控制与决策
Control and Decision

网络攻击下的信息物理系统安全状态估计研究综述

杨光红, 芦安洋, 安立伟

引用本文:
杨光红,芦安洋,安立伟. 网络攻击下的信息物理系统安全状态估计研究综述[J]. 控制与决策, 2023, 38(8): 2093−2105.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2023.0885

网络攻击下的信息物理系统安全状态估计研究综述

杨光红[1,2†]，芦安洋[1]，安立伟[1]

(1. 东北大学 信息科学与工程学院，沈阳 110819；
2. 东北大学 流程工业综合自动化国家重点实验室，沈阳 110819)

摘 要: 近年来,信息物理系统在工业界的广泛应用引起了人们对系统安全问题的极大关注. 信息物理系统对通信网络的深度依赖,使得网络攻击成为其中最为严峻的威胁之一,特别是那些能够干扰系统状态认知的攻击,因此,安全状态估计(即在遭受攻击时正确估计系统状态)已成为各界广泛关注的安全问题之一. 此文旨在总结网络攻击下信息物理系统安全状态估计研究的进展. 首先,介绍典型的网络攻击,并详细阐述在稀疏攻击下的安全状态估计问题. 其次,探讨集中式安全状态估计和分布式安全状态估计的研究现状. 在考虑稀疏攻击下安全状态估计问题的难点时,关键在于如何快速找到受到攻击的信道集合(这可能涉及到高计算复杂度). 因此,将安全状态估计方法分为遍历搜索和非遍历搜索两大类,并对现有方法的优缺点进行归纳总结和详细阐述. 然后,介绍稀疏攻击下信息物理系统安全状态能观性分析的研究现状. 现有的研究结果表明:增加检测机制或先验知识可以缓解在稀疏攻击下安全状态估计所需的基础冗余度要求;同时,通过区分攻击和故障,也能有效降低传感器冗余度要求. 最后,对信息物理系统安全状态估计仍然存在的问题进行展望,并提出一些可能的解决方向.

关键词: 信息物理系统；网络攻击；安全状态估计；分布式安全状态估计；安全状态能观性

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2023.0885

引用格式: 杨光红,芦安洋,安立伟. 网络攻击下的信息物理系统安全状态估计研究综述 [J]. 控制与决策, 2023, 38(8): 2093-2105.

A survey on secure state estimation of cyber-physical systems under cyber attacks

YANG Guang-Hong[1,2†], *LU An-yang*[1], *AN Li-wei*[1]

(1. College of Information Science and Engineering，Northeastern University，Shenyang 110819，China；2. State Key Laboratory of Synthetical Automation of Process Industries，Northeastern University，Shenyang 110819，China)

Abstract: In recent years, the widespread application of cyber-physical systems (CPS) in the industrial sphere has elicited substantial attention towards system security issues. Given the deep reliance of CPS on communication networks, cyber-attacks have emerged as one of the most severe threats, particularly those capable of disrupting system state awareness. Thus, secure state estimation — accurately gauging the system state under attack — has become a security concern of widespread interest. This paper aims to summarize the advances in secure state estimation research under cyber-attacks for CPS. Initially, we introduce typical cyber-attacks and elaborate on the secure state estimation problem under sparse attacks. Subsequently, the state of research on centralized and distributed secure state estimation is explored. When considering the difficulty of secure state estimation under sparse attacks, the crux lies in swiftly identifying the set of channels under attack — a process potentially involving high computational complexity. Therefore, we categorize secure state estimation methods into exhaustive and non-exhaustive search types, summarizing and elaborating on the strengths and weaknesses of current methods. Further, we present the status of research on observability analysis for CPS's secure state under sparse attacks. Existing studies suggest that increasing detection mechanisms or prior knowledge can alleviate the baseline redundancy requirement for secure state estimation under sparse attacks. Meanwhile, distinguishing between attacks and failures can effectively reduce sensor redundancy requirements. Finally, the paper anticipates ongoing challenges in secure state estimation for CPS and proposes potential directions for resolution.

Keywords: cyber-physical systems；cyber attacks；secure state estimation；distributed secure state estimation；secure state observability

收稿日期: 2023-05-23；**录用日期:** 2023-06-23.
基金项目: 国家自然科学基金项目(61621004, 62103091).
†通讯作者. E-mail: yangguanghong@ise.neu.edu.cn.

0 引 言

随着现代控制技术、计算机技术和通信技术的迅速发展,传统的单点技术已无法满足新时代生产装备信息化和网络化的要求[1]. 为了满足新时代生产装备的需求,信息物理系统(cyber-physical system, CPS)应运而生. 时至今日,CPS已经受到了学术界和工业界的广泛关注,成为当前控制领域的前沿研究方向. CPS采用多功能传感器和执行器以及计算和通信设备等网络组件,通过有线或无线共享通信网络紧密连接,完成数据感知、收集、处理和传输任务. 利用先进的感知、计算、通信、控制技术,实现了物理空间与信息空间中元素的相互映射、实时交互和高效协同[2].

如图1所示,CPS由感知、通信、计算、控制环节组成. 传感器对物理系统状态信号进行采集;计算处理单元对传感器采集到的物理系统状态信号进行计算和分析;通过控制执行单元得到的计算结果对物理系统进行控制;数据的流动通过通信网络实现. 然而,信息空间与物理空间的深度耦合也使得CPS面临前所未有的安全挑战. 一方面,信息系统为物理系统的动态感知、分析决策和精准控制提供了强有力的支持,从而保障系统的安全稳定运行;另一方面,CPS对通信网络的深度依赖使得网络攻击可能通过信息物理交互作用影响物理空间,导致物理故障的扩大.

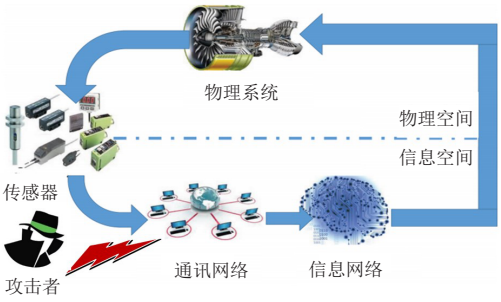

图1 信息物理系统(CPS)基本结构[3]

由于通信网络的使用,原本封闭的物理系统变得开放,从而增加了信息物理系统受到攻击的风险. 相比于传统的控制系统,CPS具有更广泛的攻击面,攻击者可以通过针对不同网络层次的恶意攻击造成巨大的破坏. 任何对CPS的成功攻击都可能导致系统灾难性的损害,并产生难以承受的损失. 例如:2003年的蠕虫病毒攻击导致美国核电站的安全监测系统遭到破坏;2010年的“震网”病毒通过U盘传播,入侵了伊朗核电站的PLC控制软件代码,造成离心机失控,最终导致设备报废[4];2014年,欧洲许多工业制造系统遭受Havex木马的攻击,通过入侵工控系统,造成水电坝失控和核电站过载等后果[5];2015年,乌克兰的电力系统遭受恶意代码攻击,导致超过一半地区断电数小时,造成无法估量的经济损失[6]. CPS的受攻击后果严重影响经济安全,甚至对人类的生命安全构成威胁. 可以明显看出,获得正确的系统状态信息是确保系统正常运行的基础,而上述恶性事件的发生往往是由系统状态监测的失败造成的. 因此,研究CPS状态监测中的安全性问题具有重大的现实意义.

针对CPS状态监测安全性的研究主要从攻击者和防守者两个方向进行考虑(见图2). 攻击者通常可以根据安全状态可观性分析的结果有选择地增强攻击能力,通过对CPS进行脆弱性分析来设计适当的攻击策略,从而破坏系统的监测性能;而防守者通常可以根据安全状态可观性分析的结果有选择地优化系统结构,采取多种安全策略来抵抗对CPS的攻击影响,以正确地监测系统状态. 近年来,从防守者角度出发的安全状态估计问题得到了深入的研究,并引起了国内外学者的广泛关注. 本文围绕安全状态估计问题,从以下4个方面对现有的研究工作进行回顾:典型的网络攻击与安全状态估计、集中式安全状态估计研究、分布式安全状态估计研究以及安全可观性分析的研究.

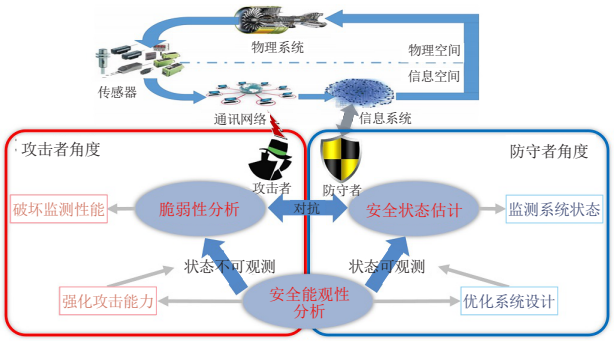

图2 信息物理系统中的安全性问题

1 典型的网络攻击与安全状态估计

1.1 典型的网络攻击

在网络环境下,虽然攻击种类多种多样,但常见的针对CPS的攻击主要可以分为两类:拒绝服务攻击(DoS attack)和错误数据注入攻击(false data injection attack).

1.1.1 拒绝服务攻击(DoS攻击)

DoS攻击是一种针对网络连通性的攻击,攻击者试图通过暂时或无限期中断互联网服务,使网络资源或机器对其用户不可用. 简单而言,DoS攻击影响了信息交换的及时性,导致数据包丢失. DoS攻击通常包括3个阶段:1) 攻击者创建了许多受其控制的攻击部队和用于执行DoS攻击的系统;2) 攻击者尽可能地增加节点或设备的数量;3) 攻击者通过发送大量

冗余的数据包来攻击目标系统[7]. 攻击者还经常使用欺骗性的IP地址来隐藏泄露节点的真实IP地址.

DoS攻击也可以表现为消耗运行不必要程序的能量或接管内存、插槽和CPU控制,这些攻击通常被视为带宽耗尽攻击和资源耗尽攻击. 如果攻击的频率和持续时间足够大,则CPS没有实时数据用于控制反馈,这可能导致CPS不稳定甚至崩溃. 然而,在实际场景中,作为攻击工具的数字设备其供能有限,攻击者的攻击能量预算也是有限的[8]. 因此,现有的研究大多考虑了攻击资源受限情况下的DoS攻击调度以及攻击对系统性能的影响. 关于最优的DoS攻击策略,文献[9]研究了具有能量约束的DoS攻击,旨在通过最大化线性二次高斯控制成本来确定最优攻击策略,并证明了在固定时间段内连续进行DoS攻击是最佳的策略. 文献[10]使用博弈论方法研究了DoS攻击下的策略问题. 因此,DoS攻击在不同应用场景下呈现出多种多样的形式.

1.1.2　错误数据注入攻击(FDI攻击)

错误数据注入攻击(FDI攻击)是一种影响传输数据可信度的网络攻击,攻击者在对系统内部模型参数和运行数据等有充分了解的基础上,通过篡改CPS的传感器和执行器信道的传输数据或数据包,从而形成更具欺骗性的FDI攻击. 需要指出的是,在不同的场景中,虚假数据注入攻击也被称为欺诈攻击(deception attack)或恶意攻击(malicious attack)[11-12]. 与DoS攻击不同,FDI攻击可以在欺骗检测器的同时尽可能地破坏CPS的稳定状态,因此,精心设计的FDI攻击往往对CPS 的可靠稳定运行构成严重威胁. 例如,攻击者可以直接向攻击节点发送虚假数据包或者篡改数据,在经过系统验证的原始数据包中注入虚假数据[13]. 与故障信号通常是随机且有界的不同,攻击者注入传感器测量中的错误数据可以是任意的,并且这些任意大小的错误数据可以通过遵循特定模型的规律来躲避异常检测.

FDI攻击不仅可以发生在传感器与执行器的信道之间,还可以发生在CPS的其他信道中. 例如,攻击者可以接管传感器节点并有意改变传感器的读数. 攻击者还可以注入虚假数据到控制程序中,从而故意误导应用程序的目标. 重放攻击也可以被视为一种特殊的错误数据注入攻击,因为接收者只能使用过去的数据进行顺序处理,而过去的数据会多次循环. 尽管检测系统能够轻松检测到这种攻击,但攻击者可以设计欺骗攻击以避开检测[7].

自从错误数据注入攻击被提出以来,已引起控制界学者的广泛关注. 文献[14]描述了系统检测和性能权衡的基本限制,并刻画了FDI攻击的效果. 文献[15]提出了由FDI攻击引起系统无限估计误差的充分必要条件. 文献[16]研究了两种高级的重置攻击,分别可以使系统估计误差无限增大和改变系统的目标状态,并给出了实现条件. 文献[17]显示,攻击者可以仅利用系统结构信息设计FDI攻击,在隐蔽性条件下破坏状态估计器的估计性能. 显然,由于其恶意性质,FDI攻击将严重干扰系统的估计性能,进而导致防守者对系统状态产生错误的认知,影响系统的正常运行. 因此,在数据被篡改的情况下如何获取正确的系统状态信息,为维持CPS的稳定运行提供信息基础,是当前CPS安全研究的重点之一.

1.2　安全状态估计

状态估计在信息物理系统的监测和控制中发挥着关键作用. 正如前文所述,CPS由处理单元组成,通过传感器和执行器网络对物理过程进行监控. 然而,由于其容易受到网络攻击的影响,特别是数据攻击,防守者可能对系统状态产生错误的认知,从而导致物理设施的损坏[18]. 因此,安全状态估计(也称为弹性状态估计),即在受到攻击干扰的情况下正确获取信息状态[19-20],成为本文讨论的重点.

安全状态估计的目标是通过设计安全状态估计策略,在部分测量信号被数据篡改的情况下消除或减弱攻击的影响,并正确地重构系统的状态,以保障系统的安全运行[21-23]. 如图3所示,远程估计器通过通信网络接收测量信号,进而获取系统状态信息. 然而,当网络受到攻击者的干扰时,测量信号可能会因为DoS攻击而中断,也可能会因为FDI攻击而传输部分错误数据. 在这种情况下,就需要设计安全状态估计策略,从受攻击干扰的数据中提取出正确的系统状态信息.

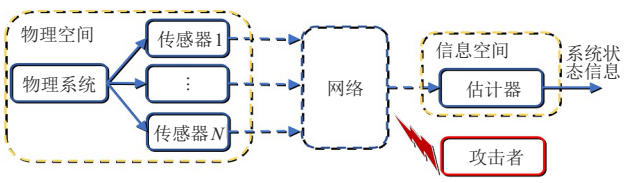

图 3　网络攻击下的系统状态估计

根据监测主体的不同,安全状态估计问题可以分为集中式安全状态估计和分布式安全状态估计两类. 在CPS中,无线传感器网络由大量传感器节点组成,这些节点负责收集物理系统的测量输出,并通过网络信道连接到估计器上[24]. 当通过单个节点收集到所有测量值并进行状态估计时,称为集中式安全状

态估计. 集中式估计器通常需要精确的协调,并且大量的通信和计算开销用于从无线传感器网络中的所有节点收集信息. 特别是当收集到的信息中的部分被篡改时,实现安全状态估计的关键在于确定受到攻击的信息传输通道(信道). 文献[25]分析了在稀疏传感器攻击下的状态可观性,结果显示,在s个传输通道受到攻击的情况下,原系统需要是$2s$稀疏可观测的(即通过移除任意$2s$个测量信道,基于剩余的测量信号,系统状态仍然可观测),以确保能够实现安全状态估计. 值得注意的是,安全状态估计本质上是一个组合优化问题,需要通过对潜在被攻击通道的排列组合来寻找正确的被攻击通道集合[26-27],其中的难点在于如何解决安全状态估计算法的高计算复杂度问题. 为此,文献[28-29]基于自适应切换策略提出了两类安全状态观测器设计方法,分别降低了安全状态估计问题的计算复杂度和存储成本. 而文献[30]针对稀疏传感器攻击,基于静态批处理方法,提出了一种低计算复杂度的受限集合划分方法.

对于带有多智能节点的CPS,在部分测量信号被攻击者篡改或部分节点被攻击者挟持的情况下,多个节点通过相互协作实现状态估计,被称为分布式安全状态估计. 由于传感单元分布在物理空间中并形成集群[31],分布式安全状态估计可以更好地探知CPS的运行状态,因而受到控制领域学者的广泛关注. 针对DoS攻击下非线性系统的安全状态估计问题,文献[32]基于切换方案和级联观测器技术,设计了一种具有切换补偿机制的新型弹性状态观测器来对抗DoS攻击. 文献[8]通过引入保持输入机制和级联观测器技术,提出了一种新的多观测器方案和切换算法,提高了状态估计的估计精度. 文献[33]针对执行器数据损坏、额外数据包注入攻击和基于集群的网络配置,开发了一种新的分布式混合伯努利随机集滤波器,用于联合攻击检测和安全状态估计. 文献[34]提出了一种同时估计状态和攻击信号的三环观测器,并设计了一种异构多智能体系统在同源攻击下的在线分布式安全状态估计. 文献[35]针对一类连续非线性系统在稀疏攻击和扰动下,构造了一种高增益K滤波器来估计未测量的状态,并引入监测功能和切换方案,以排除受攻击的测量.

在非可靠网络环境下,攻击者可能会通过劫持信息传输通道来注入攻击信号. 然而,由于资源的限制,攻击者通常无法攻击所有的传输信道. 因此,稀疏攻击得到了广泛关注,现有的研究大多假设攻击是s稀疏的,即在多个信息传输通道中最多有s个信道被攻击者攻击,或者在多个智能体中最多有s个节点受到攻击(例如,在文献[25]中,测量信号$y(t)=Cx(t)+a(t)$通过n个信道传输,则攻击信号向量$a(t)$中最多有s个非零元素). 由于稀疏攻击具有普遍性(相关假设只限制了被攻击信道的数量,对攻击信号没有任何限制),稀疏攻击下的安全状态估计问题成为近年来CPS安全问题的重点研究方向. 众多学者围绕该课题展开了深入研究,本文也将重点介绍稀疏攻击下的安全状态估计问题的研究现状.

2 集中式安全状态估计研究现状

集中式安全状态估计的目的是通过设计状态估计策略,使单个节点在攻击干扰下仍然可以从采集的测量数据中估计出可靠的系统状态[36].

以线性时不变离散系统作为CPS中的物理系统为例,其受传感器攻击时的动态可描述为

$$\begin{aligned} &x(t+1)=Ax(t)+Bu(t)+B_d d(t),\\ &y(t)=Cx(t)+a(t). \end{aligned} \tag{1}$$

其中:$x(t)\in \boldsymbol{R}^{n_x}$、$u(t)\in \boldsymbol{R}^{n_u}$、$d(t)\in \boldsymbol{R}^{n_d}$、$y(t)\in \boldsymbol{R}^{n}$分别表示系统状态、控制输入、外部扰动和传感器输出;$a(t)\in \boldsymbol{R}^{n}$表示传感器信道中的攻击信号;$A$、$B$、$B_d$、$C$是适当维度的已知矩阵,且$(A,C)$是可观测的.

实现状态估计策略的方法多种多样,常见的有基于观测器的动态估计器和基于最小二乘法的静态估计器. 基于观测器的状态估计又分为基于卡尔曼滤波器和基于隆伯格观测器两种方法. 其中,卡尔曼滤波是一种最优滤波方法,它适用于带有过程噪音和测量噪音的线性随机系统,并能得到最优的状态估计结果. 当系统没有受到攻击影响时,若$x(0)\sim\mathcal{N}(0,\Sigma)$且$d(t)\sim\mathcal{N}(0,Q)$,则可以通过经典的卡尔曼滤波器来估计系统状态,即

$$\begin{aligned} &\hat{x}(t)=\hat{x}(t|t-1)+K(t)(y(t)-C\hat{x}(t|t-1)),\\ &P(t)=P(t|t-1)-K(t)CP(t|t-1). \end{aligned} \tag{2}$$

其中

$$\begin{aligned} &\hat{x}(0|-1)=0,\ P(0|-1)=\Sigma,\\ &\hat{x}(t+1|t)=A\hat{x}(t)+Bu(t),\\ &P(t+1|t)=AP(t)A^{\mathrm{T}}+B_dQB_d^{\mathrm{T}},\\ &K(t)=P(t|t-1)C^{\mathrm{T}}(CP(t|t-1)C^{\mathrm{T}})^{-1}. \end{aligned}$$

基于观测器的状态估计器特点在于,利用上一时刻的状态估计值和当前的测量信号,可以基于动态的实时数据给出当前的状态估计值(估计误差随着时间推移逐渐收敛). 另外,连续采集τ次测量值,可得如

下的增广向量:

$$\mathcal{Y}=[y^{\mathrm{T}}(t-\tau+1)\ \ \dots\ \ y^{\mathrm{T}}(t)]^{\mathrm{T}}-F\mathcal{U}=Ox(t-\tau+1)+F_d\mathcal{D}. \tag{3}$$

其中:$\mathcal{U}=[u^{\mathrm{T}}(t-\tau+1)\ \ \dots\ \ u^{\mathrm{T}}(t-1)]^{\mathrm{T}}$, $\mathcal{D}=[d^{\mathrm{T}}(t-\tau+1)\ \ \dots\ \ d^{\mathrm{T}}(t-1)]^{\mathrm{T}}$, F_d 与 F 结构相同(B 替换为 B_d),而

$$O=\begin{bmatrix}C\\CA\\\vdots\\CA^{\tau-1}\end{bmatrix},\ F=\begin{bmatrix}0&0&\dots&0\\CB&0&\dots&0\\CAB&C&\dots&0\\\vdots&\vdots&\ddots&\vdots\\CA^{\tau-2}B&CA^{\tau-3}B&\dots&CB\end{bmatrix}.$$

进一步,可以采用最小二乘法估计系统状态

$$\hat{x}=\arg\min_{\hat{x}\in\boldsymbol{R}^{n_x}}||\mathcal{Y}-O\hat{x}||_2, \tag{4}$$

其中 $\hat{x}$ 是状态 $x(t-\tau+1)$ 的估计值. 不同于基于观测器的状态估计器,基于最小二乘法的状态估计器仅使用一段时间内测量信号而不需要历史估计值,此时估计误差仅依赖于使用的数据且仅受扰动影响(与时间无关).

在安全状态估计问题的研究中,观测器起着重要的作用,因此相关研究受到了关注. 文献[22]针对静态或时变未知传感器子集,提出了一种降低错误数据注入攻击下平均期望误差的安全估计算法和攻击检测算法. 文献[23]证明了最优卡尔曼滤波器可以分解为局部状态估计的加权和,并基于这些局部估计提出了一种基于凸优化的方法来生成更安全的状态估计. 对于确定性系统,隆伯格观测器具有更好的适用性. 文献[37]提出了一种基于高效可满足性模理论求解的多模隆伯格观测器,用于设计能够抵御传感器攻击的大规模信息物理系统. 文献[38]提出了一种新的类隆伯格观测器,并研究了在同时存在执行器和传感器攻击的情况下的安全状态估计问题.

此外,在安全状态估计问题的研究中,目前主要考虑具有一般性稀疏攻击(仅受攻击信道个数限制). 如第 1.2 节所述,稀疏攻击下的安全状态估计本质上是一个组合优化问题,其中的难点在于如何克服算法的高计算复杂度问题. 由于最小二乘法简洁且能更好地展现安全状态可观性,并突出了安全状态估计中的本质难点,基于最小二乘法的安全状态估计策略得到了广泛研究. 基于最小二乘法的思想,文献[36]提出了基于 ℓ_0 范数和 ℓ_1 范数的安全状态估计器,通过直接求解优化问题,寻找一个恰当的状态估计值,使得尽可能多的信道所对应的残差尽可能小,并在一定条件下证明了其所提出的优化问题的解就是一个可靠的状态估计值. 文献[25]将安全状态估计问题转换为类似于文献[36]中的优化问题,但不同之处在于,它在经典的梯度下降法基础上提出了事件触发投影梯度下降算法,以实现安全状态估计,并给出了算法能够得到可靠状态估计值的条件.

在稀疏攻击下的安全状态估计问题中,由于需要从多种可能性中找到正确的被攻击信道集合,问题本质上是一个组合优化问题,属于NP难问题[26]. 因此,无论是基于观测器的技术还是基于最小二乘法的技术,在实际应用中都面临着高计算复杂度的挑战. 如图4所示,稀疏攻击下的安全状态估计的核心在于移除被攻击的信息,但由于被攻击信道是未知的,需要从多种可能性中找到正确的被攻击信道集合. 这涉及到对组合数 $\mathbf{C}_s^n$ 个可能的组合进行搜索. 对于较大的通道数 n 和被攻击信道数 s,这将导致计算复杂度非常高. 因此,如何降低安全状态估计算法的计算复杂度成为稀疏攻击下安全状态估计问题的难点.

图 4　稀疏攻击下的安全状态估计难点

根据处理难点的技术不同,可将降低计算复杂度的方法分为遍历搜索方法和非遍历搜索方法. 遍历搜索方法通常是通过枚举所有可能的组合,计算每个组合对应的状态估计,并选择具有最小误差的组合作为最终结果. 然而,这种方法在组合数较大时,计算复杂度会呈指数级增长,限制了其实际应用. 非遍历搜索方法则采用一些启发式策略或优化算法,通过适当的剪枝、约束或优化策略,寻找局部最优解或近似最优解,从而降低计算复杂度. 综上所述,降低稀疏攻击下安全状态估计算法的计算复杂度是一个具有挑战性的问题,需要借助有效的搜索策略和优化算法来提高计算效率. 在表1中对比了各类搜索方法的特点和性能.

表 1　集中式安全状态估计方法对比

文献	遍历	基本思路	复杂度	保守性
[36]	是	ℓ_0 优化	高	低
[39-40]	是	可满足性模理论	中	低
[28, 41-42]	是	切换机制	中	低
[30, 43-44]	是	集合优化	中	低
[23, 25, 36, 45]	否	非凸转凸	低	高
[29, 46-48]	否	自适应机制	低	高
[26, 49-50]	否	能观性分解	低	中

2.1 遍历搜索方法

遍历搜索算法的基本思想是,枚举所有的可能性直到找到正确的攻击模式(即找到被攻击信道集合),例如文献[36]中的ℓ_0解码器就是一个典型的遍历搜索方法. 参考式(4),基于遍历搜索方法的安全状态估计可以描述为如下的优化问题:

$$\{\hat{x},\hat{\boldsymbol{A}}\}=\arg\min_{\hat{x}\in\boldsymbol{R}^{n_x},\hat{\boldsymbol{A}}\in\boldsymbol{K}_s}||\mathcal{Y}_{\hat{\boldsymbol{A}}}-O_{\hat{\boldsymbol{A}}}\hat{x}||_2. \quad (5)$$

其中:$\hat{x}$和$\hat{\boldsymbol{A}}$分别是系统状态和被攻击通道集合的估计值,$\boldsymbol{K}_s$表示s稀疏攻击下所有可能的被攻击通道集合(集合$\boldsymbol{K}_s$的元素个数为$\mathbf{C}_s^n$),$\mathcal{Y}_{\hat{\boldsymbol{A}}}$和$O_{\hat{\boldsymbol{A}}}$分别表示$\mathcal{Y}$和$O$按照$n$行为一组依次划分后每组去掉集合$\hat{\boldsymbol{A}}$所对应的行得到的向量和矩阵. 可以看出,这类方法设计简单,求解过程就是依次对$\boldsymbol{K}_s$中的候选项$\hat{\boldsymbol{A}}$求解一次最小二乘问题. 但随着n和s的增大,求解问题(5)的计算复杂度将大幅增加,特别是对于大规模CPS,其复杂度将高到难以承受.

目前,在遍历搜索方法的基础上,大量研究集中在如何减少搜索次数的问题上. 文献[39]提出了一类包含卡尔曼滤波器的安全状态估计策略,其中卡尔曼滤波器被用于搜索可靠的传感器集合,并引入基于可满足性模理论的技术来减少搜索次数. 与文献[39]类似,文献[40]也提出了一种基于可满足性模理论的安全状态估计方法,以尽可能减少搜索次数.

此外,在安全状态估计中,匹配正确的攻击模式是关键工作之一,因此,切换机制被广泛应用于提升安全状态估计方法的可靠性. 文献[41]提出了一种基于切换隆伯格观测器的安全状态估计策略,虽然没有解决高计算复杂度的问题,但该方法适用于变通道的稀疏错误数据攻击. 文献[28]提出了一种基于自适应切换机制的安全状态估计方法,通过在线观测性能指标来驱动切换机制,最终确定正确的攻击模式. 文献[42]针对遍历搜索方法中的高计算复杂度问题,提出了切换投影梯度下降算法,通过引入切换机制减少错误迭代次数,采用预处理技术提升收敛速率,并引入新的投影算子来减少搜索次数,从而提高效率.

为了降低计算复杂度并提升估计性能,还有多种方法可供选择. 其中一种方法是通过对候选项集合进行优化来减少搜索次数. 文献[43]提出了集合覆盖方法以减少待选项的数量,并证明总候选项数可以减少至少一半(少于$(1/2)\mathbf{C}_s^n$). 类似地,文献[30]从集合理论的角度出发,利用受限集合划分方法显著减少了搜索次数. 另外,文献[44]采用等价类方法对传感器进行划分,以降低算法的计算复杂度. 文献[51]提出了一类基于监测机制的非线性安全状态观测器设计方法. 文献[52]提出了一种新的最优图搜索算法,即使在建模为线性时不变系统的大规模CPS中,也能正确识别恶意攻击并安全估计状态,从而实现了最优性和较短的运行时间. 此外,文献[53]提出了一种基于正交投影的安全状态估计方法,该方法在设计安全状态估计策略的基础上,通过扰动解耦方法实现了在受到稀疏攻击和干扰的同时较好地重构系统状态.

对于遍历搜索面临的高计算复杂度问题,上述方法在一定程度上减少了搜索次数. 然而,需要指出的是,这些方法仍属于遍历搜索方法,这意味着随着信道个数n和攻击个数s的增加,计算复杂度仍将迅速增加.

2.2 非遍历搜索方法

2.2.1 凸优化方法

凸优化方法是将安全状态估计问题表述为凸优化问题,通过引入一些限制性约束为代价消除遍历搜索带来的计算复杂度爆炸性增长的问题. 例如,文献[36]将原始的安全状态估计问题转化为一个ℓ_0优化问题,进一步,通过将非凸的ℓ_0优化问题转化为凸的ℓ_1优化问题,最终得到了ℓ_1解码器. 参考文献[36]中的ℓ_1解码器和问题(5),可以得到如下的凸优化问题:

$$\{\hat{x},\hat{\boldsymbol{A}}\}=\arg\min_{\hat{x}\in\boldsymbol{R}^{n_x}}\sum_{i\in\boldsymbol{I}}||\mathcal{Y}_{\boldsymbol{I}\backslash\{i\}}-O_{\boldsymbol{I}\backslash\{i\}}\hat{x}||_2. \quad (6)$$

其中:$\boldsymbol{I}=\{1,\dots,n\}$,$\boldsymbol{I}\backslash\{i\}=\{1,\dots,i-1,i+1,\dots,n\}$,$\mathcal{Y}_{\boldsymbol{I}\backslash\{i\}}$和$O_{\boldsymbol{I}\backslash\{i\}}$的定义见式(5). 值得说明的是,式(6)的解与(5)的解并不一定相同,而是需要满足一定的条件才能保证两者的解相同. 这意味着只有在满足特定条件的情况下,利用ℓ_1解码器得到的状态估计值才能保证一定是可靠的,但凸优化的引入往往会引入较强的额外限制条件,例如文献[36]中的命题6.

当然,目前还存在一些其他的凸优化方法. 例如,文献[25]提出了基于投影梯度下降的安全状态估计策略,通过引入投影算子,将原本非凸优化的安全状态估计问题转化为凸优化问题. 类似地,文献[23]将最优卡尔曼估计分解为多个局部状态估计的叠加,并提出了一种基于凸优化的安全状态估计策略,该策略证明了所提出的安全状态估计器具有一定的抗稀疏攻击能力. 另外,文献[45]将最优卡尔曼估计分解为局部状态估计的线性组合,并提出了一种基于凸优化的方法,用于解决稀疏攻击下自动驾驶汽车的安全姿态估计问题. 然而,需要注意的是,安全状态估计问题本质上是一个非凸优化问题. 因此,引入任何凸优化方法都会引入额外的限制条件,从而导致相应方法的适用范围变小,并且保守性较强.

2.2.2 自适应方法

与遍历搜索相比,自适应方法通过引入自适应参数来限制遭受攻击的影响,同样不需要进行遍历搜索. 文献[29]提出了一类基于自适应切换策略的安全状态估计策略,通过对观测矩阵进行在线学习,降低了算法的存储负担. 文献[46]针对传感器和执行器攻击,提出了自适应逐行超扭转观测器的在线估计策略,重构信息物理系统状态并进行弹性控制. 文献[47-48]针对未知的稀疏传感器攻击干扰,提出了饱和自适应梯度下降算法,通过引入自适应饱和项来限制攻击影响,进而回避遍历搜索,实现了快速获取可靠的状态估计值. 以文献[47]饱和自适应梯度下降算法为例,考虑如下的梯度下降迭代:

$$\hat{x}_{m+1}=\hat{x}_m+\eta\sum_{i\in\boldsymbol{I}}k_{i,m}O_i^{\mathrm{T}}(\mathcal{Y}_i-O_i\hat{x}).$$

其中: O_i 和 $\mathcal{Y}_i$ 与式(3)中的 O 和 $\mathcal{Y}$ 有着相似的定义(C 替换为其第 i 行 C_i), η 是迭代步长,通过设计参数

$$\gamma_m=\rho^m\Gamma+\bar{\omega}_M$$

刻画正常信道所对应的残差2范数上界,引入的饱和自适应项

$$k_{i,m}=\min\{1,\gamma_m||\mathcal{Y}_i-O_i\hat{x}_m-F_i\mathcal{U}||^{-1}\}.$$

通过自适应调整自身数值大小,能够有效地抑制攻击信号对估计性能的影响,进而得到可靠的状态估计值.

2.2.3 能观性分解方法

近年来,随着对安全状态估计的深入研究,学者们发现在稀疏攻击下,并非始终需要通过遍历搜索或引入凸优化来获取系统状态估计值. 基于这一发现,文献[49]提出了一类能观性分解技术,以克服高计算复杂度的问题. 该技术假设存在一个适当的状态空间基,使得每个传感器的不可观测空间是该基的一个子空间,并通过选取所有估计候选者的中位数来得到每个分解元素. 在文献[49]的基础上,文献[50]进一步证明了当系统为 $2s$ 稀疏特征可观测时,系统矩阵的广义特征向量可以构成合适的基,且所有分解元素可以在多项式时间内进行估计. 然而, $2s$ 稀疏特征可观测性通常要求系统矩阵的每个特征值的几何重数为1,并且文献[50]中没有充分考虑测量矩阵的性质. 为了解决文献[49-50]中存在的问题,文献[26]提出了一类新的状态分解技术. 该技术通过对系统状态进行分解,使得每个分解元素可以通过简单的多数投票决策策略进行重构. 这种能观性分解方法的基本思路大多可以归结为对系统状态进行分解,有

$$x=\sum_{p\in\boldsymbol{I}_{n_x}}x_p=\sum_{p\in\boldsymbol{I}_{n_x}}v_p\theta_p=V\theta. \tag{7}$$

其中: $\boldsymbol{I}_{n_x}=\{1,\ldots,n_x\}$, $V=[v_1\ \ldots\ v_{n_x}]$ 表示状态的分解矩阵. 在式(7)中,系统状态 x 被按照元素分解为 n_x 部分,并进一步将 x_p 定义为向量 v_p 与系数 θ_p 的乘积. 由于分解矩阵 V 是已知且可逆的,文献[26]中将状态 x 的估计问题成功地转化为各个元素 θ_p 的估计问题. 特别地,在一定条件下,基于采集的测量信号重构单个元素 θ_p 将变得非常简单,并且通过依次基于之前的 $\theta_{p'}$ 重构当前的 θ_p ($p'\leqslant p$)可以有效地实现对向量 θ (等价于 x)的可靠估计.

总体而言,文献[49-50]中关于能观性分解的研究成功地将状态整体的估计问题转化为状态元素的估计问题,以此为基础,单个状态元素往往能够通过最多值投票得到. 受文献[49-50]的启发,文献[26]所提出的方法更具一般性. 但遗憾的是,由于对系统矩阵引入了额外的限制条件,仍然未能完全解决安全状态估计中的高计算复杂度问题.

3 分布式安全状态估计研究现状

近年来,分布式系统(多智能体系统)受到了广泛关注,其特点在于系统中多个节点之间通过信息交互实现协同合作[54-56],从而实现多智能体一致性和分布式状态估计等目标. 针对分布式状态估计问题,人们提出了多种技术. 文献[54]针对含有状态等式约束的多智能体分布式估计问题,改进了投影算子和协方差交叉融合方法,并提出了一种保证一致性的分布式卡尔曼滤波器. 文献[57]研究了基于卡尔曼滤波器的状态估计在有损网络中的随机稳定性,主要针对空间分布的信息物理系统. 文献[56]针对大规模系统的分布式参数估计问题,对循环图中基于加权最小二乘法的分布式估计算法的准确性进行了有效分析.

在分布式系统中,多个子系统通过频繁的信息交互实现协同合作,因此对网络通信有很大依赖. 考虑到分布式系统通信的特殊性,尤其是包含多个节点的分布式系统更容易受到网络攻击的影响[58]. 这种影响主要体现在两个方面: 1) 通信信道可能受到干扰而传输错误的数据; 2) 部分节点可能被攻击者挟持,向其邻节点发送错误数据. 如图5所示,在分布式系统中,传感器的攻击将导致节点接收到错误的测量信号[31],信道的攻击可能导致通信中断或数据错误[24],节点的攻击直接干扰被攻击节点的正常运行[27]. 同时,由于分布式系统中节点之间相互交互信息,如果没有适当的防护措施,则所有节点最终都将受到攻击信号的影响. 因此,确保分布式系统的安全性对于其

正常运行至关重要.

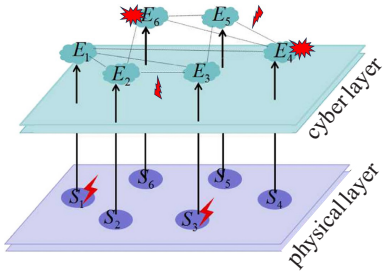

图 5 网络攻击下的分布式系统[32]

在系统受到攻击的情况下进行系统状态估计[59]、攻击检测[60]以及保护信息的完整性、可用性和隐秘性,对于确保分布式系统的安全稳定运行具有重要意义. 由于获取分布式系统各个节点的状态信息是确保系统整体安全的基础,分布式安全状态估计引起了广泛关注. 文献[61]同时考虑错误信息注入攻击和拒绝服务攻击,通过引入两个自适应分布式状态估计器实现了对篡改测量的评估,减轻了错误信息注入攻击的影响. 文献[62]研究了一类不确定非线性关联的分布式状态重构问题. 文献[33]提出了一种基于随机集理论的分布式攻击状态估计方法,通过混合伯努利随机集合滤波器,在检测攻击的基础上重构系统状态. 文献[63]开发了一种基于乘子交替方法的新算法,即使在没有局部可观测性的情况下,也可以保证其收敛,并研究了一类基于电力系统的分布式状态估计器. 受检测方法的启发,文献[64]提出了一种基于K-L散度的检测器来抵抗恶意攻击,以保证估计的准确性,并推导出最佳的估计器增益和检测器临界阈值的选择方法. 文献[65]考虑智能体受同源信号攻击的安全状态估计问题,设计了一种结合平均一致算法和递归算法的双环观测器,可以同时对系统状态和攻击信号进行估计,并给出了安全状态估计的充分必要条件. 文献[66]研究了一种攻击检测算法,通过检测算法识别受攻击的传感器信息,并设计了一个离线获取参数的分布式观测器,利用线性不等式技术给出了误差系统稳定且具有H_∞特性的充分条件.

正如第2节指出的,稀疏攻击下的安全状态估计问题本质上是一个NP难问题[26],其中最大的问题是算法计算复杂度随着信道个数n和攻击个数s的增加而呈现爆炸性增长. 与集中式安全状态估计类似,分布式安全状态估计也面临估计算法计算复杂度过高的问题. 因此,与处理高计算复杂度的方法不同,分布式安全状态估计方法可以分为遍历搜索方法和非遍历搜索方法.

3.1 遍历搜索方法

在攻击传感器、信道或节点未知的情况下,遍历搜索所有可能性以找到正确的被攻击传感器、信道或节点集合,并通过移除被篡改的信息来消除攻击的影响,是实现安全状态估计的一种直接方法.

文献[31]提出了一种基于非凸优化方法的安全状态估计算法,并利用循环投票定位机制实现了分布式安全状态估计. 该方法利用最小切换机制对攻击模态进行选择,显著减少了切换次数. 类似地,文献[67]设计了一种基于事件触发的最小切换分布式安全状态估计算法,通过最小切换机制对模态进行选择,极大地减少了切换次数. 文献[27]将分布式安全状态估计问题转化为分布式优化问题,提出了候选项移除机制,并构建了一种切换梯度下降算法,以在遭受恶意节点干扰的情况下重构可靠的状态估计. 然而,需要指出的是,遍历搜索方法通常具有较高的计算复杂度,对于多节点系统而言,实现起来可能困难. 针对双通道拒绝服务攻击和错误数据注入攻击的混合攻击,文献[24]提出了一种切换安全状态估计策略.

需要注意的是,遍历搜索方法通常具有较高的计算复杂度,在多节点系统中难以实现. 因此,在实际应用中,需要综合考虑计算复杂度和系统要求,选择合适的安全状态估计方法.

3.2 非遍历搜索方法

文献[18]针对多传感器受到攻击的情况下N个相连的单输入单输出线性子系统,设计了一个前置选择器,借鉴异常检测策略[18,66],通过前置选择器快速找到未受攻击的节点,从而实现了安全状态估计;该文还利用安全状态估计设计了虚拟分数阶分布式安全控制器. 针对无线传感器网络的分布式安全状态估计问题,文献[68-69]提出了一种“一致性+革新”算法,以保证在网络联通拓扑条件下算法估计的正确性,但该算法要求每个传感器节点都具有局部可观测性. 为了克服这一限制,文献[61,70]分别设计了基于局部滤波的安全状态估计策略,即每个智能体移除其邻居传感器接收到的极端值数据,以实现对系统状态的可靠重构,并给出了相应算法在网络拓扑联通度方面的充分必要条件. 文献[21]研究一类带有恶意传感器节点的分布式安全状态估计问题,提出了排序过滤策略,通过移除极端数值有效抑制了攻击的影响,从而在恶意节点干扰下确保正常节点正确地估计系统状态. 相较于基于遍历搜索的分布式安全状态估计策略[27],文献[21]避免了遍历搜索方法,从而减少

了计算复杂度. 针对部分传感器受到攻击的情况, 文献[48,71]分别提出了残差饱和更新策略, 通过压缩异常测量信号的幅值来抑制攻击信号对系统状态信息的影响, 以获取系统真实的状态信息. 在二跳通信的多智能体系统中, 文献[72]提出了一种基于置信度的分布式安全状态估计策略, 该策略通过引入置信度机制来刻画信息的可靠性, 并通过基于置信度的投票策略实现攻击检测, 快速确定受攻击的信道, 从而得到正确的状态估计值.

4　安全状态能观性研究现状

对于网络攻击下的CPS, 攻击者最直接的目标是破坏系统状态的可观测性. 攻击者通过攻击使得防守者无法准确获取系统的真实状态信息, 甚至将错误的信息误认为是真实的系统状态, 从而通过对控制层的反馈造成系统性能的损害. 攻击者的能力决定了他们能否成功地破坏系统状态的可观测性, 而安全状态的可观测性研究的目的在于研究在面对不同攻击强度时系统状态仍能保持可观测性, 即系统能够容忍的最大攻击强度. 在安全状态可观测性的研究中, 目前的研究主要集中在稀疏传感器攻击下的系统状态可观测性. 例如, 在第1.2节中, 文献[25]中的定理3.2给出了在s稀疏传感器攻击下系统状态可观测的充要条件: 系统需要满足$2s$稀疏可观测性(即对于任意去掉$2s$个测量信道, 基于剩余的测量信号, 系统状态仍然可观测). 举例说明, 如图6所示, 假设系统有4个测量信道用于接收高度信息, 若有2个信道被攻击(不满足$2s$稀疏可观测性), 则系统状态将不可观测(无法确定其真实值).

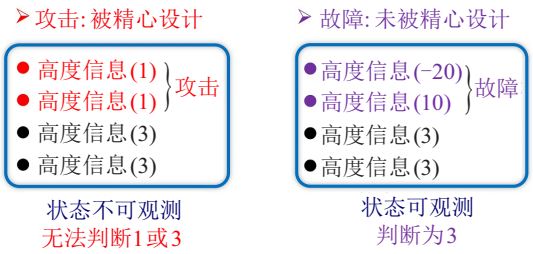

图 6　攻击和故障下的系统状态能观性

作为s稀疏攻击下安全状态可观测性的充要条件, 现有的大部分安全状态估计研究中, $2s$稀疏可观测性已成为系统的基本要求. 例如, 为了避免使用遍历搜索方法, 文献[26]提出了一种状态分解技术, 但仍需要保证$2s$稀疏可观测性以确保状态分解元素的可观测性. 文献[41]引入正交补矩阵的概念, 给出了在稀疏执行器攻击和传感器攻击下可观测性的充分必要条件, 并表明这些可观测性条件与$2s$稀疏可观测性要求是一致的. 文献[73]在基于$2s$稀疏可观测性的系统基础上, 利用分离原理和李雅普诺夫稳定性理论分析了传感器攻击下估计误差的收敛性. 然而, 值得注意的是, $2s$稀疏可观测性是一个相当严格的限制条件, 要求系统对每条信息都具备足够的冗余度. 例如, 对于一个拥有5个测量信道的系统, 若攻击强度$s = 2$(攻击者最多篡改2个信道的信号), 则系统需要能够基于任意一个信道的信息重构出系统状态. 考虑到现实世界中硬件和通信资源的有限性, 过多的冗余将导致部署成本的显著增加甚至无法实现. 因此, 近年来学者们提出了多种方法和策略来减弱这个限制.

通过对稀疏攻击下系统状态可观测性的进一步研究, 文献[74]将s稀疏攻击下的$2s$稀疏可观测性要求放宽为$2s$稀疏可检测性, 即只要系统具备$2s$稀疏可检测性, 就能够抵抗s稀疏攻击, 这达到了安全动态估计的基本极限. 此外, 在未受攻击的情况下, 该文所提出的估计方法与卡尔曼估计方法有一定的重合概率. 文献[28]引入了稀疏可检测性和稀疏强可检测性的概念, 并得出结论: 观测误差系统的s稀疏强可检测性等价于原系统的$2s$稀疏可检测性. 基于这种等价性, 文献[28]又证明了在网络攻击下观测误差系统也能达到渐近稳定.

针对未知干扰下的干扰解耦安全状态估计问题, 文献[53]引入了冗余强可观测性和冗余强可检测性(rs可观测性和rs可检测性)的概念, 解释了状态、传感器测量值、干扰与稀疏攻击信号之间的关系, 并在rs可检测性的条件下提出了扰动解耦方法. 文献[75]针对不同情况进行了系统状态的重构研究, 结果表明在安全状态估计问题中, 通过先验知识的应用能够有效增强系统对抗攻击干扰的能力, 并证明了适当增加先验知识能够使系统容忍更多的传感器攻击. 此外, 文献[76]的研究表明, 在分布式安全状态估计问题中, 通过引入稀疏的系统状态先验知识, 在给定先验信息的情况下, 即使有更多的传感器受到攻击, 系统状态仍然可以唯一地重建, 从而减少了对测量和通信网络的冗余要求.

此外, 文献[77]创新地将异常干扰分为攻击和故障两种, 并对稀疏故障和稀疏攻击下的安全状态可观测性进行了系统分析. 研究结果表明, 相对于将所有干扰都假设为攻击, 通过区分故障和攻击, 系统能够容忍更多的错误信号.

5　总结与展望

本文对近年来在网络攻击条件下的信息物理系统(CPS)安全状态估计问题的研究进行了综述. 文中着重介绍了集中式安全状态估计和分布式安全状态

估计两种方法,并根据处理高计算复杂度的问题的策略,将安全状态估计方法划分为遍历搜索和非遍历搜索两类. 最后,探讨了安全状态观察性分析的当前研究状况. 值得强调的是,随着信息技术和智能技术的快速发展,CPS安全面临着越来越多的挑战. 因此,除了需要进一步研究和发展安全状态估计技术外,还有许多其他CPS安全问题值得深入探讨.

1) 面向非线性CPS的安全状态估计技术. 虽然当前对于线性信息物理系统的安全状态估计问题已取得了一定的进展,但针对更为复杂(如非线性)的信息物理系统,其安全状态估计的研究还显得不够充分. 确实,已有一些针对非线性系统的安全状态估计研究(例如文献[32, 35]),但这些研究在系统模型和非线性项方面还存在较大的限制和要求.

2) 高效的信息和物理空间攻击定位与辨识技术. 安全状态估计的本质在于检测并识别被攻击的通道. 由于CPS是由物理设备和网络设备在分布式方式下部署,其通信网络中的任何失真信息的传播都可能影响整个系统的动态响应. 因此,在安全状态估计中,如何快速有效地定位并识别被攻击的信道成为一个关键问题.

3) 基于人工智能技术的安全状态估计技术. 安全状态估计的难点在于如何降低在识别攻击通道时的计算复杂度. 人工智能技术可以在大量数据的训练下学习到这种内在的关系,从而快速定位被攻击的通道并补偿攻击影响以获得真实的系统状态. 因此,如何为CPS设计有效的基于人工智能的安全状态估计方法,是一个需要深入研究的问题.

4) 通信、监测(包含检测和状态估计)和控制的联合设计技术. 虽然CPS安全问题的研究很多,但大多数研究都是针对特定的安全问题进行的. 如何进行多个安全问题的联合设计,目前还没有得到很好地解决. 安全状态估计的目的是在被破坏的数据中得到系统的真实运行状态. 因此,在安全状态估计的基础上,如何设计攻击检测方法、安全控制方案以及通信策略,以提高CPS的整体性能和网络安全性能,是一个值得进一步研究的问题.

5) 基于模型(自动化技术)和数据(计算机技术)的安全技术. 信息物理系统已经引起了各领域学者的广泛关注和研究. 计算机相关技术通常更注重数据的安全和利用,而自动化技术和理论则更注重基于系统模型分析数据规律. 基于模型和基于数据的技术都取得了一定的进展,但它们之间的融合程度还相对较低. 因此,如何充分利用自动化技术和计算机技术,提出更有效的基于模型+数据的CPS安全技术,是一个非常值得研究的方向. 例如,针对安全状态估计问题,可以运用神经网络学习模拟数据,构建一个能够快速找到被攻击信道的模型,然后进一步利用训练得到的模型输出来辅助基于模型的安全状态估计策略,快速估计出可靠的状态估计值.

参考文献(References)

[1] China Electron-ics Standardization Institute. White paper: Cyber-physical sysem [EB/OL]. [2023-05-23]. http://www.cesi.cn/201703/2251.html.

[2] 李洪阳, 魏慕恒, 黄洁, 等. 信息物理系统技术综述[J]. 自动化学报, 2019, 45(1): 37-50.
(Li H Y, Wei M H, Huang J, et al. Survey on cyber-physical systems[J]. Acta Automatica Sinica, 2019, 45(1): 37-50.)

[3] Guan X P, Yang B, Chen C L, et al. A comprehensive overview of cyber-physical systems: From perspective of feedback system[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(1): 1-14.

[4] Chen T M. Stuxnet, the real start of cyber warfare?[J]. IEEE Network, 2010, 24(6): 2-3.

[5] Cherdantseva Y, Burnap P, Blyth A, et al. A review of cyber security risk assessment methods for SCADA systems[J]. Computers & Security, 2016, 56: 1-27.

[6] 汤奕, 陈倩, 李梦雅, 等. 电力信息物理融合系统环境中的网络攻击研究综述[J]. 电力系统自动化, 2016, 40(17): 59-69.
(Tang Y, Chen Q, Li M Y, et al. Overview on cyber-attacks against cyber physical power system[J]. Automation of Electric Power Systems, 2016, 40(17): 59-69.)

[7] Zhang D, Wang Q G, Feng G, et al. A survey on attack detection, estimation and control of industrial cyber-physical systems[J]. ISA Transactions, 2021, 116: 1-16.

[8] Yan J J, Yang G H. Secure state estimation of nonlinear cyber-physical systems against DoS attacks: A multiobserver approach[J]. IEEE Transactions on Cybernetics, 2023, 53(3): 1447-1459.

[9] Zhang H, Cheng P, Shi L, et al. Optimal DoS attack scheduling in wireless networked control system[J]. IEEE Transactions on Control Systems Technology, 2016, 24(3): 843-852.

[10] Li Y Z, 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.

[11] Hao J P, Piechocki R J, Kaleshi D, et al. Sparse malicious false data injection attacks and defense mechanisms in smart grids[J]. IEEE Transactions on Industrial Informatics, 2015, 11(5): 1-12.

[12] Liu J L, Xia J L, Tian E G, et al. Hybrid-driven-based H_∞ filter design for neural networks subject to deception attacks[J]. Applied Mathematics and Computation, 2018,

320: 158-174.

[13] Deng R L, Xiao G X, Lu R X, et al. False data injection on state estimation in power systems — Attacks, impacts, and defense: A survey[J]. IEEE Transactions on Industrial Informatics, 2017, 13(2): 411-423.

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

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

[16] Ni Y Q, Guo Z Y, Mo Y L, et al. On the performance analysis of reset attack in cyber-physical systems[J]. IEEE Transactions on Automatic Control, 2020, 65(1): 419-425.

[17] Lu A Y, Yang G H. False data injection attacks against state estimation without knowledge of estimators[J]. IEEE Transactions on Automatic Control, 2022, 67(9): 4529-4540.

[18] Ao W, Song Y D, Wen C Y. Distributed secure state estimation and control for CPSs under sensor attacks[J]. IEEE Transactions on Cybernetics, 2020, 50(1): 259-269.

[19] Ding D R, Han Q L, Ge X H, et al. Secure state estimation and control of cyber-physical systems: A survey[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(1): 176-190.

[20] Yong S Z, Zhu M H, Frazzoli E. Resilient state estimation against switching attacks on stochastic cyber-physical systems[C]. The 54th IEEE Conference on Decision and Control. Osaka, 2016: 5162-5169.

[21] Lu A Y, Yang G H. Distributed secure state estimation for linear systems against malicious agents through sorting and filtering[J]. Automatica, 2023, 151: 110927.

[22] Chattopadhyay A, Mitra U. Security against false data-injection attack in cyber-physical systems[J]. IEEE Transactions on Control of Network Systems, 2020, 7(2): 1015-1027.

[23] Liu X H, Mo Y L, Garone E. Local decomposition of Kalman filters and its application for secure state estimation[J]. IEEE Transactions on Automatic Control, 2021, 66(10): 5037-5044.

[24] Song H Y, Yao H Y, Shi P, et al. Distributed secure state estimation of multi-sensor systems subject to two-channel hybrid attacks[J]. IEEE Transactions on Signal and Information Processing Over Networks, 2022, 8: 1049-1058.

[25] Shoukry Y, Tabuada P. Event-triggered state observers for sparse sensor noise/attacks[J]. IEEE Transactions on Automatic Control, 2016, 61(8): 2079-2091.

[26] Lu A Y, Yang G H. A polynomial-time algorithm for the secure state estimation problem under sparse sensor attacks via state decomposition technique[J]. IEEE Transactions on Automatic Control, 2023(99): 1-14.

[27] Lu A Y, Yang G H. Distributed secure state estimation in the presence of malicious agents[J]. IEEE Transactions on Automatic Control, 2021, 66(6): 2875-2882.

[28] An L W, Yang G H. Secure state estimation against sparse sensor attacks with adaptive switching mechanism[J]. IEEE Transactions on Automatic Control, 2018, 63(8): 2596-2603.

[29] An L W, Yang G H. Adaptive secure state estimation for cyber-physical systems with low memory cost[J]. IEEE Transactions on Control of Network Systems, 2020, 7(4): 1621-1632.

[30] An L W, Yang G H. State estimation under sparse sensor attacks: A constrained set partitioning approach[J]. IEEE Transactions on Automatic Control, 2019, 64(9): 3861-3868.

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

[32] Yan J J, Yang G H. Secure state estimation with switched compensation mechanism against DoS attacks[J]. IEEE Transactions on Cybernetics, 2022, 52(9): 9609-9620.

[33] Forti N, Battistelli G, Chisci L, et al. Distributed joint attack detection and secure state estimation[J]. IEEE Transactions on Signal and Information Processing Over Networks, 2018, 4(1): 96-110.

[34] Shi Y K, Wang Y Q. Online secure state estimation of multiagent systems using average consensus[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(5): 3174-3186.

[35] Guo X Y, Wang C L, Dong Z, et al. Secure state estimation for nonlinear systems under sparse attacks with application to robotic manipulators[J]. IEEE Transactions on Industrial Electronics, 2023, 70(8): 8408-8415.

[36] Fawzi H, Tabuada P, Diggavi S. Secure estimation and control for cyber-physical systems under adversarial attacks[J]. IEEE Transactions on Automatic Control, 2014, 59(6): 1454-1467.

[37] Shoukry Y, Chong M, Wakaiki M, et al. SMT-based observer design for cyber-physical systems under sensor attacks[J]. ACM Transactions on Cyber-Physical Systems, 2018, 2(1): 1-27.

[38] Lu A Y, Yang G H. Secure Luenberger-like observers for cyber−physical systems under sparse actuator and sensor attacks[J]. Automatica, 2018, 98: 124-129.

[39] Mishra S, Shoukry Y, Karamchandani N, et al. Secure state estimation against sensor attacks in the presence of noise[J]. IEEE Transactions on Control of Network Systems, 2017, 4(1): 49-59.

[40] Shoukry Y, Nuzzo P, Puggelli A, et al. Secure state estimation for cyber-physical systems under sensor

attacks: A satisfiability modulo theory approach[J]. IEEE Transactions on Automatic Control, 2017, 62(10): 4917-4932.

[41] Lu A Y, Yang G H. Secure state estimation for cyber-physical systems under sparse sensor attacks via a switched Luenberger observer[J]. Information Sciences, 2017, 417: 454-464.

[42] Lu A Y, Yang G H. Switched projected gradient descent algorithms for secure state estimation under sparse sensor attacks[J]. Automatica, 2019, 103: 503-514.

[43] Lu A Y, Yang G H. Secure switched observers for cyber-physical systems under sparse sensor attacks: A set cover approach[J]. IEEE Transactions on Automatic Control, 2019, 64(9): 3949-3955.

[44] An L W, Yang G H. Fast state estimation under sensor attacks: A sensor categorization approach[J]. Automatica, 2022, 142: 110395.

[45] Jiang R, Liu X H, Wang H, et al. Secure estimation for attitude and heading reference systems under sparse attacks[J]. IEEE Sensors Journal, 2019, 19(2): 641-649.

[46] Nateghi S, Shtessel Y, Edwards C, et al. Resilient control of cyber-physical systems using adaptive super-twisting observer[J]. Asian Journal of Control, 2023, 25(3): 1775-1790.

[47] Lu A Y, Yang G H. Secure state estimation under sparse sensor attacks via saturating adaptive technique[J]. IEEE Transactions on Control of Network Systems, 2023(99): 1-9.

[48] Chen Y, Kar S, Moura J M F. Resilient distributed parameter estimation with heterogeneous data[J]. IEEE Transactions on Signal Processing, 2019, 67(19): 4918-4933.

[49] Lee J, Kim J, Shim H. Fully distributed resilient state estimation based on distributed median solver[J]. IEEE Transactions on Automatic Control, 2020, 65(9): 3935-3942.

[50] Mao Y W, Mitra A, Sundaram S, et al. On the computational complexity of the secure state-reconstruction problem[J]. Automatica, 2022, 136: 110083.

[51] An L W, Yang G H. Supervisory nonlinear state observers for adversarial sparse attacks[J]. IEEE Transactions on Cybernetics, 2022, 52(3): 1575-1587.

[52] Luo X S, Pajic M, Zavlanos M M. An optimal graph-search method for secure state estimation[J]. Automatica, 2021, 123: 109323.

[53] Li J H, Yang G H. Disturbance decoupled secure state estimation: An orthogonal projection-based method[J]. Automatica, 2023, 147: 110740.

[54] He X K, Hu C, Hong Y G, et al. Distributed Kalman filters with state equality constraints: Time-based and event-triggered communications[J]. IEEE Transactions on Automatic Control, 2020, 65(1): 28-43.

[55] Liu Q S, Yang S F, Hong Y G. Constrained consensus algorithms with fixed step size for distributed convex optimization over multiagent networks[J]. IEEE Transactions on Automatic Control, 2017, 62(8): 4259-4265.

[56] Sui T J, Marelli D E, Fu M Y, et al. Accuracy analysis for distributed weighted least-squares estimation in finite steps and loopy networks[J]. Automatica, 2018, 97: 82-91.

[57] Deshmukh S, Natarajan B, Pahwa A. State estimation over a lossy network in spatially distributed cyber-physical systems[J]. IEEE Transactions on Signal Processing, 2014, 62(15): 3911-3923.

[58] 史雨堃. 同源攻击下的多智能体系统分布式安全状态估计[D]. 北京: 北京化工大学, 2022.
(Shi Y K. Distributed security state estimation of multi-agent system under homologous attack[D]. Beijing: Beijing University of Chemical Technology, 2022.)

[59] Ozay M, Esnaola I, Vural F T Y, et al. Sparse attack construction and state estimation in the smart grid: Centralized and distributed models[J]. IEEE Journal on Selected Areas in Communications, 2013, 31(7): 1306-1318.

[60] Pasqualetti F, Dörfler F, Bullo F. Attack detection and identification in cyber-physical systems[J]. IEEE Transactions on Automatic Control, 2013, 58(11): 2715-2729.

[61] Mitra A, Sundaram S. Byzantine-resilient distributed observers for LTI systems[J]. Automatica, 2019, 108: 108487.

[62] Ao W, Song Y D, Wen C Y. Distributed robust attack detection and reconstruction for a class of uncertain nonlinear interconnected CPSs[C]. The 12th World Congress on Intelligent Control and Automation. Guilin, 2016: 1819-1824.

[63] Kekatos V, Giannakis G B. Distributed robust power system state estimation[J]. IEEE Transactions on Power Systems, 2013, 28(2): 1617-1626.

[64] Yang W, Luo W J, Zhang X T. Distributed secure state estimation under stochastic linear attacks[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(3): 2036-2047.

[65] Shi Y K, Liu C Q, Wang Y Q. Secure state estimation of multiagent systems with homologous attacks using average consensus[J]. IEEE Transactions on Control of Network Systems, 2021, 8(3): 1293-1303.

[66] Su L, Ye D, Zhao X G. Distributed secure state estimation for cyber-physical systems against replay attacks via multisensor method[J]. IEEE Systems Journal, 2022, 16(4): 5720-5728.

[67] An L W, Yang G H. Byzantine-resilient distributed state

estimation: A min-switching approach[J]. Automatica, 2021, 129: 109664.

[68] Chen Y, Kar S, Moura J M F. Attack resilient distributed estimation: A consensus+innovations approach[C]. 2018 Annual American Control Conference. Milwaukee, 2018: 1015-1020.

[69] Chen Y, Kar S, Moura J M F. Resilient distributed estimation: Sensor attacks[J]. IEEE Transactions on Automatic Control, 2019, 64(9): 3772-3779.

[70] Su L L, Shahrampour S. Finite-time guarantees for Byzantine-resilient distributed state estimation with noisy measurements[J]. IEEE Transactions on Automatic Control, 2020, 65(9): 3758-3771.

[71] Chen Y, Kar S, Moura J M F. Resilient distributed estimation: Sensor attacks[J]. IEEE Transactions on Automatic Control, 2019, 64(9): 3772-3779.

[72] Gao R, Yang G H. Trust-based distributed secure state estimation against malicious agents via two-hop communication[J]. IEEE Transactions on Automatic Control, 2023(99): 1-8.

[73] Khan M, Iqbal N, Khan A Q. Separation principle with secure state estimation and Lyapunov stability of cyber-physical system under sensor attack[C]. 2020 International Symposium on Recent Advances in Electrical Engineering & Computer Sciences. Islamabad, 2020: 1-6.

[74] Li Z S, Mo Y L. Efficient secure state estimation against sparse integrity attack for regular linear system[J]. International Journal of Robust and Nonlinear Control, 2023, 33(1): 209-236.

[75] Shinohara T, Namerikawa T, Qu Z H. Resilient reinforcement in secure state estimation against sensor attacks with A priori information[J]. IEEE Transactions on Automatic Control, 2019, 64(12): 5024-5038.

[76] Shinohara T, Namerikawa T. Distributed secure state estimation with a priori sparsity information[J]. IET Control Theory & Applications, 2022, 16(11): 1086-1097.

[77] Lu A Y, Yang G H. Secure state estimation for multiagent systems with faulty and malicious agents[J]. IEEE Transactions on Automatic Control, 2020, 65(8): 3471-3485.

作者简介

杨光红 (1963–), 男, 教授, 博士生导师, 从事故障诊断与容错控制、信息物理系统安全性、智能无人系统等研究, E-mail: yangguanghong@ise.neu.edu.cn;

芦安洋 (1991–), 男, 副教授, 博士, 从事信息物理系统安全性、切换系统、自适应控制等研究, E-mail: luanyang@ise.neu.edu.cn;

安立伟 (1991–), 男, 副教授, 博士, 从事信息物理系统安全性、自适应控制、分布式优化等研究, E-mail: anliwei@ise.neu.edu.cn.

特邀专家　杨光红, 自 1985 年入职东北大学至今, 先后任助教、讲师、副教授、教授、博士生导师. 在此期间, 于 1995 年～2006 年为香港大学、新加坡南洋理工大学、新加坡国立大学高级访问学者. 现任东北大学信息科学与工程学院院长, 东北大学特聘教授, 享受国务院颁发的政府特殊津贴, 教育部新世纪优秀人才, 国家自然科学基金创新群体负责人, 中国自动化学会会士, 爱思唯尔 (Elsevier) 中国高被引学者 (2014～2022), 担任《控制与决策》杂志主编、国际杂志 International Journal of Systems Science 编委、IEEE 控制系统协会哈尔滨分会主席、IEEE 高级会员. 主要研究方向为故障诊断与容错控制、信息物理系统安全性、智能无人系统等. 发表学术专著 3 部, SCI 期刊论文 400 余篇, Google 引用 2 万余次. 获得 2019 年国家自然科学奖二等奖 (第 4 完成人); 2011 年教育部自然科学奖二等奖 (第 1 完成人); 辽宁省自然科学一等奖等奖项. 主持国家自然科学基金创新群体项目、重点国际合作项目、联合基金重点项目以及国家重点研发计划“战略性国际创新合作”重点专项等国家级项目.

专家寄语　百年精神引领我们, 辉煌历史激励我们. 愿东北大学再创新高, 我校科学技术领跑未来, 点亮新的百年. 在未来的岁月里, 愿你再续华章, 光芒更烁, 造就更多的精英, 为社会、为国家、为世界播撒更多的智慧与希望. 东北大学, 百年辉煌, 期待未来, 更显光华!