文章编号:2096-7543(2023)02-0038-08 DOI:10.11863/j.suse.2023.02.05

基于最大似然估计的智能电网FDIA检测

吴铭辉，高文根，华　峰，穆少鹏

(安徽工程大学检测技术与节能装置安徽省重点实验室,安徽　芜湖　241000)

摘　要:智能电网的正常运行依赖于准确反映电网物理特性的状态估计。针对虚假数据注入攻击(False Data Injection Attack,FDIA)通过向电力系统量测单元注入恶意数据来篡改状态估计结果的问题,提出了一种基于最大似然估计(Maximum Likelihood Estimation,MLE)的电网FDIA检测方法,并以此提高状态估计结果的精度。首先,基于智能电网量测向量与FDIA攻击向量服从具有不同协方差多元高斯分布的特点,通过MLE计算法求得量测数据期望与协方差,根据该协方差判断是否存在虚假量测数据。其次,若数据正常,通过加权最小二乘(Weighted Least Square,WLS)算法依据该量测数据期望进行状态估计可以得到更加优秀的系统状态结果。最后,基于IEEE-14节点系统的算例证明了该算法的可行性。

关键词:智能电网;状态估计;虚假数据注入攻击;攻击检测

中图分类号:TM744　　**文献标识码:**A

引　言

自从Schweppe等[1]在1970年首次将状态估计引入电力系统后,状态估计就一直在全世界输电网络运行的管理和控制中发挥着重要作用。状态估计是根据从智能电网监视控制和数据采集(Supervisory Control And Data Acquisition,SCADA)系统获得的不完全量测值,以此构建基于电力系统模型的非线性方程组的数学模型,为未知系统状态变量赋值的过程[2]。状态估计的准确性直接影响智能电网的性能甚至稳定性,因此状态估计对智能电网的可靠控制和运行至关重要[3]。Liu等[4]首次提出电网虚假数据注入攻击(Fake Data Injection Attack,FDIA)的概念,在这种攻击中,对手可以攻击电网系统的状态。攻击成功的FDIA会导致状态估计结果产生偏差,从而影响电力系统安全稳定地运行[5]。为了保持智能电网状态估计的准确性,电力系统一般使用不良数据检测(Bad Data Detection,BDD)来消除由于仪表故障或外部攻击导致的错误量测[6]。然而,FDIA可以绕过BDD,这是其被称为网络攻击中最危险的攻击类型之一的原因[7]。FDIA是目前研究最多的网络物理安全攻击,近年来针对此类攻击

收稿日期:2022-04-11

基金项目:国家自然科学区域创新发展联合基金项目(U21A20146);安徽省自然科学基金项目(1908085MF215);安徽省重点研究与开发计划项目(201904a05020007)

作者简介:吴铭辉(1997-),男,硕士生,研究方向为智能信息处理及应用,(E-mail)780597060@qq.com

通信作者:高文根(1973-),男,教授,博士,研究方向为智能化测控技术,(E-mail)ahpuchina@ahpu.edu.cn

进行了持续的构建和防御工作[8]。

恶意虚假数据在网络攻击节点的注入必然对电力系统的状态估计产生不利的影响[9]。而传统的检测方法不再有效量测，这表明需要一种新型的量测技术[10]。文献[11]展示了一种对智能电网实时运行过程中对FDIA和干扰攻击在线检测方法，即通过对系统状态进行估计与预测的卡尔曼滤波(Kalman Filtering，KF)算法。文献[12]提出的扩展卡尔曼滤波器(Extended Kalman Filter，EKF)以及文献[13]提出的分布式卡尔曼滤波器(Distributed Kalman Filter，DKF)都是基于KF算法的改进，可以更精确地进行状态估计和检测FDIA。文献[14]使用了感知器、k近邻(k-Nearest Neighbor，k-NN)、线性支持向量机(Linear Support Vector Machine，LSVM)、稀疏逻辑回归(Sparse Logistic Regression，SLR)和半监督支持向量机(semi-Supervised Support Vector Machine，S3VM)的统计学习方法对智能电网中的FDIA进行了检测实验。以上方法基本上都是通过预测系统状态或建立虚假数据特征集的方法来预防和检测FDIA，无法直接检测出FDIA所攻击的量测单元，且对电网数据可信度的要求较高。

由于智能量测单元的量测误差与智能电网信息传输过程中导致的误差，电网SCADA系统所获得量测数据具有高斯分布的特点。而使用FDIA的攻击者几乎不可能通过获得电网历史量测数据的方法来得到量测误差的实际分布参数，所以其模拟的虚假量测数据的误差分布很难与实际量测数据的误差分布一致。因此，在攻击者使用FDIA持续进行量测数据篡改的过程中，本文提出使用最大似然估计(Maximum Likelihood Estimation，MLE)算法对量测数据的多元高斯分布参数进行估计，通过把实时量测数据的协方差与历史量测误差分布进行比较来判断是否存在虚假量测数据；通过MLE算法得到的量测数据期望更加符合电网的真实数据，减小各种误差的影响；通过加权最小二乘(Weighted Least Square，WLS)算法使用该期望进行状态估计改善估计结果的精度。

1　系统模型

对于一个交流电力系统，其数学模型[15]可以表示为：

$$\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{x})+\boldsymbol{e} \tag{1}$$

其中：$\boldsymbol{z}=[z_1,z_2,\cdots,z_m,\cdots,z_M]^{\mathrm{T}}$为量测单元采集的原始量测向量，其包含电网节点注入的有功功率P_i和无功功率Q_i，及在电网任意两节点i、j间传输线路流过的有功功率P_{ij}和无功功率Q_{ij}；$\boldsymbol{x}=[x_1,x_2,\cdots,x_n,\cdots,x_N]^{\mathrm{T}}$为电力系统的状态变量，一般包含电网节点的电压幅值$V_i$与相角$\theta_i$；$\boldsymbol{e}=[e_1,e_2,\cdots,e_m,\cdots,e_M]^{\mathrm{T}}$为由于量测单元存在的固有噪声导致的量测误差向量，其服从均值为0、协方差为$\boldsymbol{R}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_m^2,\cdots,\sigma_M^2)$的高斯分布模型；$\boldsymbol{h}(\cdot)$表示为电力系统量测向量$\boldsymbol{z}$与状态变量$\boldsymbol{x}$之间的非线性关系，即电力系统的网络拓扑结构。

将任意一组状态变量$\boldsymbol{x}$代入$\boldsymbol{h}(\boldsymbol{x})$均可以得到一组与其对应的电网量测向量，该量测向量与通过量测单元量测获得的量测向量$\boldsymbol{z}$之间的差被称为残差，可以表示为：

$$\boldsymbol{r}(\boldsymbol{x})=\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x}) \tag{2}$$

其中$\boldsymbol{r}=[r_1,r_2,\cdots,r_m,\cdots,r_M]^{\mathrm{T}}$。

状态估计的最优结果就是根据电力系统非线性网络拓扑结构求得一组使得残差向量$\boldsymbol{r}$的各分量均为零的状态变量$\hat{\boldsymbol{x}}$，即满足

$$\boldsymbol{r}(\hat{\boldsymbol{x}})=0 \tag{3}$$

通过SCADA系统采集到的量测数据并不能完全可信，需要对其收集来的数据进行不良数据检测(Bad Data Detection，BDD)。其原理为通过残差的欧几里得范数判断是否存在不良数据[16]，定义为：

$$d=\left\|\boldsymbol{r}(\boldsymbol{x})\right\|_2=\left\|\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right\|_2 \tag{4}$$

通过阈值τ来判断是否存在不良数据。当$d<\tau$时，判断为正常数据；若$d\geqslant\tau$，则存在不良数据。

FDIA的主要思想是对电力系统量测向量$\boldsymbol{z}$添加非零攻击向量$\boldsymbol{a}=[a_1,a_2,\cdots,a_m,\cdots,a_M]^{\mathrm{T}}$的方法来

篡改电力系统量测向量[17]。假设攻击者构造的攻击向量有效,其电力系统数学模型可以表示为:

$$\boldsymbol{z}_a = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{e} + \boldsymbol{a} \tag{5}$$

其中:$\boldsymbol{z}_a$表示被FDIA注入攻击向量后的量测向量,且$\boldsymbol{a}$服从均值为$\boldsymbol{\mu}_a$、协方差为$\boldsymbol{\Sigma}_a$的高斯分布模型。

若能量管理系统(Energy Management Systems, EMS)使用$\boldsymbol{z}_a$进行状态估计得出错误状态变量$\hat{\boldsymbol{x}}_{\text{bad}}$,其与攻击前的状态变量$\hat{\boldsymbol{x}}$的关系可以表示为:

$$\hat{\boldsymbol{x}}_{\text{bad}} = \hat{\boldsymbol{x}} + \boldsymbol{c} \tag{6}$$

其中:向量$\boldsymbol{c} = [c_1, c_2, \cdots, c_n, \cdots, c_N]^{\mathrm{T}}$是添加到原始状态估计中的恶意错误数据。

若$\boldsymbol{a} = \boldsymbol{h}(\hat{\boldsymbol{x}} + \boldsymbol{c}) - \boldsymbol{h}(\hat{\boldsymbol{x}})$,则其BDD结果为:

$$\begin{aligned}\left\| \boldsymbol{z}_a - \boldsymbol{h}(\hat{\boldsymbol{x}}_{bad}) \right\|_2 &= \left\| \left(\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}}) \right) + \right. \\ &\left. \left(\boldsymbol{a} + \boldsymbol{h}(\hat{\boldsymbol{x}}) - \boldsymbol{h}(\hat{\boldsymbol{x}} + \boldsymbol{c}) \right) \right\|_2 = \left\| \boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{x}}) \right\|_2\end{aligned} \tag{7}$$

从式(7)可以看出,攻击后残差不变,骗过了BDD系统。一般来说,攻击者很难构造出满足其目标的攻击向量;即使攻击者成功构造出该攻击向量,其对智能电网持续进行量测数据篡改的过程中,也几乎不可能使得虚假量测数据的误差符合量测单元的量测误差。因此,可以从这个角度对智能电网中的FDIA进行检测。

2 FDIA检测与状态估计

2.1 基于最大似然估计的FDIA检测

对于进行状态估计前SCADA系统获得的$L(L>1)$次量测向量数据$(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_l, \cdots, \boldsymbol{z}_L)$均是独立同分布(Independently Identical Distribution, IID)的,其每次取得量测向量的概率密度函数(Probability Density Function,PDF)均可以表示为:

$$\begin{aligned}p(\boldsymbol{z}) = &\frac{1}{\sqrt{(2\pi)^M |\boldsymbol{R}|}} \times \\ &\exp\left(-\frac{(\boldsymbol{z} - \boldsymbol{\mu}_z)^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_z)}{2} \right)\end{aligned} \tag{8}$$

为了降低量测数据误差,基于其服从高斯分布的特性,可以利用MLE算法对量测向量进行预处理,求得量测数据集的期望$\boldsymbol{\mu}_z$与协方差$\boldsymbol{\Sigma}_z$。

MLE算法在数据处理的相关领域中的应用十分广泛,该方法的原理是在已知一组数据的PDF形式的情况下,针对这组数据,可以构造其对数似然函数。通过最大化该似然函数,可以求得该组数据PDF所包含矢量参数的估计值。根据本文所提出的问题和方法,样本数据的PDF属于多元高斯分布,并依据此模型对其进行参数估计。对于随机量测向量的数据样本集$(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_l, \cdots, \boldsymbol{z}_L)$,其估计参数为$\boldsymbol{\theta} = [\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z]^{\mathrm{T}}$的对数似然函数可以表示为:

$$\begin{aligned}L(\boldsymbol{\theta}) = \ln\left(\prod_{l=1}^{L} p(\boldsymbol{z}_l) \right) = &-\frac{ML}{2} \ln 2\pi - \\ &\frac{L}{2} \ln |\boldsymbol{\Sigma}_z| - \frac{1}{2} \sum_{l=1}^{L} (\boldsymbol{z}_l - \boldsymbol{\mu}_z)^{\mathrm{T}} \boldsymbol{\Sigma}_z^{-1} (\boldsymbol{z}_l - \boldsymbol{\mu}_z)\end{aligned} \tag{9}$$

为了求得参数$\boldsymbol{\theta}$,需求得使式相应估计参数一阶导函数为0的值,即求得以下函数的解:

$$\begin{cases} \dfrac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_z} = 0 \\ \dfrac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_z} = 0 \end{cases} \tag{10}$$

得到:

$$\boldsymbol{\mu}_z = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{z}_l \tag{11}$$

$$\boldsymbol{\Sigma}_z = \frac{1}{L} \sum_{l=1}^{L} (\boldsymbol{z}_l - \boldsymbol{\mu}_z)(\boldsymbol{z}_l - \boldsymbol{\mu}_z)^{\mathrm{T}} \tag{12}$$

以此可以得到更精准的量测向量作为下一步状态估计的输入。而且,根据得到的量测向量的协方差矩阵$\boldsymbol{\Sigma}_z$可以作为对量测数据是否遭受FDIA的简单判断。使用FDIA的攻击者为了达成其目标,需要对电网控制中心SCADA系统接收的量测数据进行持续的篡改,但是篡改后的虚假量测数据很难符合实际的历史量测数据误差分布:

$$\boldsymbol{\Sigma}_a \neq \boldsymbol{R} \tag{13}$$

据此,可以通过MLE算法对量测数据进行处理,根据所得误差分布模型来判断是否存在虚假数据。

若没有检测出FDIA所篡改的量测数据,说明该量测数据正常,可以进行下一步状态估计的过程。而通过MLE算法对量测数据进行预处理后得

到的数据期望$\boldsymbol{\mu}_z$大幅降低了各种因素所导致的误差,最接近智能电网相关状态量测真值。使用该量测期望进行状态估计,可以降低所得结果的估计误差。

2.2　基于加权最小二乘法的状态估计

在判断一个估计量的好坏时,为了求得一个无偏的最接近真实数据的估计量,一般选择方差作为对估计量性能好坏的度量标准。而对于使用状态估计的交流电力系统来说,其具有数据冗余和非线性的特点。因此,以牛顿法为基本原理的加权最小二乘滤波算法(Weighted Least Squares,WLS)可以达到电力系统状态估计的要求。

在WLS算法的状态估计中,为了求得状态变量的最小无偏估计,引入了量测向量误差的目标函数:

$$J(\boldsymbol{x})=\sum_{m=1}^{M}\left(z_m-h_m(x)\right)^2\Big/\sigma_m^2=\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right]^{\mathrm{T}}\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right] \tag{14}$$

为了求得使目标函数$J(\boldsymbol{x})$达到最小的$\hat{\boldsymbol{x}}$的值,通过令目标函数的一阶导为0来取得极值,其导函数表示为:

$$\begin{cases}\dfrac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}=-2\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right]=0\\ \boldsymbol{H}(\boldsymbol{x})=\dfrac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\end{cases} \tag{15}$$

其中$\boldsymbol{H}(\boldsymbol{x})$为电力系统雅克比矩阵函数。

令$f(\boldsymbol{x})=\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right]=0$,针对$\boldsymbol{h}(\boldsymbol{x})$作为非线性矢量函数的特点,使用牛顿法进行求解,其一阶泰勒展开表示为:

$$f(\boldsymbol{x})\approx f(\boldsymbol{x}_0)+\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\Delta\hat{\boldsymbol{x}}=0 \tag{16}$$

求得该函数变量$\Delta\hat{\boldsymbol{x}}$的解为:

$$\Delta\hat{\boldsymbol{x}}=\left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right]^{-1}f(\boldsymbol{x}_0)=\left[\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{x})\right]^{-1}\times\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right] \tag{17}$$

其牛顿法的第$(\eta+1)$次迭代过程可以表示为:

$$\Delta\hat{\boldsymbol{x}}^{(\eta)}=\left[\boldsymbol{H}^{\mathrm{T}}\left(\hat{\boldsymbol{x}}^{(\eta)}\right)\boldsymbol{R}^{-1}\boldsymbol{H}\left(\hat{\boldsymbol{x}}^{(\eta)}\right)\right]^{-1}\times\boldsymbol{H}^{\mathrm{T}}\left(\hat{\boldsymbol{x}}^{(\eta)}\right)\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}\left(\hat{\boldsymbol{x}}^{(\eta)}\right)\right] \tag{18}$$

$$\hat{\boldsymbol{x}}^{(\eta+1)}=\hat{\boldsymbol{x}}^{(\eta)}+\Delta\hat{\boldsymbol{x}}^{(\eta)} \tag{19}$$

直到目标函数$J(\hat{\boldsymbol{x}})$无限接近于最小值为止,其作为收敛结束标准的判据为:

$$\left|\Delta\hat{x}_n^{(\eta)}\right|_{\max_n}<\varepsilon_x \tag{20}$$

其中:$\left|\Delta\hat{x}_n^{(\eta)}\right|_{\max_n}$表示第$\eta$次迭代计算中状态修正量绝对值最大者小于给定的阈值,ε_x可取基准电压幅值的10^{-4}~10^{-6},这是实际使用中最常用的标准[15]。

WLS算法状态估计误差表示为:

$$\boldsymbol{x}-\hat{\boldsymbol{x}}=-\left[\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{x})\right]^{-1}\times\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{x})\right] \tag{21}$$

WLS算法状态估计方差阵表示为:

$$\mathrm{E}\left[(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^{\mathrm{T}}\right]=\left[\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{x})\right]^{-1} \tag{22}$$

由于真值$\boldsymbol{x}$是未知的,实际计算时使用$\hat{\boldsymbol{x}}$代替。

对量测数据进行FDIA检测与状态估计的流程如图1所示。从图1中可以看出,根据MLE算法所得的结果,可以通过$\boldsymbol{\Sigma}_z=\boldsymbol{R}$判断该组量测数据中是否存在虚假数据,进而检测出是否存在FDIA。在数据正常的情况下,通过WLS算法进行状态估计,进而得到电网的系统状态。

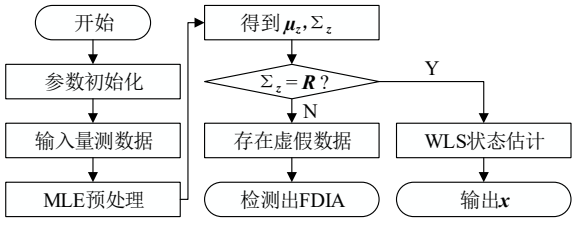


图1　基于MLE-WLS算法的工作流程

3　算法复杂度分析

算法复杂度决定了算法运行所需要消耗的硬件计算资源。一般来说,复杂度越低,算法运行速度越快,计算出错的概率越低。在复杂度分析中,本文重点关注MLE算法对数据的预处理过程以及WLS算法状态估计的迭代过程,因为它们消耗的计算量最大。复杂度用浮点运算数(Floating Point

Operations, FLOPs)来评价,一些基本操作所需的FLOPs定义[18]如下:

(1)ε_{add}为加法运算FLOPs。

(2)ε_{sub}为减法运算FLOPs。

(3)ε_{mul}为乘法运算FLOPs。

(4)ε_{div}为除法运算FLOPs。

上述操作所需的FLOPs可能因计算处理器性能的不同而异。

对于MLE算法的运算,其所需要的FLOPs为:

$$FL_{\text{MLE}} = M \times L \times \varepsilon_{\text{add}} + M \times \varepsilon_{\text{sub}} + M^2 \times \varepsilon_{\text{mul}} + (M^2 + M) \times \varepsilon_{\text{div}} \tag{23}$$

定义WLS算法一次迭代需要的FLOPs为:

$$\begin{aligned} FL_{\text{WLS}} = {} & \left(2N^3 + (M-3)N^2 + (M^2+1)N - 2\varepsilon_{\text{add}} + (N^2 + M + 1)\right)\varepsilon_{\text{sub}} + \\ & (2N^4 - 6N^3 + (M+7)N^2 + \\ & (M^2 + M - 2)\Big)N\varepsilon_{\text{mul}} + N^2\varepsilon_{\text{div}} \end{aligned} \tag{24}$$

其中:M为量测向量Z中量测量的个数;N为状态变量x中状态量的个数;L为系统中量测向量的量测次数。

本文假设WLS算法达到收敛所需要的迭代次数为$N_{\text{itr}}^{\text{WLS}}$。最终用于状态估计所需要的FLOPs约为:

$$FL(\boldsymbol{x}) \approx FL_{\text{MLE}} + N_{\text{itr}}^{\text{WLS}} FL_{\text{WLS}} \tag{25}$$

基于前面的算法介绍与分析,归纳出本文算法的详细过程见表1。

表1 基于MLE-WLS算法的FDIA检测和状态估计过程

步骤	详细过程
1	输入:量测数据$(z_1,\cdots,z_l,\cdots,z_L)$
2	初始化:设置WLS算法收敛阈值$\varepsilon_x = 1\times10^{-4}$,迭代指数$\eta = 0$
3	MLE算法: (1)根据式(11)计算$\boldsymbol{\mu}_z$; (2)根据式(12)计算$\boldsymbol{\Sigma}_z$; (3)比较$\boldsymbol{\Sigma}_z$与$\boldsymbol{R}$是否一致,若$\boldsymbol{\Sigma}_z \neq \boldsymbol{R}$,说明量测数据中存在虚假数据,终止算法;否则进行下一步。
4	WLS算法循环: (1)根据式(17)更新$\Delta\hat{\boldsymbol{x}}^{(\eta)}$,其中$\boldsymbol{z} = \boldsymbol{\mu}_z$; (2)根据式(19)更新$\hat{\boldsymbol{x}}^{(\eta+1)}$; (3)若满足收敛条件式(20),则终止WLS算法;否则,设置$\eta \leftarrow \eta+1$,继续WLS算法循环。
5	输出:$\hat{\boldsymbol{x}}$

4 算例分析

4.1 仿真系统

本文使用如图2所示的IEEE-14节点系统进行仿真实验,基于MATPOWER电力仿真包获取电网物理参数。使用MATLAB R2018b软件进行模型搭建并实验。

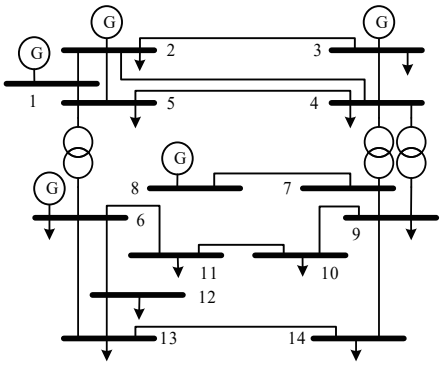


图2 IEEE-14节点系统拓扑结构

在IEEE-14节点系统状态估计过程中,功率量测个数$M = 40$,其中包含节点2、3、7、8、10、11、12、14的注入有功与无功功率,以及输电支路1-2、2-3、4-2、4-7、4-9、5-2、5-4、5-6、6-13、7-9、11-6、12-13的损耗有功与无功功率;状态估计个数$N = 28$,包含14个节点的电压幅值与相角。设置节点和输电线路有功与无功功率的量测误差见表2。

表2 电力系统量测误差

量测类型	量测误差σ/MW
P_i/Q_i	1.0
P_{ij}/Q_{ij}	0.8

4.2 FDIA检测分析

假设攻击者使用FDIA对电力系统量测单元注入的虚假数据的误差分布见表3。本文通过随机选择系统量测节点注入虚假数据验证该方法的有效性。根据蒙特卡洛方法生成1000组具有虚假数据的量测向量作为实验数据。

表3 虚假数据误差

量测类型	量测误差σ/MW
P_i/Q_i	0.5
P_{ij}/Q_{ij}	0.4

通过 MLE 对 FDIA 篡改的实验数据进行处理得到的量测误差如图 3 所示。从图 3 中可以看出，系统量测节点 2、9、10、17、18、19、29、30、31 中的量测数据经过 MLE 算法处理后得到的量测误差与理论误差不符，说明 P_3、Q_2、Q_3、$P_{1\text{-}2}$、$P_{2\text{-}3}$、$Q_{2\text{-}3}$、$Q_{1\text{-}2}$、$Q_{4\text{-}2}$ 处量测单元被注入了虚假数据。

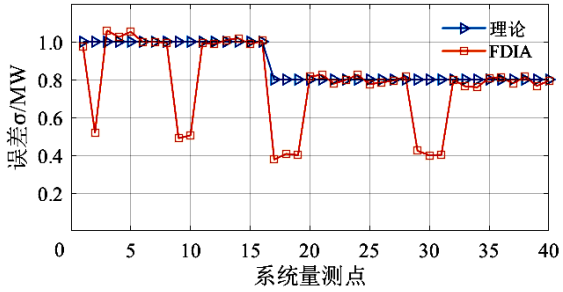


图 3　虚假量测数据误差

4.3　状态估计结果分析

根据蒙特卡洛方法依据表 2 生成 1000 组量测向量作为实验数据集。通过 MLE 算法对该数据集进行预处理得到的量测误差如图 4 所示，从图 4 中可以看出，该量测误差基本与理论误差一致，说明不存在 FDIA 所篡改的虚假数据。

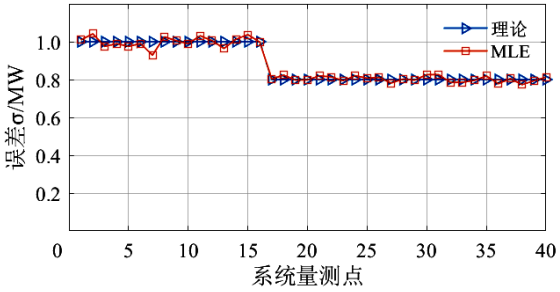
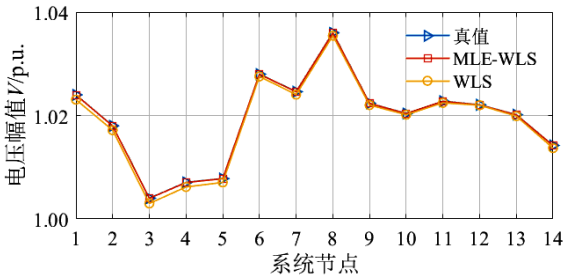
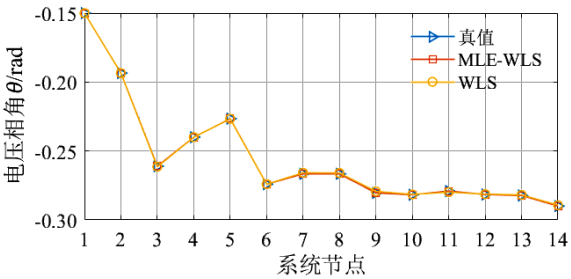


图 4　正常量测数据误差

系统节点状态估计结果如图 5 所示。从图 5 中可以看出，与原始 WLS 算法状态估计相比，经 MLE 算法对量测数据预处理后的节点电压幅值的状态估计结果更加接近状态真值。



(a) 系统节点电压幅值状态估计

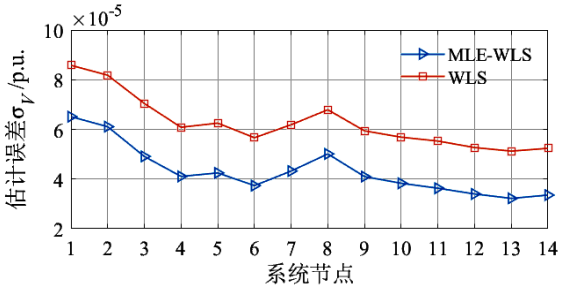


(b) 系统节点电压相角状态估计

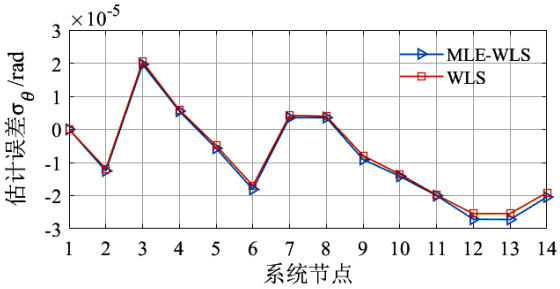
图 5　系统节点状态估计结果

系统节点状态的估计误差如图 6 所示。从图 6 中可以看出，MLE 算法与 WLS 算法的联合使用使节点电压幅值的估计误差明显降低了，且状态估计结果残差的欧几里得范数 d 经计算也从 0.0329 降到了 0.0011，说明其结果更加符合电网状态。

本文实验所用电脑 CPU 型号为 IntelR CoreTM i5-9500 CPU@3.00 GHz 8 GB RAM。基于蒙特卡洛模拟，联合 MLE 算法与 WLS 算法进行状态估计的 1000 次重复仿真实验的运行时间如图 7 所示，根据其正态分布曲线可知，对于 IEEE-14 节点测试系统来说，基本可以在 0.028 061 s 内完成状态估计过程。



(a) 系统节点电压幅值估计误差



(b) 系统节点电压相角估计误差

图 6　系统节点状态的估计误差

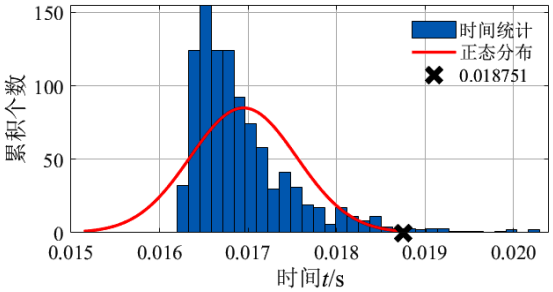


图7 1000次状态估计仿真实验运行时间直方图

5 结束语

本文通过分析电力系统控制中心SCADA系统所获得的量测数据与FDIA所篡改后的虚假数据的特点,基于IEEE-14节点系统,用MATLAB软件建立仿真模型进行实验。通过使用MLE算法对量测数据集进行预处理,得到其符合多元高斯分布模型的估计参数。其协方差矩阵可以用来检测虚假量测数据,而数据期望作为更接近电网量测真值的量测向量,在引入WLS算法状态估计的过程中可以得到更好的估计结果。因此,MLE算法与WLS算法的联合可以在通过提高状态估计精度的同时对FDIA进行检测。

参考文献:

[1] SCHWEPPE F C,WILDES J.Power system static-state estimation,Part I:exact model[J].IEEE Transactions on Power Apparatus and Systems,1970,89(1):120-125.

[2] 陈艳波,高瑜珑,赵俊博,等.综合能源系统状态估计研究综述[J].高电压技术,2021,47(7):2281-2292.

[3] 臧海祥,耿明昊,黄蔓云,等.电-热-气混联综合能源系统状态估计研究综述与展望[J].电力系统自动化,2022,46(7):187-199.

[4] LIU Y,NING P,REITER M K.False data injection attacks against state estimation in electric power grids[J].ACM Transactions on Information and System Security (TISSEC),2011,14(1):1-33.

[5] 赵俊华,梁高琪,文福拴,等.乌克兰事件的启示:防范针对电网的虚假数据注入攻击[J].电力系统自动化,2016,40(7):149-151.

[6] 杨智伟,刘灏,毕天妹,等.基于长短期记忆网络的PMU不良数据检测方法[J].电力系统保护与控制,2020,48(7):1-9.

[7] 倪明,颜诘,柏瑞,等.电力系统防恶意信息攻击的思考[J].电力系统自动化,2016,40(5):148-151.

[8] LUN Y Z,D'INNOCENZO A,SMARRA F,et al.State of the art of cyber-physical systems security:an automatic control perspective[J].Journal of Systems and Software,2019,149:174-216.

[9] 王电钢,黄林,刘捷,等.考虑负荷虚假数据注入攻击的电力信息物理系统防御策略[J].电力系统保护与控制,2019,47(1):28-34.

[10] 汤奕,王琦,倪明,等.电力信息物理融合系统中的网络攻击分析[J].电力系统自动化,2016,40(6):148-151.

[11] KURT M N,YILMAZ Y,WANG X.Real-time detection of hybrid and stealthy cyber-attacks in smart grid[J].IEEE Transactions on Information Forensics and Security,2018,14(2):498-513.

[12] KARIMIPOUR H,DINAVAHI V.Robust massively parallel dynamic state estimation of power systems against cyber-attack[J].IEEE Access,2018,6:2984-2995.

[13] KHALID H M,PENG C H.Immunity toward data-injection attacks using multisensor track fusion-based model prediction[J].IEEE Transactions on Smart Grid,2017,8(2):697-707.

[14] OZAY M,ESNAOLA I,VURAL F,et al.Machine learning methods for attack detection in the smart grid[J].IEEE Transactions on Neural Networks & Learning Systems,2015,27(8):1773-1786.

[15] 于尔铿.电力系统状态估计[M].北京:水利电力出版社,1985.

[16] 王玲,邓志,马明,等.基于状态估计残差比较的配电网故障区段定位方法[J].电力系统保护与控制,2021,49(14):132-139.

[17] 王琦,邰伟,汤奕,等.面向电力信息物理系统的虚假数据注入攻击研究综述[J].自动化学报,2019,45(1):72-83.

[18] YIN F,FRITSCHE C,GUSTAFSSON F,et al.EM- and JMAP-ML based joint estimation algorithms for robust wireless geolocation in mixed LOS/NLOS environments[J].IEEE Transactions on Signal Processing,2013,62(1):168-182.

引用格式:

中　文:吴铭辉,高文根,华峰,等.基于最大似然估计的智能电网FDIA检测[J].四川轻化工大学学报(自然科学版),2023,36(2):38-45.

英　文:WU M H,GAO W G,HUA F,et al.Detection of FDIA in smart grid based on maximum likelihood[J].Journal of Sichuan University of Science & Engineering (Natural Science Edition),2023,36(2):38-45.

Detection of FDIA in Smart Grid Based on Maximum Likelihood

WU Minghui, GAO Wengen, HUA Feng, MU Shaopeng

（Key Laboratory of Detection Technology and Energy Saving Devices of Anhui Province,
Anhui Polytechnic University, Wuhu 241000, China）

Abstract: The normal operation of smart grid depends on the state estimation that accurately reflects the physical characteristics of power grid. Aiming at the problem that false data injection attack (FDIA) tampers with the state estimation results by injecting malicious data into the power system measurement unit, a power grid FDIA detection method based on maximum likelihood estimation (MLE) has been proposed, which improves the accuracy of state estimation results. Firstly, based on the characteristics that smart grid measurement vector and FDIA attack vector obey multivariate Gaussian distribution with different covariance, the expectation and covariance of the measurement data are obtained by the MLE algorithm, and whether there is false measurement data is judged according to the covariance. Secondly, if the data is normal, better system state results can be obtained by using the weighted least square (WLS) algorithm to estimate the state according to the expectation of the measured data. Finally, the feasibility of the algorithm is proved using the example that is based on the IEEE-14 node system.

Key words: smart grid; state estimation; false data injection attack; attack detection