中国科学: 信息科学 2023 年 第 53 卷 第 2 期: 325–343
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
论文

网络攻击下产品与供应链协同演进系统数据驱动变更控制设计

李庆奎[1*], 高雪峰[1], 彭晨[2], 张蕴隆[1], 易军凯[1]

1. 北京信息科技大学自动化学院, 北京 100192
2. 上海大学机电工程与自动化学院, 上海 200072
* 通信作者. E-mail: sdlqk01@bistu.edu.cn

收稿日期: 2021–12–26; 修回日期: 2022–04–07; 接受日期: 2022–05–12; 网络出版日期: 2023–02–06

国家重点研发计划 (批准号: 2020YFB1708200) 资助项目

摘要 产品与供应链协同演进系统 (product and supply chain synchronous evolution system, PSCSE) 是一类复杂的分布式信息物理系统 (cyber-physical system, CPS), 含有大量的未建模动态与不确定性. 剧烈的需求波动及突发的网络事件, 使系统结构参数极易发生变化. 产品设计变更是维持 PSCSES 稳定、满足用户需求及保证经济效益的重要手段. 本文研究基于分布式 CPS 的 PSCSES 在受到 DoS 攻击下的应急变更控制问题. 首先, 针对 PSCSES 在网络攻击下数据包丢失问题, 利用每条子链的历史数据设计基于循环神经网络 (recurrent neural network, RNN) 的预测器以弥补因网络攻击造成的库存数据丢失; 其次, 利用博弈论思想将 H_∞ 一致性控制问题转化为多人零和图博弈问题, 提出一种应急变更补偿机制; 进而基于 Q-learning 的策略迭代技术设计了模型未知的控制器求解算法, 实现了系统的库存状态 H_∞ 一致性. 最后, 仿真实验验证了所提方法的有效性.

关键词 网络攻击, 数据驱动, 供应链, 变更控制设计, 多智能体

1 引言

产品与供应链协同演进系统是由服务于产品生产设计及流通的各个节点企业如供应商、制造商、分销商和零售商等组成, 通过对信息流、物流、资金流的控制实现将原材料生产为产品并交付给用户的分布式网络控制系统 (networked control system, NCS) [1]. 随着人工智能技术的迅速发展, 制造业正逐步由制造自动化向协同智能制造转型. 作为国家智能制造重大工程的支撑, 产品与供应链协同演进系统在电力供应、智能制造、生物医药、食品加工等领域具有广泛应用前景 [2~5].

传统的产品与供应链协同演进系统以满足生产需求为主导, 采用专用信道进行数据传输, 在运作过程中通常出现物流缓慢、信息传递不及时、无法实现定制化生产, 以及因生产设计不准确而导致库

引用格式: 李庆奎, 高雪峰, 彭晨, 等. 网络攻击下产品与供应链协同演进系统数据驱动变更控制设计. 中国科学: 信息科学, 2023, 53: 325–343, doi: 10.1360/SSI-2021-0435
Li Q K, Gao X F, Peng C, et al. Data-driven change control design for product and supply chain synchronous evolution systems under cyber-attacks (in Chinese). Sci Sin Inform, 2023, 53: 325–343, doi: 10.1360/SSI-2021-0435

© 2023《中国科学》杂志社

存积仓等问题, 致使资源利用率极低 [6]. 得益于大数据与人工智能技术的迅猛发展, 产品与供应链协同演进系统开始向数字化、智能化、可组织化转变, 成为数据驱动、综合计算和网络通信于一体的信息物理系统 (cyber-physical system, CPS) [7]. 考虑到复杂产品是由各个零部件组成的, 而复杂产品的各个零部件的生产往往需要由不同子供应链完成, 因此, 可以将产品与供应链协同演进系统看成由多个子链组成的复杂多智能体系统 [8]. 基于信息物理系统与多智能体技术, 产品与供应链协同演进系统可以实现以用户需求为导向的定制化生产, 同时对生产过程实施实时监控以及智能化仓储, 从而极大地减少商品库存过剩现象, 简化生产流程, 提高企业的经营效益.

基于信息物理系统的产品与供应链协同演进系统以可靠、高效、实时协同的优势为复杂产品的生产销售提供了极大的便利, 但大量网络设备的引入, 使得这类信息物理系统更易受到恶意网络攻击. 近年来, 网络攻击事件频发对企业造成了严重的经济损失. 如 2020 年 6 月 8 日, 日本本田汽车制造商的服务器遭受到工业型勒索软件 Ekans 攻击导致部分生产系统中断 [9]. 此类网络安全事件表明: 产品与供应链协同演进系统已经成为网络攻击的主要目标, 频繁的网络攻击会使得产品与供应链协同演进系统中的节点企业遭受严重经济损失, 甚至会危害社会稳定和国家发展. 因此, 如何提高产品与供应链协同演进系统在一类突发网络攻击事件下的应急能力和稳定性具有重要的现实意义.

基于 CPS 的产品与供应链协同演进系统的网络攻击主要分为完整性攻击和可用性攻击两类. 完整性攻击是指通过篡改传输数据包中的信息, 以降低系统可靠性和安全性为目标的攻击方式; 可用性攻击则通过阻塞 CPS 各部分之间数据与控制命令的正常传输, 使系统的某些服务被暂停甚至瘫痪 [10~12]. DoS 攻击属于可用性攻击方式, 攻击者通过占用通信资源或设备资源, 以禁止数据传输, 导致通信信道中正常传输数据包丢失 [13], 进而达到削弱或破坏网络服务目的, 破坏力极强. 因此, 如果不能有效处理 DoS 攻击下的数据包丢失问题, 则可能导致网络崩溃, 更严重者甚至会破坏物理系统 [14~18]. 当前, 针对一类如 DoS 攻击所造成的系统安全问题, 研究人员已开展大量卓有成效的工作, 这些研究主要来自以下几方面.

(1) 从攻击者角度研究最佳 DoS 攻击时刻, 如, Zhang 等 [19] 研究了 CPS 中能量约束 DoS 干扰器的最佳攻击调度, 即考虑如下情形: 传感器通过无线信道将数据包发送到远程估计器, 而 DoS 干扰器启动干扰攻击以增加无线信道数据包丢失的可能性. 在有限时间内, 能量预算有限的 DoS 干扰器只能发动 n 次攻击, 连续攻击的策略可以达到最佳攻击效果. Zhang 等 [20] 进一步研究了 DoS 干扰器的周期性攻击策略, 即在活动期间攻击无线信道 n 个单位, 然后转移到非活动时期, 以补充其在下一个攻击期间的能量, 并将最佳攻击计划扩展到具有多个子系统的 NCS.

(2) 从防御者角度研究有效控制策略, 如 Foroush 等 [21] 针对控制信道中遇到周期性 DoS 干扰攻击, 研究了 CPS 的安全控制, 提出一种能量受限的 DoS 干扰模型, 利用周期性攻击策略来破坏通信信道的质量. 为对抗 DoS 干扰攻击的影响, 在攻击策略部分已知的情形下, 给出了保证 CPS 渐近稳定的充分条件. 文献 [22] 通过选择不同的频率和持续时间值, 提出了能量预算有限的 DoS 周期攻击策略, 并研究了在 DoS 攻击下 CPS 的稳定性.

(3) 基于博弈论方法研究防御者和拒绝服务攻击者之间的对抗关系, 如文献 [23] 根据伯努利 (Bernoulli) 分布定义了一个随机变量 α_i^k 表示攻击对控制信道的影响, 利用博弈理论提出一种弹性控制方法, 以抵抗 CPS 中的 DoS 干扰攻击. Zhao 等 [24] 研究了 DoS 干扰和触发器均为能量有限的情形下, 利用完全信息零和博弈来获取双方最优策略.

由上述研究结果不难发现, 尽管基于 CPS 的安全控制问题已开展广泛研究, 但基于分布式 CPS 且模型未知的产品与供应链协同演进系统, DoS 攻击防御策略研究尚显不足. 开展基于 DoS 攻击的产品与供应链协同演进系统应急变更防御策略, 是本文的主要任务之一.

为克服控制器设计对模型的依赖, 近年来, 无模型数据驱动控制研究引起了学界的广泛关注并取得丰富成果 [25~36]. 数据驱动控制在智能交通、航空航天、智能制造等领域已得到广泛应用, 如文献 [31, 37, 38]. 在文献 [29] 中, 分布式无模型自适应迭代学习控制方法用以解决一类未知非线性多智能体系统领导跟随一致性问题. 文献 [35] 提出了一种容错无模型自适应控制方案, 用以解决一类单输入单输出非线性 NCS 在 DoS 攻击下的跟踪问题. 近年来, 随着人工神经网络技术的发展, 一种新型数据驱动技术广泛应用于系统辨识、算法设计等领域, 为无模型控制提供新的解决方案, 如文献 [28, 30] 将神经网络技术应用于系统辨识, 有效地避免了对机理模型的依赖. 然而, 基于神经网络的系统辨识所导致的误差, 为控制器设计带来了新的不确定性, 进而降低了系统性能. 为克服神经网络用于系统辨识所带来的误差, 强化 Q-learning 作为一种解决无模型问题的有效方法得到广泛应用 [25, 32, 39, 40]. 如文献 [39] 利用 Q-learning 解决了无模型 H_∞ 控制问题, 但基于多智能体及网络攻击的无模型控制亟待进一步研究. 文献 [40] 设计了基于策略迭代的 Q-learning 算法实现多智能体系统的一致性控制, 但网络攻击及扰动问题考虑不足. 值得注意的是, 作为一类重要的分布式 CPS, 产品与供应链协同演进系统存在大量未建模动态及不确定性, 剧烈的需求波动及突发的网络事件, 如贸易战导致的供应链断裂, 新冠肺炎造成的供应链阻塞, 使系统的结构与参数极易发生变化, 产品设计变更是维持产品与供应链协同演进系统稳定、满足用户需求及保证经济效益的重要手段. 如何利用数据驱动技术设计应急变更补偿机制解决 DoS 攻击下的产品与供应链系统变更控制, 是本文又一重要研究任务.

本文研究基于分布式 CPS 的产品与供应链协同演进系统在受到 DoS 攻击下的应急变更控制问题. 针对协同演进系统传感器 – 控制器通信信道受到 DoS 攻击情形, 利用数据驱动技术, 根据历史数据和参考信息设计应急防御变更机制, 通过零和微分图博弈理论与强化 Q-learning 技术设计 H_∞ 一致性控制器, 保证系统在 DoS 攻击导致设计变更情形下, 使产品与供应链协同演进系统达到领导跟随 H_∞ 一致, 并可以抑制不确定需求和变更设计带来的牛鞭效应 (即不确定用户需求信息由供应链底层向顶层逐级放大的现象 [41]), 同时保证系统具有一定的产品适应度及用户满意度. 本文的主要贡献如下: 首先, 利用历史数据和参考信息设计应急变更补偿机制, 即当 DoS 攻击造成库存状态数据丢失时, 利用缓存器中的历史数据设计基于循环神经网络 (recurrent neural network, RNN) 的库存预估器以补偿库存数据丢失; 其次, 结合零和动态图博弈、最优控制和强化学习理论获得最优一致性生产率, 即根据库存跟踪误差性能指标函数, 得出贝尔曼 (Bellman) 方程, 利用贝尔曼最优原则得到连续时间 HJI 方程, 从而得到最优生产率以及最坏情形下的不确定需求. 值得注意的是, HJI 方程通常很难得到解析解, 为此, 我们引入基于策略迭代的 Q-learning 方法求解 H_∞ 一致性控制器; 最后通过仿真算例验证了所提方法的有效性.

2 预备知识与问题描述

2.1 预备知识

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ 是一个有向图, 其中 $\mathcal{V} = \{\nu_1, \nu_2, \ldots, \nu_N\}$ 和 $\mathcal{E} = \{a_{ij} = (\nu_i, \nu_j)\} \in \mathcal{V} \times \mathcal{V}$ 分别表示有限, 非空的 N 个节点集合及一组边集. $\mathcal{A} = [a_{ij}]$ 是一个具有邻接元素 $a_{ij} \geqslant 0$ 的加权邻接矩阵, 对于图 $\mathcal{G}$ 中的第 i 个和第 j 个节点, 当 $a_{ij} = (\nu_j, \nu_i) \in \mathcal{E}$ 时, $a_{ij} > 0$ (其表示节点 i 可以从节点 j 接收信息), 否则 $a_{ij} = 0$. 若节点 i 能够接收到节点 j 发送的信息, 则称节点 j 是节点 i 的邻居, 节点 i 的邻居节点集可表示为 $N_i = \{\nu_j : (\nu_j, \nu_i) \in \mathcal{E}\}$. 若图中存在一个节点可通过有向路径到达其他任意节点, 则称图包含生成树. 用 $\mathcal{D} = \operatorname{diag}\{d_i\}$ 表示图的入度矩阵, 其中 $d_i = \sum_{j \in N_i} a_{ij}$ 表示节点 i 的入度. 图

的拉普拉斯 (Laplace) 矩阵可定义为 $L = \mathcal{D} - \mathcal{A}$.

2.2 问题描述

考虑配置某产品生产的供应链由 N 条子链组成并协同生产, 假设第 i 条子链的动力学方程为

$$\begin{cases} \dot{x}_i(t) = Ax_i(t) + Bu_i(t) + D\bar{\omega}_i(t), \\ y_i(t) = Cx_i(t), \end{cases} \tag{1}$$

其中 $x_i(t)$, $u_i(t)$, $\bar{\omega}_i(t)$ 和 $y_i(t)$ 分别为第 i 条子链中的生产库存状态、生产率、用户需求及当前成品输出. 从控制理论的角度出发, $x_i(t)$, $u_i(t)$, $\bar{\omega}_i(t)$ 和 $y_i(t)$ 分别为第 i 个子链在 t 时刻的状态变量、控制输入、外界扰动和系统输出. A, B 和 C 为系统矩阵且未知. 不失一般性, 由实践经验可将不确定用户需求 $\bar{\omega}_i(t)$ 分解为常需求 d 以及能量有限需求 $\omega_i(t)$, 即 $\omega_i(t) \in L_2[0, \infty)$.

注1 一种产品通常由若干部件组成, 每一部件都需要相应的原料供应. 因此, 某一部件的变更通常伴随相应的供应链发生改变, 另一方面, 供应链如果发生改变, 同样可能导致产品部件发生改变, 即产品设计需要变更. 为此, 我们称这样相互影响的系统为产品与供应链协同演进系统.

考虑产品与供应链协同演进系统存在一个链主, 子链需要根据链主和邻居供应链的库存状态变化调整其自身库存状态, 假设链主的动力学方程由下式给出:

$$\dot{x}_0(t) = Ax_0(t) + Dd, \tag{2}$$

其中 $x_0(t)$ 为库存需求目标, 即为子链库存状态的跟踪目标. 考虑到实际生产过程中的产能等一些限制因素, 给出如下假设.

假设1 生产率 $u_i(t)$ 会有一定的上界 $u_{\max}$, 即 $0 \leqslant u_i(t) \leqslant u_{\max}$.

假设2 实际的仓储环境下, 库存容量是有限的. 因此, 设库存水平满足 $0 \leqslant x_i(t) \leqslant x_{\max}$. 这里, $x_{\max}$ 表示第 i 条子供应链仓储能力的最大值.

在供应链系统中, 当第 i 条子链的库存状态 $x_i(t)$ 通过网络传输到分布式控制中心时, 若传感器 – 控制器 (sensor-to-controller, S-C) 通道遭受 DoS 攻击就有可能造成数据包丢失的现象, 如图 1 所示. 图 1 中, T 为极小采样时刻, $\bar{x}_i(t)$ 为变更补偿后的第 i 条子链的库存状态, 分布式控制中心可以获得的第 i 条子链的库存状态 $x_i^{\alpha}(t)$ 可以表示为

$$x_i^{\alpha}(t) = \alpha(t)x_i(t) + (1 - \alpha(t))\bar{x}_i(t), \tag{3}$$

其中 $\alpha(t) \in \{0, 1\}$, $\alpha(t) = 0$ 代表发生了网络攻击事件, 即 S-C 通道遭受了 DoS 攻击, $\alpha(t) = 1$ 代表未发生网络攻击事件, 即 S-C 通道能够进行正常的数据传输. 不失一般性, 对 DoS 攻击有如下假设.

假设3 网络攻击者连续攻击的次数有限, 即由网络攻击造成的最大连续丢包数有界, 记为 $\bar{n}_i^{\alpha}$.

假设4 网络攻击者只针对子链发起攻击.

本文从主动防控 DoS 攻击的角度出发, 设计合理的变更机制补偿因攻击造成的数据包丢失. 由于系统受到不确定性需求以及变更机制的影响将可能引发牛鞭效应, 因此需要进一步设计出基于最优控制的 H_{∞} 控制器来抑制不确定性需求及变更机制带来的牛鞭效应, 进而使得系统在遭受 DoS 攻击下子链的库存状态 $x_i(t)$ 仍然可以实现与链主的库存状态 $x_0(t)$ 保持一致. 首先由式 (1) 和 (2) 可将每条子链的局部邻域库存同步误差作如下定义:

$$\delta_i(t) = \sum_{j \in N_i} a_{ij} \left(x_i^{\alpha}(t) - x_j^{\alpha}(t)\right) + g_i \left(x_i^{\alpha}(t) - x_0(t)\right), \tag{4}$$

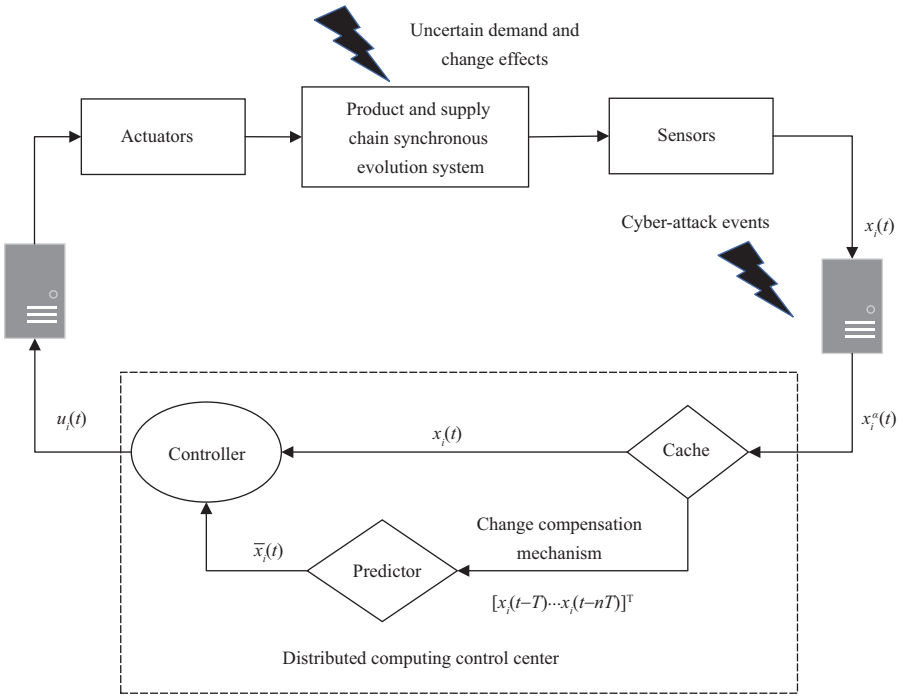

图 1 (网络版彩图) 网络攻击下变更机制示意图

Figure 1 (Color online) Diagram of change mechanism under cyber-attacks

其中 $g_i \geqslant 0$ 是子链 i 的牵引增益. 当且仅当第 i 条子链与链主之间存在有向路径时, $g_i > 0$, 否则, $g_i = 0$. 子链的全局库存同步误差向量可表示为

$$\delta(t) = ((L+G) \otimes I_n)\left(x^{\alpha}(t) - \bar{x}_0(t)\right), \tag{5}$$

其中全局库存同步误差向量为 $\delta(t) = [\,\delta_1^{\mathrm{T}}(t),\, \delta_2^{\mathrm{T}}(t),\, \ldots,\, \delta_N^{\mathrm{T}}(t)\,]^{\mathrm{T}}$, 全局库存状态向量为 $x^{\alpha}(t) = [(x_1^{\alpha}(t))^{\mathrm{T}}, (x_2^{\alpha}(t))^{\mathrm{T}}, \ldots, (x_N^{\alpha}(t))^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Nn}$, $L = [l_{ij}] \in \mathbb{R}^{N\times N}$ 是有向拓扑图的拉普拉斯矩阵; $G = [g_{ij}] \in \mathbb{R}^{N\times N}$ 是一个牵引增益作为对角线元素 ($g_{ii} = g_i$) 的对角矩阵, 其展现了供应链链主与各子链间的链接关系; I_n 是 n 维单位矩阵, $\bar{x}_0(t) = \underline{I} x_0(t)$, $\underline{I} = \underline{1} \otimes I_n$, $\underline{1}$ 为元素为 1 的 N 维向量, 即 $\bar{x}_0(t) = [\,x_0^{\mathrm{T}}(t),\, x_0^{\mathrm{T}}(t),\, \cdots,\, x_0^{\mathrm{T}}(t)\,]^{\mathrm{T}} \in \mathbb{R}^{Nn}$.

假设产品与供应链协同演进系统的拓扑结构中包含生成树, 故可令第 i 条子链的库存跟踪误差为 $\eta_i(t) = x_i^{\alpha}(t) - x_0(t)$, 进而可得到全局库存跟踪误差为 $\eta(t) = x^{\alpha}(t) - \bar{x}_0(t)$. 进一步全局库存跟踪误差向量与全局同步库存误差向量的关系可表示为

$$\delta(t) = ((L+G) \otimes I_n)\,\eta(t). \tag{6}$$

引理1 假设 $(L+G)$ 是非奇异矩阵, 全局库存跟踪误差 $\eta(t)$ 的取值范围为

$$\|\eta\| \leqslant \left(\lambda_{\min}(L+G)\right)^{-1} \|\delta\|, \tag{7}$$

其中 $\lambda_{\min}(L+G)$ 是矩阵 $(L+G)$ 的最小奇异值.

引理 1 表明, 当 $(L+G)$ 是非奇异矩阵时, 通过减小全局同步库存误差向量, 可以使得全局库存跟踪误差向量任意变小, 即 $\|\delta(t)\| \to 0 \leftrightarrow \|\eta(t)\| \to 0$, 这也表明各子链与链主的库存状态一致.

注2 如果图 $\mathcal{G}$ 包含一个生成树, 并且对于一个根节点来说其牵引增益 $g_i \neq 0$, 那么 $(L+G)$ 为非奇异矩阵.

根据式 (4), 可以得到第 i 条子链局部邻域同步库存误差的动力学方程:

$$\dot{\delta}_i(t) = \sum_{j\in N_i} a_{ij}\left(\dot{x}_i^{\alpha}(t) - \dot{x}_j^{\alpha}(t)\right) + g_i\left(\dot{x}_i^{\alpha}(t) - \dot{x}_0(t)\right). \tag{8}$$

因此, 结合式 (1) 进一步可得

$$\dot{\delta}_i(t) = A\delta_i(t) - \sum_{j\in N_i} a_{ij}Bu_j(t) + (d_i+g_i)\,Bu_i(t) - \sum_{j\in N_i} a_{ij}D\omega_j(t) + (d_i+g_i)\,D\omega_i(t). \tag{9}$$

根据式 (9) 可知每条子链的同步库存误差可以表示为由自身及其所有相邻子链的控制行为驱动的且存在不确定用户需求的动态系统. 因此, 供应链系统的一致性问题转化为设计合理的生产率可以使得在不确定用户需求下及变更机制下局部邻域同步库存误差 $\delta_i(t)$ 最小, 进而根据引理 1 可以确保链主和所有子链之间的库存趋于同步, 且生产率能使得如式 (1) 所示的每个智能体满足以下定义.

定义1 产品与供应链协同演进系统 (1) 被称为能够以 $\gamma > 0$ (γ 为待优化常数) 水平抑制牛鞭效应, 如果满足如下 H_∞ 一致性条件.

(1) 当不确定需求 $\omega_i(t) = 0$ 时, 设计合理的生产率可以使得系统满足 $\lim_{t\to\infty}\|x_0(t) - x_i(t)\| = 0$;

(2) 当不确定市场需求满足 $\omega_i(t) \neq 0$ 时, 如果每条子链都满足如下牛鞭效应抑制条件:

$$\begin{aligned}\int_0^{\mathrm{T}} \|z_i(t)\|^2\,\mathrm{d}t &= \int_0^{\mathrm{T}} \left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i + u_i^{\mathrm{T}}R_iu_i + \sum_{j\in N_i} u_j^{\mathrm{T}}R_{ij}u_j\right)\mathrm{d}t \\ &\leqslant \gamma^2\int_0^{\mathrm{T}} \left(\omega_i^{\mathrm{T}}T_{ii}\omega_i + \sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t + \beta\left(\delta_i(0)\right),\end{aligned} \tag{10}$$

则称产品与供应链协同演进系统能够以 γ 水平抑制牛鞭效应. 其中 $v_i(t) = [\omega_i^{\mathrm{T}}(t),\, \omega_j^{\mathrm{T}}(t)]^{\mathrm{T}}$; 性能输出 $z_i(t) = [\,\delta_i^{\mathrm{T}}(t),\, u_i^{\mathrm{T}}(t),\, u_j^{\mathrm{T}}(t)\,]^{\mathrm{T}}$; $Q_{ii} \geqslant 0$, $R_{ii} > 0$, $R_{ij} \geqslant 0$, $T_{ii} > 0$ 和 $T_{ij} > 0$ 为正定对称权重矩阵; β 为有界函数且满足 $\beta(0) = 0$. 令 $\gamma_{\min}$ 为 γ 最小值. 对于任意的 $\gamma \geqslant \gamma_{\min}$ 而言, 式 (10) 都成立. 为表达方便, 下文 $u_j(t)$, $\omega_j(t)$, $\delta_j(t)$ 等在不引起误解时简记为 u_j, ω_j, δ_j.

3 网络攻击下产品与供应链协同演进系统零和图博弈与无模型求解

本节首先根据具有时间特性的历史库存状态数据预测出当前时刻的库存状态来解决 DoS 攻击下传感器 – 控制器 (S-C) 网络通信信道数据包丢失导致的变更设计问题; 其次利用零和图博弈的思想解决变更以及不确定需求下的产品与供应链协同演进系统变更后的一致性问题; 最后通过无模型 Q-learning 算法求解出博弈解, 进而确定最优生产率和最坏情形不确定需求.

3.1 网络攻击下 RNN 库存状态预测器

基于 RNN 网络设计库存状态预测器, 即定义在时间 t 之前发生的连续丢包数为 $n_\alpha(t-T)$, 若在 t 时刻, $x(t)$ 传输成功, 则 $n_\alpha(t-T) = 0$. 显然, $0 \leqslant n_\alpha(t) \leqslant \bar{n}_\alpha$. 于是, 在 t 时刻, 分布式控制中心可以获得最新库存状态量为

$$x_i^{\alpha}(t) = x\left(t - n_\alpha(t-T)\right), \tag{11}$$

进而分布式控制中心可以获得的历史库存状态数据为 $X(t)=[\,x(0),\,x(T),\,\ldots,\,x_i^{\alpha}(t)\,]$. 若在 t 时刻 S-C 通信信道遭受攻击发生了丢包现象, 则根据历史的库存状态数据 X 并利用预测器预测出当前的库存状态数据 $\bar{x}_i(t)$, 即

$$\begin{aligned} &h(t)=f(U\cdot X(t)+W\cdot h(t-T)),\\ &\bar{x}_i(t)=g(V\cdot h(t)), \end{aligned} \tag{12}$$

其中 h 是隐含层, U, V 和 W 分别代表输入层、输出层和隐含层的理想权重; $f(\cdot)$ 和 $g(\cdot)$ 分别为隐含层和输出层的激活函数.

注3 循环神经网络 RNN 的学习方式采用梯度下降法 (如文献 [42]), 限于篇幅, 此处不赘.

3.2 零和图博弈和耦合 HJI 方程与纳什均衡

首先针对第 i 条子链设计如下性能指标函数:

$$\begin{aligned} J_i\left(\delta_i(0),u_i,u_{-i},\omega_i,\omega_{-i}\right)=\frac{1}{2}\int_0^{\infty}\Bigg(&\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j\\ &-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\Bigg)\,\mathrm{d}t. \end{aligned} \tag{13}$$

基于式 (13) 定义第 i 条子链的值函数为

$$V_i\left(\delta_i(t)\right)=\frac{1}{2}\int_t^{\infty}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t. \tag{14}$$

文中生产率和不确定需求可以看作博弈的双方, 那么供应链系统的一致性等价于求解下述的供应链系统零和微分图博弈问题:

$$V_i\left(\delta_i(0)\right)=\min_{u_i}\max_{\omega_i}J_i\left(\delta_i(0),u_i,u_j,\omega_i,\omega_j\right). \tag{15}$$

如果存在鞍点 (u_i^*,ω_i^*), 则该博弈存在唯一解, 即

$$V_i^*\left(\delta_i(0)\right)=\min_{u_i}\max_{\omega_i}J_i\left(\delta_i(0),u_i,u_j^*,\omega_i,\omega_j^*\right)=\max_{\omega_i}\min_{u_i}J_i\left(\delta_i(0),u_i,u_j^*,\omega_i,\omega_j^*\right), \tag{16}$$

其与以下纳什均衡条件等价:

$$J_i\left(u_i^*,u_j^*,\omega_i,\omega_j^*\right)\leqslant J_i\left(u_i^*,u_j^*,\omega_i^*,\omega_j^*\right)\leqslant J_i\left(u_i,u_j^*,\omega_i^*,\omega_j^*\right). \tag{17}$$

由莱布尼茨 (Leibniz) 公式和贝尔曼方程, 可以得到等价于式 (14) 的一个 Hamiltonian 函数 ($V_i(0)=0$):

$$\begin{aligned} &H_i\left(\delta_i,\frac{\partial V_i}{\partial\delta_i},u_i,u_j,\omega_i,\omega_j\right)\\ &\equiv\frac{\partial V_i^{\mathrm{T}}}{\partial\delta_i}\left(A\delta_i-\sum_{j\in N_i}a_{ij}Bu_j+\left(d_i+g_i\right)Bu_i-\sum_{j\in N_i}a_{ij}D\omega_j+\left(d_i+g_i\right)D\omega_i\right)\\ &\quad+\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\frac{1}{2}u_i^{\mathrm{T}}R_{ii}u_i+\frac{1}{2}\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\frac{1}{2}\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\frac{1}{2}\gamma^2\sum_{j\in N_j}\omega_j^{\mathrm{T}}T_{ij}\omega_j=0. \end{aligned} \tag{18}$$

由稳定性条件可得

$$0=\frac{\partial H_i}{\partial u_i}\Rightarrow u_i=-\left(d_i+g_i\right)R_{ii}^{-1}B^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i},\tag{19}$$

$$0=\frac{\partial H_i}{\partial \omega_i}\Rightarrow \omega_i=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i}\,.\tag{20}$$

将控制策略 (19) 和扰动策略 (20) 带入到式 (18) 中, 可以得到如下耦合 HJI 方程 ($V_i(0)=0$):

$$\begin{aligned}&\frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}A_i^c+\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\frac{1}{2}\left(d_i+g_i\right)^2\frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}BR_{ii}^{-1}B^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i}-\frac{1}{2\gamma^2}\left(d_i+g_i\right)^2\frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}DT_{ii}^{-1}D^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i}\\&+\frac{1}{2}\sum_{j\in N_i}\left(d_i+g_i\right)^2\frac{\partial V_j^{\mathrm{T}}}{\partial \delta_j}BR_{jj}^{-1}R_{ij}R_{jj}^{-1}B^{\mathrm{T}}\frac{\partial V_j}{\partial \delta_j}-\frac{1}{2\gamma^2}\sum_{j\in N_i}\left(d_i+g_i\right)^2\frac{\partial V_j^{\mathrm{T}}}{\partial \delta_j}DT_{jj}^{-1}T_{ij}T_{jj}^{-1}D^{\mathrm{T}}\frac{\partial V_j}{\partial \delta_j}=0,\end{aligned}\tag{21}$$

其中 $A_i^c=A\delta_i-\left(d_i+g_i\right)^2BR_{ii}^{-1}B^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i}+\sum_{j\in N_i}a_{ij}\left(d_i+g_i\right)BR_{jj}^{-1}B^{\mathrm{T}}\frac{\partial V_j}{\partial \delta_j}+\frac{1}{\gamma^2}\left(d_i+g_i\right)^2DT_{ii}^{-1}D^{\mathrm{T}}\frac{\partial V_i}{\partial \delta_i}-\frac{1}{\gamma^2}\sum_{j\in N_i}a_{ij}\left(d_i+g_i\right)DT_{jj}^{-1}D^{\mathrm{T}}\frac{\partial V_j}{\partial \delta_j}$.

对于一个给定的解 V_i, 定义 $u_i^*=u_i\left(V_i\right)$, $u_j^*=u_j\left(V_j\right)$, $\omega_i^*=\omega_i\left(V_i\right)$, 和 $\omega_j^*=\omega_j\left(V_j\right)$, 那么式 (21) 可以改写为

$$H_i\left(\delta_i,\frac{\partial V_i}{\partial \delta_i},u_i^*,u_j^*,\omega_i^*,\omega_j^*\right)=0,\quad V_i(0)=0.\tag{22}$$

第 i 条子链的值函数进一步写成如下二次型的形式, 即

$$V_i\left(\delta_i\right)=\frac{1}{2}\delta_i^{\mathrm{T}}P_i\delta_i,\tag{23}$$

其中 P_i 是正定对称矩阵, 进一步可得

$$\begin{aligned}&\delta_i^{\mathrm{T}}P_i\bar{A}_i^c+\left(\bar{A}_i^c\right)^{\mathrm{T}}P_i\delta_i+\sum_{j\in N_i}\left(d_i+g_i\right)^2\delta_j^{\mathrm{T}}P_jBR_{jj}^{-1}R_{ij}R_{jj}^{-1}B^{\mathrm{T}}P_j\delta_j-\gamma^2\left(d_i+g_i\right)^2\delta_i^{\mathrm{T}}P_iT_{ii}^{-1}P_i\delta_i\\&+\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\left(d_i+g_i\right)^2\delta_i^{\mathrm{T}}P_iBR_{ii}^{-1}B^{\mathrm{T}}P_i\delta_i-\gamma^2\sum_{j\in N_i}\left(d_i+g_i\right)^2\delta_j^{\mathrm{T}}P_jT_{jj}^{-1}T_{ij}T_{jj}^{-1}P_j\delta_j=0,\end{aligned}\tag{24}$$

其中 $\bar{A}_i^c=A\delta_i-\left(d_i+g_i\right)^2BR_{ii}^{-1}B^{\mathrm{T}}P_i\delta_i+\sum_{j\in N_i}a_{ij}\left(d_i+g_i\right)BR_{jj}^{-1}B^{\mathrm{T}}P_j\delta_j+\frac{1}{\gamma^2}\left(d_i+g_i\right)^2DT_{ii}^{-1}D^{\mathrm{T}}P_i\delta_i-\frac{1}{\gamma^2}\sum_{j\in N_i}a_{ij}\left(d_i+g_i\right)DT_{jj}^{-1}D^{\mathrm{T}}P_j\delta_j$.

根据式 (23), 式 (19) 和 (20) 可以进一步表示为

$$u_i=-\left(d_i+g_i\right)R_{ii}^{-1}B^{\mathrm{T}}P_i\delta_i,\tag{25}$$

$$\omega_i=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D^{\mathrm{T}}P_i\delta_i.\tag{26}$$

注4 式 (24) 高度耦合通常很难直接求解, 而式 (25) 和 (26) 的求解必须知道系统矩阵, 因此在下文中将提供基于 Q-learning 的无模型求解算法.

引理2 若相邻子链的生产策略和不确定需求策略最优, 则对于每个 u_i,ω_i, 式 (27) 都满足,

$$H_i\left(\delta_i,\frac{\partial V_i}{\partial \delta_i},u_i,u_j^*,\omega_i,\omega_j^*\right)=\frac{1}{2}\left(u_i-u_i^*\right)^{\mathrm{T}}R_{ii}\left(u_i-u_i^*\right)-\frac{1}{2}\gamma^2\left(\omega_i-\omega_i^*\right)^{\mathrm{T}}T_{ii}\left(\omega_i-\omega_i^*\right).\tag{27}$$

3.3 图博弈下的产品与供应链协同演进系统库存同步与纳什解

3.3.1 产品与供应链协同演进系统库存同步

本小节将证明通过求解耦合关联 HJI 方程所得出的生产率能够使得产品与供应链协同演进系统以 γ 水平抑制牛鞭效应并实现与链主的同步.

定理1 对于产品与供应链协同演进系统 (1), 假设 1~4 条件成立, 若存在正定函数 V_i $(i\in N)$ 满足 HJI 方程 (21), 即存在光滑正定解 $V_i^*>0$ $(i\in N)$, 使得相邻子链的生产策略和不确定需求策略最优, 那么生产率 $u_i^*=u_i\left(V_i^*\right)$ 可以保证子链 i 能够实现与链主库存同步一致并以水平 γ 抑制牛鞭效应.

证明 假设 $V_i^*\left(\delta_i(k)\right)$ 满足式 (24) 所示的耦合 HJI 方程, 且子链 i 的相邻子链满足 $u_{-i}=u_{-i}^*, \omega_j=\omega_j^*$, 由引理 2 有

$$\begin{aligned}&H_i\left(\delta_i,\frac{\partial V_i^*}{\partial\delta_i},u_i,u_j,\omega_i,\omega_j\right)\\&=\frac{\mathrm{d}V_i^*}{\mathrm{d}t}+\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\frac{1}{2}u_i^{\mathrm{T}}R_{ii}u_i+\frac{1}{2}\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\frac{1}{2}\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\frac{1}{2}\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\\&=\frac{1}{2}\left(u_i-u_i^*\right)^{\mathrm{T}}R_{ii}\left(u_i-u_i^*\right)-\frac{1}{2}\gamma^2\left(\omega_i-\omega_i^*\right)^{\mathrm{T}}T_{ii}\left(\omega_i-\omega_i^*\right).\end{aligned}\tag{28}$$

当 $u_i=u_i^*$ 时, 有

$$\frac{\mathrm{d}V_i^*}{\mathrm{d}t}+\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\frac{1}{2}u_i^{\mathrm{T}}R_{ii}u_i+\frac{1}{2}\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\frac{1}{2}\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\frac{1}{2}\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\leqslant 0.\tag{29}$$

当不确定需求满足 $\omega_i=\omega_j=0$ 时, 有

$$\frac{\mathrm{d}V_i^*}{\mathrm{d}t}\leqslant-\frac{1}{2}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j\right)\leqslant-\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i\leqslant 0.\tag{30}$$

因此, 根据李雅普诺夫 (Lyapunov) 稳定性定理可知, 子链 i 库存误差系统在最优生产率 u_i^* 下趋于渐近稳定. 当 $\omega_i\neq 0$ 时, 对式 (29) 积分可得

$$\begin{aligned}&V_i^*\left(\delta_i(t)\right)-V_i^*\left(\delta_i(0)\right)\\&+\frac{1}{2}\int_0^{\mathrm{T}}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t\leqslant 0.\end{aligned}\tag{31}$$

又由于 $V_i^*\left(\delta_i(t)\right)>0$, 进一步可以得到

$$\int_0^{\mathrm{T}}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j\right)\mathrm{d}t\leqslant\gamma^2\int_0^{\mathrm{T}}\left(\omega_i^{\mathrm{T}}T_{ii}\omega_i+\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t+V_i^*\left(\delta_i(0)\right).\tag{32}$$

故在最优生产率 u_i^* 下, 由定义 1 知第 i 条子链能以 γ 水平抑制牛鞭效应且实现与链主库存状态一致.

3.3.2 图博弈下的纳什均衡解

本小节证明在一定条件下, 最优生产率和最坏情况下的不确定需求满足纳什均衡条件 (17), 进而求得零和图博弈解.

定理2 对于产品与供应链协同演进系统 (1), 假设 1∼4 条件成立, 若存在正定函数 V_i $(i \in N)$ 满足 HJI 方程 (21), 即存在光滑正定解 $V_i^* > 0$ $(i \in N)$, 使得相邻子链的生产策略和不确定需求策略最优, 则存在图博弈纳什均衡解, 亦即鞍点 $\left(u_i^*, u_j^*, \omega_i^*, \omega_j^*\right)$ 存在, 且博弈值由 HJI 方程的解 $V_i^*\left(\delta_i(0)\right)$ 给出.

证明 对于任何光滑函数 $V_i\left(\delta_i\right)$, 式 (13) 可以进一步重写为

$$\begin{aligned}
&J_i\left(\delta_i(0), u_i, u_j, \omega_i, \omega_j\right) \\
&= \frac{1}{2}\int_0^\infty \left(\delta_i^{\mathrm{T}} Q_{ii}\delta_i + u_i^{\mathrm{T}} R_{ii} u_i + \sum_{j\in N_i} u_j^{\mathrm{T}} R_{ij} u_j - \gamma^2 \omega_i^{\mathrm{T}} T_{ii}\omega_i - \gamma^2 \sum_{j\in N_i} \omega_j^{\mathrm{T}} T_{ij}\omega_j\right) \mathrm{d}t \\
&\quad + \frac{1}{2}\delta_i^{\mathrm{T}}(0) P_i \delta_i(0) - \frac{1}{2}\delta_i^{\mathrm{T}}(\infty) P_i \delta_i(\infty) + \int_0^\infty \frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}\Bigg(A\delta_i + \left(d_i + g_i\right) B u_i \\
&\quad - \sum_{j\in N_i} a_{ij} B u_j + \left(d_i + g_i\right) D\omega_i - \sum_{j\in N_i} a_{ij} D\omega_j\Bigg)\mathrm{d}t.
\end{aligned} \tag{33}$$

假设生产率 u_i^*, u_j^* 和不确定需求 ω_i^*, ω_j^* 是由式 (25) 和 (26) 给出的最优策略, 对式 (33) 进行配方可得

$$\begin{aligned}
&J_i\left(\delta_i(0), u_i, u_j, \omega_i, \omega_j\right) \\
&= \frac{1}{2}\delta_i^{\mathrm{T}}(0) P_i^* \delta_i(0) - \frac{1}{2}\delta_i^{\mathrm{T}}(\infty) P_i^* \delta_i(\infty) + \int_0^\infty \Bigg(\frac{1}{2}\sum_{j\in N_i}\left(u_j - u_j^*\right)^{\mathrm{T}} R_{ij}\left(u_j - u_j^*\right) \\
&\quad + \frac{1}{2}\left(u_i - u_i^*\right)^{\mathrm{T}} R_{ii}\left(u_i - u_i^*\right) - \frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}\sum_{j\in N_i} a_{ij} B\left(u_j - u_j^*\right) + \sum_{j\in N_i} u_j^{*T} R_{ij}\left(u_j - u_j^*\right) \\
&\quad - \frac{1}{2}\gamma^2 \sum_{j\in N_i}\left(\omega_j - \omega_j^*\right)^{\mathrm{T}} T_{ij}\left(\omega_j - \omega_j^*\right) - \frac{1}{2}\gamma^2\left(\omega_i - \omega_i^*\right)^{\mathrm{T}} T_{ii}\left(\omega_i - \omega_i^*\right) \\
&\quad - \frac{\partial V_i^{\mathrm{T}}}{\partial \delta_i}\sum_{j\in N_i} a_{ij} D\left(\omega_j - \omega_j^*\right) - \gamma^2 \sum_{j\in N_i} \omega_j^{*T} T_{ij}\left(\omega_j - \omega_j^*\right)\Bigg)\mathrm{d}t.
\end{aligned} \tag{34}$$

由定理 1 知库存误差系统是渐近稳定的, 所以 $V_i^*\left(\delta_i(\infty)\right) = 0$. 假设 $u_j = u_j^*$, $\omega_j = \omega_j^*$, 进而可以得到

$$\begin{aligned}
&J_i\left(\delta_i(0), u_i, u_j^*, \omega_i, \omega_j^*\right) \\
&= \frac{1}{2}\delta_i^{\mathrm{T}}(0) P_i^* \delta_i(0) + \int_0^\infty \left(\frac{1}{2}\left(u_i - u_i^*\right)^{\mathrm{T}} R_{ii}\left(u_i - u_i^*\right) - \frac{1}{2}\gamma^2\left(\omega_i - \omega_i^*\right)^{\mathrm{T}} T_{ii}\left(\omega_i - \omega_i^*\right)\right)\mathrm{d}t.
\end{aligned} \tag{35}$$

由式 (35) 可得式 (17) 成立. 进一步令 $u_i = u_i^*$, $\omega_i = \omega_i^*$, 有

$$J_i^*\left(\delta_i(0), u_i^*, u_j^*, \omega_i^*, \omega_j^*\right) = V_i^*\left(\delta_i(0)\right) = \frac{1}{2}\delta_i^{\mathrm{T}}(0) P_i^* \delta_i(0), \tag{36}$$

即得到了生产率和不确定需求的最优博弈解.

3.4 基于 Q-learning 算法的无模型策略求解

由于分布式产品与供应链协同演进系统运行过程十分复杂, 通常无法建立精确的系统模型, 现有基于模型的控制设计算法往往不具可行性. 虽然基于神经网络的系统辨识一直致力于解决模型未知问

题, 但在辨识过程中因误差的存在使得控制器的有效性有所降低. 基于强化学习的 Q-learning 算法可以通过系统所产生的数据实时更新控制器而不依赖于系统模型. 本文将采用 Q-learning 算法解决系统在网络攻击导致的设计变更情形下的系统库存一致性问题, 即通过策略迭代 Q-learning 算法求解式 (25) 和 (26). 首先基于 Hamiltonian 函数 (18) 以及值函数 (23) 定义如下第 i 条子链的 Q 函数:

$$Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)=V_i\left(\delta_i\right)+H_i\left(e_i,\frac{\partial V_i}{\partial \delta_i},u_i,u_j,\omega_i,\omega_j\right), \tag{37}$$

其中 $H_i(e_i,\frac{\partial V_i}{\partial \delta_i},u_i,u_j,\omega_i,\omega_j)$ 由式 (24) 给出, 值函数 $V_i\left(\delta_i\right)$ 由式 (23) 给出, 故式 (37) 可写为

$$\begin{aligned}
&Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)\\
&\quad=\frac{1}{2}\delta_i^{\mathrm{T}}P_i\delta_i+\frac{1}{2}\delta_i^{\mathrm{T}}P_i\left(A\delta_i-\sum_{j\in N_i}a_{ij}Bu_j+\left(d_i+g_i\right)Bu_i-\sum_{j\in N_i}a_{ij}D\omega_j+\left(d_i+g_i\right)D\omega_i\right)\\
&\quad\quad+\frac{1}{2}\left(A\delta_i-\sum_{j\in N_i}a_{ij}Bu_j+\left(d_i+g_i\right)Bu_i-\sum_{j\in N_i}a_{ij}D\omega_j+\left(d_i+g_i\right)D\omega_i\right)P_i\delta_i\\
&\quad\quad+\frac{1}{2}\delta_i^{\mathrm{T}}Q_{ii}\delta_i+\frac{1}{2}u_i^{\mathrm{T}}R_{ii}u_i+\frac{1}{2}\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\frac{1}{2}\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\frac{1}{2}\gamma^2\sum_{j\in N_j}\omega_j^{\mathrm{T}}T_{ij}\omega_j.
\end{aligned} \tag{38}$$

由式 (38) 进一步可以得到

$$\begin{aligned}
&Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)\\
&=\frac{1}{2}Z_i^{\mathrm{T}}\begin{bmatrix}
P_i+Q_{ii}+P_iA+A^{\mathrm{T}}P_i & \left(d_i+g_i\right)P_iB & -\operatorname{row}\left(a_{ij}P_iB\right) & \left(d_i+g_i\right)P_iD & -\operatorname{row}\left(a_{ij}P_iD\right)\\
\left(d_i+g_i\right)B^{\mathrm{T}}P_i & R_{ii} & 0 & 0 & 0\\
-\operatorname{col}\left(a_{ij}B^{\mathrm{T}}P_i\right) & 0 & \operatorname{diag}\left(R_{ij}\right) & 0 & 0\\
\left(d_i+g_i\right)D^{\mathrm{T}}P_i & 0 & 0 & -\gamma^2T_{ii} & 0\\
-\operatorname{col}\left(a_{ij}D^{\mathrm{T}}P_i\right) & 0 & 0 & 0 & \operatorname{diag}\left(-\gamma^2T_{ij}\right)
\end{bmatrix}_{j\in N_i}Z_i,
\end{aligned} \tag{39}$$

其中 $Z_i=[\,\delta_i^{\mathrm{T}}\ u_i^{\mathrm{T}}\ u_j^{\mathrm{T}}\ \omega_i^{\mathrm{T}}\ \omega_j^{\mathrm{T}}\,]^{\mathrm{T}}$. 为避免对系统动力学信息的依赖, 用 H_i 表示式 (39) 中的矩阵, 有

$$Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)=\frac{1}{2}Z_i^{\mathrm{T}}H_iZ_i=\frac{1}{2}Z_i^{\mathrm{T}}\begin{bmatrix}
H_{\delta_i\delta_i} & H_{\delta_iu_i} & H_{\delta_iu_{N_i}} & H_{\delta_i\omega_i} & H_{\delta_i\omega_{N_i}}\\
H_{u_i\delta_i} & H_{u_iu_i} & H_{u_iu_{N_i}} & H_{u_i\omega_i} & H_{u_i\omega_{N_i}}\\
H_{u_{N_i}\delta_i} & H_{u_{N_i}u_i} & H_{u_{N_i}u_{N_i}} & H_{u_{N_i}\omega_i} & H_{u_{N_i}\omega_{N_i}}\\
H_{\omega_i\delta_i} & H_{\omega_iu_i} & H_{\omega_iu_{N_i}} & H_{\omega_i\omega_i} & H_{\omega_i\omega_{N_i}}\\
H_{\omega_{N_i}\delta_i} & H_{\omega_{N_i}u_i} & H_{\omega_{N_i}u_{N_i}} & H_{\omega_{N_i}\omega_i} & H_{\omega_{N_i}\omega_{N_i}}
\end{bmatrix}Z_i, \tag{40}$$

其中 $H_{u_iu_i}$, $H_{u_i\delta_i}$, $H_{u_i\omega_i}$, $H_{\delta_i\delta_i}$, $H_{\delta_i\omega_i}$, $H_{\omega_i\omega_i}$ 为与 Z_i 中变量 u_i, δ_i, ω_i 相乘所对应的矩阵元素的记号.

推论1 假设图 $\mathcal{G}$ 中的子链是强连接的, 相邻子链的生产策略和不确定需求策略最优, 若第 i 条子链的 Q 函数如式 (40) 所示, 则有

$$Q_i^*\left(\delta_i,u_i^*,u_j^*,\omega_i^*,\omega_j^*\right)=\min_{u_i}\max_{\omega_i}Q_i\left(e_i,u_i,u_j^*,\omega_i,\omega_j^*\right). \tag{41}$$

通过利用积分强化学习技术, 可以把第 i 条子链的值函数 (14) 写成积分形式的贝尔曼方程:

$$\begin{aligned} &V_i^*\left(\delta_i(t+T)\right)=V_i^*\left(\delta_i(t)\right)\\ &-\frac{1}{2}\int_t^{t+T}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t. \end{aligned} \tag{42}$$

那么, 根据推论 1, $Q_i^*\left(\delta_i,u_i^*,u_j^*,\omega_i^*,\omega_j^*\right)$ 在上述积分贝尔曼方程的条件下可以被重写为

$$\begin{aligned} &Q_i^*\left(\delta_i(t+T),u_i^*(t+T),u_j^*(t+T),\omega_i^*(t+T),\omega_j^*(t+T)\right)\\ &\quad=Q_i^*\left(\delta_i(t),u_i^*(t),u_j^*(t),\omega_i^*(t),\omega_j^*(t)\right)\\ &\quad\quad-\frac{1}{2}\int_t^{t+T}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t. \end{aligned} \tag{43}$$

根据式 (40), 进一步可以得到

$$\begin{aligned} &Z_i^{*\mathrm{T}}(t+T)H_iZ_i^*(t+T)\\ &=Z_i^{*\mathrm{T}}(t)H_iZ_i^*(t)-\int_t^{t+T}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i+u_i^{\mathrm{T}}R_{ii}u_i+\sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j-\gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i-\gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t, \end{aligned} \tag{44}$$

其中 $Z_i^*(t)=[\,\delta_i^{\mathrm{T}}(t)\ u_i^{*T}(t)\ u_j^{*T}(t)\ \omega_i^{*T}(t)\ \omega_j^{*T}(t)\,]^{\mathrm{T}}$. 根据第 i 条子链的 Q 函数 (39), 一个等价于式 (25) 的最优生产率 u_i^* 可以通过如下公式获得:

$$u_i^*=\arg\min_{u_i}Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)=-H_{u_iu_i}^{-1}H_{u_i\delta_i}\delta_i, \tag{45}$$

同理, 可以得到

$$\omega_i^*=\arg\max_{\omega_i}Q_i\left(\delta_i,u_i,u_j,\omega_i,\omega_j\right)=-H_{\omega_i\omega_i}^{-1}H_{\omega_i\delta_i}\delta_i. \tag{46}$$

根据式 (44)~(46) 可知, 策略评估和策略更新时不需要产品与供应链协同演进系统的模型信息. 矩阵 H_i 可以利用在采样间隔 T 内获取的数据采用最小二乘法进行求解. 将策略迭代应用于 Q-learning 算法中, 进而可以得到算法 1.

关于求解零和图博问题的 Q-learning 算法收敛性可参见文献 [40]. 通过 Q-learning 求得的零和图博弈解在不需要系统模型的情况下可以以任何精度逼近鞍点. 换句话说, H_∞ 控制器能够通过在线的基于 Q-learning 的强化学习方法求解.

定理3 对于产品与供应链协同演进系统 (1), 假设 1~4 条件成立, 子链节点及连接所构成的拓扑图 $\mathcal{G}$ 具有强连接关系, 若存在正定函数 $V_i\,(i\in N)$ 满足 HJI 方程 (24), 使得相邻子链的生产策略和不确定需求策略最优, 则产品与供应链协同演进系统在模型未知及 DoS 网络攻击下, 可以实现各子链与链主的库存状态达到一致, 且系统满足一定用户需求并可以 $\gamma\geqslant\gamma_{\min}$ 水平抑制牛鞭效应.

证明 由定理 1, 定理 2, 推论 1 易得.

注5 与已有结果如文献 [43~46] 等利用静态模型研究产品与供应链系统的变更设计不同, 本文基于数据驱动技术利用动态模型研究产品与供应链系统的变更控制问题, 所得结果更能体现系统演进

算法 1 Q-learning algorithm based on policy iterations

Input: the initial admissible production rate u_i^0, the uncertain demand ω_i^0, the bullwhip effect attenuation level γ^0, and H_i^0 are given for each subchain i, $\forall i = 1, 2, \ldots, N$.

for $r = 0, 1, \ldots$:

(1) Policy evaluation: iteratively resolving the matrix H_i for the $r+1$ time according to (44),

$$\begin{aligned} &Z_i^{\mathrm{T}}(t+T)H_iZ_i(t+T)\\ &= Z_i^{*T}(t)H_iZ_i(t) - \int_t^{t+T}\left(\delta_i^{\mathrm{T}}Q_{ii}\delta_i + u_i^{\mathrm{T}}R_{ii}u_i + \sum_{j\in N_i}u_j^{\mathrm{T}}R_{ij}u_j - \gamma^2\omega_i^{\mathrm{T}}T_{ii}\omega_i - \gamma^2\sum_{j\in N_i}\omega_j^{\mathrm{T}}T_{ij}\omega_j\right)\mathrm{d}t. \end{aligned} \tag{47}$$

(2) Policy iteration: updating the $r+1$ time production rate u_i^{r+1} and the uncertain demand ω_i^{r+1} with the obtained matrix H_i, $\forall i = 1, 2, \ldots, N$,

$$u_i^{r+1} = -H_{u_iu_i}^{-1}H_{u_i\delta_i}\delta_i,\ \ \omega_i^{r+1} = -H_{\omega_i\omega_i}^{-1}H_{\omega_i\delta_i}\delta_i,\ \ \gamma^{r+1} = \gamma^r + \epsilon_0. \tag{48}$$

until $\left\|H_i^{r+1} - H_i^r\right\| \leqslant \varepsilon_0$ (ϵ_0, ε_0 are the given constants).

end

的动力学特性. 另一方面, 基于数据驱动的 Q-learning 算法设计, 目前大多研究结果主要通过自适应动态规划来实现, 如文献 [47] 利用评价神经网络来逼近每个智能体的 Q 函数, 用执行神经网络来逼近控制器, 但这种方法不可避免地导致逼近误差, 为控制器设计带来了新的不确定性. 本文将最小二乘法与强化学习技术相结合实现 Q-learning 算法求解核心矩阵, 有效地避免神经网络逼近带来的误差.

4 算例仿真

本节以中石化汽油协同生产供应链为例, 开展仿真实验验证本文所提算法的可行性与有效性.

汽油的生产过程主要包括原油预处理、常减压蒸馏等步骤, 具体过程如下.

(1) 原油预处理. 从油田送往炼油厂的原油往往含盐 (主要是氯化物), 带水 (溶于油或呈乳化状态), 可导致设备的腐蚀, 在设备内壁结垢影响成品油的成分, 需要在加工前脱除. 常用的方法是加破乳剂和水, 使油中的水集聚, 并从油中分出, 而盐份溶于水中, 再加以高压电场配合, 使形成较大水滴顺利除去.

(2) 常减压蒸馏. 习惯上常压蒸馏和减压蒸馏合称常减压蒸馏, 基本属物理过程. 原料油在蒸馏塔里按蒸发能力分成沸点范围不同的油品 (称为馏分), 这些油有的经调合, 加添加剂后以产品形式出厂, 相当大的部分是后续加工装置的原料. 因此, 常减压蒸馏又被称为原油的一次加工. 包括 3 个工序: 原油的脱盐、脱水, 常压蒸馏, 减压蒸馏. 常减压装置产品主要作为下游生产装置的原料, 包括石脑油、煤油、柴油、蜡油、渣油, 以及汽油等.

以中石化汽油生产供应链的 4 个炼油企业为协同场景, 考虑每个企业包含两个主要生产流程且结构相似, 从而构成 4 条子链, 如图 2 所示. 设 4 条子链的两个主要生产流程的初始库存量 (单位: 千吨) 分别为 $x_1 = [0.3, 0.2]^{\mathrm{T}}$, $x_2 = [0.7, 0.5]^{\mathrm{T}}$, $x_3 = [0.6, 0.7]^{\mathrm{T}}$, $x_4 = [0.8, 0.55]^{\mathrm{T}}$. 为验证所采用预测算法的有效性, 在如图 3 所示网络攻击造成的数据丢包情况下, 用第一条子链 200 天的历史库存数据预测未来 50 天的库存数据, 得到如图 4(a) 和 (b) 所示库存量预测结果.

下面考虑对汽油生产供应链系统进行变更补偿设计以确保系统库存状态一致且有效抑制牛鞭效应. 选择合适的权值矩阵并通过策略迭代 Q-learning 算法求解最小生产率以及最大不确定需求, 进一步可得各个子链的库存状态和跟踪误差状态曲线, 如图 5∼8 所示. 为验证变更补偿措施的有效性, 本

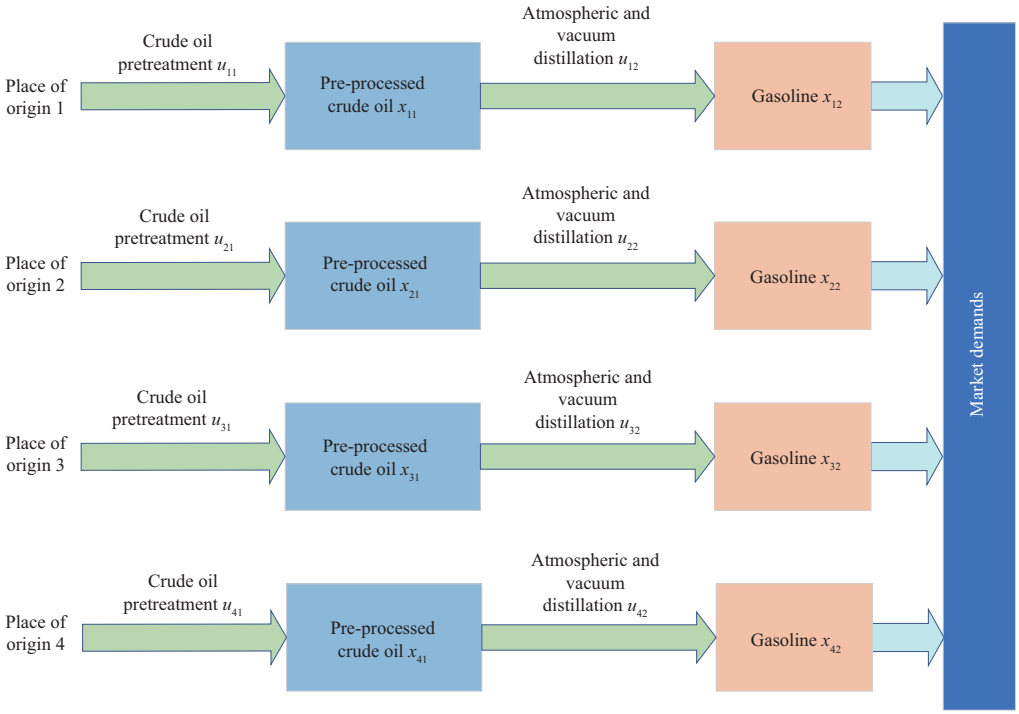

图 2　(网络版彩图) 汽油生产流程示意图

Figure 2　(Color online) Diagram of gasoline production process

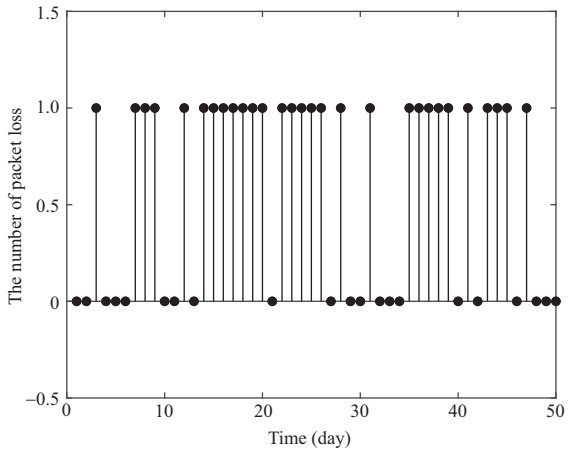

图 3　S-C 通道丢包情况

Figure 3　S-C channel packet loss

文对引入变更前后的各子供应链库存情况以及跟踪误差进行对比. 图 5 和 7 分别展示了 DoS 攻击下在引入变更补偿机制前后各个子链库存水平的变化情况. 从图中仿真结果可以看出, 在系统引入变更补偿机制后, 系统在第 19 天左右可以有效地跟踪协同演进系统设定的用户需求, 即达到一致. 图 6 和 8 展示了引入变更补偿机制前后各个子链与链主的库存跟踪误差水平变化, 即在系统引入变更补偿机制后 4 条子链的库存状态和链主库存状态之间的误差逐渐趋于 0, 这也验证了子链的库存状态能够与

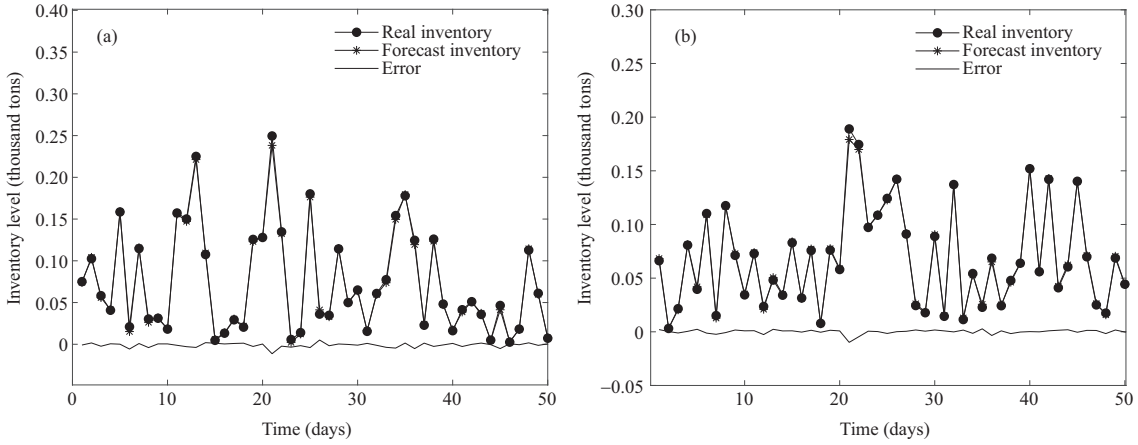

图 4 原油预处理过程 (a) 和常减压蒸馏过程 (b) 库存预测图

Figure 4 Forecast inventories of the crude oil pretreatment process (a) and the atmospheric and vacuum distillation process (b)

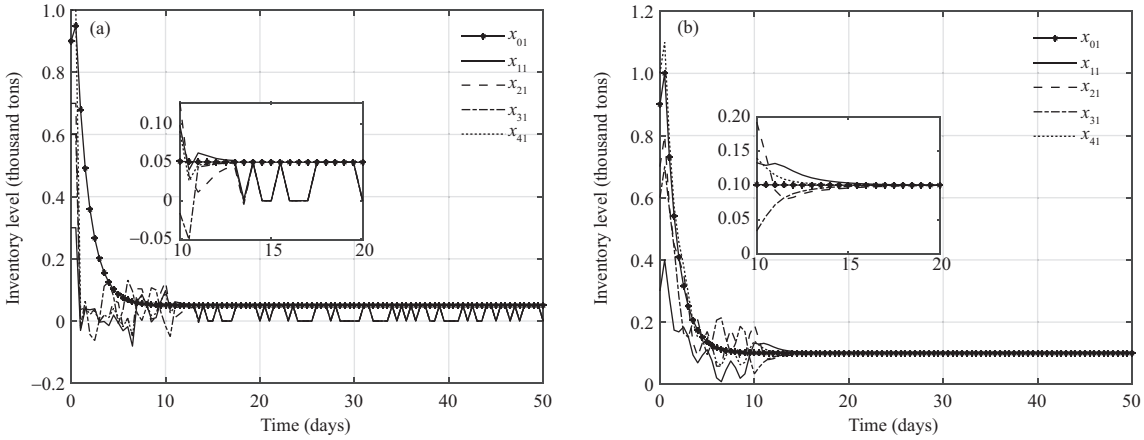

图 5 变更前 (a) 和变更后 (b) 原油预处理过程库存状态变化曲线

Figure 5 Inventory status curves of the crude oil pretreatment process before (a) and after (b) the proposed change control method

用户需求同步一致. 为了进一步验证变更补偿机制对产品适应度以及用户满意度的影响, 本文考虑如表 1 所列指标值, 即网络攻击下交货时间以及库存平均跟踪误差. 由表 1 可以看出, 在网络攻击下通过引入变更补偿机制可以有效提升系统的交货时间 (即系统实现库存一致达到满足合同规定将产品交付客户的时间) 且降低平均库存跟踪误差. 综上, 本文提出的变更机制下基于 Q-learning 的 H_∞ 控制器能够使得产品与供应链协同演进系统在抵制 DoS 攻击的同时还可以有效地抑制牛鞭效应并实现领导 – 跟随一致性.

5 结论

针对基于分布式 CPS 的产品与供应链协同演进系统, 本文研究了系统模型未知并在 DoS 攻击下的应急变更控制问题. 利用数据驱动技术, 根据历史数据和参考信息设计应急防御变更机制应对系统

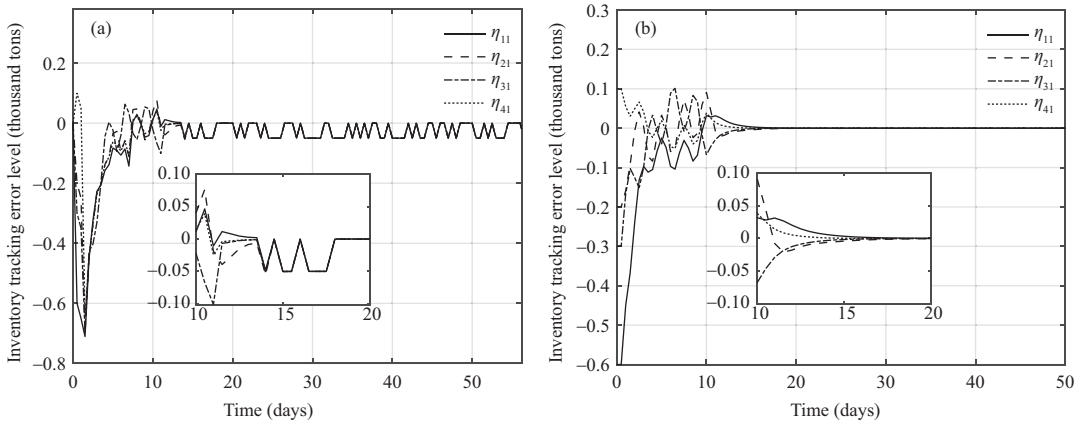

图 6　变更前 (a) 和变更后 (b) 原油预处理过程库存跟踪误差变化曲线

Figure 6　Inventory tracking error curves of the crude oil pretreatment process before (a) and after (b) the proposed change control method

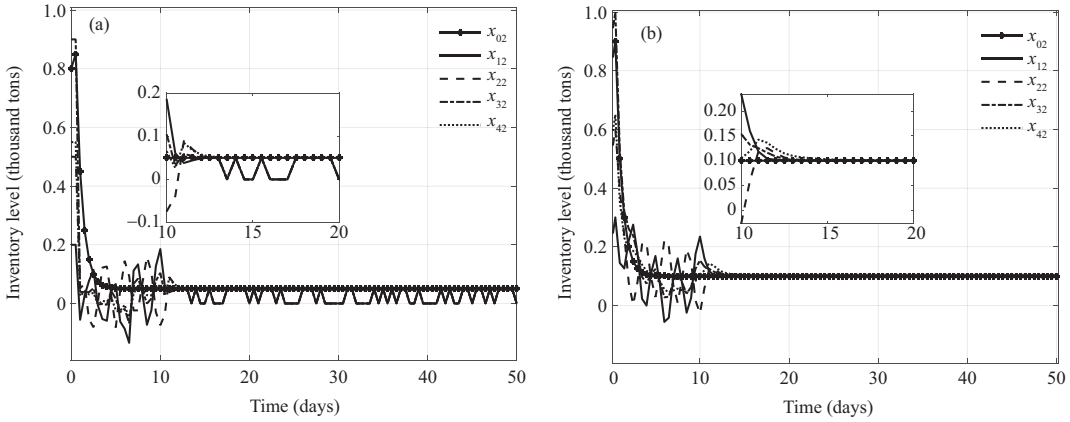

图 7　变更前 (a) 和变更后 (b) 常减压蒸馏过程库存状态变化曲线

Figure 7　Inventory status curves of the atmospheric and vacuum distillation process before (a) and after (b) the proposed change control method

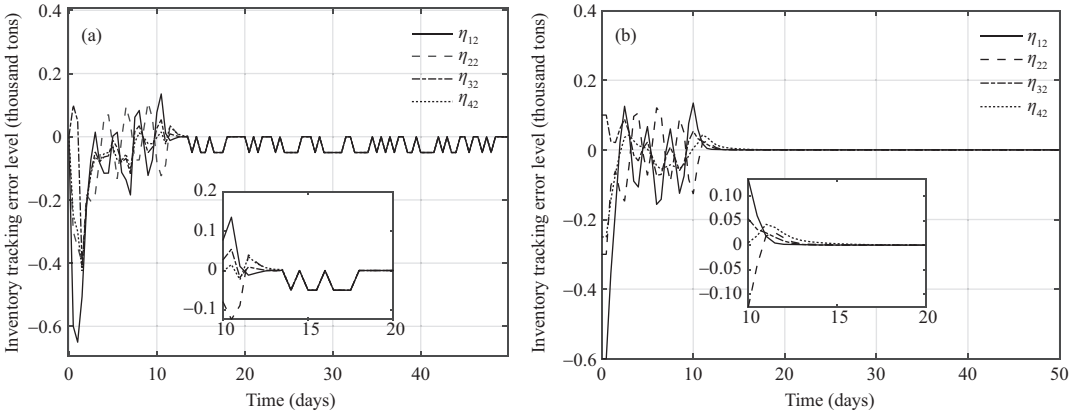

图 8　变更前 (a) 和变更后 (b) 常减压蒸馏过程库存跟踪误差变化曲线

Figure 8　Inventory tracking error curves of the atmospheric and vacuum distillation process before (a) and after (b) the proposed change control method

表 1 变更前后指标对比

Table 1 Comparison of indicators before and after the proposed change control method

Index (under cyber-attacks)	Before design	After design
Delivery time t	∞	19
Average tracking error of inventory level $\eta(t)$	0.042164	0.017594

传感器 – 控制器通信信道受到 DoS 攻击情形, 通过零和微分图博弈理论与强化 Q-learning 技术设计 H_∞ 一致性控制器, 保证系统在 DoS 攻击导致设计变更情形下, 使产品与供应链协同演进系统达到领导跟随 H_∞ 一致, 并可以抑制不确定需求和变更设计带来的牛鞭效应, 同时保证系统具有一定的产品适应度及用户满意度. 仿真算例验证了文中所提方法的正确性与有效性. 开展基于数据驱动的变更效应传播分析与多智能体系统仿真及数据可视化系统研究, 将是产品与供应链协同演进系统变更控制设计需要进一步研究的重要课题.

参考文献

1 Li Q K, Lin H, Tan X, et al. H_∞ consensus for multiagent-based supply chain systems under switching topology and uncertain demands. IEEE Trans Syst Man Cybern Syst, 2020, 50: 4905–4918

2 Khosrojerdi A, Zegordi S H, Allen J K, et al. A method for designing power supply chain networks accounting for failure scenarios and preventive maintenance. Eng Optimization, 2016, 48: 154–172

3 Yang Y, Huang Z, Zhang B, et al. Closed-loop supply chain network static equilibrium and dynamics under pollution permits system. Environ Prog Sustain Energy, 2019, 38: e13021

4 Gui W H, Zeng Z H, Chen X F, et al. Knowledge-driven process industry smart manufacturing. Sci Sin Inform, 2020, 50: 1345–1360 [桂卫华, 曾朝晖, 陈晓方, 等. 知识驱动的流程工业智能制造. 中国科学: 信息科学, 2020, 50: 1345–1360]

5 Chitrakar B, Zhang M, Bhandari B. Improvement strategies of food supply chain through novel food processing technologies during COVID-19 pandemic. Food Control, 2021, 125: 108010

6 Agarwala N, Chaudhary R D. 'Made in China 2025': poised for success? India Q, 2021, 77: 424–461

7 Wang Z, Xie W, Wang B, et al. A survey on recent advanced research of CPS security. Appl Sci, 2021, 11: 3751

8 Zhang D, Feng G, Shi Y, et al. Physical safety and cyber security analysis of multi-agent systems: a survey of recent advances. IEEE CAA J Autom Sin, 2021, 8: 319–333

9 Al-Hawawreh M, Sitnikova E, Aboutorab N. X-IIoTID: a connectivity-agnostic and device-agnostic intrusion data set for industrial Internet of Things. IEEE Internet Things J, 2022, 9: 3962–3977

10 Ge H, Yue D, Xie X, et al. A unified modeling of muti-sources cyber-attacks with uncertainties for CPS security control. J Franklin Inst, 2021, 358: 89–113

11 Zhao Y, Zhu F. Security control of cyber-physical systems under denial-of-service sensor attack: a switching approach. In: Proceedings of IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), 2021. 1112–1117

12 Yan S, Gu Z, Fei S M, et al. Memory-based event-triggered secure state estimation of cyber-physical systems. Sci Sin Inform, 2021, 51: 1302–1315 [严沈, 顾洲, 费树岷, 等. 基于记忆型事件触发的信息物理系统的安全状态估计. 中国科学: 信息科学, 2021, 51: 1302–1315]

13 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. Neurocomputing, 2019, 338: 101–115

14 Chen X, Wang Y, Hu S. Event-based robust stabilization of uncertain networked control systems under quantization and denial-of-service attacks. Inf Sci, 2018, 459: 369–386

15 Hu S, Yue D, Xie X, et al. Resilient event-triggered controller synthesis of networked control systems under periodic DoS jamming attacks. IEEE Trans Cybern, 2018, 49: 4271–4281

16 Li M Y, Koutsopoulos I, Poovendran R. Optimal jamming attack strategies and network defense policies in wireless sensor networks. IEEE Trans Mobile Comput, 2010, 9: 1119–1133

17 Osanaiye O, Alfa A S, Hancke G P. A statistical approach to detect jamming attacks in wireless sensor networks.

Sensors, 2018, 18: 1691
18 Wang M, Xu B. Guaranteed cost control of cyper-physical systems under periodic DoS jamming attacks. In: Proceedings of the 37th Chinese Control Conference (CCC), 2018. 6241–6246
19 Zhang H, Cheng P, Shi L, et al. Optimal denial-of-service attack scheduling with energy constraint. IEEE Trans Automat Contr, 2015, 60: 3023–3028
20 Zhang H, Cheng P, Shi L, et al. Optimal DoS attack scheduling in wireless networked control system. IEEE Trans Contr Syst Technol, 2015, 24: 843–852
21 Foroush H S, Martinez S. On event-triggered control of linear systems under periodic denial-of-service jamming attacks. In: Proceedings of IEEE 51st IEEE Conference on Decision and Control (CDC), 2012. 2551–2556
22 de Persis C, Tesi P. Input-to-state stabilizing control under denial-of-service. IEEE Trans Automat Contr, 2015, 60: 2930–2944
23 Yuan Y, Yuan H H, Guo L, et al. Resilient control of networked control system under DoS attacks: a unified game approach. IEEE Trans Ind Inf, 2016, 12: 1786–1794
24 Zhao Y, He X, Zhou D. Optimal joint control and triggering strategies against denial of service attacks: a zero-sum game. IET Control Theor Appl, 2017, 11: 2352–2360
25 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-infinity control. Automatica, 2007, 43: 473–481
26 Solomatine D P, See L M, Abrahart R J. Data-Driven Modelling: Concepts, Approaches and Experiences. Berlin: Springer, 2009. 17–30
27 Hou Z S, Jin S T. Model Free Adaptive Control. Beijing: Science Press, 2013 [侯忠生, 金尚泰. 无模型自适应控制. 北京: 科学出版社, 2013]
28 Wei Q, Liu D, Lin H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. IEEE Trans Cybern, 2016, 46: 840–853
29 Bu X, Yu Q, Hou Z, et al. Model free adaptive iterative learning consensus tracking control for a class of nonlinear multiagent systems. IEEE Trans Syst Man Cybern Syst, 2019, 49: 677–686
30 Sun J, Qi G, Zhu Z. A sparse neural network based control structure optimization game under DoS attacks for DES frequency regulation of power grid. Appl Sci, 2019, 9: 2217
31 Brunton S L, Kutz J N. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge: Cambridge University Press, 2022
32 Qiu X, Wang Y, Xie X, et al. Resilient model-free adaptive control for cyber-physical systems against jamming attack. Neurocomputing, 2020, 413: 422–430
33 Liu Y Y, Wang Z S, Shi Z. H_∞ tracking control for linear discrete-time systems via reinforcement learning. Int J Robust Nonlinear Control, 2020, 30: 282–301
34 Peng Z N, Zhang J F, Hu J P, et al. Optimal containment control of continuous-time multi-agent systems with unknown disturbances using data-driven approach. Sci China Inf Sci, 2020, 63: 209205
35 Su M Y, Che W W. Fault-tolerant control for model-free networked control systems under DoS attacks. J Franklin Inst, 2021, 358: 9023–9033
36 Zhang Y J, Niu H, Tao J M, et al. Virtual unmodeled dynamic compensation and data-driven nonlinear generalized predictive control. Sci Sin Inform, 2021, 51: 1146–1155 [张亚军, 牛宏, 陶金梅, 等. 虚拟未建模动态补偿与数据驱动的非线性广义预测控制. 中国科学: 信息科学, 2021, 51: 1146–1155]
37 Zhang J, Wang F Y, Wang K, et al. Data-driven intelligent transportation systems: a survey. IEEE Trans Intell Transp Syst, 2011, 12: 1624–1639
38 Tao F, Qi Q, Liu A, et al. Data-driven smart manufacturing. J Manuf Syst, 2018, 48: 157–169
39 Peng Y, Chen Q, Sun W. Reinforcement Q-learning algorithm for H_∞ tracking control of unknown discrete-time linear systems. IEEE Trans Syst Man Cybern Syst, 2020, 50: 4109–4122
40 Mu C, Zhao Q, Gao Z, et al. Q-learning solution for optimal consensus control of discrete-time multiagent systems using reinforcement learning. J Franklin Inst, 2019, 356: 6946–6967
41 Lee H L, Padmanabhan V, Whang S. The bullwhip effect in supply chains. IEEE Eng Manag Rev, 2015, 43: 108–117
42 Baldi P. Gradient descent learning algorithm overview: a general dynamical systems perspective. IEEE Trans Neural Netw, 1995, 6: 182–195

43 Wright I C. A review of research into engineering change management: implications for product design. Des Studies, 1997, 18: 33–42
44 Jarratt T A W, Eckert C M, Caldwell N H M, et al. Engineering change: an overview and perspective on the literature. Res Eng Des, 2011, 22: 103–124
45 Zheng Y J, Yang Y, Zhang N, et al. Dynamic majorization scheme for engineering change propagation paths in complex product. Comput Integr Manuf Syst, 2018, 24: 474–483 [郑玉洁, 杨育, 张娜, 等. 复杂产品工程变更传播路径动态优化. 计算机集成制造系统, 2018, 24: 474–483]
46 Reitsma E, Hilletofth P, Johansson E. Supply chain design during product development: a systematic literature review. Production Planning Control, 2023, 34: 1–18
47 Vamvoudakis K G. Q-learning for continuous-time graphical games on large networks with completely unknown linear system dynamics. Int J Robust Nonlinear Control, 2017, 27: 2900–2920

Data-driven change control design for product and supply chain synchronous evolution systems under cyber-attacks

Qingkui LI[1*], Xuefeng GAO[1], Chen PENG[2], Yunlong ZHANG[1] & Junkai YI[1]

1. *School of Automation, Beijing Information Science and Technology University, Beijing* 100192*, China*;
2. *School of Mechanical Engineering and Automation, Shanghai University, Shanghai* 200072*, China*
* Corresponding author. E-mail: sdlqk01@bistu.edu.cn

Abstract A product and supply chain synchronous evolution system (PSCSES) is a kind of complex distributed cyber-physical system (CPS), which contains several unmodeled dynamics and uncertainties, and the system structure and parameters are vulnerable to changes due to drastic demand fluctuations and unexpected network events. Thus, a product design change is an important way to maintain the stability of the synchronous evolution system to meet customers' demands and ensure economic efficiency. In this paper, the emergency change control problem of a distributed CPS-based PSCSES under DoS attacks is studied. First, an emergency change compensation mechanism in which a predictor based on recurrent neural networks is proposed to compensate for the lost inventory data caused by cyber-attacks. Second, the H_∞ consensus control problem is transformed into a multiplayer zero-sum graph game using game theory, and a Q-learning-based H_∞ consensus control protocol is designed, such that sufficient conditions that guarantee the consensus of the system can be derived. The bullwhip effects caused by the customers' demand and design change are analyzed thereby. Third, using the strategy iteration algorithm, the online solution is resolved for the controller design under the unknown model. Finally, the effectiveness of the proposed method is verified through the simulation of a collaborative production supply chain.

Keywords cyber-attacks, data-driven, supply chain, change control design, multi-agent