

文章编号: 1006-3080(2025)05-0671-10

DOI: 10.14135/j.cnki.1006-3080.20241227001

基于水印的参数不确定性电力系统攻击检测方法

陈佳玲, 杨 文

(华东理工大学能源化工过程智能制造教育部重点实验室, 上海 200237)

摘要: 针对具有参数不确定性的负载频率控制 (LFC) 系统中虚假数据注入 (FDI) 攻击检测问题, 基于卡尔曼滤波法构建了 LFC 系统的状态估计模型。为了提高现有 χ^2 检测器的检测率, 考虑了一种添加伪随机数水印矩阵的数据传输策略。与传统方法相比, 在没有攻击的情况下, 该策略不会增加计算复杂度或牺牲估计性能, 以保障电力系统在复杂环境下可靠运行。此外, 考虑到 LFC 系统存在参数不确定性问题, 采用了粒子群优化 (PSO) 算法进行参数辨识, 以提高估计模型与实际系统的匹配度。最后, 提出一种基于伪随机数水印和 PSO 算法相结合的在线辨识-检测方法, 并运用二区域 LFC 系统开展仿真实验, 证实了上述方法的可行性与有效性。

关键词: 状态估计; 攻击检测; 卡尔曼滤波; 协同估计; 网络化

中图分类号: TP301

文献标志码: A

电力系统作为现代社会的关键基础设施, 其稳定运行关系着社会经济发展与民生保障。在信息技术深度融入电力系统的进程中, 网络安全威胁相伴而生且呈现出日益严峻的态势, 其中系统参数不确定性更增添了系统的复杂与脆弱性。为确保电力系统稳定运行, 亟需对精准且高效的攻击检测机制展开研究。基于水印的检测方法在电力系统安全领域日益凸显其重要性与其独特价值, 它通过嵌入水印来破坏攻击者精心设计的虚假信息的隐秘性, 从而提升攻击检测的准确性, 助力构建高效的检测体系。

受益于先进传感器和通信技术的发展与应用, 传统电网正逐渐向智能化方向演进。然而, 高效的通信网络虽为智能电网带来诸多便利, 却也引发了新的安全隐患^[1]。2015 年, 乌克兰电网遭遇了一次严重的网络攻击, 此次事故致使该地区约 23 万居民断电, 停电时间持续 1~6 h 不等。2019 年, 委内瑞拉电网多地发生停电事故, 仅有 5 个州未受影响。2024 年, 日本某大型电网中的 800 余台远程监控设备遭受到了恶意攻击者挟持, 尽管此次攻击暂未威胁到电力系统的运行, 但潜在风险巨大, 若黑客目标转向破坏电网设备, 极有可能造成严重后果。可见,

智能电网领域所面临的网络攻击威胁已经对国家和公共安全产生了重大且深远的影响。

负载频率控制 (LFC) 机制是构成电力系统^[2] 稳定运行的关键要素之一。由于 LFC 系统需要在短时间内进行响应, 在这种情况下, 复杂的数据验证算法通常难以被应用^[3], 这为潜在攻击者提供了一个篡改系统信息的机会, 从而对 LFC 系统的正常运行产生负面影响。虚假数据注入 (FDI) 攻击往往具有高度破坏性和隐蔽性, 其目的是操纵系统信息, 向通信链路注入虚假信息, 以降低系统性能。Amulya 等^[4] 从攻击者的角度分析了 LFC 系统中的 FDI 攻击建模问题, 并研究了其对系统运行的影响。Tan 等^[5] 发现, 凭借窃听传感器数据和部分系统参数, 攻击者可以改进攻击模型并实现最优攻击效果。Ye 等^[6] 考虑到外部干扰的存在, 对针对 LFC 系统的 3 种 FDI 攻击进行了建模和分析。以上事实表明, LFC 系统在面对 FDI 攻击时具有一定的脆弱性。鉴于此, 确保 LFC 系统对 FDI 攻击具备高效的检测能力以及稳定的估计性能更为迫切且意义重大。

近年来, 许多学者提出了针对 FDI 攻击的检测方法, 可大致分为两类: 数据驱动型和模型驱动型。

收稿日期: 2024-12-27

作者简介: 陈佳玲 (2000—), 女, 福建人, 硕士生, 主要研究方向为状态估计和信息物理系统安全。E-mail: jllc@mail.ecust.edu.cn

通信联系人: 杨 文, E-mail: weny@ecust.edu.cn

引用本文: 陈佳玲, 杨 文. 基于水印的参数不确定性电力系统攻击检测方法 [J]. 华东理工大学学报 (自然科学版), 2025, 51(5): 671-680.

Citation: CHEN Jialing, YANG Wen. Watermarking-Based Attack Detection Method for Power System with Parameter Uncertainty[J]. Journal of East China University of Science and Technology, 2025, 51(5): 671-680.

数据驱动型检测方法主要基于历史测量数据展开训练,并构建经验模型。Chen 等^[7]使用多层感知器分类器提取被攻击和正常区域控制误差 (ACE) 之间的差异。Abbaspour 等^[8]提出了一种由龙伯格 (Luenberger) 观测器和人工神经网络 (ANN) 组成的弹性控制机制。然而,这些数据驱动方法存在结构和计算复杂、参数变化鲁棒性弱、需要大量的训练数据等问题。此外,在 FDI 攻击下,电力系统实际运行数据的采集也存在一定困难与挑战。

模型驱动型检测方法依赖于建立一个表示系统预期行为的数学模型,然后将真实测量与预期测量进行比较以检测 FDI 攻击。Liu 等^[9]将攻击检测转化为系统低秩矩阵分解和核范数最小化问题,提出了一种基于异常隔离的安全响应机制。Deng 等^[10]研究了系统方和攻击者的相互作用,从防御预算和混合非线性规划出发制定了最优部署策略。常用的攻击检测方法包括 χ^2 检测、Kullback-Leibler(KL) 散度检测和未知输入观测器 (UIO)^[11]。Guo 等^[12]和 Mo 等^[13]的工作揭示了 χ^2 检测器的一处盲点,攻击者可以利用此盲点实施攻击而不被检测器发现。在这种情况下, χ^2 检测、KL 散度等基于统计分布的传统检测方法无法准确地识别攻击。因此,汪迪等^[14]、Liu 等^[15]和 Zhou 等^[16]分别提出了基于水印的攻击检测方法,主要利用水印矩阵的独立同分布性质,当系统受到 FDI 攻击时,可以提高攻击检测器的检测率。

在建立 LFC 系统模型时,为了便于分析,通常只考虑其中的主要部分,忽略了发电机的铁芯饱和、绕组的涡流损耗等次要因素^[17],这些被忽略的因素会导致模型参数和实际系统参数有不同程度的偏差。此外,系统组件老化、运行环境变化、测量误差和信号干扰等因素也会导致理论模型和实际系统之间存在差异,从而产生参数不确定性。准确的状态估计是 LFC 系统进行有效控制的前提。由于 LFC 系统的状态估计依赖于系统参数,因此参数的不确定性会使估计结果产生偏差,进而影响电力系统的安全运行。Feng 等^[18]为了减小孤立微电网中的参数不确定性影响,借助最小二乘有理函数近似法和 Gerschgorin 圆定理提出一种具有弹性的控制方法。Ramlal 等^[19]设计了一个模糊的迭代学习控制器 (FILC),展现出比 PI 控制器更优的性能。Mu 等^[20]将扰动与参数不确定性纳入 LFC 系统建模中,设计了一种基于粒子群优化 (PSO) 的改进滑模控制器。这些研究成果为解决 LFC 系统中的参数不确定性问题提供了思路和方法,有助于提升电力系统的控制效果和安全性。

1 问题描述

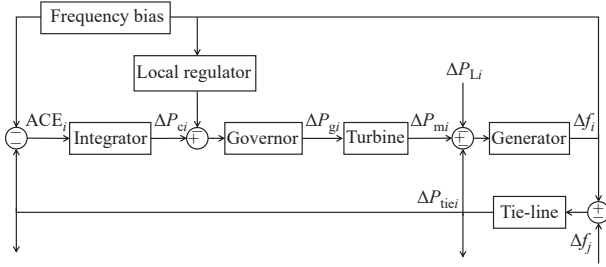
1.1 LFC 系统模型

假设 LFC 系统由 N 个区域组成,其中第 i 个区域的控制模型如图 1 所示。当调节负载频率时,数据采集设备以相应的采样周期测量本区域状态信息,然后通过无线网络将测量信息发送至控制中心。离散时间 LFC 系统模型如式(1)所示:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \end{cases} \quad (1)$$

其中, $\mathbf{x}_k$ 、 $\mathbf{u}_k$ 、 $\mathbf{y}_k$ 分别为系统状态、控制输入和测量输出, $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{C}$ 是已知的状态空间矩阵。过程噪声 $\mathbf{w}_k \sim N(0, \mathbf{R})$ 和测量噪声 $\mathbf{v}_k \sim N(0, \mathbf{Q})$ 是独立的高斯白噪声, $\mathbf{R}$ 和 $\mathbf{Q}$ 分别是过程噪声和测量噪声的协方差矩阵。各变量符号表达式如下:

$$\begin{aligned} \mathbf{x}_k &= [\mathbf{x}_{1,k} \quad \mathbf{x}_{2,k} \quad \cdots \quad \mathbf{x}_{N,k}]^T \\ \mathbf{x}_{i,k} &= [\Delta f_{i,k} \quad \Delta P_{mi,k} \quad \Delta P_{tiei,k} \quad \Delta P_{gi,k}]^T \\ \mathbf{y}_k &= [\mathbf{y}_{1,k} \quad \mathbf{y}_{2,k} \quad \cdots \quad \mathbf{y}_{N,k}]^T, \mathbf{y}_{i,k} = [\Delta f_{i,k} \quad \Delta P_{tiei,k}]^T \\ \mathbf{u}_k &= [\mathbf{u}_{1,k} \quad \mathbf{u}_{2,k} \quad \cdots \quad \mathbf{u}_{N,k}]^T, \mathbf{u}_{i,k} = [\Delta P_{ci,k} \quad \Delta P_{Li,k}]^T \\ \mathbf{A} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1N} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{N1} & \mathbf{A}_{N2} & \cdots & \mathbf{A}_{NN} \end{bmatrix} \\ \mathbf{A}_{ii} &= \begin{bmatrix} \frac{-D_i}{2H_i} & \frac{1}{2H_i} & 0 & \frac{-1}{2H_i} \\ 0 & \frac{-1}{T_{ii}} & \frac{1}{T_{ii}} & 0 \\ \sum_{j=1, j \neq i}^N P_{si,j} & 0 & 0 & 0 \\ \frac{-1}{G_i T_{gi}} & 0 & \frac{-1}{T_{gi}} & 0 \end{bmatrix} \\ \mathbf{A}_{ij} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\sum_{j=1, j \neq i}^N P_{si,j} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathbf{B} &= \text{diag}[\mathbf{B}_1 \quad \mathbf{B}_2 \quad \cdots \quad \mathbf{B}_N] \\ \mathbf{B}_i &= \begin{bmatrix} 0 & \frac{-1}{2H_i} \\ 0 & 0 \\ 0 & 0 \\ \frac{1}{T_{gi}} & 0 \end{bmatrix} \end{aligned}$$

图1 第*i*个区域控制模型Fig. 1 Control model of the *i*th area

$$\mathbf{C} = \text{diag}[\mathbf{C}_1 \quad \mathbf{C}_2 \quad \cdots \quad \mathbf{C}_N]$$

$$\mathbf{C}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

上式中, $\Delta f_{i,k}$ 、 $\Delta P_{tie,i,k}$ 、 $\Delta P_{mi,k}$ 、 $\Delta P_{gi,k}$ 、 $\Delta P_{ci,k}$ 和 $\Delta P_{Li,k}$ 分别表示LFC系统第*i*个区域的频率变化、联络线有功功率偏差、机械功率变化、调速器输出变化、控制信号和负载变化。 H_i 、 $P_{si,j}$ 、 T_{gi} 、 T_{ti} 、 β_i 、 D_i 和 G_i 分别代表第*i*个区域的惯性常数、同步功率系数、调速器时间常数、涡轮机时间常数、频率偏差因子、频率灵敏度负载系数和速度下降特性。

为了实现系统中有效的LFC机制,考虑以下输出反馈控制:

$$\begin{cases} \text{ACE}_{i,k} = \beta_i \Delta f_{i,k} + \Delta P_{tie,i,k} \\ \Delta P_{ci,k} = -\alpha_i \times \text{ACE}_{i,k} \end{cases} \quad (2)$$

其中, α_i 为需要设计的控制增益, $\text{ACE}_{i,k}$ 是第*i*个区域的控制误差。

1.2 攻击模型及其影响

在多区域互联电力系统中,因电力通信具有开放互联的特性,使得控制中心与数据采集设备之间的通信链路存在一定风险,极易遭受FDI攻击。攻击者可以通过攻击通信网络或控制中心来干扰测量设备传输的数据,导致控制中心难以生成精准的调整指令,进而影响整个电力系统的运行。

要发起FDI攻击,攻击者必须对系统信息有一定的了解,如系统测量值和拓扑结构等。为精确评估FDI攻击对系统运行的干扰,使用 $\mathbf{a}_k$ 来表示注入测量值的攻击信号量,将被攻击后接收到的系统测量值建模为:

$$\mathbf{y}_{a,k} = \mathbf{y}_k + \mathbf{a}_k \quad (3)$$

一般而言,一个随机的攻击 $\mathbf{a}_k$ 往往存在较大可能性暴露于系统的不良数据检测机制。然而,一些精心设计的攻击向量,如 $\mathbf{a}_k = \mathbf{C}\phi_k$,可以绕过不良数据检测^[21],其中 ϕ_k 是一个随机向量,可以通过计算测量残差来验证:

$$\mathbf{r}_{a,k} = \mathbf{y}_{a,k} - \hat{\mathbf{y}}_{a,k} = \mathbf{y}_k + \mathbf{C}\phi_k - \mathbf{C}(\hat{\mathbf{x}}_k + \phi_k) = \mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k = \mathbf{r}_k \quad (4)$$

其中, $\hat{\mathbf{x}}_k$ 表示状态估计值, $\mathbf{r}_k$ 为正常情况下的系统测量残差, $\mathbf{r}_{a,k}$ 为受攻击后的测量残差, $\hat{\mathbf{y}}_{a,k}$ 为受攻击后的系统测量估计值。当攻击 $\mathbf{a}_k$ 符合上述条件,此时的FDI攻击就能够避免不良数据检测。

1.3 基于卡尔曼滤波的系统状态估计

在LFC系统中,测量数据需依靠无线网络传输至控制中心。然而,由于设备自身难以避免的测量噪声,致使传感器所获取的数据无法达到理想的精准水平。卡尔曼滤波能够依据系统的动态特性和测量数据,以递归的方式不断优化系统状态估计,故而在电力系统状态估计领域得到了广泛的应用。卡尔曼滤波的递归公式如下:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} \quad (5)$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \quad (6)$$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}^T (\mathbf{C}\mathbf{P}_k^- \mathbf{C}^T + \mathbf{R})^{-1} \quad (7)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-) \quad (8)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}) \mathbf{P}_k^- \quad (9)$$

其中, $\hat{\mathbf{x}}_k^-$ 是先验估计值, $\hat{\mathbf{x}}_k$ 是后验估计值, $\mathbf{P}_k^-$ 是先验估计协方差, $\mathbf{P}_k$ 是后验估计协方差, $\mathbf{K}_k$ 是卡尔曼增益。将Lyapunov算子和Riccati算子分别定义为: $f_1(\mathbf{x}) = \mathbf{A}\mathbf{x}\mathbf{A}^T + \mathbf{Q}$ 和 $f_2(\mathbf{x}) = \mathbf{x} - \mathbf{x}\mathbf{C}^T(\mathbf{C}\mathbf{x}\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{x}$ 。卡尔曼滤波在多数初始条件下能够以较快的速度进入稳态^[22],因此假设系统在时间为0时已达到稳态,即 $\hat{\mathbf{x}}_0 = 0$, $\mathbf{P}_0 = \bar{\mathbf{P}}$,卡尔曼滤波的固定增益为 $\mathbf{K} = \bar{\mathbf{P}}\mathbf{C}^T(\mathbf{C}\bar{\mathbf{P}}\mathbf{C}^T + \mathbf{R})^{-1}$,其中 $\bar{\mathbf{P}}$ 是 $f_1 \circ f_2(\mathbf{x}) = \mathbf{x}$ 的唯一正半定解。

2 水印分析与攻击检测

2.1 水印的添加与设计

精心设计的FDI攻击可以欺骗不良数据检测器,使其无法判断系统是否受到攻击。受密码学的启发,考虑在传输的数据中预先添加一个水印矩阵,以打破注入的虚假数据隐秘性。具体来说,系统测量数据 $\mathbf{y}_k$ 在传输到控制中心之前会被加密为数据 $\mathbf{z}_k$:

$$\mathbf{z}_k = \mathbf{W}_k^{-1} \mathbf{y}_k \quad (10)$$

其中, $\mathbf{W}_k$ 是具备可逆性的水印矩阵。

在控制中心的接收端,需要执行以下操作来重建原始测量值:

$$\mathbf{y}_k^* = \mathbf{W}_k \mathbf{z}_k^* \quad (11)$$

其中, $\mathbf{y}_k^*$ 为去除水印后的系统测量值, $\mathbf{z}_k^*$ 为控制中

心接收到的数据,在通信过程未遭受攻击的情况下,控制中心接收到的测量值不受水印影响,即 $z_k^* = z_k$, $y_k^* = W_k z_k^* = y_k$ 。

FDI 攻击下基于水印的数据传输机制如图 2 所示,如果系统测量值在传输过程中受到攻击, $z_k^* = z_k + a_k$,去除水印后的重构测量值 $y_{a,k}^*$ 将以如下形式表示:

$$y_{a,k}^* = W_k(z_k + a_k) = y_k + W_k a_k \quad (12)$$

遭受攻击后的系统残差值可以表示为:

$$r_{a,k}^* = y_{a,k}^* - C\hat{x}_k^- = y_k + W_k a_k - C\hat{x}_k^- = r_k + W_k a_k \quad (13)$$

经上述分析可知,通过添加水印矩阵 W_k ,可以使 $r_{a,k}^*$ 与 r_k 的统计特征产生明显差异,这有助于 χ^2 检测器检测到攻击的存在。为了防止攻击者在一段时间内通过观察数据的传输过程而推断出水印的存在,采用下述随机变量来设计水印矩阵:

$$W_k = pI + q_k q_k^T \quad (14)$$

其中, p 为一个正实数, $q_k = [q_{1,k} \ q_{2,k} \ \cdots \ q_{n,k}]$ 是一个 n 维向量, $q_{r,k} \sim N(0, \Sigma_q), \forall r \in n$ 是一个零均值、协方差为 Σ_q 的高斯随机变量,与 y_k 、 w_k 、 v_k 都无关,显然, pI 是一个正定矩阵,而 $q_k q_k^T$ 是一个正半定矩阵。基于正定矩阵和正半定矩阵的基本性质,可以很容易地推断出 W_k 是一个可逆矩阵^[23]。 q_k 由伪随机数生成器给出,通过使用适当的随机数种子和密钥管理技术,可以保证发送方和接收方生成的水印序列保持一致^[24]。鉴于水印矩阵 W_k 具有保密性和随机性的特质,攻击者难以确定其具体值。

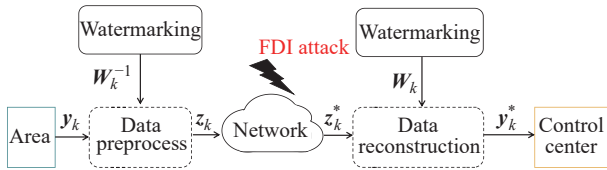


图 2 基于水印的数据传输过程

Fig. 2 Watermarking-based data transmission process

在数据传输前,数据发送方基于水印对需要传输的测量数据预先进行矩阵乘法运算,控制中心接收到数据后进行相应的矩阵逆运算,上述矩阵运算并不涉及复杂的迭代和优化步骤。在控制中心的状态估计和攻击检测部分,核心计算主要集中在卡尔曼滤波算法,水印的添加和去除并未增加卡尔曼滤波的计算步骤或复杂度。与 KL 散度、UIO 等基于模型的传统攻击检测方法相比,基于水印的传输策略并不会增加计算复杂度。

2.2 基于水印的攻击检测可行性分析

为了保障电力系统的稳定运行,需要及时判断

系统是否遭受攻击,因此配备了基于残差的 χ^2 检测器。系统残差可以定义为:

$$r_k = y_k - C\hat{x}_k^- \quad (15)$$

在系统未遭受攻击的情形下,从统计学角度而言,系统残差服从零均值的高斯分布,其协方差为 $\Sigma_r = C\bar{P}C^T + R$ 。基于系统残差及其协方差构造一个 χ^2 检验统计量,并参考标准 χ^2 表给定检测阈值 δ ,一旦测试统计量超过此阈值,就将触发警告,该 χ^2 检验量可以描述为:

$$g_k = r_k^T \Sigma_r^{-1} r_k \begin{cases} \leq \delta, & H_0 \\ > \delta, & H_1 \end{cases} \quad (16)$$

其中,在每个时间步长 k 处, χ^2 检测器将 g_k 与阈值 δ 进行比较 (δ 由置信系数决定)。

假定 k 时刻,若系统没有受到攻击,根据式 (10)、(11) 和式 (15) 易得到系统残差值 $r_k^* = r_k$, χ^2 检测量的期望为:

$$E\{g_k^*\} = E\{r_k^T \Sigma_r^{-1} r_k\} = \Psi \quad (17)$$

当系统遭受攻击后,依据式 (13) 可知,系统残差值 $r_{a,k}^* = r_k + W_k a_k$,可推得受攻击后 χ^2 检测量的期望值为:

$$\begin{aligned} E\{g_{a,k}^*\} &= E\{r_{a,k}^{*T} \Sigma_r^{-1} r_{a,k}^*\} = \\ &= E\{(r_k + W_k a_k)^T \Sigma_r^{-1} (r_k + W_k a_k)\} = \\ &= E\{r_k^T \Sigma_r^{-1} r_k + r_k^T \Sigma_r^{-1} W_k a_k + (W_k a_k)^T \Sigma_r^{-1} r_k + \\ &\quad (W_k a_k)^T \Sigma_r^{-1} W_k a_k\} = \\ &= \Psi + E\{(W_k a_k + 2r_k)^T \Sigma_r^{-1} W_k a_k\} \end{aligned} \quad (18)$$

其中, Ψ 为无攻击时 χ^2 检测量的期望值,当水印矩阵 $W_k \rightarrow \infty$ 时,易知 $E\{(W_k a_k + 2r_k)^T \Sigma_r^{-1} W_k a_k\} \gg 0$,受攻击后 χ^2 检测量的期望 $E\{g_{a,k}^*\}$ 远大于无攻击时 χ^2 检测量的期望 $E\{g_k^*\}$,因此 $E\{g_{a,k}^*\}$ 的值很容易超过预先设定的检测阈值,即 $\Pr(g_{a,k}^* > \delta) \rightarrow 1$,系统能够有效检测到攻击行为的存在。对于攻击者来说,存在着一种特殊情形,当所设计的注入信息 $a_k = -2W_k^{-1} r_k = -2W_k^{-1} (y_k - C\hat{x}_k^-)$ 时,攻击可以规避 χ^2 检测器的监测。然而,攻击者无法得知 y_k 的值,因为在信息传输过程中实际发送的是经过水印处理后的 z_k 。因此,攻击者注入的虚假数据原本所具有的隐秘性被破坏,其与正常系统数据的差异被更加清晰地凸显出来。

着眼于电力系统的安全运行,基于水印矩阵的信息传输机制可以破坏攻击信号的隐秘性,促使攻击快速暴露。理论上,水印矩阵 W_k 参数数值越大,攻击检测效果就越理想,但实际应用中,若参数设置过大,传输信息中水印占比偏大,将使其中真实测量

数据的占比过小。在这种情况下, 攻击者截获传输的数据, 通过统计发现其不符合常理, 就有可能察觉到水印的存在。因此, 综合实际考虑选取适宜的参数会有更好的效果。

3 考虑参数不确定性的辨识-检测方法

3.1 参数不确定性对状态估计的影响

为了定量描述系统参数矩阵的不确定性, 将真实参数 $\mathbf{A}_r$ 的不确定状态定义为:

$$\mathbf{A}_r = \mathbf{A}_d + \mathbf{A}_e \quad (19)$$

其中, $\mathbf{A}_e$ 为真实参数 $\mathbf{A}_r$ 和仿真参数 $\mathbf{A}_d$ 之间的偏差, 控制参数 $\mathbf{B}_r$ 的定义与 $\mathbf{A}_r$ 类似。由式 (1) 能够推得存在不确定参数时的系统状态:

$$\mathbf{x}_{k+1} = \mathbf{A}_r \mathbf{x}_k + \mathbf{B}_r \mathbf{u}_k + \mathbf{w}_k = (\mathbf{A}_d + \mathbf{A}_e) \mathbf{x}_k + (\mathbf{B}_d + \mathbf{B}_e) \mathbf{u}_k + \mathbf{w}_k \quad (20)$$

LFC 系统中的测量值经由通信网络向控制中心传输, 然而因设备固有的噪声干扰, 测量数据在准确性方面仍有待考量。为了便于讨论, 先假设 $\mathbf{B}_r = \mathbf{B}_d$, 以下分析以状态矩阵 $\mathbf{A}_r$ 为例, 在考虑系统参数不确定性的情况下, 卡尔曼估计的一步估计误差如下:

$$\Delta \hat{\mathbf{x}}_r^- = (\mathbf{A}_r - \mathbf{A}_d) \hat{\mathbf{x}}_{k-1} = \mathbf{A}_e \hat{\mathbf{x}}_{k-1} \quad (21)$$

考虑 s 步估计, 设真实系统的估计值为 $\hat{\mathbf{x}}_{r,k}^-$, 控制中心模拟的估计值为 $\hat{\mathbf{x}}_{d,k}^-$, $k=1, 2, \dots, s$, 初始值 $\hat{\mathbf{x}}_{r,0} = \hat{\mathbf{x}}_{d,0} = \hat{\mathbf{x}}_0$, 通过迭代可得到多步估计偏差:

$$\Delta \hat{\mathbf{x}}_s^- = [(\mathbf{A}_d + \mathbf{A}_e)^s - \mathbf{A}_d^s] \hat{\mathbf{x}}_0 + \sum_{i=0}^{s-1} [(\mathbf{A}_d + \mathbf{A}_e)^{s-1-i} - \mathbf{A}_d^{s-1-i}] \mathbf{B} \mathbf{u}_i \quad (22)$$

根据矩阵范数的三角不等式性质, 可以得到:

$$\|(\mathbf{A}_d + \mathbf{A}_e)^s - \mathbf{A}_d^s\| \leq \|(\mathbf{A}_d + \mathbf{A}_e)^s\| + \|\mathbf{A}_d^s\| \quad (23)$$

基于矩阵幕次的范数不等式可知 $\|\mathbf{A}^s\| \leq \|\mathbf{A}\|^s$, 设 $\|\mathbf{A}_d\| = M_1$, $\|\mathbf{A}_e\| = M_2$, 则:

$$\begin{aligned} \|(\mathbf{A}_d + \mathbf{A}_e)^s - \mathbf{A}_d^s\| &\leq \|(\mathbf{A}_d + \mathbf{A}_e)^s\| + \|\mathbf{A}_d^s\| \leq \\ &(\|\mathbf{A}_d\| + \|\mathbf{A}_e\|)^s + \|\mathbf{A}_d\|^s = (M_1 + M_2)^s + M_1^s \end{aligned} \quad (24)$$

因此, 只要 $\mathbf{A}_d$ 和 $\mathbf{A}_e$ 是有界矩阵 (在实际系统中, 系统矩阵的元素通常是有限值), $(\mathbf{A}_d + \mathbf{A}_e)^s - \mathbf{A}_d^s$ 在给定的情况下就是有界的。

在实际系统中, 输入通常是有界的物理量, 即 $\|\mathbf{B} \mathbf{u}_i\| \leq M_3$, 又因为 $(\mathbf{A}_d + \mathbf{A}_e)^s - \mathbf{A}_d^s$ 有界, 因此式 (22) 中 $\sum_{i=0}^{s-1} [(\mathbf{A}_d + \mathbf{A}_e)^{s-1-i} - \mathbf{A}_d^{s-1-i}] \mathbf{B} \mathbf{u}_i$ 也是有界的, 易得多步估计偏差 $\Delta \hat{\mathbf{x}}_s^-$ 是有界的。然而, 系统参数的不确定性会使得状态估计受到偏差 $\mathbf{A}_e$ 干扰, 这种偏差会

一直随着时间积累, 最终导致了系统的估计状态与实际状态逐步偏离。

系统参数不确定时的误差协方差偏差计算如下:

$$\begin{aligned} \Delta \mathbf{P}_k^- &= \mathbf{P}_{r,k}^- - \mathbf{P}_{d,k}^- = (\mathbf{A}_d + \mathbf{A}_e) \mathbf{P}_{k-1} (\mathbf{A}_d + \mathbf{A}_e)^T - \mathbf{A}_d \mathbf{P}_{k-1} \mathbf{A}_d^T = \\ &\mathbf{A}_d \mathbf{P}_{k-1} \mathbf{A}_e^T + \mathbf{A}_e \mathbf{P}_{k-1} \mathbf{A}_d^T + \mathbf{A}_e \mathbf{P}_{k-1} \mathbf{A}_e^T \end{aligned} \quad (25)$$

由于 $\Delta \mathbf{P}_k^-$ 的存在, 使得计算出的卡尔曼增益不能准确反映系统的实际情况, 设真实卡尔曼增益为 $\mathbf{K}_{r,k}$, 控制中心模拟的卡尔曼增益为 $\mathbf{K}_{d,k}$, 易得卡尔曼增益偏差:

$$\Delta \mathbf{K}_k = \mathbf{K}_{r,k} - \mathbf{K}_{d,k} = \frac{(\mathbf{P}_{d,k}^- + \Delta \mathbf{P}_{d,k}^-) \mathbf{C}^T}{\mathbf{C} \mathbf{P}_{d,k}^- \mathbf{C}^T + \mathbf{C} \Delta \mathbf{P}_{d,k}^- \mathbf{C}^T} - \frac{\mathbf{P}_{d,k}^- \mathbf{C}^T}{\mathbf{C} \mathbf{P}_{d,k}^- \mathbf{C}^T} \quad (26)$$

由于卡尔曼增益存在偏差, 在更新状态时, 难以处理测量值和估计值的权重问题, 这种偏差也会在迭代过程中不断累积, 使得估计值逐步远离实际值, 最后对系统整体性能产生不利影响。

3.2 参数不确定性对 χ^2 检测器的影响

当系统存在参数不确定性却错误地使用未考虑参数不确定性的系统矩阵 $\mathbf{A}_d$ 进行状态估计时, 会产生估计偏差 $\Delta \hat{\mathbf{x}}_k^-$, 进而使系统残差中也带上偏差 $\Delta \mathbf{r}_k$:

$$\Delta \mathbf{r}_k = \mathbf{y}_k - \mathbf{C} \hat{\mathbf{x}}_{r,k}^- - (\mathbf{y}_k - \mathbf{C} \hat{\mathbf{x}}_{d,k}^-) = \mathbf{C} (\hat{\mathbf{x}}_{d,k}^- - \hat{\mathbf{x}}_{r,k}^-) = \mathbf{C} \Delta \hat{\mathbf{x}}_k^- \quad (27)$$

由于残差偏差的存在, 残差的均值不再为 0, 协方差也发生变化。考虑估计偏差后的系统其真实残差为:

$$\mathbf{r}_{r,k} = \mathbf{r}_{d,k} + \Delta \mathbf{r}_k \quad (28)$$

易知, 其均值 $E(\mathbf{r}_{r,k}) = E(\Delta \mathbf{r}_k)$, 此时, χ^2 检测统计量 $\mathbf{g}_{r,k} = \mathbf{r}_{r,k}^T \boldsymbol{\Sigma}_r^{-1} \mathbf{r}_{r,k}$ 的概率分布不再是原假设下的概率分布。如果仍按照原定概率分布来设定 χ^2 检测的阈值, 将致使判定结果也出现偏差。因为残差偏差改变了 χ^2 检测量的概率分布, 使用不考虑偏差的 χ^2 检测量进行检测, 可能会出现将常规系统工况错误地判断为异常工况的情形。同时, 如果因攻击或异常情况导致的残差变化被残差偏差所掩盖, 使得 χ^2 检测量未超过阈值, 也会出现漏警现象。

3.3 基于 PSO 算法的参数辨识方法

为了弥补在系统参数不确定时估计不准确的问题, 可以将利用 PSO 算法辨识得到的系统参数代入卡尔曼滤波和 χ^2 检测器进行计算。PSO 是一种模拟鸟群觅食的群智能优化算法^[25]。粒子依据自身经验 (个体最优位置), 参考了群体中其他粒子的经验 (全局最优位置), 实现对位置和速度的动态调整。在参数偏差有界的情况下, 利用 PSO 算法搜索最优解就可以确定一个合理的搜索空间。粒子 o 在 V 维空间

中的最优个体位置和整个种群的最优速度可以由以下方程得到:

$$\begin{cases} v_{o,k+1} = \varpi v_{o,k} + \tau_1 u_1 (L_{o,k}^{\text{best}} - L_{o,k}) + \tau_2 u_2 (L_k^{\text{best}} - L_{o,k}) \\ L_{o,k+1} = L_{o,k} + v_{o,k+1} \end{cases} \quad (29)$$

其中, 粒子 o 的位置 $L_o = (L_{o1}, L_{o2}, \dots, L_{oV})$, 粒子的速度 $v_o = (v_{o1}, v_{o2}, \dots, v_{oV})$ 。设粒子个体的最优位置为 $L_o^{\text{best}} = (L_{o1}^{\text{best}}, L_{o2}^{\text{best}}, \dots, L_{oV}^{\text{best}})$, 整个种群的全局最优位置为 $L^{\text{best}} = (L_1, L_2, \dots, L_V)$ 。 ϖ 是惯性权重, τ_1 和 τ_2 均是加速度系数, u_1 和 u_2 是介于 0 和 1 之间的随机数。

PSO 算法的另一个关键指标是适应度函数, 它限制了各粒子的运动趋向, 也是算法停止迭代更新的依据原则。如上所述, 优化的目标是在更新仿真系统参数的同时, 尽量减少仿真系统与实际系统之间的误差。首先将相同的扰动输入到模拟系统和实际系统中, 然后将两个输出之间的差值设为适应度, 采用均方根误差 (RMSE) 作为评估 PSO 性能的指标, 即适应度函数为:

$$\text{RMSE}_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{act},k}^i - y_{\text{ps},k}^i)^2} \quad (30)$$

其中, RMSE_k 是系统状态真实值与仿真值之间的误差值, $y_{\text{act},k}^i$ 是真实值的第 i 个分量, $y_{\text{ps},k}^i$ 是仿真值的第 i 个分量。

在满足终止条件的情形下, 输出全局最优位置所对应的参数值, 即为通过 PSO 算法所辨识得到的最优系统参数。参数识别过程可建模如下:

$$\begin{cases} x_{\text{ps},k+1} = A_{\text{ps},k} x_{\text{ps},k} + B_{\text{ps},k} u_k + w_k \\ y_{\text{ps},k} = C x_{\text{ps},k} + v_k \\ y_{\text{act},k} = F(x_{\text{act},k}, u_k, w_k) \end{cases} \quad (31)$$

$$\text{RMSE}_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{act},k}^i - y_{\text{ps},k}^i)^2}$$

3.4 在线辨识-检测方法

卡尔曼滤波最优性能的前提是基于系统参数已知且准确, 从上述分析可知, 当系统存在参数不确定性且未对卡尔曼滤波算法进行修正时, 如式 (25) 和式 (26) 所示, 系统的估计误差协方差和卡尔曼增益都受到参数不确定性的影响, 系统的估计性能会降低。根据式 (16) 和式 (27) 可知, 参数不确定性导致了残差偏差的存在, χ^2 检测量的概率分布受到残差偏差的影响, 这使得攻击检测的可靠性难以保证, 增加了误报、漏报情况出现的可能性。通过 PSO 算法进行参数辨识, 根据辨识得到更准确的系统参数来计算检测指标, 使得基于水印的攻击检测方法能更准确地判断系统是否遭受攻击, 从而有效避免了因

参数不确定性导致的误报和漏报, 提高了系统在复杂场景下的攻击检测能力。

将在线参数辨识和攻击检测算法相结合, 能够为 LFC 系统的性能和安全提供双重保障, 及时察觉系统参数变化和潜在的安全威胁, 有效增强系统的可靠性, 如图 3 所示。在线辨识-检测算法可以描述如下:

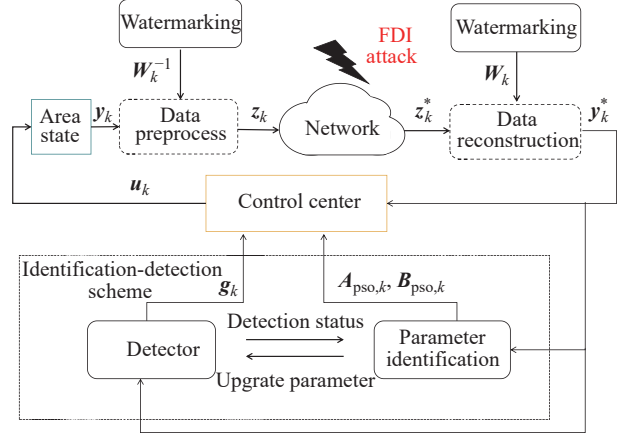


图 3 在线辨识-检测算法

Fig. 3 Online identification-detection algorithm

步骤 1: 给定噪声协方差矩阵 Q 和 R , 系统状态参数 A 、 B 、 C , 系统初始状态 x_0 , 初始估计误差协方差 P_0 , 检测阈值 δ , 时间步长 T , 水印参数 p 和 Σ_q , 设定粒子的初始位置 L_0 、速度 v_0 、惯性权重参数 ϖ 、加速度系数 τ_1 和 τ_2 等, 参数辨识的初始值 $A_{\text{ps},0}$ 和 $B_{\text{ps},0}$;

步骤 2: 当 $k \geq 1$ 时, 使用式 (10) 在要发送的测量值 y_k 中添加水印得到加密后的测量值 z_k ;

步骤 3: 控制中心接收到测量值 z_k^* 后, 基于式 (11) 重建测量值 y_k^* ;

步骤 4: 基于 PSO 算法, 使用式 (29) 和式 (31) 计算辨识得到的系统参数 $A_{\text{ps},k}$ 和 $B_{\text{ps},k}$;

步骤 5: 将辨识后的参数 $A_{\text{ps},k}$ 和 $B_{\text{ps},k}$ 分别代入式 (5)~式 (9) 更新时间 k 时的系统估计值;

步骤 6: 使用式 (15) 和式 (16) 计算 χ^2 检验统计量 g_k ;

步骤 7: 如果 $g_k > \delta$, 则说明系统受到攻击, 触发警报;

步骤 8: 设置 $k = k + 1$, 然后返回到步骤 2。

4 实验仿真

为了验证先前推导的理论结果, 将基于一个二区域电力 LFC 系统给出一些数值模拟。首先, 二区域系统的参数设置如表 1 所示, D_i 和 G_i 分别为区域

表1 二区域电力系统的相关参数值

Table 1 Parameter values of two-area power system

Parameter	G_i	D_i	H_i	T_{gi}	T_{ti}	$P_{si,j}$	α_i	β_i	N
Area1	0.0500	0.6	5	0.2	0.5	2	0.3	$\frac{1}{G_1} + D_1$	2
Area2	0.0625	0.9	4	0.3	0.6	2	0.3	$\frac{1}{G_2} + D_2$	2

i 的频率灵敏度负载系数和速度下降特性。定义过程噪声 $\mathbf{w}_k$ 的协方差矩阵 $\mathbf{Q} = 2.5E^{-3}\mathbf{I}$ 和测量噪声 $\mathbf{v}_k$ 的协方差矩阵 $\mathbf{R} = 8E^{-5}\mathbf{I}$, 其中 E 代表科学计数法中 10 的幂次, 区域负载变化在整个仿真时间内被建模为常数, $\Delta P_{L1} = 0.08$, $\Delta P_{L2} = 0.12$ 。

图4展示了LFC机制对区域频率和连接线功率带来的影响。在 $t < 300$ s 时, LFC 机制未启动。由于负载的变化, 测量值稳定在一个偏移量附近。当 LFC 机制在 $t = 300$ s 启动后, 这个偏移量被解决, 系统处于稳定工作状态, $\lim_{k \rightarrow \infty} \Delta f_{i,k}, \Delta P_{tiei,k} = 0$ 。

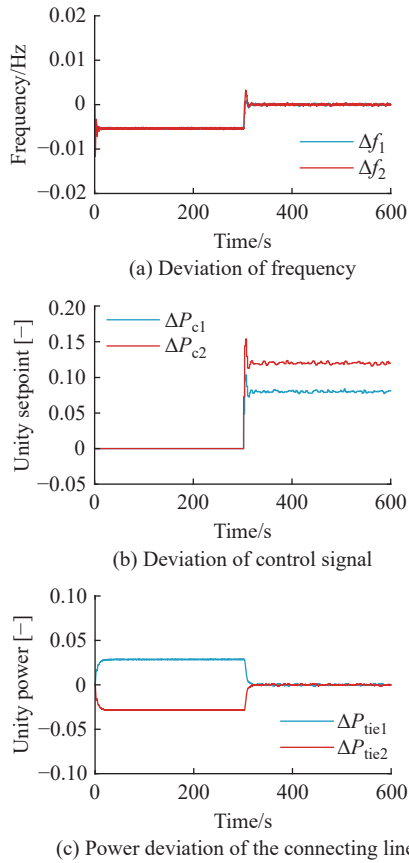
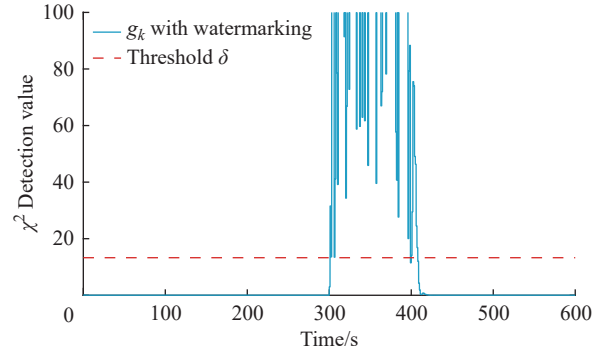


图4 LFC 机制对二区域系统的影响

Fig. 4 Impact of the LFC mechanism on the two-area system

考虑对于 FDI 攻击的检测, 将显著性水平设置为 0.01, 即 χ^2 检测器的阈值 δ 为 13.277。水印矩阵参数设为 $p = 3.58$, $\Sigma_q = 4.72$ 。FDI 攻击的设计可以参考文献 [26], 将攻击持续时间范围设定为 300 ~ 400 s。如图 5 所示, 在传输测量中添加有水印信号, 当

图5 添加水印时的 χ^2 检测结果Fig. 5 χ^2 Detection result with watermarking

$t < 300$ s 时, 系统正常运行, 没有攻击发生, 检测函数值在允许区间内波动。当 $t \geq 300$ s 时, 攻击者发动攻击, χ^2 检测量的值将显著增加, 超过了设定的阈值, 进而触发异常警报。当攻击者停止攻击后, 检测功能恢复正常, 将不会触发异常报警。

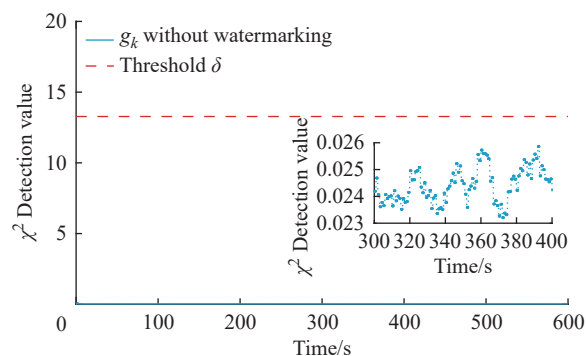
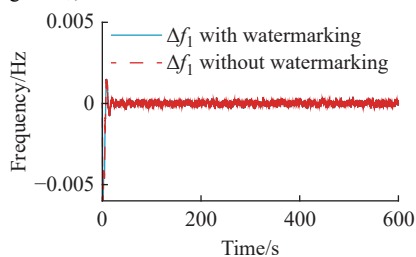
为进一步说明添加水印的传输机制对 FDI 攻击检测的有效性, 考虑当没有添加水印信号时, 采取与上述相同的攻击策略, 如图 6 所示。当没有攻击发生时, 检测函数值在一定范围内有所波动, 同样地, 在 300 s 后, 攻击者发动攻击, 系统检测量的值只发生了少许增加, 但没有超过给定的阈值, 因此, 注入的虚假信号将不会触发异常警报。

当没有攻击发生时, 以控制区域 1 为例, 由图 7 可以推断水印的添加并不会影响系统的稳定运行和估计效果。从以上的数值模拟结果可以看出, 水印的存在可以检测到上述提出的 FDI 攻击, 水印参数不需要设定得很大, 攻击仍然可以暴露。同时, 当没有攻击发生时, 水印对系统估计没有影响, 这表明基于伪随机数水印和 χ^2 检测器的检测方法是有效且可靠的。

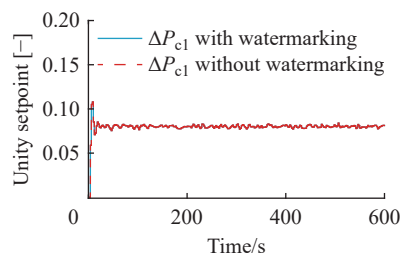
当系统参数发生变化时, 为提供更好的展示, 以区域 1 的识别结果作为代表。PSO 算法中的参数设定如下: 粒子数量为 50, 加速度系数 $\tau_1 = 1.25$, $\tau_2 = 1.5$, 惯性权重参数 $w = 0.9$, 最大迭代次数为 300 [25]。图 8 显示了识别过程中的频率误差曲线, 其中频率误差越小, 辨识结果就越准确。

当系统参数在运行期间存在不确定性时, 例如在 $t = 300$ s 时, 阻尼系数 D 从 [0.6, 0.9] 变为 [0.7, 0.8], FDI 攻击和 χ^2 检测阈值设置同上, 如图 9 所示。在 300 ~ 400 s 间, χ^2 检测量超出既定阈值, 将触发系统警报。

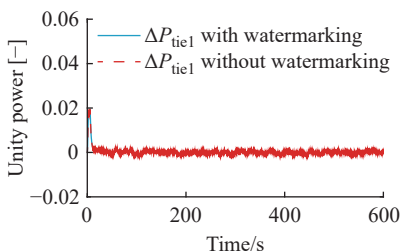
而当系统参数发生变化但没有攻击发生时, 如图 10 所示, 在 $t = 300$ s 时, 阻尼系数 D 同上发生变化。只基于水印的检测方法会在参数发生变化时产

图 6 没有添加水印时的 χ^2 检测结果Fig. 6 χ^2 Detection result without watermarking

(a) Deviation of frequency



(b) Deviation of control signal



(c) Power deviation of the connecting line

图 7 没有攻击时添加水印对区域 1 的影响

Fig. 7 Impact of watermarking on the area 1 when no attacks occur

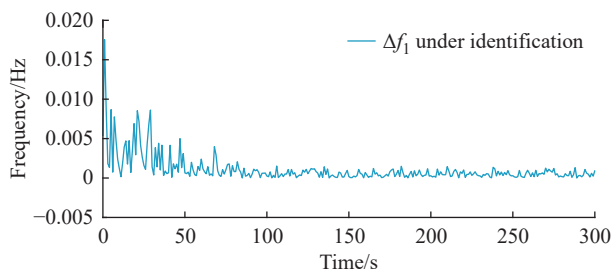
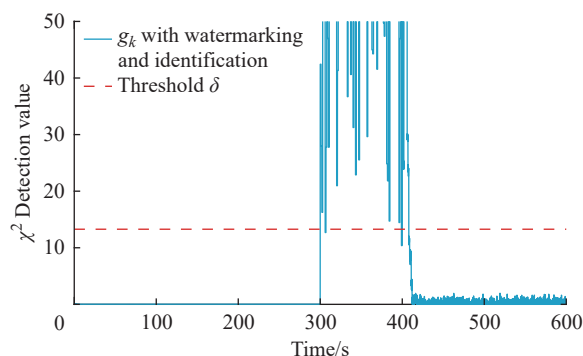
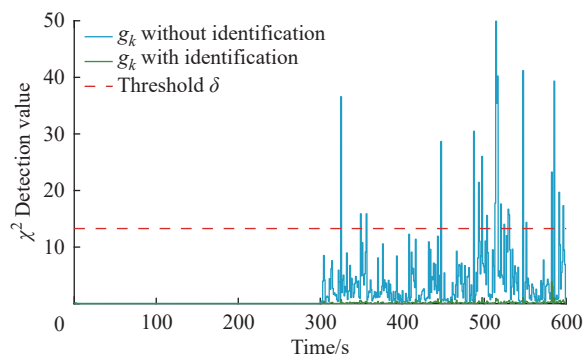


图 8 采用 PSO 算法进行辨识时的频率误差

Fig. 8 Frequency error with PSO algorithm during identification

生虚警,且这种误差会一直累积,但在线辨识-检测机制下区域 1 的 χ^2 检测量几乎没有发生变化,这说明

图 9 攻击下参数不确定系统 χ^2 检测结果Fig. 9 χ^2 Detection result with uncertain parameters under attacks图 10 没有攻击时参数不确定系统 χ^2 检测结果Fig. 10 χ^2 Detection result with uncertain parameters when no attacks occur

基于水印和 PSO 算法的在线检测-辨识机制是有效的,并不会引起无攻击时的误报。

此外,为深入探究在线辨识-检测方法与其他检测方法(如 ANN^[8]、UIO^[27])在性能上的差异,设计了对比实验并重复进行了 1000 次,系统参数和攻击设置同上,对比结果如表 2 所示。以在线辨识-检测方法的运行时间为基准(100%),UIO 算法的检测时间最短,为在线辨识-检测方法的 87.61%。ANN 算法因其结构复杂,检测时间是在线辨识-检测方法的 137.48%。ANN 算法和在线辨识-检测方法在检测准确率上表现较好且较为接近,而 UIO 算法的检测准确率仅有 78.49%,明显低于前两者。在误报率方面,在线辨识-检测方法仅为 0.39%,ANN 算法为 0.71%,UIO 算法

表 2 相同设置下不同方法的对比结果

Table 2 Comparison results of different methods under the same settings

Method	Time/%	Detection accuracy/%	False alarm rate/%
Online identification and detection	100.00	94.37	0.39
ANN	137.48	94.51	0.71
UIO	87.61	78.49	10.35

的误报率则高达 10.35%。综合评估可得,在线辨识-检测方法能够以较高的准确率和极低的误报率完成检测任务,在计算复杂度和检测时间上都具有优势。

5 结 论

本文主要聚焦 LFC 系统在 FDI 攻击下的安全问题。为有效检测 LFC 系统中的 FDI 攻击,提出了在信息传输过程中添加一个伪随机数水印矩阵的传输策略,通过水印加密和解密的过程来恢复传输的真实数据。在无 FDI 攻击情况下,经理论推导证实该水印策略不会影响系统正常的状态估计。当系统遭受 FDI 攻击时,攻击者对真实数据进行篡改,致使解密后的测量数据中存在水印矩阵和注入攻击信号的耦合量,很容易被 χ^2 检测器识别,进而确保了系统的高效稳定运行。此外,针对 LFC 系统中可能存在的参数不确定性问题,采用了 PSO 算法进行在线参数辨识,减小参数不确定性对系统状态估计和攻击检测的影响。最后,通过在双区域 LFC 系统上的仿真,验证了结合水印和 PSO 算法这一在线辨识-检测方法的有效性,保障了 χ^2 检测器在 LFC 系统存在参数不确定性时的攻击检测准确性。

参考文献:

- [1] 董亦, 邹媛媛, 牛玉刚. 基于分层内嵌预测控制的多微电网能量优化管理 [J]. 华东理工大学学报 (自然科学版), 2018, 44(4): 463-473.
- [2] WANG X L, DING D R, DONG H L, *et al.* PI-Based security control against joint sensor and controller attacks and applications in load frequency control[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(2): 970-980.
- [3] AZARBAHRAM A, AMINI A, SOJOODI M. Resilient fixed-order distributed dynamic output feedback load frequency control design for interconnected multi-area power systems[J]. *IEEE/CAA Journal of Automatica Sinica*, 2019, 6(5): 1139-1151.
- [4] AMULYA A, SWARUP K S. Analysis of false data injection attacks on multi-area load frequency control[C]//2019 8th International Conference on Power Systems (ICPS). Jaipur, India: IEEE, 2019: 1-6.
- [5] TAN R, NGUYEN H H, FOO E Y S, *et al.* Optimal false data injection attack against automatic generation control in power grids[C]//2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS), Vienna, Austria: IEEE, 2016: 1-10.
- [6] YE J, YU X. Detection and estimation of false data injection attacks for load frequency control systems[J]. *Journal of Modern Power Systems and Clean Energy*, 2022, 10(4): 861-870.
- [7] CHEN C Y, ZHANG K F, YUAN K, *et al.* Novel detection scheme design considering cyber attacks on load frequency control[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(5): 1932-1941.
- [8] ABBASPOUR A, SARGOZLAIE A, FOROUZAN-NESZHAD P, *et al.* Resilient control design for load frequency control system under false data injection attacks[J]. *IEEE Transactions on Industrial Electronics*, 2020, 67(9): 7951-7962.
- [9] LIU L C, ESMALIFALAK M, DING Q F, *et al.* Detecting false data injection attacks on power grid by sparse optimization[J]. *IEEE Transactions on Smart Grid*, 2014, 5(2): 612-621.
- [10] DENG R L, XIAO G X, LU R X. Defending against false data injection attacks on power system state estimation[J]. *IEEE Transactions on Industrial Informatics*, 2017, 13(1): 198-207.
- [11] 王彩云, 李芳菲. 基于 KL 散度检测器下的最优线性欺诈攻击 [J]. 华东理工大学学报 (自然科学版), 2021, 47(5): 627-634.
- [12] GUO Z Y, SHI D W, JOHANSSON K H, *et al.* Worst-case innovation based integrity attacks with side information on remote state estimation[J]. *IEEE Transactions on Control of Network Systems*, 2019, 6(1): 48-59.
- [13] MO Y L, SINOPOLI B. On the performance degradation of cyber physical systems under stealthy integrity attacks[J]. *IEEE Transactions on Automatic Control*, 2016, 61(9): 2618-2624.
- [14] 汪迪, 李芳菲, 许思遥, 等. 针对信息物理系统线性欺诈攻击的水印加密策略 [J]. 华东理工大学学报 (自然科学版), 2020, 46(6): 828-832.
- [15] LIU H, LI Y Z, HAN Q L, *et al.* Watermark-based proactive defense strategy design for cyber-physical systems with unknown-but-bounded noises[J]. *IEEE Transactions on Automatic Control*, 2023, 68(6): 3300-3315.
- [16] ZHOU J Y, YANG W, DING W J, *et al.* Watermarking-based protection strategy against stealthy integrity attack on distributed state estimation[J]. *IEEE Transactions on Automatic Control*, 2023, 68(1): 628-635.
- [17] 陈亦平, 皮杰明, 李崇涛, 等. 考虑 AGC 采样保持特性的时滞电力系统频率稳定性分析 [J]. 电网技术, 2023, 47(8): 3335-3342.
- [18] FENG Y W, JIA W. Resilient H_∞ load frequency control for uncertain isolated microgrid under stealth attack[C]// 2022 12th International Conference on Power and Energy Systems (ICPES). Guangzhou, China, IEEE: 2022: 393-400.
- [19] RAMLAL C J, SINGH A, ROCKE S, *et al.* Decentralized fuzzy H_∞ iterative learning LFC with time-varying communication delays and parametric uncertainties[J]. *IEEE*

- Transactions on Power Systems, 2019, 34(6): 4718-4727.
- [20] MU C X, TANG Y F, H H B. Improved sliding mode design for load frequency control of power system integrated an adaptive learning strategy[J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(8): 6742-6751.
- [21] LIU S C, LUO W S, WU L G. Co-design of distributed model-based control and event-triggering scheme for load frequency regulation in smart grids[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, 50(9): 3311-3319.
- [22] 蒋志凯. 数字滤波与卡尔曼滤波 [M]. 北京: 中国科学技术出版社, 1993.
- [23] HOM R A, JOHNSON R. Matrix Analysis[M]. Cambridge: Cambridge University Press, 2013.
- [24] AHMED R E. Efficient pseudo-random number generators for wireless sensor networks[C]// 2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS). Abu Dhabi, United Arab Emirates, IEEE: 2016: 1-4.
- [25] JABER A S, AHMAD A Z, ABDALLA A N. A new parameters identification of single area power system based LFC using segmentation particle swarm optimization (SePSO) algorithm[C]// 2013 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). Hong Kong, China, IEEE: 2013: 1-6.
- [26] MO Y L, GARONE E, CASAVOLA A, *et al.* False data injection attacks against state estimation in wireless sensor networks[C]// 49th IEEE Conference on Decision and Control (CDC). Atlanta, USA, IEEE: 2010: 5967-5972.
- [27] HASSANI H, RAZANI-FAR R, SAIF M, *et al.* Unknown input observers design for real-time mitigation of the false data injection attacks[C]// 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Toronto, Canada, IEEE: 2020: 3612-3617.

Watermarking-Based Attack Detection Method for Power System with Parameter Uncertainty

CHEN Jialing, YANG Wen

(Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

Abstract: A novel online identification and detection method based on watermarking and particle swarm optimization (PSO) algorithm is proposed to address the issue of false data injection (FDI) attack detection in the load frequency control (LFC) system under parameter uncertainty. Firstly, a state estimation model for the LFC system is established based on Kalman filtering. To assist the χ^2 detector in identifying FDI attacks, a transmission strategy of adding a pseudo-random watermark matrix during information transmission is proposed. Prior to transmission, the transmitting terminal multiplies the measurement to be transmitted by a pseudo-random matrix. Subsequently, the receiving terminal decrypts the received data to retrieve the genuine information. Compared with traditional methods, this strategy neither increases computational complexity nor sacrifices estimation performance in the absence of attacks. When an FDI attack occurs, the system's measurement signals are maliciously tampered with by attackers. During the decryption process at the receiver end, a unique coupling relationship emerges between the pre-embedded watermark matrix and the injected attack signal. This coupling feature is highly distinguishable by the χ^2 detector, enabling rapid and accurate identification of FDI attacks and thus ensuring the efficient and stable operation of the LFC system. Furthermore, to tackle the parameter uncertainty inherent in LFC systems (a factor that may degrade the accuracy of state estimation and attack detection), the PSO algorithm is introduced for online parameter identification. This algorithm optimizes the key parameters of the LFC system in real time, significantly improving the matching degree between the established state estimation model and the actual operating system. Finally, by fusing the watermarking-based attack detection strategy and the PSO-based parameter identification technique, a complete online FDI attack identification and detection method for LFC systems is formed. The effectiveness and feasibility of the proposed method are verified through simulations on a two-area LFC system.

Key words: state estimation; attack detection; Kalman filter; cooperative estimation; networking

(责任编辑: 张欣)