [9] Yang C, Yang W, Shi H. DoS attack in centralised sensor network against state estimation [J]. IET Control Theory & Applications, 2018, 12(9): 1244-1253.
- [10] Li Q, Wang Z, Sheng W, et al. Dynamic event-triggered mechanism for  $H_\infty$  non-fragile state estimation of complex networks under randomly occurring sensor saturations [J]. Information Sciences, 2020, 509: 304-316.
- [11] Gan R, Xiao Y, Shao J, et al. An analysis on optimal attack schedule based on channel hopping scheme in cyber-physical systems [J]. IEEE Transactions on Cybernetics, 2019, 51(2): 994-1003.
- [12] Hu S, Yue D, Xie X, et al. Resilient event-triggered controller synthesis of networked control systems under periodic dos jamming attacks [J]. IEEE transactions on cybernetics, 2018, 49(12): 4271-4281.
- [13] DeBruhl B, Tague P. Digital filter design for jamming mitigation in 802.15. 4 communication [C]. 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN). IEEE, 2011: 1-6.
- [14] Foroush H S, Martínez S. On event-triggered control of linear systems under periodic denial-of-service jamming attacks [C]. 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). IEEE, 2012: 2551-2556.
- [15] Zhang H, Qi Y, Wu J, et al. DoS attack energy management against remote state estimation [J]. IEEE Transactions on Control of Network Systems, 2016, 5(1): 383-394.
- [16] Guo Z, Wang J, Shi L. Optimal denial-of-service attack on feedback channel against acknowledgment-based sensor power schedule for remote estimation [C]. 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, 2017: 5997-6002.
- [17] Li H, Lai L, Qiu R C. A denial-of-service jamming game for remote state monitoring in smart grid [C]. 45th Annual Conference on Information Sciences and Systems, CISS 2011, The John Hopkins University, Baltimore, MD, USA, 23-25 March 2011. IEEE, 2011: 1-6.
- [18] Li Y, Quevedo D E, Dey S, et al. SINR-Based DoS attack on remote state estimation: a game-theoretic approach [J]. IEEE Transactions on Control of Network Systems, 2017, 4(3): 632-642.
- [19] Ding K, Li Y, Quevedo D E, et al. A multi-channel transmission schedule for remote state estimation under dos attacks [J]. Automatica, 2017, 78: 194-201.
- [20] Zhang J, Sun J. A game theoretic approach to multi-channel transmission scheduling for multiple linear systems under dos attacks [J]. Systems & Control Letters, 2019, 133(1): 104546.
- [21] Yuan H, Xia Y, Yang H. Resilient State Estimation of Cyber-Physical System With Multichannel Transmission Under DoS Attack[J]. IEEE Transactions

- on Systems, Man, and Cybernetics: Systems, 2021, 51(11):6926-6937.
- [22] Zhou H, Chen J, Zheng H, et al. Energy efficiency and contact opportunities tradeoff in opportunistic mobile networks [J]. IEEE Transactions on Vehicular Technology, 2016, 65(5): 3723-3734.
- [23] Yuan Yuan, Fuchun Sun, Quanyan Zhu. Resilient control in the presence of DoS attack: Switched system approach [J]. International Journal of Control, Automation and Systems, 2015, 13(6): 1423-1435.
- [24] Ding K, Ren X, Qi H, Shi G, Wang X, Shi L. Interference Game for Intelligent Sensors in Cyber-physical Systems[J]. Automatica, 2021, 129: 109668.
- [25] Wang K, Liu W, Lim T J, Deep Reinforcement Learning for Joint Sensor Scheduling and Power Allocation under DoS Attack [C], ICC 2022-IEEE International Conference on Communications, pp.1968-1973, 2022.
- [26] 杨晓峰,谢巍,张浪文. 信息物理环境下不确定系统的随机分布式预测控制[J]. 控制与决策, 2020, 35(8): 1895-1901.  
(Yang X F, Xie W, Zhang L W. A stochastic distributed predictive control algorithm for uncertain systems under cyber-physical system environment[J]. Control and Decision, 2020, 35(8): 1895-1901.)
- [27] Nash J F. Non-Cooperative games [J]. Annals of Mathematics, 1951, 54: 286-295.
- [28] Hu J, Wellman M P. Nash Q-learning for general-sum stochastic games [J]. Journal of Machine Learning Research, 2003, 4(4): 1039-1069.

### 作者简介

徐鑫(1997-), 男, 硕士研究生, 从事基于博弈论的信息物理系统安全研究, E-mail: 923992491@qq.com;

王慧敏(1985-), 女, 副教授, 博士, 从事复杂动态系统故障诊断、信息物理系统安全状态估计、安全控制等研究, E-mail: wanghuimin@ise.neu.edu.cn.