信息物理融合系统综合安全威胁与防御研究

刘 烃[1] 田 决[1] 王稼舟[1] 吴宏宇[2] 孙利民[3] 周亚东[1] 沈 超[1] 管晓宏[1,4]

摘 要 信息物理融合系统 (Cyber-physical system, CPS) 是计算单元与物理对象在网络空间中高度集成交互形成的智能系统. 信息系统与物理系统的融合在提升系统性能的同时, 信息系统的信息安全威胁 (Security) 与物理系统的工程安全问题 (Safety) 相互影响, 产生了新的综合安全问题, 引入严重的安全隐患. 本文介绍了 CPS 的概念与安全现状, 给出了 CPS 综合安全的定义; 在对现有安全事件进行分析的基础上, 提出了 CPS 的综合安全威胁模型; 从时间关联性和空间关联性的角度, 对现有 CPS 攻击和防御方法进行了分类和总结, 并探讨 CPS 综合安全的研究方向.

关键词 信息物理融合系统, 综合安全, 威胁模型, 综合防御

引用格式 刘烃, 田决, 王稼舟, 吴宏宇, 孙利民, 周亚东, 沈超, 管晓宏. 信息物理融合系统综合安全威胁与防御研究. 自动化学报, 2019, **45**(1): 5−24

DOI 10.16383/j.aas.2018.c180461

Integrated Security Threats and Defense of Cyber-physical Systems

LIU Ting[1] TIAN Jue[1] WANG Jia-Zhou[1] WU Hong-Yu[2] SUN Li-Min[3]
ZHOU Ya-Dong[1] SHEN Chao[1] GUAN Xiao-Hong[1,4]

Abstract Cyber-physical system (CPS) is an intelligent system consisting of computing units and physical objects that highly interact with each other through the network. The combination of cyber system and physical system is promoting the performance of the CPS. At the same time, it also introduces a new integrated security threat into CPS, which combines engineering safety threat of physical system and information security threat of cyber system. In this paper, concepts and security status of CPS are introduced, and the concept of CPS integrated security is defined. After analyzing several existing security accidents in CPS, a new threat model of CPS integrated security is proposed. Existing CPS attacks and defense methods are classified and summarized from a perspective of temporal and spatial correlation. The future research of CPS integrated security is discussed.

Key words Cyber-physical system (CPS), integrated security, threat model, integrated defense

Citation Liu Ting, Tian Jue, Wang Jia-Zhou, Wu Hong-Yu, Sun Li-Min, Zhou Ya-Dong, Shen Chao, Guan Xiao-Hong. Integrated security threats and defense of cyber-physical systems. *Acta Automatica Sinica*, 2019, **45**(1): 5−24

近年来, 随着信息科学与技术的快速发展和信息计算与数据处理能力的不断提升, 工程系统和信息计算高度融合的趋势十分明显. 信息物理融合系统 (Cyber-physical system, CPS) 是计算单元和物理对象在网络环境中高度集成交互而成的智能系统[1], 如图 1 所示. CPS 包括物联网、信息物理融合能源系统或能源互联网、智能电网、智能交通系统、智能制造系统、智能物流系统等, 已成为支撑和引领新一轮产业变革的核心技术.

互联网是信息社会的关键基础设施, 其问世目的主要是为了信息共享, 因而被设计成匿名开放系统, 较少考虑安全的因素. 互联网的发展和应用完全超出了任何人预料, 网络信息安全已成为互联网发展和应用的关键问题之一. 互联网面临的安全威胁包括网络攻击 (如拒绝服务攻击、缓存区溢出、SQL 注入等)、恶意代码 (如病毒、木马、蠕虫等)、数据欺诈 (如身份伪造、数据篡改等)、网络窃听、在线社交网络攻击等. 物理系统特别是工程系统由于自身的封闭性和预设的安全保障机制, 被认为可以有效

收稿日期 2018-07-10 录用日期 2018-09-03
Manuscript received July 10, 2018; accepted September 3, 2018
国家重点研发计划 (2016YFB0800202), 国家自然科学基金 (61472318, 61632015, 61772408, U1766215, U1736205), 中央高校基本科研专项资金, 霍英东教育基金会 (151067) 资助
Supported by National Key Research and Development Program of China (2016YFB0800202), National Natural Science Foundation of China (61472318, 61632015, 61772408, U1766215, U1736205), Fundamental Research Funds for the Central Universities, and Fok Ying-Tong Education Foundation, China (151067)
本文责任编委 吕宜生
Recommended by Associate Editor LV Yi-Sheng
1. 西安交通大学智能网络与网络安全教育部重点实验室 西安 710049 中国 2. 堪萨斯州立大学电气和计算机工程系 曼哈顿 堪萨斯州 66506 美国 3. 中国科学院信息工程研究所 北京 100093 中国 4. 清华大学智能与网络化系统研究中心 北京 100084 中国
1. Key Laboratory of Intelligent Network and Network Security, Xi′an Jiao Tong University, Xi′an 710049, China 2. ECE Department, Kansas State University, Manhattan, Kansas State 66506, USA 3. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China 4. Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China

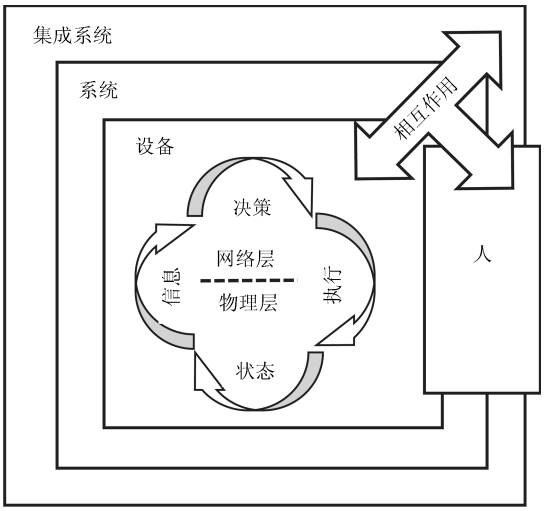


图 1 CPS 概念模型[1]

Fig. 1 CPS conceptual model[1]

抵御各类攻击, 进而保证 CPS 的安全. 然而近年来, Stuxnet 安全事件表明, 信息系统安全与物理系统工程安全相互影响, 使得网络攻击能够直接影响 CPS 安全, 对国家安全和人民生活造成严重威胁.

首先, 传统物理系统 (如工业控制系统、驾驶系统、医疗设备等) 由于其相对隔离的运行环境, 且多采用专用信道进行通信, 使得攻击者相对难接入、渗透和实施攻击. 由于信息化和智能化的需要, 传统物理系统向 CPS 演化过程中, 系统的运行环境由封闭和隔离变得开放和互联[2]. 这在提高运行效率的同时, 同样为攻击者提供了新的攻击渠道, 使得 CPS 更有可能面临来自内部或外部的攻击. 如 Stuxnet 攻击通过 U 盘摆渡侵入核设施控制系统、WindShark 则直接通过物理接入无人值守的风电场控制系统. 攻击者一旦突破 CPS 网络边界进入内部网络, 攻击成功率可能比互联网攻击更高, 威胁可能更大.

其次, 大多物理系统从工程安全的角度, 设计了故障诊断和安全应急措施, 如电力系统采用状态估计来检测和消除错误数据, 采用继电保护检测、隔离、切除系统故障等. 然而, 这些安全保护机制主要针对自然发生的工程故障. 当攻击者通过多点协同攻击时, 可以躲避相关检测. 如 Stuxnet 攻击不断伪造上报的系统量测数据, 使得控制中心无法检测到系统异常; BlackEnergy3 通过破坏系统通信模块并采用拒绝服务攻击干扰电网电话服务系统, 使得系统预设的应急响应和防御策略无法有效实施. 这些攻击的共性特点体现在: 针对特定物理系统的业务流程和安全预案, 制定相应的攻击策略, 利用网络攻击技术对多个目标实施协同攻击, 绕过物理安全防护体系, 破坏系统的正常运行, 甚至损毁物理设备.

综上所述, 在 CPS 中信息系统的信息安全 (Security) 与物理系统的工程安全 (Safety) 深度关联, 信息系统安全或物理系统安全都不能完整地定义 CPS 安全特性. 本文提到的信息安全包括数据安全、网络安全、软件和硬件安全、通信安全等与信息系统相关的安全威胁, 工程安全包括物理设备安全、系统运行安全、产品质量等与物理系统相关的安全问题. 美国国家标准与技术研究院在《CPS 架构白皮书》中指出 “安全已经成为 CPS 的主要关注点, 与网络安全只关注减轻网络攻击的影响不同, CPS 安全需要考虑到物理和网络脆弱性的协同作用”, “CPS 的本质特点不仅放大了漏洞的危害性, 还将引入新型的安全漏洞”[1]. 作者曾指出 CPS 安全正在从工程故障为主的物理安全问题, 变成同时考虑物理系统安全和信息系统安全的综合安全问题; CPS 的安全控制正从面向物理系统的安全防护走向涉及自然和社会因素的综合灾变防御[3].

基于以上分析, 本文将 CPS 综合安全的攻击定义为: 攻击者利用信息系统与物理系统之间的紧密耦合关系, 采用网络攻击或者物理攻击技术, 造成 CPS 系统故障或诱导故障在系统中传播, 破坏 CPS 完整性、可用性或保密性的行为.

1 CPS 定义与安全案例

CPS 的概念最早在 2006 年由美国国家科学院和国家科学基金会提出: CPS 是计算与物理过程的集成. 2007 年, 美国总统科学技术顾问委员会报告《挑战下的领先 — 竞争世界中的信息技术》, 将 CPS 列为未来 8 项网络与信息技术之首[4]. 2016 年, 美国国家标准技术研究院 CPS 工作组对 CPS 给出更具体的定义: CPS 是计算单元和物理对象在网络环境中高度集成交互而成的智能系统.

欧盟在 2007 年至 2013 年的 “嵌入智能与系统的先进研究与技术” (ARTMEIS) 以及 2014 年起的 “地平线 2020” (Horizon 2020) 等研究计划中投入数十亿欧元, 发展包括 CPS 在内的多项先进科技研究[5]. 德国在 2013 年 4 月的汉诺威工业博览会上推出的 “工业 4.0” 项目中, 将 CPS 列为首位[6]. 德国科学工程院强调了 CPS 在制造业中的作用, 认为 CPS 由智能机器、存储系统和生产设备组成, 能够自主地交换信息、触发动作并相互独立控制.

中国从 2009 年起开始关注 CPS, “国家自然科学基金”、“863 计划” 和 “973 计划” 等均设立课题对其进行支持, 党的十八大和十九大相继提出了 “信息化和工业化深度融合”、“互联网 +”、“中国制造 2025” 等国家战略, 将 CPS 列为支撑新一轮产业变革的核心技术[7]. 2017 年, 工信部和国家标准化管理委员会发布的《信息物理系统白皮书》给出定义: CPS 通过集成先进的感知、计算、通信、控制等信

息技术和自动控制技术, 构建了物理空间与信息空间中人、机、物、环境、信息等要素相互映射、适时交互、高效协同的复杂系统, 实现系统内资源配置和运行的按需响应、快速迭代、动态优化[8].

国内外对 CPS 的定义都强调 CPS 的核心在于信息系统与物理系统的融合, 而融合引入的 CPS 综合安全问题也一直受到世界各国的重视. 2006 年, 美国美国国家科学基金会 (National Science Foundation) 将 CPS 安全列为科学研究重要领域, 国土安全部制订了 "国家基础设施保护计划". 2010 年英国发布国家安全策略, 开展 "国家网控安全项", 支持工业控制系统安全的相关研究. 2013 年, 欧洲网络与信息安全局发布《工业控制系统网络安全白皮书》. 在中国, 2011 年 4 月工信部等五部委联合发布《关于加快推进信息化与工业化深度融合的若干意见》, 指出 "发展完善面向工业行业的安全可靠的信息化服务平台"; 2012 年 6 月国务院要求 "保障工业控制系统的安全, 加强重要领域工业控制系统, 以及物联网应用、数字城市建设中的安全防护和管理"; 2016 年和 2018 年, 国家重点研发计划先后设立两项 CPS 安全的项目《内生安全的主动防御工控系统防护技术研究》和《工业控制系统安全保护技术应用示范项目》.

为了分析 CPS 综合安全的特征, 本文回顾近年来全球影响广泛的 4 起 CPS 安全事件: 2010 年针对伊朗核设施的 Stuxnet 攻击、2014 年针对欧洲工业制造系统的 Havex 攻击, 2015 年针对乌克兰电网的 BlackEnergy3 攻击以及 2017 年美国风电场安全实验, 并对其攻击过程进行分析.

2010 年伊朗政府承认, 伊朗位于纳坦兹的铀浓缩工厂遭受 Stuxnet 蠕虫攻击. 据报导, 这次攻击导致伊朗近千台离心机损毁, Bushehr 核电站也遭 Stuxnet 病毒感染, 最终 Bushehr 核电站被迫关闭、铀浓缩计划停滞和伊朗核计划推迟[9]. Stuxnet 蠕虫是首次出现以核电站和核设施为目标的网络攻击. Stuxnet 蠕虫利用了 4 个 Zero-day 漏洞入侵核电站信息系统, 躲过入侵检测系统等信息网络安全监控, 随后采用 Rootkit 技术入侵 PLC 控制器, 控制离心机异常运行, 同时伪造离心机的运行数据, 欺骗数据采集与监控系统 (Supervisory control and data acquisition, SCADA) 的故障诊断功能, 直到离心机损坏[10−11].

2014 年, 欧洲大量工业制造系统遭受 Havex 木马袭击, 其专门针对 SCADA 和工控系统 (Industrial control system, ICS) 中的工业控制软件, 进行远程控制, 可造成水电坝失控、核电站过载、电网断路等后果[12].

2015 年 12 月 23 日, 乌克兰电力部门遭受 BlackEnergy3 攻击, 造成了 7 座 110 KV 和 23 座 35 KV 变电站断电长达 3 个小时, 使得 3 个不同区域大约 22 万人失去电力供应. BlackEnergy3 是全球首个导致电力系统瘫痪的网络攻击. BlackEnergy3 利用 Office 的漏洞入侵电力部门办公系统, 再通过 VPN 和 ICS 的远程管理功能入侵电网控制系统, 下达断路器断开指令, 导致输电网断路. 为阻断电网保护和恢复机制, 攻击者篡改日志文件、破坏数据存储系统、关闭监控系统, 甚至对客服电话系统发动电话拒绝服务攻击[13].

2017 年 7 月, 美国塔尔萨大学 Staggs 博士团队公布了 3 个针对风电场的攻击. Windshark 向联网的涡轮机发送命令, 禁用或者反复制动急停, 以造成磨损和破坏. Windworm 利用 Telnet 和 FTP 在可编程自动化控制器间扩散, 感染整个风电场的计算机. Windpoison 利用 ARP 缓存病毒, 发现和定位控制系统的网络组件漏洞, 并伪造涡轮机发回的信号, 隐瞒机组遭攻击破坏的事实. 研究团队在美国中部的风电场撬开风力发电设备的服务器机柜, 将通信设备接入风电控制系统, 实现远程控制风力发电机[14].

从上述真实案例可以看出, CPS 攻击过程一般由两部分组成. 1) 利用网络攻击技术对系统信息网络进行探测、入侵、提权和控制, 探知目标系统拓扑结构、运行模式等, 获得目标系统量测或控制数据的修改权限, 为后续攻击提供基础; 2) 利用 CPS 设计和业务流程实施攻击, 包括篡改控制指令造成系统异常运行, 阻断或篡改系统量测数据以阻止控制系统的安全响应.

第一部分与现有网络攻击技术相似, 第二部分体现出 CPS 攻击的特殊性, 因为物理系统的状态变化有一定限制 (如电力系统中发电机出力的提升有爬升约束限制), 且物理系统都有安全应急机制和保护措施. 因此攻击者往往结合物理系统的业务逻辑和保护机制, 设计攻击策略, 一方面通过持续的攻击使得系统达到特定状态; 另一方面隐藏自身的攻击行为, 躲避系统的异常检测和保护机制. 上述 CPS 安全事件都有信息物理耦合与攻击隐蔽两个特点.

1) 信息物理耦合

由于信息与物理系统耦合, 为使得攻击效果最大化, 针对 CPS 的攻击必须考虑系统业务流程和物理约束. 攻击的构建会受到物理系统本身相应条件的约束, 而攻击的达成及其巨大的破坏力又依赖于这些条件.

2) 攻击隐蔽

攻击者往往长期潜伏, 以获得足够的物理系统知识特别是安全约束信息和控制权限, 在安全监控系统未察觉的情况下进行攻击. 从开始接入探测到

完成攻击目标, 一般需要数日至数周的潜伏期, 在这一阶段保持隐蔽.

这两个特点在已知的 CPS 攻击中普遍存在, 也是区别于互联网攻击最为明显的差异. 本文基于信息系统与物理系统的耦合特征, 构建 CPS 安全威胁模型, 并针对 CPS 攻击的隐蔽性, 对 CPS 安全威胁和防御方法进行分类和总结.

2 CPS 综合安全威胁模型

抵御信息网络攻击的能力是衡量 CPS 性能的一项重要指标. 近年来的安全事件显示, CPS 面临的攻击手段在不断更新, 其多样性和隐蔽性极大地增加了防御代价与成本. 成功的系统防御通常建立在充分了解系统结构及攻击的基础上. 因此, 需要充分了解 CPS 攻击特点和现有异常检测机制, 制订针对性防御策略, 以增强系统安全防护能力.

信息系统与物理系统的耦合性是建立 CPS 综合安全威胁模型的难点, 不仅要分析不同系统自身的拓扑结构、安全监控机制等, 还要考虑两类系统间的信息交换模式、安全漏洞等. 本文首先总结现有 CPS 系统模型和安全监控机制, 结合现有威胁模型, 重点针对控制信号和量测信号篡改攻击的隐蔽性, 提出一种 CPS 综合安全威胁模型.

2.1 CPS 系统模型

典型的 CPS 由控制中心、物理设备、执行器与传感器构成, 如图 2 所示. 当系统状态在一定范围内变化时, CPS 可近似为线性系统. 由于线性系统理论和方法较为成熟, 本文以离散时间的线性时不变 (Linear time invariant, LTI) 系统为例, 阐述建模思想与方法. 具体来说, LTI 模型可描述如下:

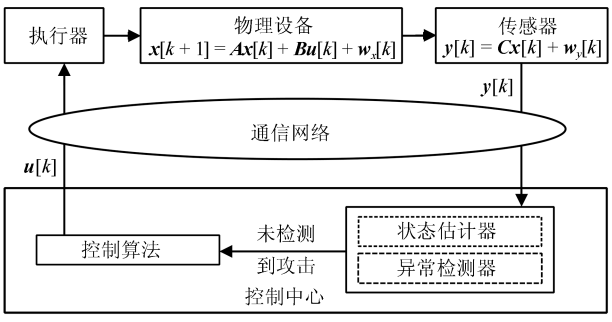


图 2 CPS 控制模型

Fig. 2 CPS control model

$$\boldsymbol{x}[k+1] = A\boldsymbol{x}[k] + B\boldsymbol{u}[k] + \boldsymbol{w}_x[k] \tag{1}$$

$$\boldsymbol{y}[k] = C\boldsymbol{x}[k] + \boldsymbol{w}_y[k] \tag{2}$$

其中, $\boldsymbol{x}[k] \in \mathbf{R}^n$ 与 $\boldsymbol{y}[k] \in \mathbf{R}^m$ 分别表示系统状态与传感器量测值; $\boldsymbol{u}[k] \in \mathbf{R}^l$ 为控制信号; $A \in \mathbf{R}^{n\times n}$、$B \in \mathbf{R}^{n\times l}$ 与 $C \in \mathbf{R}^{m\times n}$ 分别为系统矩阵、控制矩阵与量测矩阵; $\boldsymbol{w}_x[k] \in \mathbf{R}^n$ 与 $\boldsymbol{w}_y[k] \in \mathbf{R}^m$ 分别表示过程噪声与量测噪声. 假设 $\boldsymbol{w}_x[k]$ 与 $\boldsymbol{w}_y[k]$ 均服从零均值多元高斯分布, 其协方差矩阵分别为 Q 与 R, 式中, $Q \in \mathbf{R}^{n\times n}$ 且 $R \in \mathbf{R}^{m\times m}$. 控制信号 $\boldsymbol{u}[k]$ 与传感器量测值 $\boldsymbol{y}[k]$ 由通信网络进行传输.

工业控制系统对于传感器的量测数据需要经过预处理, 主要包括状态估计器与异常检测器, 其检测原理可分为两类: 一类是利用系统过去时间的状态预测未来的状态, 并根据预测结果与量测结果进行异常数据检测, 本文将其命名为时间相关型检测; 另一类是通过系统不同传感器间的关联关系, 进行交叉验证, 以实现检测目的, 本文将其命名为空间相关型检测.

2.2 时间相关型检测

2.2.1 基本模型

时间相关型检测机制中, 系统当前状态估计值应与之前的估计值、量测值以及控制信号相关. 具体来讲, 用 $\hat{\boldsymbol{x}}$ 表示系统状态估计值. 在第 $k+1$ 时刻, 状态估计器可计算系统状态估计值如下:

$$\hat{\boldsymbol{x}}[k+1] = L_1(\hat{X}[k], U[k], Y[k+1]) \tag{3}$$

其中, $\hat{X}[k] = [\hat{\boldsymbol{x}}[k], \cdots, \hat{\boldsymbol{x}}[0]] \in \mathbf{R}^{n\times(k+1)}$; $U[k] = [\boldsymbol{u}[k], \cdots, \boldsymbol{u}[0]] \in \mathbf{R}^{l\times(k+1)}$; $Y[k] = [\boldsymbol{y}[k], \cdots, \boldsymbol{y}[0]] \in \mathbf{R}^{m\times(k+1)}$, $L_1(\cdot)$ 为抽象函数. 为了恰当地初始化状态估计器, 本文假设系统操作者知晓系统初始状态 $\boldsymbol{x}[0]$. 结合 k 时刻预测的传感器量测值, 可计算系统量测值的残差:

$$\hat{\boldsymbol{y}}[k+1] = L_2(\hat{X}[k], U[k], Y[k]) \tag{4}$$

$$\boldsymbol{\epsilon}[k+1] := \boldsymbol{y}[k+1] - \hat{\boldsymbol{y}}[k+1] \tag{5}$$

其中, $\hat{\boldsymbol{y}}[k+1]$ 为第 $(k+1)$ 时段量测预测值; $L_2(\cdot)$ 为抽象函数. $\boldsymbol{\epsilon}$ 是系统量测值残差, 异常检测器基于残差可检测各种数据异常或错误数据注入攻击. 若 $-\boldsymbol{\epsilon}_0 \le \boldsymbol{\epsilon} \le \boldsymbol{\epsilon}_0$, 则说明未检测到异常, 即不存在数据异常或攻击; 反之, 则说明存在数据异常或攻击. $\boldsymbol{\epsilon} \le \boldsymbol{\epsilon}_0$ 表示向量 $\boldsymbol{\epsilon}$ 中每一个元素均不超过 $\boldsymbol{\epsilon}_0$ 中对应元素; $\boldsymbol{\epsilon}_0$ 为预设的阈值向量以保证一定 α 等级的检测率.

为保证 CPS 的安全稳定运行, 控制算法模块计算系统的控制信号, 以使得系统状态维持在期望目标状态 $\boldsymbol{x}_0$ 附近. 考虑通用控制算法如下:

$$\boldsymbol{u}[k+1] = L_3(\hat{X}[k+1], U[k], Y[k+1], \boldsymbol{x}_0) \tag{6}$$

其中, $L_3(\cdot)$ 为抽象函数.

2.2.2 基于卡尔曼滤波的状态估计器与异常检测器

本节以文献研究与真实系统中广泛采用的基于卡尔曼滤波的状态估计器和卡方异常检测器[15−18]为实例, 分析 CPS 的时间相关型检测理论方法.

基于卡尔曼滤波的状态估计器将式 (3) 实例化为

$$\begin{aligned}\hat{\boldsymbol{x}}[k+1] &= A\hat{\boldsymbol{x}}[k] + B\boldsymbol{u}[k] + FD\\ F &= PC^{\mathrm{T}}(CPC^{\mathrm{T}} + R)^{-1}\\ D &= \boldsymbol{y}[k+1] - CA\hat{\boldsymbol{x}}[k] - CB\boldsymbol{u}[k]\\ P &= APA^{\mathrm{T}} + Q - APC^{\mathrm{T}}(CPC^{\mathrm{T}} + R)^{-1}\end{aligned} \tag{7}$$

其中, $F \in \mathbf{R}^{n\times m}$ 为卡尔曼增益矩阵. 式 (4) 转化为

$$\hat{\boldsymbol{y}}[k+1] = CA\hat{\boldsymbol{x}}[k] + CB\boldsymbol{u}[k] \tag{8}$$

卡方检测器计算标量残差如下:

$$\boldsymbol{r}[k] = \boldsymbol{\epsilon}[k]^{\mathrm{T}} G^{-1}\boldsymbol{\epsilon}[k] \tag{9}$$

其中, $G \in \mathbf{R}^{m\times m}$ 为 $\boldsymbol{\epsilon}[k]$ 的协方差矩阵. 当系统采用卡尔曼估计器, 则有

$$G = CPC^{\mathrm{T}} + R \tag{10}$$

且 $r[k]$ 服从 χ^2_m 分布. 若 $\boldsymbol{r}[k+1] > \chi^2_{m,\alpha}$, 卡方检测器结果呈阳性, 式中, $\chi^2_{m,\alpha}$ 表示卡方分布在自由度为 m、概率值 P 为 $1-\alpha$ 的卡方值.

基于卡尔曼滤波的状态估计器和卡方异常检测器为时间相关型检测提供了理论和方法支持, 已经广泛应用于电力系统、生产制造等系统. 但与现有大部分时间相关型检测方法类似, 该类方法对于运行状态不稳定的系统, 难以进行有效监控.

2.3 空间相关型检测

2.3.1 基本模型

空间相关型检测的基本原理是根据系统在某个时间断面上, 不同节点量测值之间的耦合关系对系统当前状态进行估计, 此时要求系统的量测信号维数 m 大于等于系统状态维数 n, 同时量测矩阵 C 通常列满秩, 即 $\mathrm{rank}(C) = n$. 具体来讲, 用 $\hat{\boldsymbol{x}}$ 表示系统状态估计值. 在第 k 时段, 状态估计器可计算系统状态估计值如下:

$$\hat{\boldsymbol{x}}[k] = L_4(\boldsymbol{y}[k]) \tag{11}$$

其中, $L_4(\cdot)$ 为抽象函数. 基于状态估计值与式 (2), 可得传感器量测估计值:

$$\hat{\boldsymbol{y}}[k] = C\hat{\boldsymbol{x}}[k] \tag{12}$$

其中, $\hat{\boldsymbol{y}}[k]$ 为第 k 时段量测估计值. 在时间相关型异常检测器下, 其系统残差设定式与式 (5) 一致. 值得注意的是, 在时间相关型状态估计器中, 式 (4) 为系统对于下一时刻量测量的预测值, 而残差定义为实际量测量与预测值的偏差; 而在空间相关型检测器中, 式 (12) 为系统根据实际量测量对于当前时刻量测量的估计值, 而残差定义为实际量测量与估计值的偏差. 此外, 系统控制算法如下:

$$\boldsymbol{u}[k] = L_5(\hat{\boldsymbol{x}}[k], \boldsymbol{x}_0) \tag{13}$$

其中, $L_5(\cdot)$ 为抽象函数.

2.3.2 电力系统异常数据检测

现有文献中, 空间相关型状态估计器与异常检测器通常特指电力系统下状态估计与错误数据检验, 本文从电力系统的角度详细描述其模型. 在电力系统中, 潮流计算用于研究稳态情况下电力系统的运行情况, 其任务为根据电力系统运行条件与网络结构参数计算电力系统的运行情况, 如母线节点电压 (包括幅值与相角)、网络潮流分布等.

电力系统潮流模型通常分为交流潮流模型与直流潮流模型. 在交流潮流模型中, 电表量测值包含母线节点注入有功功率与无功功率、各传输线始端与末端有功功率与无功功率以及各节点电压幅值; 系统状态包含节点电压幅值与相角. 假设电力系统具有 m 个量测值 $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)^{\mathrm{T}}$ 以及 n 个状态量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$, 通常情况下 $m \geq n$, 则电表量测值 $\boldsymbol{y}$ 与系统状态 $\boldsymbol{x}$ 关系如下式:

$$\boldsymbol{y} = c(\boldsymbol{x}) + \boldsymbol{w} \tag{14}$$

式中, $c(\cdot)$ 为非线性函数, 由系统拓扑结构、线路阻抗、线路对地电纳与变压器变比等参数决定; $\boldsymbol{w} \in \mathbf{R}^m$ 表示系统噪声. 通常假设 $\boldsymbol{w}$ 服从零均值的高斯分布, 其协方差矩阵为 $R \in \mathbf{R}^{m\times m}$.

对于大规模电力系统, 由于交流潮流模型的非线性会导致计算复杂度极度增加, 甚至求解过程无法收敛. 在一些情况下, 电力系统会采用线性化的直流潮流模型以近似交流模型. 相比于交流模型, 直流模型计算精度较低, 但具有更强的鲁棒性. 直流模型通常用于电力系统实时运算, 如计算节点边际电价等. 直流模型对交流模型主要进行了如下简化: 1) 假设各节点电压相同; 2) 假设各线路无损, 即线路电阻为 0; 3) 假设各线路始末两端相角差很小, 因此可用 θ 近似代替 $\sin(\theta)$, 式中, $\theta \approx 0$. 在上述假设下, 全电网无功功率为 0, 各传输线始端与末端有功功率相同. 因此电表量测值包含母线节点注入有功功率、各传输线始端或末端有功功率, 系统状态为各节点电压相角. 假设电力系统量测值 $\boldsymbol{y}$ 维数为 m、状态量 $\boldsymbol{x}$ 维数为 n, 且 $m \geq n$. 系统状态 $\boldsymbol{x}$ 由节点注入功率与节点电纳矩阵决定, 具体模型如下:

$$\boldsymbol{p} = \mathcal{B}\boldsymbol{x} \tag{15}$$

式中, $\boldsymbol{p}$ 为节点注入功率向量, 且 $\boldsymbol{p} \in \mathbf{R}^n$; $\mathcal{B}$ 为节点电纳矩阵, 包含了系统拓扑与线路电纳, 且 $\mathcal{B} \in \mathbf{R}^{n\times n}$. 对于连通的电力系统, 矩阵 $\mathcal{B}$ 是非奇异的. 因而, $\boldsymbol{x} = \mathcal{B}^{-1}\boldsymbol{p}$. 此外, 直流模型下电表量测值 $\boldsymbol{y}$ 与系统状态 $\boldsymbol{x}$ 关系如下式:

$$\boldsymbol{y} = C\boldsymbol{x} + \boldsymbol{w} \tag{16}$$

式中, $C \in \mathbf{R}^{m\times n}$ 表示系统量测矩阵, 通常情况下, 该矩阵列满秩, 即 $\mathrm{rank}(C) = n$. 用 (i, j) 表示连接节点 i 与 j 的线路, y_{ij} 表示线路 (i, j) 的电力潮流量测值, w_{ij} 表示 y_{ij} 中所含量测噪声, x_i 表示 $\boldsymbol{x}$ 中与节点 i 相关的元素, 则有 $y_{ij} = -b_{ij}(x_i - x_j) + w_{ij}$, 式中, b_{ij} 表示线路电纳. 因而, 量测矩阵 C 中与线路 (i, j) 对应的行向量 (记为 $\boldsymbol{C}_{ij}$) 为[19]

$$\boldsymbol{C}_{ij} = \big[0 \ \cdots \ 0 \quad \underbrace{-b_{ij}}_{\text{列}i} \quad 0 \ \cdots \ 0 \quad \underbrace{b_{ij}}_{\text{列}j} \quad 0 \ \cdots \ 0\big] \tag{17}$$

为提高电力系统数据质量, 在硬件层面上可增加高量测精度、高量测速度以及高可靠性量测设备; 在软件层面上, 可采用状态估计等技术对数据进行实时处理. 状态估计的实质为对系统量测值进行滤波, 利用量测系统的冗余度来提高系统的量测精度、降低随机噪声对系统的干扰, 以估计或预测系统的运行状态. 电力系统周期性地进行状态估计, 在离线状态下, 电力系统只通过 SCADA 进行估计, 其周期与 SCADA 周期一致, 为 20 秒至几分钟; 而处于在线估计时, 需要使用尽量准确实时的数据分析系统当前的情景, 周期为 10 分钟至数小时. 本文将此周期称为状态估计周期.

在交流模型下, 基于加权最小二乘 (Weighted least-squares, WLS) 的状态估计为寻找具有最小 WLS 误差的状态估计值 $\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} [\boldsymbol{y} - c(\boldsymbol{x})]^{\mathrm{T}} R^{-1} [\boldsymbol{y} - c(\boldsymbol{x})] \tag{18}$$

由于交流模型的非线性, 通常采用高斯牛顿 (Gausson-Newton) 迭代法或牛顿−拉夫逊 (Newton-Raphson) 法进行求解. 然而, 当系统维数较大时, 迭代法的时间开销很大, 同时迭代法可能无法获得全局最优解甚至无法收敛. 在直流模型下, 若量测噪声是独立同分布的高斯噪声, 则具有最小均方误差 (Mean squared error, MSE) 的系统状态估计值:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} [\boldsymbol{y} - C\boldsymbol{x}]^{\mathrm{T}} R^{-1} [\boldsymbol{y} - C\boldsymbol{x}] = (C^{\mathrm{T}} R^{-1} C)^{-1} C^{\mathrm{T}} R^{-1} \boldsymbol{y} = C^{+}\boldsymbol{y} \tag{19}$$

其中, $C^{+} := (C^{\mathrm{T}} R^{-1} C)^{-1} C^{\mathrm{T}} R^{-1}$ 表示矩阵 C 的广义逆矩阵. 在电网系统中, 状态估计残差 (用 r 表示) 定义为量测量真实值与估计值之间的偏差. 特别的, 若 R 为对角矩阵 (即系统噪声独立), 则有:

$$r = \left\| \sqrt{R^{-1}} (\boldsymbol{y} - \hat{\boldsymbol{y}}) \right\|_2 \tag{20}$$

其中, $\hat{\boldsymbol{y}}$ 为量测量估计值; $\sqrt{R^{-1}}$ 为对矩阵 R^{-1} 对角线上所有元素进行开方所得的对角矩阵. 交流模型下, $\hat{\boldsymbol{y}} = c(\hat{\boldsymbol{x}})$; 直流模型下, $\hat{\boldsymbol{y}} = CC^{+}\boldsymbol{y}$. 错误数据检验将该残差与阈值 η 比较, 以判断系统中是否存在错误数据. 若 $r > \eta$, 则检验结果呈阳性, 即系统中存在错误数据. 若系统噪声服从正态分布, 由于 r^2 服从卡方分布 $\chi^2_{(m-n)}$, 则阈值 η 可设定为

$$\eta = \sqrt{\chi^2_{(m-n),\alpha}} \tag{21}$$

式中, $\chi^2_{(m-n),\alpha}$ 表示卡方分布在自由度为 $(m-n)$、概率值 P 为 $1-\alpha$ 的卡方值.

电力系统状态估计与错误数据检验可以有效检测系统中的量测误差和错误, 在实际电网中已有广泛应用. 然而类似空间相关型检测方法都面临两大挑战: 1) 计算复杂度问题. 状态估计的计算复杂度不低于 $\mathrm{O}(N^2)$, 大规模系统的状态估计本身就是挑战性难题; 2) 系统结构的动态变化. 大规模系统的拓扑结构可能经常变化, 系统参数随着环境、运行状态发生发生改变, 上述变化将导致系统矩阵和量测矩阵的改变, 严重影响状态估计和异常检测的精度.

2.4 CPS 综合安全威胁模型分析

本节分析攻击场景下的 CPS 综合安全威胁模型, 为了保证分析的普适性, 考虑了控制信号与量测值同时被篡改的攻击场景, 如图 3 所示. 沿用第 2.1 节定义的符号表示正常状态下的物理量, 并定义系统遭受的攻击量如下 (为方便本节定义中省略时间索引):

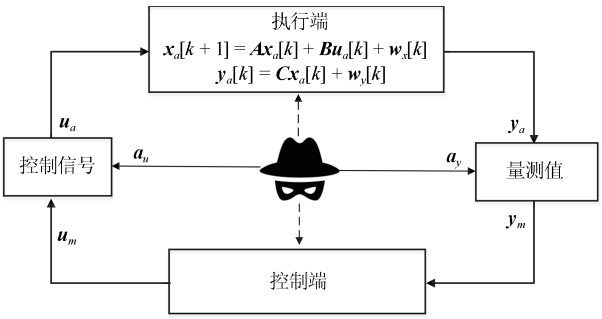


图 3 CPS 综合安全威胁模型

Fig. 3 Integrated security model of CPS

1) 用 $\boldsymbol{x}_a$ 与 $\boldsymbol{y}_a$ 分别表示实际系统状态与传感器量测值.

2) 用 $\hat{\boldsymbol{x}}_m$、$\hat{\boldsymbol{y}}_m$、$\boldsymbol{\epsilon}_m$ 与 $\boldsymbol{u}_m$ 分别表示控制中心的

状态估计值、传感器量测的预测值、异常检测器残差以及确定的控制信号.

3) 用 $\boldsymbol{u}_a$ 表示执行器接收的实际控制信号, 该信号已经被攻击者篡改. 具体来讲, $\boldsymbol{u}_a = \boldsymbol{u}_m + \boldsymbol{a}_u$, 式中, $\boldsymbol{u}_m$ 表示控制中心下达的控制指令; $\boldsymbol{a}_u \in \mathbf{R}^l$ 表示 $\boldsymbol{u}_m$ 中注入的恶意控制信号.

4) 用 $\boldsymbol{y}_m$ 表示控制中心接收的传感器量测值, 该信号也已被攻击者篡改. 具体来讲, $\boldsymbol{y}_m = \boldsymbol{y}_a + \boldsymbol{a}_y$, 式中, $\boldsymbol{a}_y \in \mathbf{R}^m$ 表示注入传感器实际量测值 $\boldsymbol{y}_a$ 的恶意信号.

从被控制的物理系统来看, 物理系统接收到的控制指令已被攻击者篡改, 物理系统基于该指令做出错误的决策, 同时, 对当前状态的量测值也会被攻击者篡改. 在 k 时刻, 其接受的控制指令、系统状态方程和量测结果可表示为

$$\boldsymbol{u}_a = \boldsymbol{u}_m + \boldsymbol{a}_u \tag{22}$$

$$\boldsymbol{x}_a[k+1] = A\boldsymbol{x}_a[k] + B\boldsymbol{u}_a[k] + \boldsymbol{w}_x[k] \tag{23}$$

$$\boldsymbol{y}_a[k] = C\boldsymbol{x}_a[k] + \boldsymbol{w}_y[k] \tag{24}$$

从控制中心来看, 其接收到的量测值 $\boldsymbol{y}_m$ 已经被攻击者篡改为正常状态下的数据, 控制中心对所接收到的量测值进行异常检测, 判定为正常后, 会根据控制策略发送控制信号 $\boldsymbol{u}_m$. 在 k 时刻, 被篡改的传感器量测值将影响式 (3) 与式 (4) 中所示的系统状态估计过程、式 (5) 中所示的异常检测以及式 (6) 中所示的控制算法可以表示为

$$\hat{\boldsymbol{x}}_m[k+1] = L_1(\hat{X}_m[k], U_m[k], Y_m[k+1]) \tag{25}$$

$$\hat{\boldsymbol{y}}_m[k+1] = L_2(\hat{X}_m[k], U_m[k], Y_m[k]) \tag{26}$$

$$\boldsymbol{\epsilon}_m[k+1] = \boldsymbol{y}_m[k+1] - \hat{\boldsymbol{y}}_m[k+1] \tag{27}$$

$$\boldsymbol{u}_m[k+1] = L_3(\hat{X}_m[k+1], U_m[k], Y_m[k+1], \boldsymbol{x}_0) \tag{28}$$

式中, $\hat{X}_m[k] = [\hat{\boldsymbol{x}}_m[k] \cdots \hat{\boldsymbol{x}}_m[0]]$; $Y_m[k] = [\boldsymbol{y}_m[k] \cdots \boldsymbol{y}_m[0]]$; $U_m[k] = [\boldsymbol{u}_m[k] \cdots \boldsymbol{u}_m[0]]$. 注意 $\hat{\boldsymbol{y}}_m[0] = \hat{\boldsymbol{y}}[0]$.

本文分析具体以时间相关型攻击下的威胁模型为例, 此外, 由于空间相关型数据完整性攻击下的威胁模型与时间相关型攻击下的威胁模型类似, 在此不再进行赘述.

本文设计的 CPS 综合安全威胁模型的核心思想在于从系统管理者的角度, 将 CPS 攻击可能发生的情况, 抽象为系统控制流程中, 对量测值与控制指令的篡改; 将 CPS 攻击策略构建的策略, 描述为在安全防御约束下的策略求解问题. 为研究人员和管理人员理解 CPS 综合安全威胁, 提供了一种可行的理论方法.

3　CPS 综合安全攻击研究现状

3.1　攻击分类

通过调研现有文献和报导, 已知的 CPS 攻击有十余种, 如数据重放攻击、数据放大攻击、零动态攻击、零状态诱导攻击以及系统模拟攻击等, 大部分属于数据完整性攻击, 即通过篡改系统的量测量或者控制量, 躲避 CPS 安全监控和实现攻击目的. 因此, 本章结合前文总结的空间相关性和时间相关性检测机制, 从空间隐蔽和时间隐蔽的角度将已知的 CPS 攻击分为 4 类, 如图 4 所示.

空间隐蔽型攻击典型代表为最早由 Liu 等[20−21] 在 2009 年提出的电力系统错误数据注入攻击. 该类攻击主要针对基于多维数据间的数学耦合关系的错误数据检测机制, 构造错误控制指令和量测数据并实施攻击, 使得系统对当前的运行状态或者拓扑结构产生错误的估计; 但该类攻击不能躲避

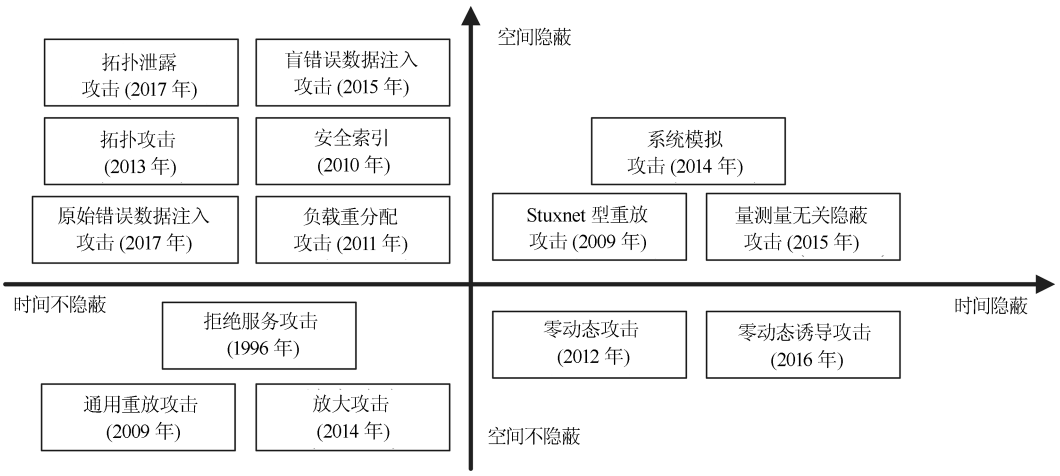


图 4　CPS 攻击分类

Fig. 4　Taxonomy of CPS attack

时间相关型检测方法. 这类攻击对系统的完备度以及获取的数据维数有一定要求, 目前主要以电力系统为目标, 相关研究包括原始错误数据注入攻击、拓扑攻击、盲错误数据注入攻击、负载重分配攻击、拓扑泄漏攻击以及安全索引等.

时间隐蔽型攻击包括零动态攻击和零状态诱导攻击, 该类攻击主要针对利用系统状态在连续时间上的动态规律, 来预测和检测错误量测的安全机制, 构造错误控制指令和量测数据并实施攻击, 使得系统对当前的运行状态或者拓扑结构产生错误的估计; 但该类攻击不能躲避空间相关型检测方法.

空间−时间隐蔽型攻击在已知 CPS 攻击中, 已引发多起重大安全事件, 如 Stuxnet 攻击等. 这类攻击既考虑了各个量测数据之间耦合关系, 也考虑了量测数据本身在时间上的变化规律与约束, 成功躲避了时间相关型和空间相关型的异常检测方法.

部分 CPS 攻击未考虑时间和空间相关型的异常检测方法. 虽然这类攻击容易被系统检测到, 难以造成严重的后果, 但其思想和技术可能被攻击者利用, 值得考虑和借鉴, 本章最后对其进行统一分析.

3.2 空间隐蔽型攻击

本节首先分析空间隐蔽型攻击, 已有研究中该类攻击主要针对电力系统状态估计与错误数据检验. 由于空间隐蔽型攻击仅和当前系统结构知识有关, 本文将空间隐蔽型攻击分为两类场景: 攻击者拥有完备系统信息与信息受限.

3.2.1 信息完备条件下空间隐蔽型攻击

假设攻击者将系统量测量篡改为 $\boldsymbol{y}_m$, 即 $\boldsymbol{y}_m = \boldsymbol{y} + \boldsymbol{a}$, 式中, $\boldsymbol{y}_m \in \mathbf{R}^m$; $\boldsymbol{a} \in \mathbf{R}^m$ 表示攻击者对量测值的注入量.

2009 年, Liu 等[20] 首次提出了直流模型下隐蔽错误数据注入攻击 (False data injection, FDI). 该攻击假设攻击者能获取当前电力系统拓扑结构与配置信息, 并具有篡改电表量测值的能力. 攻击者通过精心构造虚假电力量测数据, 绕过电力系统的错误数据检验, 并诱导电力系统状态估计对系统状态值产生错误估计[20−21]. FDI 针对电力系统直流模型设计, 其攻击量满足:

$$\boldsymbol{a} = C\dot{\boldsymbol{x}} \tag{29}$$

式中, $\dot{\boldsymbol{x}}$ 为任意向量, 且 $\dot{\boldsymbol{x}} \in \mathbf{R}^n$. 则有:

$$\begin{aligned}\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m =& (I_n - CC^+)\boldsymbol{y}_m = \\ & (I_n - CC^+)\boldsymbol{y} + (I_n - CC^+)\boldsymbol{a} = \\ & (I_n - CC^+)\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}}\end{aligned}$$

因此, 可得 $r_m = r$, 式中, r_m 表示攻击后的系统残差. 这意味着攻击前后, 系统残差不发生变化, 即攻击是隐蔽的. 这是因为, 当攻击发生后, 系统状态估计值 $\hat{\boldsymbol{x}}_m$ 变为

$$\hat{\boldsymbol{x}}_m = C^+\boldsymbol{y}_m = C^+(\boldsymbol{y} + C\dot{\boldsymbol{x}}) = \hat{\boldsymbol{x}} + \dot{\boldsymbol{x}} \tag{30}$$

这意味着攻击量额外注入量 $\boldsymbol{a}$ 完全模拟了系统状态为 $\dot{\boldsymbol{x}}$ 时的系统量测量.

Hug 等将直流模型下的隐蔽攻击推广至交流模型[22], 其中攻击量可设定为

$$\boldsymbol{a} = c(\hat{\boldsymbol{x}} + \dot{\boldsymbol{x}}) - c(\hat{\boldsymbol{x}}) \tag{31}$$

进而可以计算攻击实施时, 系统残差为

$$\begin{aligned}r_m = & \left\| \sqrt{R^{-1}}(\boldsymbol{y}_m - c(\hat{\boldsymbol{x}}_m)) \right\|_2 \leq \\ & \left\| \sqrt{R^{-1}}(\boldsymbol{y}_m - c(\hat{\boldsymbol{x}} + \dot{\boldsymbol{x}})) \right\|_2 = \\ & \left\| \sqrt{R^{-1}}(\boldsymbol{y} - c(\hat{\boldsymbol{x}})) \right\|_2 = r\end{aligned}$$

即 $r_m \leq r$. 这意味着攻击后的系统残差不大于攻击前的系统残差, 即攻击是隐蔽的. 值得注意的是, 交流模型下的隐蔽攻击需要攻击者获得系统状态的估计量 $\hat{\boldsymbol{x}}$.

2011 年, Yuan 等根据电力系统的特点, 将 Liu 等提出的攻击增加了如下约束限制以使得攻击在真实情况下更为合理, 该攻击被称为负载重分配攻击[23]. 约束包括: 发电机量测值不可修改; 电网中零注入节点的功率不可被修改 (零注入节点不与任意负载或发电机相连); 负载量测值可被修改; 传输线潮流量测值可被修改.

2013 年, Kim 等将错误数据注入攻击的目标由电力系统对状态的错误估计延伸至对系统当前拓扑结构的错误估计, 称为拓扑攻击[24]. 与 FDI 相比, 拓扑攻击中攻击量 $\boldsymbol{a}$ 不仅与系统量测矩阵相关, 同时与系统当前状态 (或量测值) 相关. 因此, 将 $\boldsymbol{a}$ 重新表示为 $\boldsymbol{a}(\boldsymbol{y})$. Kim 等主要研究了拓扑攻击下的状态保持攻击, 即攻击前后系统的估计状态不变, 估计拓扑发生变化. 在直流模型下, 攻击量为

$$\boldsymbol{a}(\boldsymbol{y}) = (\bar{C} - C)\boldsymbol{x} \tag{32}$$

C 为攻击前系统真实量测矩阵, $\bar{C}$ 为攻击者的目标量测矩阵, $\boldsymbol{x}$ 为系统的真实状态. 当系统无量测噪声时, 有 $\boldsymbol{x} = \hat{\boldsymbol{x}} = (C^{\mathrm{T}}R^{-1}C)^{-1}C^{\mathrm{T}}R^{-1}\boldsymbol{y}$; 当系统存在量测噪声时, 需要用系统的状态估计值来代替真实状态值. 与直流模型类似, 可得交流模型时拓扑攻击为

$$\boldsymbol{a}(\boldsymbol{y}) = \bar{c}(\boldsymbol{x}) - c(\boldsymbol{x}) \tag{33}$$

式中, $c(\cdot)$ 为系统真实潮流函数; $\bar{c}(\cdot)$ 为攻击者的目标潮流函数.

2010 年, Sandberg 等研究了错误数据注入攻击下安全索引的概念, 即以最小的攻击代价执行隐

蔽攻击[25−26]. 直流条件下, 当攻击者篡改节点数最小时, 安全索引可表示为

$$\alpha_i = \min_{C} \|C\dot{\boldsymbol{x}}\|_0, \quad \text{s.t. } a_i = C_i\dot{\boldsymbol{x}} = 1 \tag{34}$$

a_i 表示 $\boldsymbol{a}$ 的第 i 个元素; C_i 为 C 的第 i 行; $\|C\dot{\boldsymbol{x}}\|_0$ 为攻击向量 $\boldsymbol{a}$ 中非零元的个数, 即 α_i 代表了执行隐蔽攻击时最少的攻击节点数. 通常情况下, 0-范数的求解比较困难, 同时 0-范数最优问题与 1-范数最优问题等价, 因此问题通常会转化为更容易求解的 1-范数问题:

$$\beta_i = \min_{C} \|C\dot{\boldsymbol{x}}\|_1, \quad \text{s.t. } a_i = C_i\dot{\boldsymbol{x}} = 1 \tag{35}$$

Teixeira 等进一步将安全索引由 0-范数、1-范数推广至 p-范数问题[27,28], 即

$$\gamma_i = \min_{C} \|C\dot{\boldsymbol{x}}\|_p, \quad \text{s.t. } a_i = C_i\dot{\boldsymbol{x}} = 1 \tag{36}$$

3.2.2　信息受限条件下空间隐蔽型攻击

由于电力系统的拓扑结构、线路参数、实时状态难以获得, 因此如何在信息受限情况下构建 FDI 成为众多研究者关注的热点[29−31]. Liu 等研究了攻击者拥有全部拓扑信息以及部分线路参数时, 直流模型下 FDI 构建方法[30]. 如图 5 所示, 电力系统被分为攻击区域 $Q1$ 与非攻击区域 $Q2$, 攻击者知道攻击区域 $Q1$ 以及区域连接线的线路参数, 但不知道非攻击区域 $Q2$ 的线路参数. 攻击者可以仅修改区域连接线的潮流量测值使得攻击保持隐蔽. 用 Ω_{Q1}、Ω_{Q2} 分别表示 $Q1$、$Q2$ 中节点集合, 用 Ω_{T12} 表示 $Q1$ 与 $Q2$ 连接线集合 (方向为从 $Q1$ 至 $Q2$), 则攻击向量为

$$\begin{aligned} &\boldsymbol{a} = C\dot{\boldsymbol{x}} \\ &\text{s.t. } \dot{x}_i = \Delta\dot{x}, \forall i \in \Omega_{Q1} \\ &\dot{x}_i = 0, \forall i \in \Omega_{Q2} \end{aligned} \tag{37}$$

$\Delta\dot{x} \in \mathbf{R}$ 为任意常数. 可以证明上式等价于

$$\begin{aligned} &a_{ij} = -b_{ij}\Delta\dot{x}, \forall (i,j) \in \Omega_{T12} \\ &a_{ij} = 0, \forall (i,j) \notin \Omega_{T12} \end{aligned} \tag{38}$$

b_{ij} 为线路 (i,j) (即连接节点 i 与节点 j 的传输线) 的电纳, a_{ij} 为 $\boldsymbol{a}$ 中对线路 (i,j) 量测值的攻击量. 从上式可以看出, 攻击者可以在仅了解区域连接线电纳的情况下进行隐蔽的错误数据注入攻击. 类似的, 攻击者可以仅影响攻击区域 $Q1$ 的系统状态而不影响非攻击区域 $Q2$ 的系统状态, 其攻击向量为

$$\begin{aligned} &\boldsymbol{a} = C\dot{\boldsymbol{x}} \\ &\text{s.t. } \dot{x}_i = 0, \forall i \in \Omega_{Q2} \end{aligned} \tag{39}$$

式中, 约束 $\dot{x}_i = 0, \forall i \in \Omega_{Q2}$ 保证了非攻击区域 $Q2$ 的系统状态不发生变化, 而攻击区域 $Q1$ 中系统状态可以任意变化. 用 Ω_{T2} 表示非攻击区域 $Q2$ 内全部传输线集合, 则可以证明上式等价于

$$\begin{aligned} &a_{ij} = -b_{ij}(\dot{x}_i - \dot{x}_j), \quad \forall (i,j) \notin \Omega_{T2} \\ &a_{ij} = 0, \quad \forall (i,j) \in \Omega_{T2} \end{aligned} \tag{40}$$

攻击者可以在不知道区域 $Q2$ 传输线参数的情况下进行隐蔽攻击.

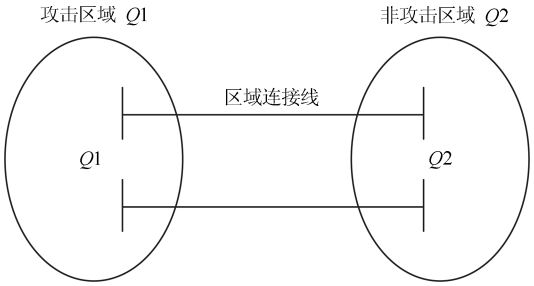


图 5　攻击区域与非攻击区域

Fig. 5　Attack area and non-attack area

Liu 等将信息受限时的隐蔽攻击拓展至交流模型[31]. 在交流模型下, 系统状态为电压幅值与电压相角. 将系统状态重表示为 $\langle \boldsymbol{v}, \boldsymbol{\theta} \rangle$, 潮流模型表示为 $C(\boldsymbol{v}, \boldsymbol{\theta})$. 交流模型下的攻击场景与直流模型类似, 场景如图 5 所示, 攻击者知道攻击区域 $Q1$ 以及区域连接线的线路参数, 而不知道非攻击区域 $Q2$ 的线路参数. 首先, 攻击者可以仅修改区域连接线的潮流量测值使得攻击保持隐蔽, 其攻击向量为

$$\begin{aligned} &\boldsymbol{a} = C(\hat{\boldsymbol{v}}, \hat{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}) - C(\hat{\boldsymbol{v}}, \hat{\boldsymbol{\theta}}) \\ &\text{s.t. } \Delta\theta_i = \Delta\dot{x}, \forall i \in \Omega_{Q1} \\ &\Delta\theta_i = 0, \forall i \in \Omega_{Q2} \end{aligned} \tag{41}$$

$\hat{\boldsymbol{v}}$、$\hat{\boldsymbol{\theta}}$ 为系统状态估计量; $\Delta\boldsymbol{\theta}$ 为攻击者对相角的修改量; $\Delta\theta_i$ 为 $\Delta\boldsymbol{\theta}$ 的第 i 个元素; $\Delta\dot{x} \in \mathbf{R}$ 为任意常数.

其次, 攻击者可以仅影响攻击区域 $Q1$ 的系统状态而不影响非攻击区域 $Q2$ 的系统状态, 其攻击向量为

$$\begin{aligned} &\boldsymbol{a} = c(\hat{\boldsymbol{v}} + \Delta\boldsymbol{v}, \hat{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}) - c(\hat{\boldsymbol{v}}, \hat{\boldsymbol{\theta}}) \\ &\text{s.t. } \Delta v_i = \Delta\theta_i = 0, \forall i \in \Omega_{Q2} \end{aligned} \tag{42}$$

$\Delta\boldsymbol{v}$ 为攻击者对电压幅值的修改量; Δv_i 为 $\Delta\boldsymbol{v}$ 的第 i 个元素.

Kim 等分析了攻击者信息受限时拓扑攻击方法, 该方法适用于直流与交流模型[24, 32]. 以交流模型为例进行说明, 如图 6 所示, 左图显示了攻击前系统状况, 右图显示了攻击者的目标, 即希望控制中心误认为线路 (i,j) 已经断开. 图中, p_{ij} 与 q_{ij} 分别代表了线路 (i,j) 在节点 i 的有功功率与无功功率量

测值, p_i 与 q_i 代表了节点 i 的注入功率. 攻击方法为

$$\begin{aligned}&\bar{p}_{ij}=\bar{q}_{ij}=0,\bar{p}_i=p_i-p_{ij},\bar{q}_i=q_i-q_{ij}\\&\bar{p}_{ji}=\bar{q}_{ji}=0,\bar{p}_j=p_j-p_{ji},\bar{q}_j=q_j-q_{ji}\end{aligned}\tag{43}$$

参数上短线代表攻击后的量测值, 如 $\bar{p}_{ij}$ 代表攻击后 p_{ij} 的数值. 值得注意的是, 攻击者进行上述攻击并不需知道系统任何线路参数. 可以证明, 在系统无噪声时, 攻击是隐蔽的; 而系统存在噪声时, 攻击能几乎保持隐蔽. 在直流模型下, 攻击方法类似, 只需忽略上式中所有无功功率相关部分 (因为直流模型下全电网无功功率为 0).

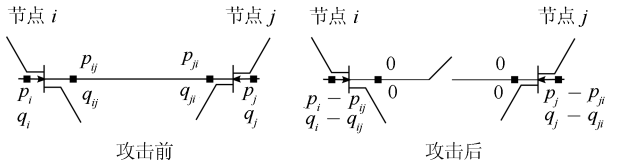


图 6 信息受限下的拓扑攻击

Fig. 6 Topology attacks with limited information

2015 年, Yu 等分析了攻击者在完全不知道系统拓扑与线路参数信息时隐蔽错误数据注入攻击的构建方法, 该方法被称为盲错误数据注入攻击[33]. 盲错误数据注入攻击要求攻击者具有足够时间段内电力系统全部量测值, 其核心思想为根据历史量测数据构建出与系统真实量测矩阵等价 (即矩阵列向量张成的子空间相同) 的估计量测矩阵, 他们采用的方法为主成分分析法 (PCA). 在直流模型下, 攻击者首先收集一定量的历史量测数据, 记为 $\boldsymbol{y}$, 并对 $\boldsymbol{y}$ 进行 PCA 提取:

$$[C_{\rm PCA},\boldsymbol{x}_{\rm PCA}]={\rm PCA}(\boldsymbol{y},n)\tag{44}$$

$C_{\rm PCA}\in\mathbf{R}^{m\times n}$ 为 $\boldsymbol{y}$ 的主成分矩阵; $\boldsymbol{x}_{\rm PCA}\in\mathbf{R}^n$ 且 $\boldsymbol{y}\approx C_{\rm PCA}\boldsymbol{x}_{\rm PCA}$. PCA 的具体过程可参见文献 [34]. 若系统无噪声, 则 $\boldsymbol{y}=C_{\rm PCA}\boldsymbol{x}_{\rm PCA}$. 可证明 $\boldsymbol{x}=P_x\boldsymbol{x}_{\rm PCA}$, 式中, $P_x=C^+C_{\rm PCA}$. 因此, $C_{\rm PCA}=CP_x$. 攻击者可根据 $C_{\rm PCA}$ 构造隐蔽攻击, 即

$$\boldsymbol{a}=C_{\rm PCA}\dot{\boldsymbol{x}}\tag{45}$$

此外, 当系统有噪声时, 可证明攻击是几乎隐蔽的.

2017 年, Markwood 等分析了当攻击者拥有历史节点注入有功功率与节点电压相角时, 对系统矩阵的估计方法, 该方法被称为拓扑泄露攻击[34]. 用 $Y\in\mathbf{R}^{m\times n}$ 表示历史节点注入有功功率, 用 $X\in\mathbf{R}^{n\times n}$ 表示历史节点电压相角, 则

$$Y=CX+W\tag{46}$$

式中, $W\in\mathbf{R}^{m\times n}$ 为历史量测噪声. 因此, 可得

$$C\approx YX^{-1}\tag{47}$$

3.3 时间隐蔽型错误数据注入攻击

现有时间隐蔽型攻击包括零动态攻击、局部零动态攻击和零状态诱导攻击, 本节对其攻击原理和相应权限逐一进行分析.

3.3.1 零动态攻击

2016 年, Chen 等研究了零动态攻击, 攻击者可以通过仅修改控制量保持隐蔽, 同时攻击量的构造与系统的在线知识 (如系统控制量、系统量测量或控制算法) 无关, 即攻击可以完全离线构造, 因此被命名为零动态攻击[35]. 零动态攻击模型如下:

$$\begin{aligned}&\boldsymbol{a}_u[k]=\lambda^k\boldsymbol{g}\\&\boldsymbol{a}_y[k]=0,\forall k\end{aligned}\tag{48}$$

$\boldsymbol{g}\neq0$; $\lambda\in\mathbf{C}$ ($\mathbf{C}$ 表示复数集合) 满足

$$\begin{bmatrix}\lambda I-A & -B\\ C & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}'[0]\\ \boldsymbol{g}\end{bmatrix}=0$$

$\boldsymbol{x}'[0]$ 为攻击者伪造的系统初始状态偏移, 且 $\boldsymbol{x}'[0]\in\mathbf{R}^n$. 注意攻击者仅需要知晓系统的动态模型 (即矩阵 A、B 与 C) 发起攻击. 当矩阵 B 为单射 (即矩阵 B 的核空间为 $\{0\}$ 或 ${\rm rank}(B)=l$) 时, 由于 $\boldsymbol{g}\neq0$, 可得 $\boldsymbol{x}'[0]\neq0$. 这意味着, 当防御者知晓系统初始状态时, 零动态攻击不存在非零解. 因此, 零动态攻击通常需要在系统开始运行时即发起攻击.

3.3.2 局部零动态攻击

Teixeira 等研究了局部零动态攻击, 该场景下攻击者仅拥有部分系统的信息[36]. 具体来说, 系统动态方程可表示为

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}_1[k+1]\\ \boldsymbol{x}_2[k+1]\end{bmatrix}=&\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1[k]\\ \boldsymbol{x}_2[k]\end{bmatrix}+\\&\begin{bmatrix}B_1\\ 0\end{bmatrix}\boldsymbol{u}[k]+\boldsymbol{w}_x[k]\\ \boldsymbol{y}[k]=&\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1[k]\\ \boldsymbol{x}_2[k]\end{bmatrix}+\boldsymbol{w}_y[k]\end{aligned}\tag{49}$$

式中, $\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}$; $A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}$; $B=\begin{bmatrix}B_1\\ 0\end{bmatrix}$; $C=\begin{bmatrix}C_1 & C_2\end{bmatrix}$. 假设攻击者仅知晓 A_{11}、A_{21}、B_1 以及 C_1, 攻击者可根据知晓部分的系统信息构造攻

击如下:

$$\begin{aligned}&\boldsymbol{a}_u[k]=\lambda^k\boldsymbol{g}\\&\boldsymbol{a}_y[k]=0,\forall k\end{aligned}\tag{50}$$

式中, $\boldsymbol{g}\neq 0$; $\lambda\in\mathbf{C}$ 满足

$$\begin{bmatrix}\lambda I-A_{11} & -B_1\\ C_1 & 0\\ A_{21} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1'[0]\\ \boldsymbol{g}_1\end{bmatrix}=0$$

3.3.3　零状态诱导攻击

Chen 等提出了零状态诱导攻击的概念, 并分析了该攻击与零动态攻击的区别. 相比于零动态攻击, 零状态诱导攻击中攻击者不诱导控制中心对系统初始状态产生错误认知[35]. 为了保证攻击隐蔽, 需要保证攻击导致的系统实际状态偏移无法被传感器观测. 本文中, 将零状态诱导攻击建模如下:

$$\begin{aligned}&\boldsymbol{e}[k+1]=A\boldsymbol{e}[k]+B\boldsymbol{a}_u[k],\boldsymbol{e}[0]=0\\&\quad -C\boldsymbol{e}[k]=0\\&\boldsymbol{a}_y[k]=0,\forall k\end{aligned}\tag{51}$$

式中, $\boldsymbol{e}[k]=\boldsymbol{x}[k]-\boldsymbol{x}_a[k]$ 表示攻击导致的系统实际状态偏移; $\boldsymbol{e}[0]=0$ 表示针对系统初始状态进行攻击. 由于 $C\boldsymbol{e}[k]=0$, 若攻击者要发起可影响系统的攻击 (即 $\boldsymbol{e}[k]\neq 0$), 则有 C 的核空间非零, 即 $\mathrm{rank}(C)<n$.

从攻击者先验知识、读写权限与攻击隐蔽性对三类时间隐蔽型攻击进行总结对比, 如表 1 所示.

3.4　空间 – 时间隐蔽型攻击

3.4.1　Stuxnet 型重放攻击

Stuxnet 通过篡改控制指令破坏物理系统, 同时监听和重放系统正常状态下的量测数据, 使得控制中心无法感知物理系统的异常状态[11−12]. 本文将该类攻击命名为 Stuxnet 型重放攻击. Stuxnet 型重放攻击对被攻击系统和攻击手段有限定要求: 被攻击系统存在确定的稳定状态且已发生攻击时工作在稳点状态, 攻击者需要重放全部的量测信号.

为了简化分析, 假设系统从 t 时刻系统状态已经稳定, 系统中无噪声 ($\boldsymbol{w}_x=0$ 且 $\boldsymbol{w}_y=0$), 重点讨论 Stuxnet 型重放攻击的隐蔽性. 由于从 t 时刻系统状态已经稳定, 则有: 1) $\boldsymbol{x}[s]=\boldsymbol{x}[t],\forall s\geq t$; 2) $\boldsymbol{u}[s]=\boldsymbol{u}[t],\forall s\geq t$; 3) $\boldsymbol{y}[s]=\boldsymbol{y}[t],\forall s\geq t$. 注意到若控制中心接收到的量测信号不发生变化, 则会认为系统无攻击. 因此攻击者可以注入任意的控制信号, 并重放事先收集的无攻击时的历史量测信号来保持隐蔽, 具体来说:

$$\begin{aligned}&\boldsymbol{a}_u[k]\in\mathbf{R}^l\\&\boldsymbol{a}_y[k]=\boldsymbol{y}[s]-\boldsymbol{y}_a[k],\quad \forall k\geq k_0\end{aligned}\tag{52}$$

3.4.2　量测量无关隐蔽攻击

2015 年, Vu 和 Weerakkody 等[37−38] 提出了一种隐蔽攻击, 该攻击的构建可以不需要具体系统状态知识. 该攻击的实现原理为系统的线性性质, 攻击者模拟了任意无噪声的系统动态过程. 将该系统与原系统进行叠加, 并移除模拟系统的量测量, 则攻击后控制中心收到的量测量依然符合系统动态, 即攻击是隐蔽的[37−38]. 攻击模型如下:

$$\begin{aligned}&\boldsymbol{e}[k+1]=A\boldsymbol{e}[k]+B\boldsymbol{a}_u[k],\boldsymbol{a}_u[k]\in\mathbf{R}^l\\&\boldsymbol{a}_y[k]=-C\boldsymbol{e}[k],\forall k\end{aligned}\tag{53}$$

式中, $\boldsymbol{e}[k]=\boldsymbol{x}[k]-\boldsymbol{x}_a[k]+A^k\boldsymbol{e}[0]$ 代表了攻击导致系统真实状态的偏差 (即 $\boldsymbol{x}[k]-\boldsymbol{x}_a[k]$) 与攻击诱导的系统初始状态偏移 (即 $A^k\boldsymbol{e}[0]$) 之和.

该攻击的实现原理为系统的线性性质, 攻击者模拟了任意无噪声的系统动态过程. 将该系统与原系统进行叠加, 并移除模拟系统的量测量, 则攻击后控制中心收到的量测量依然符合系统动态, 即攻击是隐蔽的. 注意由于叠加的为无噪声系统, 攻击前后系统噪声将不发生变化. 由于系统噪声为异常检测器残差的主要决定因素, 噪声不变时系统攻击前后检测器残差也不变, 本文中将其命名为量测量无关隐蔽攻击.

3.4.3　系统模拟攻击

2014 年, Weerakkody 等提出攻击者可根据系统的动态模型以及控制中心决策的控制信号, 来模拟无攻击下系统的演变过程, 以此伪造符合系统动态的量测数据, 进而实施攻击[17, 39], 本文将其命名为系统模拟攻击. 该攻击可分为两种场景:

表 1　时间隐蔽型攻击对比

Table 1　Time concealment attack contrast

攻击类别	攻击者先验知识	读写权限	隐蔽性
零动态攻击	矩阵 A、B 与 C	大部分控制指令读权限与写权限	隐蔽
局部零动态攻击	A_{11}、A_{21}、B_1 以及 C_1	大部分控制指令读权限与写权限	隐蔽
零状态诱导攻击	矩阵 A、B 与 C	大部分控制指令读权限与写权限	隐蔽

攻击场景 1. 该场景下, 攻击者拥有所有控制指令的读权限. 攻击者可根据获取的控制指令进行系统模拟, 具体攻击模型如下

$$\begin{aligned}&\boldsymbol{x}'_a[k+1]=A\boldsymbol{x}'_a[k]+B\boldsymbol{u}_m[k]+\boldsymbol{w}'_x[k]\\&\boldsymbol{y}'_a[k]=C\boldsymbol{x}'_a[k]+\boldsymbol{w}'_y[k]\\&\boldsymbol{a}_u[k]\in\mathbf{R}^l\\&\boldsymbol{a}_y[k]=\boldsymbol{y}'_a[k]-\boldsymbol{y}_a[k],\forall k\end{aligned}\tag{54}$$

$\boldsymbol{x}'_a[k]$ 为攻击者模拟的无攻击时系统状态, 且 $\boldsymbol{x}'_a[0]$ 为攻击者设定的系统初始状态; $\boldsymbol{y}'_a[k]$ 为攻击者模拟的无攻击时系统量测值; $\boldsymbol{u}_m[k]$ 为控制中心实际设定的控制量; $\boldsymbol{w}'_x[k]$ 为攻击者任意选定的系统过程噪声, 且与 $\boldsymbol{w}_x[k]$ 具有相同分布; $\boldsymbol{w}'_y[k]$ 为攻击者任意选定的系统量测噪声, 且与 $\boldsymbol{w}_y[k]$ 具有相同分布.

注意当防御者不知晓系统真实的初始状态时, $\boldsymbol{x}'_a[0]$ 可以任意设定; 而当防御者知晓系统初始状态时, 为了保持攻击隐蔽, 攻击者需要尽可能将 $\boldsymbol{x}'_a[0]$ 与控制中心获取的系统初始状态一致. 一种简单方法为, 攻击者根据一定时间内的系统量测值, 自行对系统初始状态进行估计.

攻击场景 2. 该场景下, 攻击者仅有部分或没有任何控制指令的读权限. 注意到攻击者若知晓控制决策函数, 则可以通过所获取的量测量来估计系统的控制指令. 具体来说, 用 $\hat{\boldsymbol{u}}_m[k]=\begin{bmatrix}\boldsymbol{u}_{m1}[k]\\ \hat{\boldsymbol{u}}_{m2}[k]\end{bmatrix}$ 表示攻击者对控制指令的估计量, 其中, $\boldsymbol{u}_{m1}[k]\in\mathbf{R}^{l_1}$ 为攻击者可直接获取的控制指令部分 (即攻击者拥有此部分控制指令的读权限); $\hat{\boldsymbol{u}}_{m2}[k]\in\mathbf{R}^{l-l_1}$ 为攻击者估计的控制指令部分. 对这三种攻击从攻击者所需要的先验知识、读写权限、隐蔽性等方面总结如表 2.

3.5 非隐蔽型攻击

除了上述的攻击, 部分学者提出利用各种网络攻击技术对 CPS 进行攻击, 这类研究较少考虑 CPS 系统中对于系统状态的分析和异常检测, 因此无法躲避时间相关型和空间相关型的异常检测.

3.5.1 拒绝服务攻击

在拒绝服务攻击中, 攻击者通过向目标主机或终端中发送大量无用的数据包, 使得服务器的可用网络资源耗尽, 从而停止服务. 在 CPS 中, 拒绝服务攻击的目标为使得控制中心与执行器/传感器间通信无法正常进行. 因此, 当拒绝服务攻击发动成功时, 问题可等效为, 攻击者将控制中心发送的控制信号或将控制中心接收的量测信号修改为 0.

由于拒绝服务攻击的实现方式通常为阻塞通信信道, 因此不需要与系统的先验知识, 也不需要控制信号或量测信号的读写权限. 攻击者实施拒绝服务攻击时将造成量测数据缺失, 防御者可以快速感知到系统异常; 但由于控制中心即无法获得系统的量测信息, 也无法下达控制指令给受控系统, 因此在发生拒绝服务攻击时防御者没有更好的办法来保证系统的安全. 在 2015 年乌克兰电网攻击事件中, 攻击者在篡改控制指令的同时, 就利用拒绝服务攻击方式使得电网管理者不能及时感知电网状态并采取防御策略.

3.5.2 通用重放攻击

通用重放攻击是指攻击者收集一定时间的历史控制信号或量测信号, 并将其重放或延迟发送. 具体来说, 假设攻击者从 k_0 时刻开始进行重放攻击, 其通用攻击模型可表示为

$$\begin{aligned}&\boldsymbol{a}_u[k]=S_u[k](\boldsymbol{u}_m[k-\tau(k)]-\boldsymbol{u}_m[k])\\ \text{或}\quad&\boldsymbol{a}_y[k]=S_y[k](\boldsymbol{y}_a[k-\tau(k)]-\boldsymbol{y}_a[k]),\forall k\geq k_0\end{aligned}\tag{55}$$

$\mathbf{B}$ 为布尔对角矩阵, $S_u[k]\in\mathbf{B}^{l\times l}$ 且 $S_y[k]\in\mathbf{B}^{m\times m}$. $S_{u,ii}[k]$ 表示攻击者是否对第 i 个控制量发起重放攻击; $S_y[k]$ 表示攻击者是否对第 i 个量测量发起重放攻击; $\tau(k)\in\mathbf{N}$ 为重放攻击的时间延迟.

与第 3.4.1 节中 Stuxnet 型重放攻击相比, 通用重放攻击仅需要部分控制信号或部分量测信号的读写权限, 而且对被攻击对象的运行状态没有要求. 因此, 通用重放攻击的实施要求相对较低, 但不能保证攻击的隐蔽性, 比如控制中心预期系统的运行状态与攻击者重放的状态不相符, 则该攻击将被检测.

3.5.3 放大攻击

Sridhar 等提出了放大攻击, 攻击者将控制信号或量测信号按照正常信号的 λ_k 倍进行篡改[40]. 具体来说.

表 2 空间–时间隐蔽型攻击对比

Table 2 Space–time stealth attack contrast

攻击类别	攻击者先验知识	读写权限	隐蔽性
Stuxnet 重放攻击	不需要	部分控制指令写权限, 全部量测信号的读写权限	隐蔽
量测量无关攻击	矩阵 A、B 与 C	部分或全部控制指令读写权限, 全部量测信号的读写权限	隐蔽
系统模拟攻击	矩阵 A、B、C、Q 与 R, 系统初始状态 $\boldsymbol{x}[0]$	部分或没有控制指令的读写权限, 全部量测值的读写权限	隐蔽

$$\boldsymbol{a}_u[k] = \lambda_k S_u[k]\boldsymbol{u}_m[k]$$
$$\text{或}\quad \boldsymbol{a}_y[k] = \lambda_k S_y[k]\boldsymbol{y}_a[k], \forall k \tag{56}$$

式中, $\lambda_k \in \mathbf{R}$ 为放大系数; $S_u[k] \in \mathbf{B}^{l\times l}$ 与 $\mathbf{S}_y[k] \in \mathbf{B}^{m\times m}$ 为布尔对角矩阵, 分别表示攻击者是否对控制量或量测量发起放大攻击. 注意若 $\lambda_k = -1$, 则其攻击效果与拒绝服务攻击类似. 放大攻击是重放攻击的一种特例, 只需要部分控制信号或部分量测信号的读写权限. 此外, 由于其相当于只对信号进行放大, 因此不需要关于系统的先验知识. 这种攻击思路在针对电力系统的攻击中能快速地找到攻击的可行解, 但由于攻击过程和攻击思路较为简单, 攻击过程不能保证隐蔽, 当放大倍数较大或异常数据检测装置容错率较小时, 将被检测.

4　CPS 综合安全防御方法

在 CPS 安全防御方面, 网络安全专家将网络安全技术应用于各类物理系统, 以防御各类 CPS 攻击, 如通过信息加密以及相关的密钥分配与管理等技术, 对系统信息和隐私数据进行加密, 以抵御窃听和数据篡改等攻击; 物理系统工程安全专家通过改进物理系统工程安全防护机制提升 CPS 的安全防御能力, 如通过部署测量单元 (Phasor measurement unit, PMU) 等新型量测设备, 提升对 FDI 攻击的检测能力; 近年来, 越来越多研究机构和学者开始关注通过结合网络安全技术与物理系统工程安全理论方法, 实现 CPS 综合安全防御. 本文对现阶段的防御方法列举分类如图 7 所示.

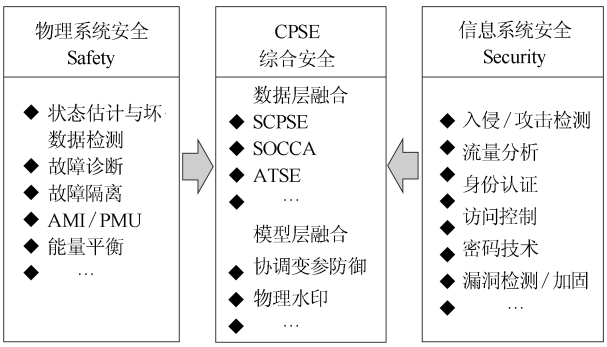


图 7　CPS 防御分类

Fig. 7　Taxonomy of CPS defense

4.1　信息系统网络安全防御方法

针对 CPS 面临的安全威胁, 研究者将网络安全技术应用于 CPS 安全防御. 一类方法是采用防篡改的通信系统、先进的认证协议、加密和非对称加密机制、访问控制等方法, 以阻断攻击者进入系统. Wang 等提出了智能电网中时间有效的一次性签名 (Time valid one-time signature, TV-OTS) 模式, 以改善传统一次性签名机制的效率[41]. 基于 TV-OTS 模式, 文中设计了多播认证方案 TV-HORS, 可实现数据的快速签名与认证, 并具有较高的丢包容错率以及攻击抵抗能力. Cao 等设计了一种基于哈希链技术的分层加密机制以保护系统中的敏感数据, 该机制同时提供了轻量级的密钥管理[42]. Saxena 等设计了 CPS 下的签名方案以保证通信网络中仅可传输合法指令, 该方案通过哈希方法对控制指令进行签名, 可抵御恶意指令攻击、重放攻击、伪装攻击等[43]. Kumar 等提供了智能家居环境中的匿名安全框架 (Anonymous secure framework, ASF), 该框架可提供有效的认证与密钥协商, 可保证家庭各设备的匿名性与不可链接性[44].

另一类方法是采用流量分析、入侵检测等方法, 以实现攻击的实时检测以及恶意行为的判断. Zhang 等在智能电网环境中设计了分布式入侵检测系统, 在智能电网家庭局域网 (HAN)、邻域网 (NAN) 以及广域网 (WAN) 环境中部署大量的分析模块, 并采用支持向量机 (Support vector machine, SVM) 以及人工免疫系统 (Artificial immune system, AIS) 以检测网络中的恶意数据以及各类攻击[45]. Baig 提供了基于异常流量的轻量级模式匹配方法, 可检测系统异常设备行为, 并具有较小的通信与存储开销[46]. Palácios 等提供了信息物理融合系统环境中两类入侵检测系统: 第一类入侵检测系统基于自组织映射网络 (Self-organizing map, SOM), 第二类基于 K-Means 算法. 这两类入侵检测系统可改善高度异构网络中的攻击检测速率、精度以及适应性, 并可方便地部署于 CPS 环境中[47].

然而, 上述方法应用于 CPS 的安全防御时遇到了诸多困难: 1) CPS 对信息的实时性有极高的要求, 而受限于 CPS 通信节点的计算与存储能力, 先进、复杂的加密认证机制等方法将会影响 CPS 的运行效率, 增加网络延时; 2) 流量分析与入侵检测可以发现网络中的异常行为, 而无法校验传输的控制指令以及量测信号的合法性与完整性; 3) CPS 对可用性与安全性的高标准, 现有技术很难同时达到异常检测时低漏报率与误报率需求.

4.2　物理系统工程安全防御方法

物理系统工程安全根据物理系统运行机制、量测数据在时间和空间上的关联等特性, 实现 CPS 的异常检测, 但如第 3 节中介绍, 攻击者可以通过设计特定攻击策略, 躲避现有工程安全机制. 对此, 研究者提出一系列改进的物理系统工程安全防御方法, 提升 CPS 的安全防御能力.

一类方法是防御者有策略地选择部分控制指令或量测信号进行保护, 使得攻击者无法构建完全隐蔽的攻击策略, 进而保证全部数据的完整性. 近年

来, 此类方法主要集中在对于电力系统空间相关型 FDI 攻击的检测中. Bobba 等提出了两类方法以检测 FDI 攻击: 一类方法是通过有策略性地选择一组受保护的传感器量测值使得 Liu 等[20] 提到的隐蔽 FDI 攻击不存在可行解; 另一类方法是验证一定数量的独立状态变量的完整性[48]. 在电力系统中, 随着相量测量单元的逐步应用, 通过 PMU 数据以辅助检测 FDI 攻击受到了广泛关注. Giani 等分析了在系统恰当地方部署 $p+1$ 个相量测量单元, 可使得 p 个不可约 FDI 攻击无效化[49]. Nuqui 和 Qi 等研究了 PMU 的优化部署问题, 设计使用最少的 PMU 实现攻击者无法构造有效的 FDI 攻击策略[50−51]. 由于不同变电站 PMU 同步需要借助 GPS 信号, 而 GPS 信号可能被篡改, 使得 PMU 的量测信号也可能受到攻击[52−54]. 对此, Jafarnia-Jahromi 等提出了电力系统中针对 GPS 欺骗攻击的相应策略[55−56]. Dán 等提出了贪婪算法以对 FDI 攻击提供完全的或部分的防御, 其中完全防御意味着 FDI 攻击不存在可行解[57].

对于大型系统, 防御者通常需要确保足够多数据 (控制指令、量测信号或辅助数据) 的可信性, 才可为系统提供完全的保护, 这需要防御者投入大量成本. 此外, 对于电力系统之外的其他信息物理融合系统, 如何将此类思想应用于 FDI 攻击检测仍然是一个开放性的研究问题.

另一类方法是研究更好的攻击检测算法, 以实现异常数据或攻击的快速精准识别, 减少系统噪声或其他干扰对检测性能的影响, 降低检测的漏报率和误报率. Huang 等使用自适应累积和 CUSUM 的方法, 在实现 FDI 攻击快速检测的同时, 可保持高检测精度与低误报率[58−59]. Vu 等将基于耗散理论的故障检测器应用于 FDI 攻击检测, 从理论上分析了故障检测器最少部署方案以完全检测控制信号或量测信号篡改攻击[37]. Liu 等将错误数据检验转化为矩阵分离问题, 并设计了基于正常与异常电力系统状态分离的新型检测器, 检测器的核心算法包括核范数最小化和低阶矩阵分解两类方法, 可以显著得提升检测器对 FDI 攻击的检测性能[60]. Liu 等使用有色排水网描述智能电表信息流以检测高级电表架构 (Advanced metering infrastructures, AMI) 下的 FDI 攻击[61]. 此外, 部分学者将不同时段、不同设备的数据结合起来, 提升 FDI 攻击的检测能力. Gu 等使用历史数据追踪量测值变化的动态, 利用 KL 散度 (Kullback-Leibler distance, KLD) 来度量历史数据概率分布与可疑量测值数据概率分布的差异, 以判断 FDI 攻击的存在[62]. Ashok 等提出了在线异常检测算法, 通过电力系统负载预测值、发电机调度情况以及同步发电机相角等数据对电力系统实际情况进行预测, 通过预测值与实际量测值的比较来检测 FDI 攻击[63].

上述方法大都是利用现有 FDI 攻击中未考虑的因素, 设计针对性的数据一致性检测方法, 本质上是假设攻击者不能获取特定资源或检测方法. 近年来, 研究人员提出一系列 CPS 攻击方法, 通过设定针对性的攻击构造条件, 以躲避新提出的检测技术. 因此, 防御方法对特定资源的依赖程度越低, 对攻防双方能力的假定越均衡, 就越能使用实际 CPS 安全防御的需求.

4.3 基于信息物理融合的 CPS 综合安全防御

从前两节的分析可以看出, 从信息系统网络安全或者物理系统工程安全单一角度, 都难以全面描述、检测和防御 CPS 攻击. 安全研究者希望结合信息系统网络安全和物理系统工程安全方法, 设计面型 CPS 的新型攻击检测架构. 已有的工作可分为两类: 基于信息与物理数据融合的检测和基于信息与物理模型融合的检测.

4.3.1 基于数据融合的 CPS 综合安全防御

信息层面的融合方法为将信息系统数据 (例如网络异常流量) 与物理系统数据进行关联以检测 CPS 异常事件和攻击行为. Zonouz 等提出了安全导向的信息物理关联状态估计 (Security-oriented cyber-physical state estimation, SCPSE) 以识别信息网络中恶意主机以及物理网络中可疑传感器集合, SCPSE 根据信息网络异常数据构建攻击图, 通过在状态估计中移除可疑节点的量测值, 以检测系统中存在的不良数据, 并提供对系统真实状态的可靠估计[64]. Zonouz 等还提出了一种信息物理安全评估技术 (Security-oriented cyber-physical contingency analysis, SOCCA) 以分析电力系统意外事故, SOCCA 提供了信息物理融合系统中信息元素与物理元素统一的形式化描述方法, 根据信息物理意外事故的威胁程度来评估意外事件可能产生的影响[65]. Wang 等建立了信息物理融合系统 SCADA 信息流的状态转移图, 提出了基于关系流程图的入侵检测思想以检测错误数据注入攻击[66].

除了直接关联分析信息系统和物理系统的数据, 研究者还利用信息系统的异常检测结果引导物理系统异常检测, 提升 CPS 安全防御能力. Liu 等提出了一种电力系统异常数据检测结果和信息网络异常通信检测结果相融合的方法, 报警数据融合是基于物理系统拓扑与信息网络的关联性, 将信息系统报警与物理系统的标准化残差检测的结果进行综合评估. 通过数据融合将实际系统安全进行融合, 得到系统层面的分析结果, 从而消除了单纯物理层检测方法或单纯信息层检测方法的偏差, 可以有效降低信

息系统与物理系统检测方法的漏报率与误报率[67]. Liu 等提出了一种基于状态估计的智能电网动态加密和认证方法, 对智能电网中的控制中心和远程终端装置之间的的通信进行保护. 该方法对两种信号都生成加密密钥, 随着电力系统的变化, 每个远程终端都会定期更新其密钥, 而控制中心也会动态同步地估计所有远程终端的新密钥[68].

然而, 网络安全技术主要针对程序行为、身份权限、网络流量等信息系统对象的异常检测, 工程安全技术主要针对物理系统运行状态的异常检测, 两者在判别规则、报警格式、报警内容等方面都存在显著差异. 基于信息和物理系统数据融合的检测思路在实际应用中存在诸多挑战, 包括信息和物理异构数据的同构化问题、多源数据的关联分析问题、不同安全需求的关联建模问题. 因此, 基于信息和物理系统数据融合的检测方法仍处于探索研究中.

4.3.2 基于系统模型融合的 CPS 综合安全防御

模型层面的融合的含义是将一些网络安全防御的思想和技术应用于物理系统的安全监控, 构建出新的 CPS 综合安全防御方法. 本文重点介绍两类代表性方法: 协调变参防御 (Coordinate parameter variation defense, CPVD) 和物理系统水印.

协调变参防御源自网络安全技术中的移动目标防御 (Moving target defense, MTD). 2010 年美国在 “网络与信息技术研发计划”(Networking and information technology research and development) 中提出了 MTD, 通过 “改变游戏规则” 的革命性技术手段以积极主动地实施网络防御[69]. 移动目标防御的核心思想是, 通过可管理的方式持续不断地改变系统, 以减少系统攻击面, 增加系统的不确定性, 从而降低攻击者的攻击成功率, 增加攻击者探测系统漏洞或实施攻击的代价[69].

移动目标防御最初用于网络安全防御, 可分为软件变换机制[70]、动态平台技术[71]、网络地址更换[72] 等. 国内学者进一步提出了网络空间拟态安全防御[73]. 近年来, 学者们将移动目标防御或拟态安全防御这些网络空间防御的思想应用于信息物理融合系统安全防御与攻击检测中. 值得注意的是, 网络空间防御中的移动目标防御或拟态安全防御更强调信息层面或逻辑层面上的变换; 而 CPS 中的 “移动目标防御” 则要求系统的物理环节 (包括控制环节、量测环节甚至是物理设备本身) 可受调节或变换. 因此, 本文中将 CPS 中的 MTD 命名为协调变参防御.

Morrow 等提出了主动改变电力系统配置参数的方法, 以检测空间相关型错误数据注入攻击[74−76]. 近 20 年里, 通过调整电力系统配置以保持电力系统潮流稳定运行已经受到广泛研究[77−80]. 新兴的分布式柔性交流输电 (Distributed flexible ac transmission system, D-FACTS) 设备可以改变电力系统线路阻抗值, 由于其较低的部署成本可更广泛部署于电力系统中, 以此提高系统操作者对电力系统配置的调节能力. 由于这项新增的调节能力, 该方法可用在智能电网阻碍攻击者实施隐蔽 FDI 攻击. 具体来说, Morrow 等提出了事后型 CPVD 方法以检测正在进行中的 FDI 攻击[74−75]. 防御者主动改变线路阻抗并观测系统潮流改变量, 若攻击存在, 则对系统配置采用已知改变时, 观测的潮流改变量将与期望改变量不同. 然而, 该方法对于攻击的检测与防御相对滞后, 无法及时有效地发现并阻止攻击的进行.

Rahman 等提出了事前型协调变参防御方法, 防御者通过随机选取一组传输线并随机改变传输线阻抗, 以及随机选取用于状态估计的传输线集合, 可使得攻击者无法获取最新的系统信息, 从而阻碍 FDI 攻击的执行[76]. 本文将此类 CPVD 方法称为任意型协调变参防御方法. Teixeira 等将协调变参防御概念应用于零动态攻击的检测, 并分别给出了当仅改变物理系统、控制环节与量测环节时所有零动态攻击可被检测的充分必要条件[81−82]. Gyugyi 等提出了通过可切换的混合控制器, 防御者可在多个子控制环节中相互切换以防御各种类型的攻击[77]. Hoehn 等在系统控制环节之前加入可控调制环节以改变系统控制过程的规律, 以此来检测零动态攻击与隐蔽攻击[78]. Weerakkody 等提出在系统原始结构不变的情况下额外增加动态物理设备, 可在实现协调变参防御的同时使得原始系统正常操作保持不变[83], 并进一步将该方法应用于 CPS 中攻击的鉴别问题, 并给出了相应的系统参数配置策略以鉴别可疑传感器量测值[84]. 然而这两种方法方法可能增加系统的操作与运行开销, 并可能会给系统带来新的风险 (例如不稳定性).

Tian 等提出了攻击者可以通过窃听到的电网系统量测值, 来检测系统是否采用 CPVD, 并将这种威胁被命名为参数确定优先的错误数据注入 (PCF-FDI)[85]. 在这种情况下现有的 CPVD 策略被此方法轻易检测到, 而攻击者发现系统采用 CPVD 可放弃原先的攻击计划, 等待获取最新的电网参数配置信息后再发起攻击. 为了改善 CPVD 的隐蔽性, Tian 等提出了隐蔽型 CPVD 策略, 该策略不会被 PCF-FDI 攻击检测[86]. 具体来讲, 当隐蔽型 CPVD 策略激活时, 攻击者基于原始系统结构的错误数据检验不会产生报警. 同时, 攻击者基于原始系统结构构造的错误数据注入攻击, 将以大概率被 CPVD 策略下的错误数据检验检测到, 从而被防御者丢弃或重定向至蜜罐等监控系统, 为进一步分析攻击行

为和攻击取证提供支持. 由此可见, 隐蔽型 CPVD 能引导攻击者进行无效攻击并增加其暴露的可能性. 在隐蔽性防御的实现问题上, Tian 等定义了潮流不变型 CPVD, 并证明了隐蔽型 CPVD 和潮流不变型 CPVD 的等价性, 将抽象的隐蔽性转化为可以明确达到的控制潮流的现实目标, 为隐蔽型 CPVD 策略的计算提供了理论依据.

物理系统水印来自电子产品的知识产权保护. 水印的概念最初被应用于数字图像、音频、视频等媒体产品以实现产权保护并校验信息完整性, 具有鲁棒性、不可见性、安全性、可证明性等特征[87]. 近年来, 研究者将水印法应用于 CPS 攻击检测中[15, 39, 88−92]. 具体来说, 防御者在控制信号中主动添加可控和保密的激励信号, 该信号随着物理系统的执行, 将影响系统的状态和量测值. 若攻击者无法获取水印信号, 则攻击者将难以在伪造的量测数据中包含由水印信号造成的系统状态和量测值变化. Mo 等首次提出了 CPS 中水印的概念, 并将其应用于重放攻击的检验中[89]. Mo 和 Weerakkody 等将物理系统水印应用于检测更一般性的信息攻击[89−92]. Satchidanandan 等将物理系统水印拓展为动态水印, 在假设控制信号不被篡改的前提下, 通过 L 组关联检测器 (L 为系统控制信号数, 每组包含两个检测器) 可保证仅有零均值异变能量的攻击可绕过该检测器[39]. 然而, 由于物理系统水印需要在系统控制量中添加扰动信号, 扰动信号的引入会影响系统的性能; 而且, 物理系统水印的检测精度与水印信号的强度相关. 这意味着防御者需要以牺牲系统性能为代价, 提高攻击的检测性能.

5 结论与展望

随着 CPS 近年来的不断发展, 研制具有高实时性、自动化和智能化的 CPS 逐渐成为了各个行业的共性需求. 各类 CPS 应用传感、嵌入式处理、数字通信、人工智能等信息网络技术, 在提升现有系统性能和效率的同时, 信息网络中的各种漏洞和脆弱性已经严重影响到 CPS 的安全运行, 并造成了巨大的损失. 近年来的攻击事件显示, 攻击者正在针对物理系统的业务流程和安全防护机制, 设计出隐蔽性更好和破坏力更强的 CPS 攻击策略, 并结合网络攻击技术对智能电网、智能交通、智慧医疗等国家基础设施, 以及可穿戴设备、自动驾驶、智能家居等个人安全系统进行攻击, 严重威胁国防、政治、经济和人民日常生活.

本文主要从信息安全与工程安全融合的角度对 CPS 面临的综合安全威胁进行了探讨. 随着人−机−物的深度融合, CPS 综合安全除了信息安全与物理系统工程安全的结合, 还包括信息安全与人类社会安全问题的结合, 如网络舆情安全、身份安全等. 人的不确定因素将进一步提高 CPS 综合安全的复杂性, 值得研究人员更深入的探讨.

有效提升 CPS 综合防御能力已成为了当务之急. 本文通过分析国内外防御方法研究发现, 仅仅依靠物理系统的工程安全防护机制和网络系统的信息安全技术, 都难以达到 CPS 综合安全的要求. 而两类方法由于数据来源、建模方法、技术思路存在较大差异, 如何有效结合已经成为 CPS 综合防御的难题. 本文分析探讨了现有的 CPS 综合防御技术, 并从数据融合、系统模型融合和防御思想相互借鉴等角度进行分类总结, 为未来 CPS 综合安全防御方法的研究提供了参考.

References

1 National Institute of Standards and Technology, *Framework for Cyber-Physical Systems*. Cyber Physical Systems PWG, 2015.

2 State Council of the People′s Republic of China. *Made in China 2025 Strategy*. Beijing: State Council of the PRC, 2015.
(中华人民共和国国务院. 中国制造 2025. 北京: 中华人民共和国国务院, 2015)

3 Guan Xiao-Hong, Zhao Qian-Chuan, Jia Qing-Shan, Wu Jiang, Liu Ting. *Cyber-Physical Energy System*. Beijing: Science Press, 2016.
(管晓宏, 赵千川, 贾庆山, 吴江, 刘烃. 信息物理融合能源系统. 北京: 科学出版社, 2016.)

4 Marburger J H, Kvamme E F, Scalise G, Feed D A. Leadership under Challenge: Information Technology R&D in A Competitive World. An Assessment of the Federal Networking and Information Technology R&D Program. Washington, DC, USA: PCAST, 2007.

5 Khaitan S K, McCalley J D. Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal*, 2015, **9**(2): 350−365

6 Kagermann H, Wahlster W, Helbig J. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group. Germany: Forschungsunion, 2013.

7 Wang Zhong-Jie, Xie Lu-Lu. Cyber-physical systems: a survey. *Acta Automatization Sinica*, 2011, **37**(10): 1157−1166
(王中杰, 谢璐璐. 信息物理融合系统研究综述. 自动化学报, 2011, **37**(10): 1157−1166)

8 China Cyber-Physical System Development Forum. Cyber-Physical System White Paper[Online], available: https://www.innovation4.cn/library/r14012, May 25, 2018
(中国信息物理系统发展论坛. 信息物理系统白皮书 [Online]. 获取自: https://www.innovation4.cn/library/r14012, 2018 年 5 月 25 日)

9 Shakarian P. Stuxnet: cyberwar revolution in military affairs. 2011, DTIC Document.

10 Chen T M. Stuxnet, the real start of cyber warfare? [Editor′s Note]. *IEEE Network*, 2010, **24**(6): 2−3

11 Falliere N, Murchu L O, Chien E. *W32.Stuxnet Dossier (Version 1.4)*. Symantec Security Response, 2011.

12 Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, et al. A review of cyber security risk assessment methods for SCADA systems. *Computers & Security*, 2016, **56**: 1−27

13 Tang Yi, Chen Qian, Li Meng-Ya, Wang Qi, Ni Ming, Liang Yun. Overview on cyber-attacks against cyber physical power system. *Automation of Electric Power Systems*, 2016, **40**(17): 59−69
(汤奕, 陈倩, 李梦雅, 王琦, 倪明, 梁云. 电力信息物理融合系统环境中的网络攻击研究综述. 电力系统自动化, 2016, **40**(17): 59−69)

14 Staggs J, Ferlemann D, Shenoi S. Wind farm security: attack surface, targets, scenarios and mitigation. *International Journal of Critical Infrastructure Protection*, 2017, **17**: 3−14

15 Mo Y L, Sinopoli B. Secure control against replay attacks. In: Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing. Monticello, USA: IEEE, 2009. 911−918

16 Lakshminarayana S, Teng T Z, Yau D K Y, Tan R. Optimal attack against cyber-physical control systems with reactive attack mitigation. In: Proceedings of the 8th ACM International Conference on Future Energy Systems. Hong Kong, China: ACM, 2017. 179−190

17 Weerakkody S, Mo Y L, Sinopoli B. Detecting integrity attacks on control systems using robust physical watermarking. In: Proceedings of the 53rd Annual Conference on Decision and Control (CDC). Los Angeles, USA: IEEE, 2014. 3757−3764

18 Simon D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications*, 2010, **4**(8): 1303−1318

19 Liang G Q, Weller S R, Zhao J H, Luo F J, Dong Z Y. The 2015 Ukraine blackout: implications for false data injection attacks. *IEEE Transactions on Power Systems*, 2017, **32**(4): 3317−3318

20 Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids. In: Proceedings of the 16th ACM Conference on Computer and Communications Security. Chicago, USA: ACM, 2009. 21−32

21 Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 2011, **14**(1): Article No. 13

22 Hug G, Giampapa J A. Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks. *IEEE Transactions on Smart Grid*, 2012, **3**(3): 1362−1370

23 Yuan Y L, Li Z Y, Ren K. Modeling load redistribution attacks in power systems. *IEEE Transactions on Smart Grid*, 2011, **2**(2): 382−390

24 Kim J, Tong L. On topology attack of a smart grid: undetectable attacks and countermeasures. *IEEE Journal on Selected Areas in Communications*, 2013, **31**(7): 1294−1305

25 Sandberg H, Teixeira A, Johansson K H. On security indices for state estimators in power networks. In: Proceedings of the 1st Workshop on Secure Control Systems. Stockholm, Sweden: CPSWEEK, 2010.

26 Hendrickx J M, Johansson K H, Jungers R M, Sandberg H, Sou K C. Efficient computations of a security index for false data attacks in power networks. *IEEE Transactions on Automatic Control*, 2014, **59**(12): 3194−3208

27 Teixeira A, Amin S, Sandberg H, Johansson K H, Sastry S S. Cyber security analysis of state estimators in electric power systems. In: Proceedings of the 49th IEEE Conference on Decision and Control (CDC). Atlanta, GA, USA: IEEE, 2010. 5991−5998

28 Teixeira A, Dán G, Sandberg H, Johansson K H. A cyber security study of a SCADA energy management system: stealthy deception attacks on the state estimator. *IFAC Proceedings Volumes*, 2011, **44**(1): 11271−11277

29 Rahman M A, Mohsenian-Rad H. False data injection attacks with incomplete information against smart power grids. In: Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM). Anaheim, USA: IEEE, 2012. 3153−3158

30 Liu X, Li Z Y. Local load redistribution attacks in power systems with incomplete network information. *IEEE Transactions on Smart Grid*, 2014, **5**(4): 1665−1676

31 Liu X, Li Z Y. False data attacks against ac state estimation with incomplete network information. *IEEE Transactions on Smart Grid*, 2017, **8**(5): 2239−2248

32 Liu X, Li Z Y. Local topology attacks in smart grids. *IEEE Transactions on Smart Grid*, 2017, **8**(6): 2617−2626

33 Yu Z H, Chin W L. Blind false data injection attack using PCA approximation method in smart grid. *IEEE Transactions on Smart Grid*, 2015, **6**(3): 1219−1226

34 Markwood I, Liu Y, Kwiat K, Kamhoua C. Electric grid power flow model camouflage against topology leaking attacks. In: Proceedings of the 2017 Conference on Computer Communications (INFOCOM). Atlanta, USA: IEEE, 2017. 1−9

35 Chen Y, Kar S, Moura J M F. Dynamic attack detection in cyber-physical systems with side initial state information. *IEEE Transactions on Automatic Control*, 2017, **62**(9): 4618−4624

36 Teixeira A, Shames I, Sandberg H, Johansson K H. A secure control framework for resource-limited adversaries. *Automatica*, 2012, **51**: 135−148

37 Vu Q D, Tan R, Yau D K Y. On applying fault detectors against false data injection attacks in cyber-physical control systems. In: Proceedings of the 35th IEEE International Conference on Computer Communications (INFOCOM). San Francisco, USA: IEEE, 2016. 1−9.

38 Weerakkody S, Sinopoli B. Detecting integrity attacks on control systems using a moving target approach. In: Proceedings of the 54th IEEE Conference on Decision and Control (CDC). Osaka, Japan: IEEE, 2015. 5820−5826

39 Satchidanandan B, Kumar P R. Dynamic watermarking: active defense of networked cyber-physical systems. *Proceedings of the IEEE*, 2017, **105**(2): 219−240

40 Sridhar S, Govindarasu M. Model-based attack detection and mitigation for automatic generation control. *IEEE Transactions on Smart Grid*, 2014, **5**(2): 580−591

41 Wang Q, Khurana H, Huang Y, Nahrstedt K. Time valid one-time signature for time-critical multicast data authentication. In: Proceedings of the 2009 IEEE Conference on Computer Communications (INFOCOM). Rio de Janeiro, Brazil: IEEE, 2009. 1233−1241

42 Cao H Y, Zhu P D, Lu X C, Gurtov A. A layered encryption mechanism for networked critical infrastructures. *IEEE Network*, 2013, **27**(1): 12−18

43 Saxena N, Grijalva S. Efficient signature scheme for delivering authentic control commands in the smart grid. *IEEE Transactions on Smart Grid*, 2018, **9**(5): 4323−4334

44 Kumar P, Braeken A, Gurtov A, Iinatti J, Ha P H. Anonymous secure framework in connected smart home environments. *IEEE Transactions on Information Forensics and Security*, 2017, **12**(4): 968−979

45 Zhang Y C, Wang L F, Sun W Q, Green II R C, Alam M. Distributed intrusion detection system in a multi-layer network architecture of smart grids. *IEEE Transactions on Smart Grid*, 2011, **2**(4): 796−808

46 Baig Z A. On the use of pattern matching for rapid anomaly detection in smart grid infrastructures. In: Proceedings of the 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). Brussels, Belgium: IEEE, 2011. 214−219

47 Palácios J M B, Garay J R B, Oliveira A M, Kofuji S T. Intrusion detection system: A hybrid approach for cyber-physical environments. *Ciencia & Saude Coletiva*, 2013, **269**(2): 2019−2021

48 Bobba R B, Rogers K M, Wang Q Y, Khurana H, Overbye T J. Detecting false data injection attacks on DC state estimation. In: Proceedings of the 1st Workshop on Secure Control Systems. Stockholm, Sweden: University of Illinois Urbana-Champaign, 2010.

49 Giani A, Bitar E, Garcia M, McQueen M, Khargonekar P, Poolla K. Smart grid data integrity attacks: characterizations and *countermeasures*$^{\Pi}$. In: Proceedings of the 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). Brussels, Belgium: IEEE, 2011. 232−237

50 Nuqui R F, Phadke A G. Phasor measurement unit placement techniques for complete and incomplete observability. *IEEE Transactions on Power Delivery*, 2005, **20**(4): 2381−2388

51 Qi J J, Sun K, Kang W. Optimal PMU placement for power system dynamic state estimation by using empirical observability Gramian. *IEEE Transactions on Power Systems*, 2015, **30**(4): 2041−2054

52 Yang Q Y, An D, Yu W. On time desynchronization attack against IEEE 1588 protocol in power grid systems. In: Proceedings of the 2013 IEEE Energytech. Cleveland, USA: IEEE, 2013. 1−5

53 Zhang Z H, Gong S P, Dimitrovski A D, Li H S. Time synchronization attack in smart grid: impact and analysis. *IEEE Transactions on Smart Grid*, 2013, **4**(1): 87−98

54 Jiang X C, Zhang J M, Harding B J, Makela J J, Domínguez-García A D. Spoofing GPS receiver clock offset of phasor measurement units. *IEEE Transactions on Power Systems*, 2013, **28**(3): 3253−3262

55 Jafarnia-Jahromi A, Broumandan A, Nielsen J, Lachapelle G. GPS vulnerability to spoofing threats and a review of antispoofing techniques. *International Journal of Navigation and Observation*, 2012, **2012**: Article No. 127072

56 Fan Y W, Zhang Z H, Trinkle M, Dimitrovski A D, Song J B, Li H S. A cross-layer defense mechanism against GPS spoofing attacks on PMUs in smart grids. *IEEE Transactions on Smart Grid*, 2015, **6**(6): 2659−2668

57 Dán G, Sandberg H. Stealth attacks and protection schemes for state estimators in power systems. In: Proceedings of the 1st IEEE International Conference on Smart Grid Communications. Gaithersburg, USA: IEEE, 2010. 214−219

58 Huang Y, Li H S, Campbell K A, Han Z. Defending false data injection attack on smart grid network using adaptive CUSUM test. In: Proceedings of the 45th Annual Conference on Information Sciences and Systems (CISS). Baltimore, USA: IEEE, 2011. 1−6

59 Murguia C, Ruths J. CUSUM and chi-squared attack detection of compromised sensors. In: Proceedings of the 2016 IEEE Conference on Control Applications (CCA). Buenos Aires, Argentina: IEEE, 2016. 474−480

60 Liu L C, Esmalifalak M, Ding Q F, Emesih V A, Han Z. Detecting false data injection attacks on power grid by sparse optimization. *IEEE Transactions on Smart Grid*, 2014, **5**(2): 612−621

61 Liu X X, Zhu P D, Zhang Y, Chen K. A collaborative intrusion detection mechanism against false data injection attack in advanced metering infrastructure. *IEEE Transactions on Smart Grid*, 2015, **6**(5): 2435−2443

62 Gu C J, Jirutitijaroen P, Motani M. Detecting false data injection attacks in ac state estimation. *IEEE Transactions on Smart Grid*, 2015, **6**(5): 2476−2483

63 Ashok A, Govindarasu M, Ajjarapu V. Online detection of stealthy false data injection attacks in power system state estimation. *IEEE Transactions on Smart Grid*, 2018, **9**(3): 1636−1646

64 Zonouz S, Rogers K M, Berthier R, Bobba R B, Sanders W H, Overbye T J. SCPSE: Security-oriented cyber-physical state estimation for power grid critical infrastructures. *IEEE Transactions on Smart Grid*, 2012, **3**(4): 1790−1799

65 Zonouz S, Davis C M, Davis K R, Berthier R, Bobba R B, Sanders W H. SOCCA: a security-oriented cyber-physical contingency analysis in power infrastructures. *IEEE Transactions on Smart Grid*, 2014, **5**(1): 3−13

66 Wang Y, Xu Z Y, Zhang J L, Xu L, Wang H P, Gu G F. SRID: State relation based intrusion detection for false data injection attacks in SCADA. In: Proceedings of the 19th European Symposium on Research in Computer Security. Wroclaw, Poland: Springer, 2014. 401−418

67 Liu T, Sun Y N, Liu Y, Gui Y H, Zhao Y C, Wang D, et al. Abnormal traffic-indexed state estimation: a cyber physical fusion approach for Smart Grid attack detection. *Future Generation Computer Systems*, 2015, **49**: 94−103

68 Liu T, Tian J, Gui Y H, Liu Y, Liu P F. SEDEA: state estimation-based dynamic encryption and authentication in smart grid. *IEEE Access*, 2017, **5**: 15682−15693

69 Chong F, Lee R B, Vishik C, Acquisti A, Horne W, Palmer C, et al. National Cyber Leap Year Summit 2009: Co-chairs′ Report. Arlington, Virginia, USA: Networking and Information Technology Research and Development Program (U.S.), 2009.

70 Jajodia S, Ghosh A K, Swarup V, Wang C, Wang X S. *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats: Volume 54*. New York, USA: Springer Science & Business Media, 2011.

71 Okhravi H, Hobson T, Bigelow D, Streilein W. Finding focus in the blur of moving-target techniques. *IEEE Security & Privacy*, 2014, **12**(2): 16−26

72 Carroll T E, Crouse M, Fulp E W, Berenhaut S. Analysis of network address shuffling as a moving target defense. In: Proceedings of the 2014 IEEE International Conference on Communications (ICC). Sydney, Australia: IEEE, 2014. 701−706

73 Wu Jiang-Xing. Research on cyber mimic defense. *Journal of Cyber Security*, 2016, **1**(4): 1−10
(邬江兴. 网络空间拟态防御研究. 信息安全学报, 2016, **1**(4): 1−10)

74 Morrow K L, Heine E, Rogers K M, Bobba R B, Overbye T J. Topology perturbation for detecting malicious data injection. In: Proceedings of the 45th Hawaii International Conference on System Sciences. Maui, USA: IEEE, 2012. 2104−113

75 Davis K R, Morrow K L, Bobba R, Heine E. Power flow cyber attacks and perturbation-based defense. In: Proceedings of the 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm). Tainan, China: IEEE, 2012. 342−347

76 Rahman M A, Al-Shaer E, Bobba R B. Moving target defense for hardening the security of the power system state estimation. In: Proceedings of the 1st ACM Workshop on Moving Target Defense. Scottsdale, USA: ACM, 2014. 59−68

77 Gyugyi L, Schauder C D, Sen K K. Static synchronous series compensator: a solid-state approach to the series compensation of transmission lines. *IEEE Transactions on Power Delivery*, 1997, **12**(1): 406−417

78 Hoehn A, Zhang P. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In: Proceedings of the 2016 IEEE American Control Conference (ACC). Boston, USA: IEEE, 2016. 302−307

79 Edris A, Adapa R, Baker M, Bohmann L, Clark K, Habashi K, et al. Proposed terms and definitions for flexible AC transmission system (FACTS). *IEEE Transactions on Power Delivery*, 1997, **12**(4): 1848−1853

80 Divan D M, Brumsickle W E, Schneider R S, Kranz B, Gascoigne R W, Bradshaw D T, et al. A distributed static series compensator system for realizing active power flow control on existing power lines. *IEEE Transactions on Power Delivery*, 2006, **22**(1): 642−649

81 Teixeira A, Shames I, Sandberg H, Johansson K H. Revealing stealthy attacks in control systems. In: Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Monticello, USA: IEEE, 2012. 1806−1813

82 Kwon C, Hwang I. Hybrid robust controller design: cyber attack attenuation for cyber-physical systems. In: Proceedings of the 52nd IEEE Conference on Decision and Control (CDC). Florence, Italy: IEEE, 2013. 188−193

83 Weerakkody S, Ozel O, Griffioen P, Sinopoli B. Active detection for exposing intelligent attacks in control systems. In: Proceedings of the 2017 IEEE Conference on Control Technology and Applications (CCTA), Mauna Lani, HI, USA: IEEE, 2017. 1306−1312

84 Weerakkody S, Sinopoli B. A moving target approach for identifying malicious sensors in control systems. In: Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Monticello, USA: IEEE, 2016. 1149−1156

85 Tian J, Tan R, Guan X H, Liu T. Hidden moving target defense in smart grids. In: Proceedings of the 2nd Workshop on Cyber-Physical Security and Resilience in Smart Grids. Pittsburgh, USA: IEEE, 2017. 21−26

86 Tian J, Tan R, Guan X H, Liu T. Enhanced hidden moving target defense in smart grids. *IEEE Transactions on Smart Grid*, 2018, DOI: 10.1109/TSG.2018.2791512

87 Yin Hao, Lin Chuang, Qiu Feng, Ding Rong. A survey of digital watermarking. *Journal of Computer Research and Development*, 2005, **42**(7): 1093−1099
(尹浩, 林闯, 邱锋, 丁嵘. 数字水印技术综述. 计算机研究与发展, 2005, **42**(7): 1093−1099)

88 Hespanhol P, Porter M, Vasudevan R, Aswani A. Dynamic watermarking for general LTI systems. In: Proceedings of the 56th Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, 2017. 1834−1839

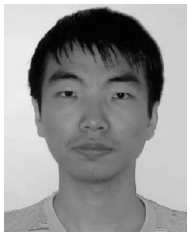
89 Mo Y L, Garone E, Casavola A, Sinopoli B. False data injection attacks against state estimation in wireless sensor networks. In: Proceedings of the 49th IEEE Conference on Decision and Control (CDC). Atlanta, USA: IEEE, 2010. 5967−5972

90 Mo Y L, Chabukswar R, Sinopoli B. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 2014, **22**(4): 1396−1407

91 Weerakkody S, Ozel O, Sinopoli B. Bernoulli-Gaussian physical watermark for detecting integrity attacks in control systems. In: Proceedings of the 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL,USA: IEEE, 2017. 966−973

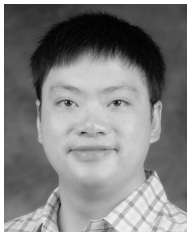
92 Mo Y L, Weerakkody S, Sinopoli B. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine*, 2015, **35**(1): 93−109

刘 烃 西安交通大学网络空间安全学院副教授. 2010 年于西安交通大学获得系统工程专业博士学位. 主要研究方向包括软件安全和智能电网安全. 本文通信作者.
E-mail: tingliu@mail.xjtu.edu.cn
(**LIU Ting** Associate professor at the School of Cyber Security, Xi′an Jiaotong University. He received his Ph. D. degree in systems engineering from Xi′an Jiaotong University in 2010. His research interest covers software security and smart grids security. Corresponding author of this paper.)



田 决 中国香港理工大学电机工程学院研究助理. 2018 年于西安交通大学获得网络空间安全专业博士学位. 主要研究方向为信息物理融合系统安全.
E-mail: juetian@sei.xjtu.edu.cn
(**TIAN Jue** Research assistant in the Department of Electrical Engineering, The Hong Kong Polytechnic University, China. He received his Ph. D. degree in cyberspace security from Xi′an Jiaotong University in 2018. His research interest covers cyber-physical systems security.)

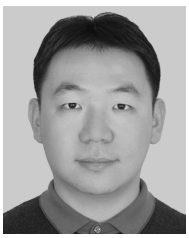
王稼舟 西安交通大学电子与信息工程学院硕士研究生. 2017 年于华北电力大学获得学士学位. 主要研究方向为智能电网安全. E-mail: wjz_98@163.com
(**WANG Jia-Zhou** Master student at the School of Electronic and Information Engineering, Xi′an Jiaotong University. He received his bachelor degree from North China Electric Power University in 2017. His main research interest is smart grids security.)



吴宏宇 美国堪萨斯州立大学电气与计算机工程学院助理教授. 2011 年于西安交通大学获得控制科学与工程专业博士学位. 主要研究方向为智能电网的安全与防御. E-mail: hongyuwu@ksu.edu
(**WU Hong-Yu** Assistant professor in the Department of Electrical and Computer Engineering, Kansas State University, USA. He received his Ph. D. in control science and engineering from Xi′an Jiaotong University, China in 2011. His main research interest is cyber security in smart grids.)

孙利民 中国科学院信息工程研究所研究员, 中国科学院大学教授. 1998 年于国防科学技术大学计算机学院获得博士学位. 主要研究方向为工控系统安全以及物联网安全.
E-mail: sunlimin@iie.ac.cn
(**SUN Li-Min** Research fellow at the Institute of Information Engineering, Chinese Academy of Sciences. Professor at the University of Chinese Academy of Sciences. He received his Ph. D. degree from the School of Computer Science, National University of Defense Technology in 1998. His research interest covers the safety of industrial control systems and IoT security.)

周亚东 西安交通大学电子与信息工程学院副教授. 2011 年于西安交通大学获得控制科学与工程专业博士学位. 主要研究方向为数据驱动的网络行为与内容安全, 网络科学理论及其应用.
E-mail: ydzhou@xjtu.edu.cn
(**ZHOU Ya-Dong** Associate professor at Xi′an Jiaotong University. He received his Ph. D. degree in control science and engineering from Xi′an Jiaotong University in 2011. His research interest covers data driven network security, network science and its applications.)



沈 超 西安交通大学电子与信息工程学院、网络空间安全学院副教授. 主要研究方向为数据驱动的网络行为与内容安全, 人工智能安全, 工控系统与网络安全.
E-mail: chaoshen@xjtu.edu.cn
(**SHEN Chao** Associate professor at the School of Electronic and Information Engineering and the School of Cyber Security, Xi′an Jiaotong University. His research interest covers data driven network behavior and content security, security in artificial intelligent, and industrial system and network security.)

管晓宏 中国科学院院士, 西安交通大学和清华大学教授. 1993 年获得美国康涅狄格大学博士学位. 主要从事能源电力系统优化与安全理论与应用研究.
E-mail: xhguan@mail.xjtu.edu.cn
(**GUAN Xiao-Hong** Member of the Chinese Academy of Sciences, professor at Xi′an Jiaotong University and Tsinghua University. He received his Ph. D. degree from the University of Connecticut in 1993. His research interest covers energy and power system optimization and security theory and application research.)