基于智能电表量测的配电网非侵入式窃电检测

韩建富[1] 肖 春[2] 贾探喜[3] 张建民[2] 王飞飞[1] 谭沛然[2]

[1](国网山西省电力公司 山西 太原 030000)

[2](国网山西省电力公司营销服务中心 山西 太原 030000)

[3](国网山西省电力公司吕梁供电公司 山西 太原 030000)

摘 要 先进的传感技术和通信功能在改善量测和控制功能的同时,也使得配电网面临着各种可能的网络攻击。当攻击者通过特定手段篡改其他用户和自身量测记录,从而实现在维持总电量数据不变情况下的窃电行为,对于这种情况传统基于电量平衡的窃电监测方法难以有效识别。因此,基于智能电表量测数据的高阶统计信息,提出一种基于智能电表量测的配电网非侵入式窃电检测方法,实现对攻击发起方和受害方进行检测。实验证明,所提方法通过利用更高阶的用电量统计信息高效地检测到非侵入式攻击,并能够识别出所涉及的用户(攻击者和受害者),获得了理想的效果。

关键词 错误数据注入机制 高阶信息 智能电表 阈值协方差矩阵

中图分类号 TP3 TM761 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2024.09.018

NON-INVASIVE DETECTION METHOD FOR ELECTRIC LARCENY IN DISTRIBUTION NETWORK BASED ON SMART METER MEASUREMENT

Han Jianfu[1] Xiao Chun[2] Jia Tanxi[3] Zhang Jianmin[2] Wang Feifei[1] Tan Peiran[2]

[1](*State Grid Shanxi Electric Power Company*, *Taiyuan 030000*, *Shanxi*, *China*)

[2](*Marketing Service Center of State Grid Shanxi Electric Power Company*, *Taiyuan 030000*, *Shanxi*, *China*)

[3](*Lüliang Power Supply Company of State Grid Shanxi Electric Power Company*, *Taiyuan 030000*, *Shanxi*, *China*)

Abstract While the advanced sensing and communication functions improve the measurement and control functions, the distribution network is also faced with various possible network attacks. When the attacker tampers with other users and their own measurement records by specific means, so as to achieve the electricity stealing behavior under the condition of keeping the total electricity data unchanged, the traditional electricity stealing monitoring method based on power balance is difficult to effectively identify. Therefore, on the basis of the high-order statistical information of smart meter measurement data, this paper proposes a non-invasive detection method of power larceny in distribution network based on smart meter measurement. It can detect the attacker and the victim. Experiments show that the proposed method can effectively detect non-invasive attacks by using higher-order statistics of power consumption, and can identify the users (attackers and victims) involved.

Keywords False data injection mechanism Higher-order information Smart meter Threshold covariance matrix

0 引 言

窃电行为一直是世界范围内电网运营商面临的主要问题,它给电力公司造成了巨大的收入损失[1]。智能电表的引入使得可以通过高频监控功能来降低窃电的风险,并通过负荷预测技术来对故障情况和网络可控性进行及时地故障排除[2]。但与此同时,智能电表也为网络攻击者通过本地网络攻击,采用错误数据注入的方式来远程篡改电网运行数据提供新机会,其后

收稿日期:2020-12-18。**韩建富**,高工,主研领域:电力计量技术。**肖春**,工程师。**贾探喜**,工程师。**张建民**,高工。**王飞飞**,工程师。**谭沛然**,工程师。

果将会损害目标台区的需求响应方案、危害电网状态估计结果甚至导致系统停电[3]。

目前对于窃电问题的研究主要有两大类,一类是致力于根据智能电表的历史数据,使用机器学习和数据挖掘技术来检测用户的异常用电模式。这类方法包括在训练数据中利用标签样本(已知错误数据注入与非错误数据注入)的有监督方法,以及试图从正常用电模式中识别异常的无监督方法[4]。有监督的方法可能很有效,但是错误标签注入样本的可用性仍然是一个巨大的挑战,而无监督的方法容易受到季节变化、设备变化和用户变化等改变用电模式的非恶意因素的影响[5]。

另一类方法则是利用智能配电网中邻域网络架构的有关信息[6]。具体而言,将每个台区看作是“电力路由器”,变电站借助于“电力路由器”将电力分配给台区内的用户,由一个主智能电表(称为收集器)在一定时间间隔内采集总供电量,同时对于台区内每个用户也利用其所安装的智能电表在相同的时间间隔内记录其用电量。文献[7]提出了一种利用这种量测方法以及有关将消耗点连接到配电变压器线路的电阻信息方法,以估算由于低压电力线造成的技术损失以及变压器的固有效率低下。文献[8]采用线性回归框架来识别窃电,其中因变量对应于收集器的总量测值,而预测变量对应于用户智能电表的量测值。但是,这种思路实际上是假设预测变量之间是不相关的,这与实际窃电行为并不相符。

在本文中,考虑窃电者(攻击者)试图通过向自身的智能电表注入错误量测值来降低实际用电量,但为避开电力公司的电量检测,需要向邻域网中的另一台智能电表进行补偿注入。因此,在考虑来自中心节点的总量测值或用于终端用户智能电表时,尤其是如果攻击者注入了小幅度的虚假数据,各种基于机器学习的监控方案(专注于平均用电模式的变化)将无法检测到此类攻击。由此,本文在考虑终端用户间用电模式相关联情况下,通过检查量测值之间的相关性来检测窃电行为的攻击者和受害者。

1 问题建模

令 $\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_P$ 对应于一个时间间隔内智能电表的量测数据,并假设 $\boldsymbol{Y}_i=\rho_i\boldsymbol{W}+\boldsymbol{U}_i$,其中 $E(\boldsymbol{W})=\mu_w$,$Var(\boldsymbol{W})=\sigma_w,\rho\in(-1,1)$。从而表示智能电表的量测数是非独立的,即 $Cor(\boldsymbol{Y}_i,\boldsymbol{Y}_j)=sqrt(\rho_i\rho_j),\forall i=j$。每个量测度量 $\boldsymbol{Y}_i$ 的独立项 $\boldsymbol{U}_i$ 服从分布 F,其一阶和二阶矩分别为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\sigma}$。用 $\boldsymbol{Y}=(\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_P)'$ 表示量测值的协方差矩阵。令 $\boldsymbol{Z}$ 表示由电力公司控制的收集器节点(例如配电变压器智能电表)处的量测数据,当没有由于功率分配和传输问题造成的技术损失,根据定义有 $\boldsymbol{Z}=\mathrm{sum}(\boldsymbol{Y}_i),i=1,2,\cdots,P$。由于窃电者要在保持台区用电总量不变的情况下降低自身的用电量数值,因此势必会将其他用户的用电量数值增加,为了简化,本文中将智能电表的电量量测结果降低定义为攻击者,将用电量量测结果增加定义为受害者。

1.1 独立攻击事件识别

首先考虑智能电表的量测是完全独立的这一特殊情况,则有 $E(\boldsymbol{Y}_i)=\boldsymbol{\mu},Var(\boldsymbol{Y}_i)=\boldsymbol{\sigma}$,协方差矩阵 $\boldsymbol{\Sigma}$ 为对角矩阵通过检查涉及单个受害者节点的攻击场景来开始分析。

(1) 成对攻击情况。该情况下仅涉及一名攻击者和一名受害者,即节点 i(攻击者)和节点 j(受害者),其对应智能电表量测值分别为 $\boldsymbol{Y}_i-\boldsymbol{\alpha}$ 和 $\boldsymbol{Y}_j+\boldsymbol{\alpha}$。检测攻击以及识别所涉及的节点的关键是检查功耗量测的较高矩信息。在成对攻击情况下,让节点 i 与受害节点 j 发起大小为 $\boldsymbol{\alpha}_e$ 的攻击,类似地让攻击者节点 k 和受害节点 l 发起另一大小为 $\boldsymbol{\alpha}_f$ 的攻击。用 $Var(\boldsymbol{\alpha}_e)=\boldsymbol{\sigma}_{\alpha e}$ 和 $Var(\boldsymbol{\alpha}_f)=\boldsymbol{\sigma}_{\alpha f}$ 表示,则协方差矩阵 $\boldsymbol{\Sigma}$ 的将从对角矩阵变为块对角矩阵:

$$\boldsymbol{\Sigma}'=\begin{bmatrix}\boldsymbol{\sigma}+\boldsymbol{\sigma}_\alpha & -\boldsymbol{\sigma}_\alpha & 0 & \cdots & 0\\ -\boldsymbol{\sigma}_\alpha & \boldsymbol{\sigma}+\boldsymbol{\sigma}_\alpha & 0 & \cdots & 0\\ 0 & 0 & \boldsymbol{\sigma} & \cdots & 0\\ \vdots & \vdots & \vdots & & 0\\ 0 & 0 & 0 & \cdots & \boldsymbol{\sigma}\end{bmatrix} \tag{1}$$

类似的,如果存在涉及不同的两节点成对攻击,第一个被指定为攻击者,第二个被指定为受害者,则智能电表量测的相应协方差矩阵可表示为:

$$\boldsymbol{\Sigma}'=\begin{bmatrix}\boldsymbol{B}_1 & & & & & & \\ & \boldsymbol{B}_2 & & & & & \\ & & \ddots & & & & \\ & & & \boldsymbol{B}_s & & & \\ & & & & \boldsymbol{\sigma} & & \\ & & & & & \ddots & \\ & & & & & & \boldsymbol{\sigma}\end{bmatrix} \tag{2}$$

式中:$\boldsymbol{B}_h$ 是攻击 $\boldsymbol{\alpha}_h$ 的子协方差块矩阵,$h=1,2,\cdots,s$。

由此,如果 $Cor(\boldsymbol{Y}_i,\boldsymbol{Y}_j)<0,i\neq j$,则节点 i 和 j 参与同一攻击,它们属于同一攻击组;如果 $Cor(\boldsymbol{Y}_i,\boldsymbol{Y}_j)=0,i\neq j$,则不存在涉及节点 i 和 j 的攻击,它们属于不同的攻击组。为了进一步识别出在成对攻击中哪一方是攻

击者哪一方是受害者,需要有关智能电表量测的第三时刻信息。在成对攻击情况下,如果终端节点 i 和 j 处于同一攻击组,且幅度为 $\boldsymbol{\alpha}$,则 $E(\boldsymbol{Y}_i-\boldsymbol{\alpha})^3-E(\boldsymbol{Y}_j+\boldsymbol{\alpha})^3>0$。这表明在同一攻击组中,受害者节点的三阶矩严格大于攻击者节点的三阶矩。

(2) 一对多攻击情况。当攻击者节点旨在使用较大的 $\boldsymbol{\alpha}$,则可以使用更改点分析技术来标记此动作,因为受害者节点的用电方式就会发生,或者受到攻击的用户会反过来向公用事业公司抱怨其电费急剧增加。在这种情况下,攻击者可能希望将攻击分散到更大的用户节点组中,以免引起怀疑。这导致了一个更复杂的情况,即攻击者节点将节点 i 处的智能电表量测值减少一个 $\boldsymbol{\alpha}$,并以相等的量累计增加受害者组智能电表的量测。具体而言,当节点 i 发起大小为 $\boldsymbol{\alpha}$ 的攻击时,即 $\boldsymbol{Y}_i-\boldsymbol{\alpha}$,而对于多个受害者节点,则有 $\boldsymbol{Y}_{j1}+k_1{}^{\alpha}\boldsymbol{\alpha}$,$\boldsymbol{Y}_{j2}+k_2{}^{\alpha}\boldsymbol{\alpha},\cdots,\boldsymbol{Y}_{jl}+k_l{}^{\alpha}\boldsymbol{\alpha}$,其中 $\mathrm{sum}(k_j{}^{\alpha})=1$。

同理,协方差矩阵为块对角矩阵:

$$\boldsymbol{\Sigma}'=\begin{bmatrix}\boldsymbol{B}_1 & & & & & \\ & \ddots & & & & \\ & & \boldsymbol{B}_s & & & \\ & & & \boldsymbol{\sigma} & & \\ & & & & \ddots & \\ & & & & & \boldsymbol{\sigma}\end{bmatrix}_{P\times P} \tag{3}$$

式中:$\boldsymbol{B}_h$ 是对应于第 h 次攻击的协方差矩阵,$h=1,2,\cdots,s$,每个块 $\boldsymbol{B}_h$ 具有以下形式:

$$\boldsymbol{B}_h=\boldsymbol{\Sigma}+\boldsymbol{\sigma}_{\alpha_h}\begin{bmatrix}1 & -k_1^{\alpha_h} & \cdots & -k_{d_\mathrm{h}}^{\alpha_h}\\ -k_1^{\alpha_h} & (k_1^{\alpha_h})^2 & \cdots & k_1^{\alpha_h}k_{d_\mathrm{h}}^{\alpha_h}\\ \vdots & \vdots & & \vdots\\ -k_{d_\mathrm{h}}^{\alpha_h} & k_1^{\alpha_h}k_{d_\mathrm{h}}^{\alpha_h} & \cdots & (k_{d_\mathrm{h}}^{\alpha_h})^2\end{bmatrix} \tag{4}$$

式中:$\boldsymbol{\Sigma}$ 为 $(d_\mathrm{h}+1)\times(d_\mathrm{h}+1)$ 维矩阵,d_h 是该组中受害者的数量,$\mathrm{sum}(d_\mathrm{h}+1)=m$。

因此,在多对一攻击情况下,如果 $Cor(\boldsymbol{Y}_i,\boldsymbol{Y}_j)\neq0$,$i\neq j$,则终端节点 i 和 j 属于同一攻击组;如果 $Cor(\boldsymbol{Y}_i,\boldsymbol{Y}_j)=0$,$i\neq j$,则节点 i 和 j 属于不同的攻击组。在这种情况下,攻击机制的参与程度更高,一旦确定了攻击组,就可以很容易地将攻击者节点与受害者节点分开。也即只有攻击者节点会与同一攻击组中的所有其他节点显示负协方差值;即 $Cor(\boldsymbol{Y}_{i0},\boldsymbol{Y}_j)<0$,$j\neq i_0$;另一方面,同一攻击组中的所有受害者节点将彼此具有正协方差值。由此,对于每个攻击组,要识别出攻击者节点和受害者节点只需要先识别块内仅具有负协方差值的节点,将其作为攻击者节点,再将块中彼此具有正协方差值的其余节点标记为受害者节点即可。

1.2　非独立攻击事件识别

更一般的,当智能电表量测值表示为 $\boldsymbol{Y}_i=\rho_i\boldsymbol{W}+\boldsymbol{U}_i$,定义 $\boldsymbol{X}_{ij}=\boldsymbol{Y}_i-\boldsymbol{Y}_j$,$i,j=1,2,\cdots,P$ 且 $i\neq j$。通过这组新的 $(P-1)^2$ 个量测变量,借助于 1.1 节的结论,可以得到当 $i\neq j\neq k\neq l$ 时,$Cov(\boldsymbol{X}_{ij},\boldsymbol{X}_{kl})<<0$。这表明原始量测值集的差分变换会在很大程度上降低它们的相关性,从而可以通过这一判据来识别攻击。为了说明这一点,从最一般的情况开始。假设对于任意 $i\neq j\neq k\neq l$,有四种不同的攻击 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、$\boldsymbol{\alpha}_3$ 和 $\boldsymbol{\alpha}_4$ 作用于 $\boldsymbol{Y}_i$、$\boldsymbol{Y}_j$、$\boldsymbol{Y}_k$ 和 $\boldsymbol{Y}_l$,且 $\boldsymbol{X}_{ij}=\boldsymbol{Y}_i-\boldsymbol{Y}_j\pm(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)$、$\boldsymbol{X}_{kl}=\boldsymbol{Y}_k-\boldsymbol{Y}_l\pm(\boldsymbol{\alpha}_3+\boldsymbol{\alpha}_4)$取决于他们是攻击者还是受害者。那么,当且仅当来自同一攻击组的攻击被分别作用时,才满足 $Cor(\boldsymbol{X}'_{ij},\boldsymbol{X}'_{ikl})\neq0$,$i\neq j\neq k\neq l$。即一个攻击是作用在节点 i(或 j 或 i 和 j)上并且另一个在节点 k(或 l 或 k 和 l)上。也就是说只有 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_3$(或 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_4$,或 $\boldsymbol{\alpha}_2$ 和 $\boldsymbol{\alpha}_3$,或 $\boldsymbol{\alpha}_2$ 和 $\boldsymbol{\alpha}_4$)来自同一攻击组时,$Cor(\boldsymbol{X}'_{ij},\boldsymbol{X}'_{ikl})\neq0$。由此,利用这一判据可以根据节点所属的攻击组来识别成对攻击情况和一对多攻击情况下的攻击者节点和对应的受害者节点。

2　算例仿真与验证

算例仿真在 MATLAB 2016 下展开,通过仿真模拟上文提到针对原始数据的不同攻击类型。在理想情况下,当人们完全了解量测数据特征参数后,就可以通过第 1 节的方法来识别检测攻击者和对应的受害者。但实际上,则需要通过实际的样本数据来获取。因此,协方差矩阵的估计将不可避免地存在噪声。另一方面,先前的分析确定了真正的协方差矩阵是稀疏的,因此也应致力于稀疏其样本类似物。为此,在运行检测和识别算法之前,采用通用阈值方法[9]来对样本相关矩阵进行正则化,以便获得其稀疏估计。给定一个独立且分布均匀的随机样本 $\{X_1,X_2,\cdots,X_n\}$,其协方差估计值为:

$$\hat{\boldsymbol{\sigma}}_{ij}=n^{-1}\sum_{t=1}^{n}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{jt}-\bar{\boldsymbol{x}}_j)\quad i,j=1,2,\cdots,P \tag{5}$$

由此,$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的样本相关系数表示为:

$$\hat{p}_{ij}=\frac{\hat{\boldsymbol{\sigma}}_{ij}}{\sqrt{\hat{\boldsymbol{\sigma}}_{ii}\hat{\boldsymbol{\sigma}}_{jj}}} \tag{6}$$

则 $\boldsymbol{R}=(\rho_{ij})$ 的估计值为:

$$\bar{p}_{ij}=\hat{p}_{ij}I\left[\,|\hat{p}|>n^{-\frac{1}{2}}\boldsymbol{\Phi}^{-1}\left(1-\frac{q}{P(P-1)}\right)\right] \tag{7}$$

$$i=1,2,\cdots,P-1,j=i+1,i+2,\cdots,P$$

式中:q 为多重假设检验程序的显著水平。

最后,$\boldsymbol{\Sigma}$ 的估计量为:

$$\tilde{\boldsymbol{\Sigma}} = \hat{\boldsymbol{D}}^{\frac{1}{2}} \tilde{\boldsymbol{R}} \hat{\boldsymbol{D}}^{\frac{1}{2}} \tag{8}$$

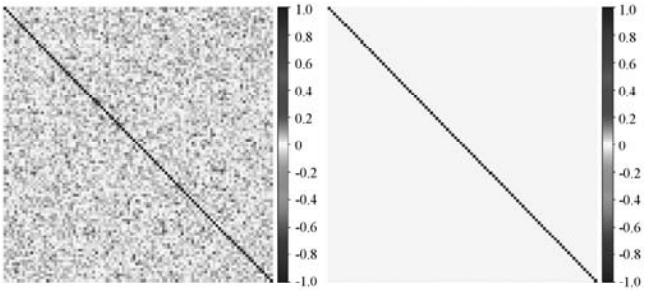
式中:$\hat{\boldsymbol{D}} = \mathrm{diag}(\hat{\boldsymbol{\sigma}}_{11}, \hat{\boldsymbol{\sigma}}_{22}, \cdots, \hat{\boldsymbol{\sigma}}_{PP})$。

首先提供方差比的定义,假设在大小为 $\boldsymbol{\alpha}$ 的任意攻击组中有 l 个受害者,定义该组的方差比(Variance Ratio,VR)为:

$$VR = \frac{Var\left(\frac{\boldsymbol{\alpha}}{l}\right)}{Var(\boldsymbol{Y})} = \frac{1}{l^2}\frac{Var(\boldsymbol{\alpha})}{Var(\boldsymbol{Y})} \tag{9}$$

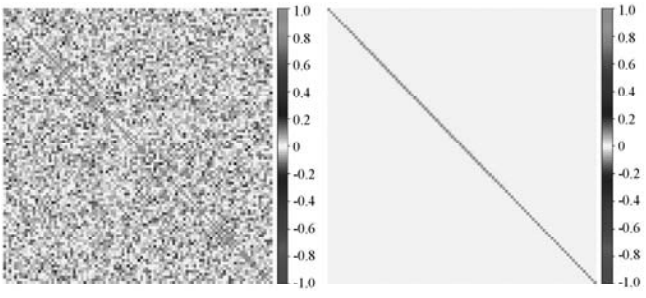
(1) 独立案例:$\boldsymbol{Y}_i = \boldsymbol{U}_i$,而 $\boldsymbol{U}_i^{i.i.d} \sim F(\boldsymbol{\mu}, \boldsymbol{\sigma})$。

考虑 $P = 100$ 个智能电表节点。此外,通用阈值方法中的显著性水平 q 设置为 0.1。生成 $n = 200$ 组独立的量测数据向量,分别来自以下两个分布:均匀分布 $U(625,675)$ 和伽马分布 $G(400,1.5)$。均匀分布将用电量限制在预先指定的范围内,这是大多数用户的情况,而伽马分布呈现出较长的尾部,因此使得检测问题更具挑战性。图 1 和图 2 充分说明了当获得样本协方差矩阵时,所提出的方法在过滤噪声信息方面表现良好。



(a) 未处理前的协方差矩阵 (b) 过滤噪声后的协方差矩阵

图 1 独立均匀分布下不考虑网络攻击的原始数据及过滤后数据的相关矩阵热力图



(a) 未处理前的协方差矩阵 (b) 过滤噪声后的协方差矩阵

图 2 独立伽马分布下不考虑网络攻击的原始数据及过滤后数据的相关矩阵热力图

接下来考虑存在网络攻击的场景。对于所有攻击变量,设 $E(\boldsymbol{\alpha}) = 130$,即平均消耗水平的 20%,每个攻击组设置相同的变化率。因此,从不同的均匀分布生成 $\boldsymbol{\alpha}$,其参数是根据预先指定的平均值和虚拟现实 = 0.1 的要求计算的。此外,对于单个攻击者多受害者攻击,让同一攻击组中的所有受害者节点增加相同的数量,即受害节点的攻击变量权重相等。为了计算检测的概率,从各自的均匀分布和伽马分布中生成了 50 个数据集,对于成对和单个攻击者,生成了许多受害者案例。

图 3 所示的检测概率对应于在 50 个数据集中检测和识别攻击者-受害者组的相对频率。基于图 3 中的结果,可以发现随着样本量的增加,检测的概率也增加,并且当有更复杂类型的攻击时,被发现的概率会更低。如果能够对每个智能电表进行足够数量的重复量测,则攻击的检测概率会收敛到 1。

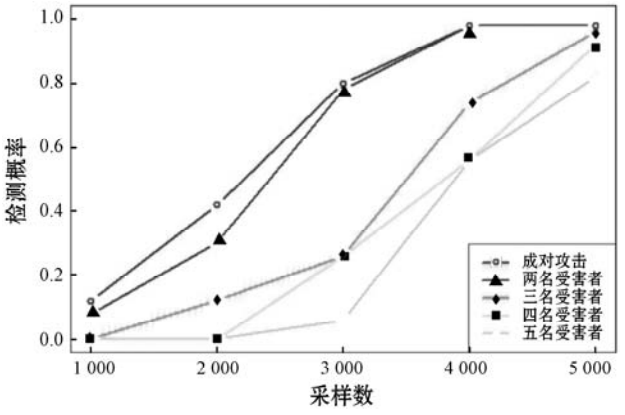


图 3 服从均匀分布 $U(625,675)$ 的非关联数据检测概率

在下一个场景中,智能电表数据是在方差高于前一个场景的均匀分布 $U(600,700)$ 下生成的。此外,平均攻击设置为 $E(\boldsymbol{\alpha}) = 80$,小于量测范围。使用相同的机制生成攻击变量,并考虑相同的攻击场景。图 4 描述了基于 50 个模拟数据集计算的检测概率,与之前的场景相比,仅观察到检测概率的轻微恶化。

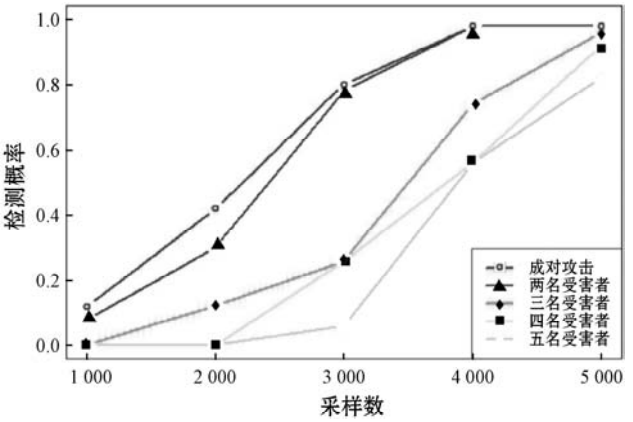


图 4 服从均匀分布 $U(600,700)$ 的非关联数据检测概率

接下来,考虑从伽马分布 $G(625,675)$ 生成的 50 个数据集,同时相应的攻击值 $\boldsymbol{\alpha}$ 也服从伽马分布的,从而满足预先指定的均值(130)和 $VR(0.10)$ 要求。图 5 显示了伽马分布量测的检测概率,较长的右尾会导致更高的样本量要求,以获得与均匀情况下相同的检测概率。

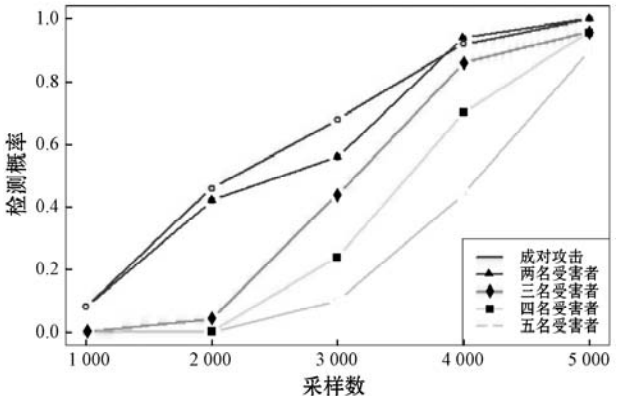


图 5　服从伽马分布的非关联数据检测概率

考虑成对的多个攻击和一对多攻击情况,量测结果从不同的分布生成。具体来说,量测数据服从均匀分布 $U(625,675)$,考虑 10 对成对攻击,5 对 1 个攻击者 -2 个受害者攻击,3 对 1 个攻击者 -3 个受害者攻击服从均匀分布 $U(106.3,152.7)$ 分布,最小的 $VR=0.1$。此外,将样本大小设置为 $n=5\ 000$,图 6 显示了相关矩阵估计的结果热力图。

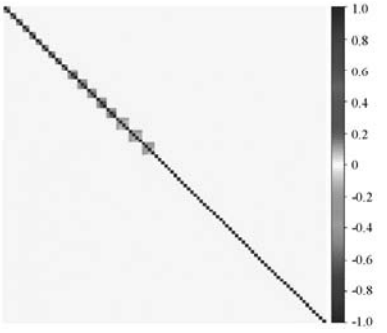


图 6　均匀分布 $U(625,675)$"混合攻击"情况下的相关矩阵热力图

对于服从均匀分布 $U(600,700)$ 分布,遵循相同的实验设置,并根据均匀分布 $U(33,127)$ 生成攻击变量,图 7 显示了相应的结果。对于服从伽马分布 $G(400,1.5)$,选择相同的设置,根据伽马分布 $G(17.78,6.75)$ 生成所有攻击变量,图 8 描述了结果。表 1 包含基于 50 次重复攻击的检测结果。此表显示了每种攻击的平均成功检测次数,括号中的数字是检测到的标准偏差在 50 次复制中进行攻击,这意味着括号中的数字越小越好。

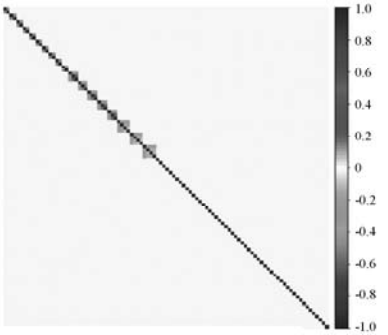


图 7　均匀分布 $U(600,700)$"混合攻击"情况下的相关矩阵热力图

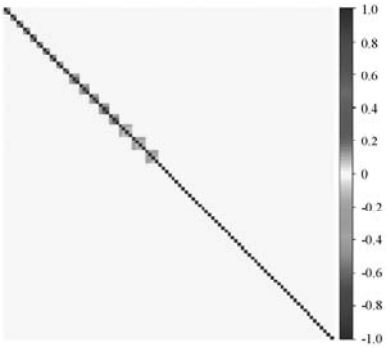


图 8　独立伽马分布"混合攻击"情况下相关矩阵的热力图

表 1　重复测试 50 次的平均成功检测次数

分布	攻击次数	2 名受害者	3 名受害者
均匀分布(625,675)	10(0.000)	5(0.000)	1.76(0.431)
均匀分布(600,700)	10(0.000)	5(0.000)	1.82(0.388)
伽马分布	9.98(0.141)	4.96(0.198)	1.84(0.370)

可以看出,对于成对设置,不论是在均匀分布还是在伽马分布下,所提出的方法总是能够成功检测到所有 10 对攻击,平均 1.76(1.84)个受害者被成功地检测到。因此,即使在"混合攻击"场景中,论文所提出的检测算法也能检测到所有的窃电活动。当 $VR=0.2$ 时,对于所有攻击场景和数据生成分布,当样本大小在 2 000 左右时,检测概率接近 1。当 $VR=0.4$,500 个样本就能够以 0.95 的概率检测出成对攻击。

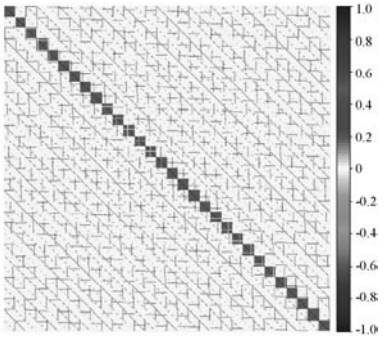


图 9　服从均匀分布在非独立无攻击情况下的相关矩阵热力图

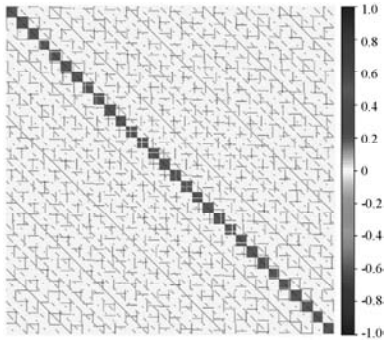
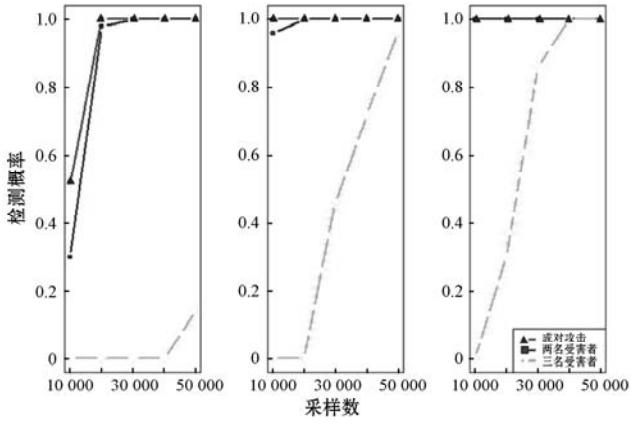


图 10　服从伽马分布在非独立无攻击情况下的相关矩阵热力图

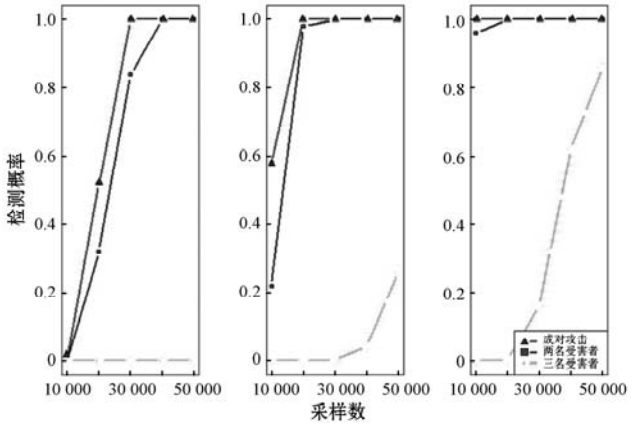
(2) 非独立案例。将 P 取为 30,并将噪声过滤中的显著性水平设置为 $q=0.1$。n 设为 10 000,分别按 $\boldsymbol{W}\sim U(625,675)$ 和 $\boldsymbol{W}\sim G(400,1.2)$ 生成量测数据。

此外,设置 $\boldsymbol{U}_i \sim N(0,100)$ 和 $\rho_i \sim U(0.80,0.85)$。图 9 和图 10 显示了当没有针对不同生成过程的攻击时的相关矩阵估计值热力图。当 $\boldsymbol{W} \sim U(625,675)$,分别发起不同类型的攻击,为 $E(\boldsymbol{\alpha}) = 130$,即大约 $E(\boldsymbol{Y})$ 的 20%,并且每个攻击组的方差比都相同。与独立情况类似,从不同的均匀分布中生成 $\boldsymbol{\alpha}$,其参数基于预先指定的平均值和 $VR = 0.1$(或 0.2)。生成 50 个数据集来计算检测概率,图 11 描述了不同类型攻击之间的检测概率、样本大小和 VR 的关系。对于 $\boldsymbol{W} \sim G(400, 1.5)$,选择 $E(\boldsymbol{\alpha}) = 120$,约为 $E(\boldsymbol{Y})$ 的 20%。图 12 描述了在此设置下不同类型攻击的检测概率、样本大小和 VR 之间的关系,由于伽马分布的长尾性质,其检测概率低于均匀分布场景。



(a) $VR = 0.1$ (b) $VR = 0.15$ (c) $VR = 0.2$

图 11 服从均匀分布非独立数据的各攻击场景下的检测概率



(a) $VR = 0.1$ (b) $VR = 0.15$ (c) $VR = 0.2$

图 12 服从伽马分布非独立数据的各攻击场景下的检测概率

3 结 语

目前有较多的机器学习方法来解决检测问题。然而,识别“攻击者”及其相应的“受害者”是一个更具挑战性的问题,这些方法很少能够解决。这就是所提方法的一个重要特征。本文考虑独立和非独立的智能电表数据生成机制来解决协调的电力盗窃活动检测问题。在每种情况下,模拟了单一成对攻击、多对成对攻击和一对多攻击等不同场景。应用所提出的基于智能电表量测的配电网非侵入式窃电检测方法,利用了正则化的协方差估计器,并对结果矩阵进行模态分析,成功检测出不同攻击场景中的攻击者和相应的受害者。基于大量数值结果说明了所提出方法的优越性能。在实际工程应用中可以通过提取智能电表量测数据,并获得统计信息,针对统计结果用所提方法进行检测识别。

此外,在未来研究中会进一步探究包括涉及多个攻击者和多个受害者在内的其他场景。然而,由于从攻击者的角度来看,它们需要更高的复杂性,所以这种协同攻击更难发起,具有一定的挑战性。

参 考 文 献

[1] 张承智,肖先勇,郑子萱. 基于实值深度置信网络的用户侧窃电行为检测[J]. 电网技术,2019,43(3):1083 - 1091.

[2] 陈卫东,饶军鹏,谢晓帆,等. 敏感台区反窃电的智能诊断分析技术研究与开发应用[J]. 电子世界,2019(19):183 - 184.

[3] 薛峪峰,张俊超,马晓琴. 基于大数据高维随机矩阵的反窃电识别定位[J]. 自动化与仪器仪表,2020(11):220 - 222,226.

[4] Jokar P, Arianpoo N, Leung V C. Electricity theft detection in AMI using customers' consumption patterns[J]. IEEE Transactions on Smart Grid,2015,7(1):216 - 226.

[5] Messinis G M, Hatziargyriou N D. Review of non-technical loss detection methods[J]. Electric Power Systems Research,2018,158:250 - 266.

[6] Li F X, Qiao W, Sun H B, et al. Smart transmission grid: Vision and framework[J]. IEEE Transactions on Smart Grid,2010,1(2):168 - 177.

[7] Yan Y, Qian Y, Sharif H, et al. A survey on smart grid communication infrastructures: Motivations requirements and challenges[J]. IEEE Communications Surveys & Tutorials, 2012,15(1):5 - 20.

[8] Yip S C, Wong K, Hew W P, et al. Detection of energy theft and defective smart meters in smart grids using linear regression[J]. International Journal of Electrical Power & Energy Systems,2017,91:230 - 240.

[9] Bailey N, Pesaran M H, Smith L V. A multiple testing approach to the regularisation of large sample correlation matrices[J]. Journal of Econometrics,2019,208(2):507 - 534.