A Statistical Framework for Detecting Electricity Theft Activities in Smart Grid Distribution Networks

Jin Tao and George Michailidis, *Member, IEEE*

***Abstract*— Electricity distribution networks have undergone rapid change with the introduction of smart meter technology, that have advanced sensing and communications capabilities, resulting in improved measurement and control functions. However, the same capabilities have enabled various cyber-attacks. A particular attack focuses on electricity theft, where the attacker alters (increases) the electricity consumption measurements recorded by the smart meter of other users, while reducing her own measurement. Thus, such attacks, since they maintain the total amount of power consumed at the distribution transformer are hard to detect by techniques that monitor mean levels of consumption patterns. To address this data integrity problem, we develop statistical techniques that utilize information on higher order statistics of electricity consumption and thus are capable of detecting such attacks and also identify the users (attacker and victims) involved. The models work both for independent and correlated electricity consumption streams. The results are illustrated on synthetic data, as well as emulated attacks leveraging real consumption data.**

***Index Terms*— False data injection mechanism, anomaly detection and diagnosis, higher order information, smart grid, inverse problem, thresholded covariance matrix.**

I. Introduction

ELECTRICITY theft has been a major concern worldwide and costs utility companies significant revenue losses [1], [2]. It takes various forms, ranging from physical interventions through illegal connections and meter tampering, to billing irregularities and unpaid bills by customers. The introduction of advanced metering infrastructure has the potential to reduce the risk of electricity theft through its increasing frequency monitoring capabilities. In addition, smart meter technology can lead to effective and accurate load forecasting and on-time troubleshooting for outage remediation and network controllability (see, e.g., [3]–[5]). At the same time, it offers new opportunities for tampering with operations of the power grid through cyber-attacks both locally and remotely, that take the form of false data injections. The consequences range from compromising demand response schemes for selected targeted areas, to endangering the power grid's state estimation process or even inducing power outages [6].

There is a growing literature on false data injection (FDI) attack activities (a brief summary is given in [7]). A lot of attention has been paid to the impact of FDI on the grid's state estimation problem [8] and how coordinate attacks can occur [9], [10]. [11] proposes an adaptive procedure to test whether there is a data attack activity combined with a multivariate hypothesis testing method in order to avoid the wrong grid-state estimate. Addressing the problem from a different angle, [12] attempts to prevent the state estimation from being compromised, by approaching the problem from a graph theoretic method aiming to design an optimal set of meter measurements. Reference [13] considers a setting where multiple simultaneous nefarious data attacks are launched and proposes a game theoretic framework to build a defense system.

Another thrust has focused on the electricity theft problem and there are two general streams in the literature. One of them focuses on using machine learning and data mining techniques to detect anomalies in the consumption patterns of a household or business, based on smart meters' historical data -see e.g. [14]–[20]- potentially augmented with information about the consumer type [20]. These methods can be further subdivided to supervised ones that leverage labels (known FDI vs non-FDI) samples in the training data, and unsupervised ones that try to identify abrupt changes from normal consumption patterns. Supervised methods can be powerful, but availability of labeled FDI samples remains a big challenge. Unsupervised methods are susceptible to the impact of non-malicious factors that alter consumption patterns; e.g. seasonality, change of appliances, change of occupants, and so forth [21].

A different stream in the literature utilizes information about the architecture of a neighborhood area network in the smart grid [22]–[26]. Specifically, it assumes that the electricity provider builds a distribution station within every neighborhood that acts as an "electricity router" to distribute power from the substation to all consumers, A master smart meter (known as the *collector*) measures aggregate power supply from the power provider to all consumers within a certain time interval. Further, smart meters installed at each

Manuscript received June 8, 2019; revised September 15, 2019; accepted October 2, 2019. Date of publication November 7, 2019; date of current version January 31, 2020. This work was supported in part by the NSF under Grant DMS 1830175. *(Corresponding author: George Michailidis.)*

J. Tao is with the Department of Statistics, University of Florida, Gainesville, FL 32611 USA (e-mail: jtao@ufl.edu).

G. Michailidis is with the Department of Statistics and Computer Science, University of Florida, Gainesville, FL 32611 USA, and also with the Informatics Institute, University of Florida, Gainesville, FL 32611 USA (e-mail: gmichail@ufl.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2019.2952181

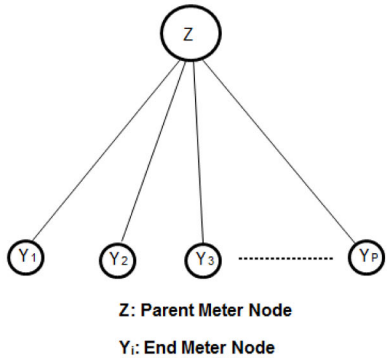


Fig. 1. Structure of the neighborhood area network, with a central smart meter node (collector) for the distribution transformer and smart meters at consumption points.

consumer (households or businesses) record their corresponding energy consumption for the same time interval. Reference [24] proposed a method that utilizes such measurements, together with information about the resistances of lines connecting the consumption points to the distribution transformers to estimate technical losses due to low voltage power lines, as well as intrinsic inefficiencies in the transformers. References [25], [26] employed such measurements and a linear regression framework to identify electricity theft, wherein the dependent variable corresponds to the aggregate measurement by the collector, and the predictor variables to the household/business smart meter measurements. However, for this approach to work, it is assumed that the predictor variables are uncorrelated, an assumption that is automatically violated when theft occurs, as technically demonstrated in Section II below. Note that this regression framework would work to identify faulty individual smart meters, since their measurements will most likely be random and hence uncorrelated.

In this paper, we adopt the architecture of the neighborhood area network, as previously described. In this setting, electricity theft involves an attacker who attempts to lower her energy bill by injecting false measurements to her own smart meter, but to avoid detectability by the utility company compensates with another false injection to smart meters within the neighborhood area network. Specifically, consider a set of smart meters in a neighborhood under a common distributional transformer, as depicted in Figure 1. If the attacker alters (increases) the electricity consumption measurements recorded by the smart meters of other users, while reducing her own measurement, the total consumption reported at the collector is not altered. Hence, various machine learning based monitoring schemes that focus on alterations in mean consumption patterns will fail to detect such an attack when considering measurements from the central node, or when used in end user smart meters, especially if the attacker injects small magnitude false data[1]

On the other hand, examining correlations between the measurements (and in certain cases, information encapsulated in 3rd moments of the data distribution) proves a powerful strategy, not only to detect an attack, but also identify the attacking node, as well as the "victim" nodes. Further, the proposed strategy works even when the consumption patters amongst end users are correlated. The key message of the work is that examining correlation patterns can be a powerful approach for the electricity theft problem.

The remainder of this paper is organized as follows: Section II introduces the modeling framework for the problem at hand. Section III describes the detection and identification/diagnosis strategies, while Section IV discusses implementation issues and evaluates the strategies based on synthetic data, as well as emulated attacks based on real consumption data. Finally, some concluding remarks are drawn in Section V.

II. Model Description and Problem Formulation

Let $Y_1, Y_2, \ldots, Y_P$ correspond to the smart meter[2] variables that measure electricity consumption over a time interval, and further assume that $Y_i = \rho_i W + U_i$, where $\mathbb{E}(W) = \mu_w$, $Var(W) = \sigma_w$, $\rho \in (-1, 1)$ and $U_i \overset{i.i.d.}{\sim} F(u)$. In words, the smart meter measurements are correlated, namely $\text{Corr}(Y_i, Y_j) = \sqrt{\rho_i \rho_j}$, $\forall\ i \neq j$. This is a reasonable assumption, since neighboring households can exhibit a certain degree of similarity in their electricity consumption patterns [27]. The idiosyncratic component U_i of each measurement Y_i (that captures the heterogeneity among electricity consumers) has a distribution F, whose first two moments are denoted by μ and σ, respectively. Finally, denote the $P \times P$ covariance matrix of the measurements by $\mathbf{Y} = (Y_1, \cdots, Y_P)'$ by

$$\mathbf{\Sigma} = Var(\mathbf{Y}) = \mathbb{E}[(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))']. \quad \text{(II.1)}$$

Let Z denote the measurement variable at the collector node (e.g. distributional transformer smart meter) controlled by the power utility company, where the smart meter measurements are communicated to; hence, assuming absence of technical losses due to power distribution and transmission issues (see discussion in [24]), we have by definition that $Z = \sum_{i=1}^{P} Y_i$.[3]

An electricity theft attacker aims to distort the measurements recorded by the end node smart meters, while not changing their sum. For example, if the attacker can lower the measurement of meter i by an amount α and increase that of meter j by an equal amount, then the attacker can benefit financially. We coin the smart meter (end node), whose electricity consumption measurement is decreased as the *"Attacker Node"*, and the end node whose electricity consumption measurement is increased as the *"Victim Node"*.

Next, we impose a number of assumptions on $\mathbf{Y}$ and α that are used in future technical developments. We start by defining the key variables.

Notation:

Y_i: value of i^{th} smart meter measurement;

W: the variable of the common component for each smart meter measurement;

[1]Note that this attack mechanism violates the assumption of uncorrelated predictors in the regression framework proposed by [25], and thus renders it inapplicable.

[2]End nodes of the distribution network; for example, deployed at households or businesses.

[3]In the presence of technical losses, the techniques in [24] can be used to adjust the controller and individual smart meter measurements, so that equality holds.

ρ_i: the relative contribution of W for Y_i;

U_i: idiosyncratic component for each measurement Y_i;

α: data attack variable;

$\asymp$: approximate to or close to. For example, $a-b \asymp 0$ means the value of $a-b$ is close to 0.

Assumptions:

Assumption 1: Assume that all smart meter measurements are non-negative, i.e., $Y_i \geq 0, \forall\ i$ and in addition that their 3rd moments exist; i.e, $\mathbb{E}(Y_i^3) \leq \infty$.

Assumption 2: $\mid \rho_i - \rho_j \mid \asymp 0$ for any $1 \leq i \neq j \leq P$; namely that the magnitude of ρ_i's is similar.

Assumption 3: $Cov(U_i, W) = 0$ for any $1 \leq i \leq P$; namely, the common and idiosyncratic components of each smart meter measurement are uncorrelated.

Assumption 4: Each smart meter measurement is uncorrelated with any attack variable, i.e., $\text{Cov}(Y_i, \alpha_j) = 0$ for $1 \leq i \leq P$ and $j \in \mathbb{N}$, the latter being the set of nodes involved in the attack either as attacker or victims.

Assumption 5: All attack variables are positive, i.e., $\alpha > 0$, and independently distributed with $Var(\alpha) = \sigma_\alpha$. In addition, we assume that $\mathbb{E}(\alpha^3) \leq \infty$.

Assumption 6: The attacker manages to coordinate the attacks, so that there is a single one during a reporting period by the smart meters to the utility company.

Remark (Discussion of Assumptions): Assumption 2.1 is mild, since it is easily satisfied by distributions for electricity consumption; only, fairly heavy tailed distributions are excluded. The same holds for Assumption 2.5. Assumptions 2.2 and 2.3 posit a mechanism that induces correlations amongst the end node smart meter measurements that can be leveraged by the proposed approach for both electricity theft detection and identification of the attacker and the victim nodes. At the same time, it is a very general mechanism that allows for both strongly and weakly correlated data, depending on the magnitude of the ρ_i's and the variance of W.[4] Finally, Assumption 2.4 posits that the magnitude of the attack is uncorrelated with any of the smart meter measurements. This is reasonable in practice; otherwise, the attacker would need to continuously adjust the amount of electricity to that of the measurement, which implies a high level of sophistication on the attacker's part.

Finally, note that the proposed model assumes that the variances of all variables involved remain constant over time. Hence, the model can naturally accommodate shifts in the mean electricity consumption.

A. Identifying Attacks: The Independent Case

To gain insight into the issue of *whether* and *how* an attack can be identified, we first consider the special case where $\rho_i = 0, \forall\ 0 \leq i \leq P$; namely, that the smart meter measurements are *independent* since $Y_i = U_i$. Consequently, we have that $Y_1, Y_2, \ldots, Y_P$ are i.i.d with $\mathbb{E}Y_i = \mu$ and $Var(Y_i) = \sigma$. Therefore, the covariance matrix $\mathbf{\Sigma}$ becomes a diagonal matrix, i.e., $\mathbf{\Sigma} = \sigma\mathbf{I}$.

We start our analysis by examining an attack scenario involving a single Victim Node.

[4]The case of uncorrelated data requires special treatment and is examined in Section II.A

A Pairwise Attack Scenario: This setting involves nodes i (Attacker) and j (Victim) with respective measurements $Y_i - \alpha$ and $Y_j + \alpha$. Then, the following result can be easily established.

Theorem 7: The pairwise attack is *undetectable* by only monitoring the mean levels of electricity consumption at the smart meters (end nodes) and the central node.

Proof: If there is no attack, we have that $\mathbb{E}(Z) = \sum_{l=1}^{P} \mathbb{E}(Y_l)$. In the pairwise attack scenario, we obtain $\mathbb{E}(Z) = \sum_{l \neq i,j} \mathbb{E}(Y_l) + \mathbb{E}(Y_i - \alpha) + \mathbb{E}(Y_j + \alpha) = \sum_{l=1}^{P} \mathbb{E}(Y_l)$. Consequently, the meter in the central node Z has measurements that agree with the sum of those obtained at the smart meters. ■

The key to *detect* the attack, as well as *identify* the nodes involved is to examine higher moment (variance, etc.) information of the consumption measurements.

Under a pairwise attack mechanism, let an attack of magnitude α_e be launched by node i with victim node j, and similarly let another attack of magnitude α_f be launched with Attacker node k and Victim node l. Denote by $Var(\alpha_e) = \sigma_{\alpha_e}$ and $Var(\alpha_f) = \sigma_{\alpha_f}$; then, the following relationships hold:

- $Var(Y_i - \alpha_e) = \sigma + \sigma_{\alpha_e}$, $Var(Y_j + \alpha_e) = \sigma + \sigma_{\alpha_e}$, $Var(Y_k - \alpha_f) = \sigma + \sigma_{\alpha_f}$, $Var(Y_l + \alpha_f) = \sigma + \sigma_{\alpha_f}$.
- $Cov(Y_i - \alpha_e, Y_j + \alpha_e) = -\sigma_{\alpha_e}$, $Cov(Y_k - \alpha_f, Y_l + \alpha_f) = -\sigma_{\alpha_f}$.

Some straightforward algebra shows that

$$Var(Y_i \pm \alpha_e) = Var(Y_i) + Var(\alpha_e) \pm 2Cov(Y_i, \alpha_e)$$

Then, by Assumption 4, we obtain $Cov(Y_i, \alpha_e) = 0$. Hence,

$$Var(Y_i \pm \alpha_e) = Var(Y_i) + Var(\alpha_e) = \sigma + \sigma_{\alpha_e},$$

and

$$\begin{aligned} Cov(Y_i - \alpha_e, Y_j + \alpha_e) = {} & Cov(Y_i, Y_j) + Cov(Y_i, \alpha_e) \\ & - Cov(\alpha_e, Y_j) - Cov(\alpha_e, \alpha_e) \end{aligned}$$

Since $Cov(Y_i, Y_j) = 0$ and $Cov(Y_i, \alpha_j) = 0$ for $1 \leq i \leq P$ and $j \in \mathbb{N}$, we have

$$Cov(Y_i - \alpha_e, Y_j + \alpha_e) = -Var(\alpha_e) = -\sigma_{\alpha_e}$$

Similarly,

$$Cov(Y_i - \alpha_e, Y_k - \alpha_f) = \text{Cov}(\alpha_e, \alpha_f)$$

By Assumption 5, we then get $Cov(\alpha_i, \alpha_j) = 0$ for $\forall\ i,\ j \in \mathbb{N}$, and thus

$$\text{Cov}(Y_i - \alpha_e, Y_k - \alpha_f) = 0$$

The other relationships can be obtained analogously. These calculations imply that in the presence of an attack of magnitude α involving nodes 1 and 2, the covariance matrix $\mathbf{\Sigma}$ will change its pattern from a diagonal matrix to a *block diagonal* matrix given by

$$\mathbf{\Sigma}' = \begin{bmatrix} \sigma + \sigma_\alpha & -\sigma_\alpha & 0 & \cdots & 0 \\ -\sigma_\alpha & \sigma + \sigma_\alpha & 0 & \cdots & 0 \\ 0 & 0 & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma \end{bmatrix}$$

Analogously, if there are s different pairwise attacks involving the first $2s$ end nodes with the first one designated as the Attacker and the second as the Victim, the corresponding covariance matrix of the smart meter measurements will have the following form

$$\boldsymbol{\Sigma}' = \begin{bmatrix} B_1 & & & & & & \\ & B_2 & & & & & \\ & & \ddots & & & & \\ & & & B_s & & & \\ & & & & \sigma & & \\ & & & & & \ddots & \\ & & & & & & \sigma \end{bmatrix}_{P\times P}$$

where B_h, $h = 1, \ldots, s$, is the sub-covariance block matrix for attack α_h, and $B_h = \begin{bmatrix} \sigma + \sigma_{\alpha_h} & -\sigma_{\alpha_h} \\ -\sigma_{\alpha_h} & \sigma + \sigma_{\alpha_h} \end{bmatrix}$.

This explicit pattern dictates the following algorithmic strategy to detect an electricity theft attack and identify the pairs of nodes involved in it.

- If $\text{Cov}(Y_i, Y_j) < 0$, $i \neq j$, we can conclude that end nodes i and j are involved in the same attack; i.e. they belong to the same attack group;
- Similarly, if $\text{Cov}(Y_i, Y_j) = 0$, $i \neq j$, we can conclude that there is no attack involving nodes i and j; rather, they belong to different attack groups.

Hence, the above simple strategy detects electricity theft and the nodes involved as Attacker and Victim in it.

Remark: In practice, the s attacks will involve random pairs of nodes and not the first $2s$ ones. Then, one needs to reorder the rows and columns of the covariance matrix to obtain the desired structure previously discussed.

The previously outlined strategy identifies the s attack groups, but not which node in the pair is the Attacker and which is the Victim. To address this issue, information involving the 3rd moment of the smart meter measurements is required, as the following result shows.

Theorem 8: Under the pairwise attack scenario, if end nodes i and j are in the same attack group with magnitude α, then $\mathbb{E}(Y_i + \alpha)^3 - \mathbb{E}(Y_j - \alpha)^3 > 0$.

Proof: Note that

$$\mathbb{E}(Y_i + \alpha)^3 = \mathbb{E}(Y_i)^3 + 3\mathbb{E}(Y_i)^2\mathbb{E}(\alpha) + 3\mathbb{E}(Y_i)\mathbb{E}(\alpha)^2 + \mathbb{E}(\alpha)^3 \quad \text{(II.2)}$$

Similarly,

$$\mathbb{E}(Y_j - \alpha)^3 = \mathbb{E}(Y_j)^3 - 3\mathbb{E}(Y_j)^2\mathbb{E}(\alpha) + 3\mathbb{E}(Y_j)\mathbb{E}(\alpha)^2 - \mathbb{E}(\alpha)^3, \quad \text{(II.3)}$$

where nodes i and j are the Victim and Attacker nodes, respectively. Further, since $\alpha > 0$, we obtain

$$\mathbb{E}\alpha > 0, \mathbb{E}\alpha^3 > 0$$

A subtraction of the the last two relationships [i.e., (II.2) − (II.3)] yields

$$\begin{aligned}\mathbb{E}(Y_i + \alpha)^3 - \mathbb{E}(Y_j - \alpha)^3 = {} & \mathbb{E}(Y_i)^3 - \mathbb{E}(Y_j)^3 \\ & + 3\mathbb{E}(\alpha)(\mathbb{E}Y_i^2 + \mathbb{E}Y_j{}^2) \\ & + 3\mathbb{E}(\alpha)^2(\mathbb{E}Y_i - \mathbb{E}Y_j) \\ & + 2\mathbb{E}(\alpha)^3\end{aligned}$$

It is easy to check that

$$\begin{aligned}\mathbb{E}(Y_i)^3 - \mathbb{E}(Y_j)^3 = {} & (\rho_i^3 - \rho_j^3)\mathbb{E}W^3 \\ & + 3(\rho_i^2\mathbb{E}U_i - \rho_j^2\mathbb{E}U_j)\mathbb{E}W^2 \\ & + 3(\rho_i\mathbb{E}U_i^2 - \rho_j\mathbb{E}U_j^2)\mathbb{E}W \\ & + \mathbb{E}U_i^3 - \mathbb{E}U_j^3\end{aligned}$$

and

$$\mathbb{E}Y_i - \mathbb{E}Y_j = (\rho_i - \rho_j)\mathbb{E}W + \mathbb{E}U_i - \mathbb{E}U_j$$

By Assumption 3, we have $\mathbb{E}(Y_i)^3 - \mathbb{E}(Y_j)^3 \asymp 0$ and $\mathbb{E}Y_i - \mathbb{E}Y_j \asymp 0$. Since $\mathbb{E}\alpha > 0$ and $\mathbb{E}\alpha^3 > 0$, then

$$\mathbb{E}(Y_i + \alpha)^3 - \mathbb{E}(Y_j - \alpha)^3 > 0$$

This proves the result. ■

A direct consequence of Theorem 8 is that the 3rd moment of the Victim node is strictly larger than that of the Attacker node within the same attack group. Hence, leveraging the results of Theorem 7, the block diagonal structure of the covariance matrix and Theorem 8, give rise to the following *detection* and *identification* algorithm for the pairwise attack scenario.

Algorithm 1 Pairwise Attack Detection and Identification

```
i ← 1
while i < P do
  while i < j ≤ P do
    if Cov(y_i, y_j) < 0 then
      Δ = E(y_i)^3 − E(y_i')^3
      if Δ > 0 then
        label y_i as the Victim node for Attack Group i
        and label y_i' as the Attacker node for the same
        Group;
      else[Δ < 0]
        label y_i as the Attacker for Attack Group i
        and label y_i' as the Victim for the same Group;
      end if
    else
    end if
    j = j + 1
  end while
  i = i + 1
end while
```

A Single Attacker-Many Victims Scenario: The pairwise scenario is the simplest one to execute, since only one Victim node is involved. However, if the Attacker node aims to use a large α, this action may be flagged by either using change point analysis techniques, since a sharp change in

the electricity consumption pattern of the Victim node would occur, or by the consumer under attack, who may in turn complain to the utility company for a sharp and unexpected increase in her/his electricity bill. In that case, the Attacker may want to spread the attack among a larger group of nodes, so as not to raise such suspicions. This leads to a more involved setting, where the Attacker node decreases the smart meter measurement at node i by an amount α, and increases cumulatively the measurements of the Victim group smart meters by an equal amount. Specifically, when a magnitude α attack is launched by node i, we have $Y_i - \alpha$; further, for the l Victim Nodes we also have $Y_{j_1}+k_1^{\alpha}\alpha$, $Y_{j_2}+k_2^{\alpha}\alpha$,...,$Y_{j_l}+k_l^{\alpha}\alpha$, where $\sum_j k_j^{\alpha} = 1$.

Analogously to the pairwise attack scenario, the single Attacker-Many Victims case is *undetectable* by monitoring discrepancies in the measurements at the controller and the end node smart meters, since by construction the sum of the latter measurements agree with the former; i.e. Z.

A similar analysis to the pairwise scenario shows that the resulting covariance matrix exhibits again a block diagonal pattern; namely,

$$\mathbf{\Sigma}' = \begin{bmatrix} B_1 & & & & & & \\ & B_2 & & & & & \\ & & \ddots & & & & \\ & & & B_s & & & \\ & & & & \sigma & & \\ & & & & & \ddots & \\ & & & & & & \sigma \end{bmatrix}_{P\times P}$$

where B_h, $h = 1, \ldots, s$, is the block of the covariance matrix corresponding to the h-th attack. Each block B_h has the following form:

$$B_h = \mathbf{\Sigma} + \sigma_{\alpha_h} \begin{bmatrix} 1 & -k_1^{\alpha_h} & \cdots & -k_{d_h}^{\alpha_h} \\ -k_1^{\alpha_h} & (k_1^{\alpha_h})^2 & \cdots & k_1^{\alpha_h} k_{d_h}^{\alpha_h} \\ \vdots & \vdots & \ddots & \vdots \\ -k_{d_h}^{\alpha_h} & k_{d_h}^{\alpha_h} k_1^{\alpha_h} & \cdots & (k_{d_h}^{\alpha_h})^2 \end{bmatrix}$$

where $\mathbf{\Sigma}$ has $(d_h + 1) \times (d_h + 1)$ dimensions and d_h is the number of Victims in the α_h attack group, $\mathbf{\Sigma}_{(d_h+1)\times(d_h+1)}$ is the original block of the covariance matrix of the end nodes in the α_h attack group, and $\sum_{h=1}^{s}(d_h + 1) = m$.

Thus, the same broad strategy to the pairwise attack scenario is applicable. Specifically, under the single Attacker-Many Victims attack mechanism,

- If $\text{Cov}(Y_i, Y_j) \neq 0$, $i \neq j$, we conclude that end nodes i and j belong to the same attack group;
- If $\text{Cov}(Y_i, Y_j) = 0$, $i \neq j$, we conclude that nodes i and j belong to different attack groups.

Hence, a close examination of such patterns in the covariance structure of the smart meter measurements leads to detecting such attacks.

Interestingly, even though in this scenario the attack mechanism is more involved, once an attack group has been identified, it is straightforward to separate the Attacker node from the Victim ones. A close examination of the B_h block shows, that under this scenario only the Attacker node will exhibit negative covariance values with *all* other nodes in the same attack group; i.e., $Cov(Y_{i_0}, Y_j) < 0$, $\forall j \neq i_0$. On the other hand, all the Victim nodes in the same attack group will have positive covariance values with each other. Hence, for each attack group, labeling the Attacker and the Victim nodes requires

1) Identify the node who has only negative covariance values within the B_h sub-block and laebl it is as the Attacker node.
2) Label the remaining nodes in the block as the Victim ones.

The following algorithm summarizes the detection and node identification strategy.

Algorithm 2 Single Attacker-Many Victims Attack Detection

$i \leftarrow 1$, $g \leftarrow 1$, $l_g \leftarrow 1$
while $i < P$ **do**
  while $i < j \leq P$ **do**
    if $Cov(y_i, y_j) \neq 0$ **then**
      denote $y_g^1 = y_i$ and let $y_g^1 \in Group_g$, and for each j, then let $l_g = l_g + 1$, denote $y_g^{l_g} = y_j$ and let $y_g^{l_g} \in Group_g$;
      for $Group_g$, find the node y_g^a s.t. $Cov(y_g^a, y_g^b) < 0$ for all $b \neq a$ and $b \in (1, 2, \ldots, l_g)$. Then label y_g^a as the Attacker for $Group_g$ and label the remaining nodes as Victims for $Group_g$. Then $g = g + 1$, $l_g = 1$.
    else
    end if
    $j = j + 1$
  end while
  $i = i + 1$
end while

B. Identifying Attacks: The Dependent Case

Recall that in the general case, the smart meter measurements are generated according to $Y_i = \rho_i W + U_i$. Note that $\text{Cov}(Y_i, Y_j) = \rho_i \rho_j \sigma_w$, which complicates detection and attacker-victim(s) identification strategies. We start by defining $X_{ij} = Y_i - Y_j$ *for* $i, j = 1, 2, \ldots, P$ *and* $i \neq j$. By using this new set of $(P-1)^2$ measurement variables, we show next that their covariance exhibits patterns that lead to detection and identification.

Denote by $\mathbf{X} = \left(X_{12}, \ldots, X_{1p}, X_{21}, X_{23}, \ldots, X_{(P-1)P}\right)^T$; we then can obtain the following result.

Theorem 9: $\text{Cov}(X_{ij}, X_{kl}) \asymp 0$ if $i \neq j \neq k \neq l$.

Proof: Since U_i are i.i.d,, we get

$$\text{Cov}(X_{ij}, X_{kl}) = (\rho_i - \rho_j)(\rho_k - \rho_l)\sigma_w$$

Since by Assumption 2 $\mid \rho_i - \rho_j \mid \asymp 0$, we get $(\rho_i - \rho_j)(\rho_k - \rho_l) \asymp 0$. Therefore, $\text{Cov}(X_{ij}, X_{kl}) \asymp 0$. ■

Note that Theorem 9 implies that the differencing transformation of the original set of measurements leads to reducing their correlation to a large extent, which proves key to our detection and identification strategy.

To illustrate, we start with the most general case. Without loss of generality, we assume that there are four different attack

variables $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ applied to Y_i, Y_j, Y_k, Y_l, respectively for any $i \neq j \neq k \neq l$. Then, $X'_{ij} = Y_i - Y_j \pm (\alpha_1 + \alpha_2)$ and $X'_{kl} = Y_k - Y_l \pm (\alpha_3 + \alpha_4)$, depending on whether they are attackers or victims.

Theorem 10: $\text{Cov}(X'_{ij}, X'_{kl}) \neq 0$, for $i \neq j \neq k \neq l$, only if attack variables from the same attack group are separately applied, i.e., one is on node i (or j or $i\&j$) and the other is on node k (or l or $k\&l$).

Proof: First, calculate

$$\begin{aligned}\text{Cov}(X'_{ij}, X'_{kl}) &= \text{Cov}((\rho_i - \rho_j)W + U_i - U_j \pm (\alpha_1 + \alpha_2),\\ &\qquad (\rho_k - \rho_l)W + U_k - U_l \pm (\alpha_3 + \alpha_4))\\ &= (\rho_i - \rho_j)(\rho_k - \rho_l)\sigma_w\\ &\quad \pm (\text{Cov}(\alpha_1, \alpha_3) + \text{Cov}(\alpha_1, \alpha_4)\\ &\quad + \text{Cov}(\alpha_2, \alpha_3) + \text{Cov}(\alpha_2, \alpha_4)) \qquad \text{(II.4)}\end{aligned}$$

Since $(\rho_i - \rho_j)(\rho_k - \rho_l) \asymp 0$, we have

$$\begin{aligned}\text{Cov}(X'_{ij}, X'_{kl}) &= \pm (\text{Cov}(\alpha_1, \alpha_3) + Cov(\alpha_1, \alpha_4)\\ &\quad + \text{Cov}(\alpha_2, \alpha_3) + \text{Cov}(\alpha_2, \alpha_4))) \qquad \text{(II.5)}\end{aligned}$$

It can then be easily seen that only if α_1 and α_3 (or α_1 and α_4, or α_2 and α_3, or α_2 and α_4) are from the same attack group, $\text{Cov}(X'_{ij}, X'_{kl}) \neq 0$. ∎

Based on the nature of this property, we could develop an easy-to-implement algorithm to identify and group nodes for both the pairwise attack and the single Attacker-Many Victims scenarios according to which attack groups they belong to.

Algorithm 3 Detection Algorithm for Dependent Case

```
while 1 ≤ i ≠ j ≠ k ≠ l ≤ P do
  if △1 = Cov(X_ij, X_kl) ≠ 0 then
    while 1 ≤ l' ≤ P & l' ≠ i ≠ j ≠ k ≠ l do
      if △2 = Cov(X_ij, X_kl') = 0 then
        while 1 ≤ j' ≤ P & j' ≠ l' ≠ i ≠ j ≠ k ≠ l do
          if △3 = Cov(X_ij', X_kl) = 0 then
            conclude j & l belong to the same attack group
          else
            conclude i & l belong to the same attack group
          end if
        end while
      end if
    end while
  end if
  update i, j, k, l
end while
```

Remark: For example, if the output of Algorithm II-B is $\{(1,2),(1,3),(2,3)\}$, we will conclude that $(1,2,3)$ are within the same attack group. The same logic applies to more complicated outcomes.

III. Implementation Issues and Numerical Results

The previous results discuss detection and identification strategies, in the ideal case, when one has full knowledge of population parameters (e.g. the true variances σ_w, σ). However, in practice one needs to replace them with their *sample* counterparts. Thus, the estimate of the covariance matrix $\mathbf{\Sigma}$ would be *noisy*. On the other hand, the previous analysis established that the true covariance matrix is *sparse* and hence we should aim to sparsify its sample analogue as well. To that end, we employ the Universal Thresholding Method [28] to regularize the sample correlation matrix in order to obtain a sparse estimate of it, before running our detection and identification algorithms.

We provide the necessary details of our estimation strategy next. Let $\mathbf{X} = (X_1, \ldots, X_P)^T$ be a p-variate random vector with covariance matrix $\mathbf{\Sigma}$. Given an independent and identically distributed random sample $\{\mathbf{X_1}, \ldots, \mathbf{X_n}\}$, we can calculate the covariance estimator as

$$\hat{\sigma}_{ij} = n^{-1} \sum_{t=1}^{n} (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j), \quad i, j = 1, \ldots, P \qquad \text{(III.1)}$$

where $\bar{x}_i = n^{-1} \sum_{t=1}^{n} x_{it}$.

Then, the sample correlation coefficient of x_i and x_j is given by $\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$. Thus, the estimator of $\mathbf{R} = (\rho_{ij})$, denoted by $\tilde{\mathbf{R}} = (\tilde{\rho}_{ij})$, is given by

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I\left[|\hat{\rho}| > n^{-\frac{1}{2}} c_q(P)\right], \quad i = 1, 2, \ldots, P-1, j = i+1, \ldots, P,$$

where $c_q(P) = \Phi^{-1}(1 - \frac{q}{2f(P)})$ and q is the significant level for this multiple hypothesis testing procedure. We select $f(P) = \frac{P(P-1)}{2}$.

Finally, the estimator of $\mathbf{\Sigma}$, denoted by $\tilde{\mathbf{\Sigma}}$, is given by

$$\tilde{\mathbf{\Sigma}} = \hat{\mathbf{D}}^{1/2} \tilde{\mathbf{R}} \hat{\mathbf{D}}^{1/2} \qquad \text{(III.2)}$$

where $\hat{\mathbf{D}} = diag(\hat{\sigma}_{11}, \hat{\sigma}_{22}, \ldots, \hat{\sigma}_{PP})$.

A. Simulation Studies Results

We start by providing the definition of the Variance Ratio, an important quantity in the sequel. Assume there are l victims in the group of an arbitrary attack of magnitude α; we then define the Variance Ratio (VR) for that group to be

$$VR = \frac{Var(\frac{\alpha}{l})}{Var(Y)} = \frac{1}{l^2} \frac{Var(\alpha)}{Var(Y)} \qquad \text{(III.3)}$$

The quantity VR can be thought of a *signal-to-noise* measure for the problem at hand.

1) Independent Case: Recall that in this case, the model reduces to $Y_i = U_i$, where $U_i \overset{i.i.d.}{\sim} F(\mu, \sigma)$.

To illustrate the detection algorithms, we consider $P = 100$ smart meter nodes; $\mathbf{Y} = (Y_1, \ldots, Y_{100})^T$. Further, the significance level in the Universal Thresholding Method is set to be $q = 0.1$. In the first simulation setting, we generate $n = 200$ independent sets of smart meter vectors, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, from the following two distributions: (i) Uniform$(625, 675)$ and (ii) Gamma$(400, 1.5)$, respectively. The Uniform distribution limits electricity consumption within a prespecified range, which is the case for most consumers, while the Gamma

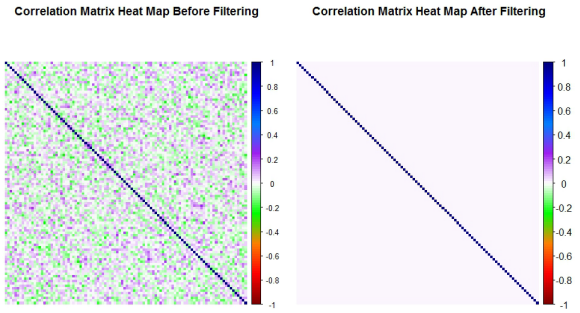


Fig. 2. Heat maps of the correlation matrix and its filtered version in the independent Uniform, no attack case.

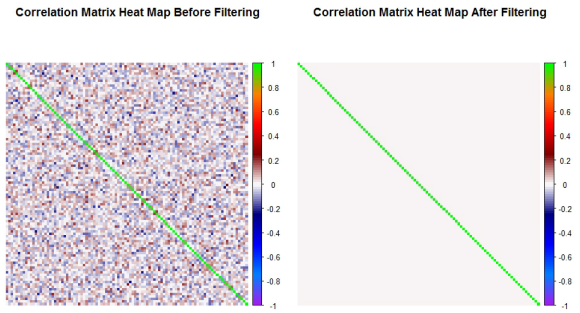


Fig. 3. Heat maps of the correlation matrix and its filtered version in the independent Gamma, no attack case.

distribution exhibits a longer tail and hence makes the detection problem more challenging. We employ the Universal Thresholding Method to estimate the covariance matrix of $\mathbf{Y}$. Figures 2 & 3 amply illustrate that the method performs well in filtering noisy information when obtaining the sample covariance matrix.

Next, we describe the attack scenarios considered. For all attack variables, let $\mathbb{E}(\alpha) = 130$, i.e., 20% of the mean consumption level $\mathbb{E}(Y)$; further, set the *Variance Ratio* to be the same for each attack group. As a consequence, we generate α from different Uniform Distributions, whose parameters are calculated based on the prespecified mean and VR=0.1 requirement. In addition, for the single attacker-many victims attack, we let all the victim nodes in the same attack group to be increased by the same amount; i.e., the attack variable for the victim nodes are equally weighted. To calculate the probability of detection, 50 data sets from the respective Uniform and Gamma distributions, and for both pairwise and single attacker-many victims cases were generated. The probability of detection shown in Figure 4 corresponds to the relative frequency of both detecting and identifying the attacker-victim groups in the 50 data sets.

We draw the following conclusions based on the results depicted in Figure 4: (i) As the sample size increases, the probability of detection also increases; (ii) The probability of detection will get lower when we have more complicated types of attacks; (iii) If we are able to have adequate number of repeated measurements from each smart meter, the detection probability of an attack converges to one.

In the next setting, the smart meter data are generated from a Uniform$(600, 700)$ that has higher variance than the previous

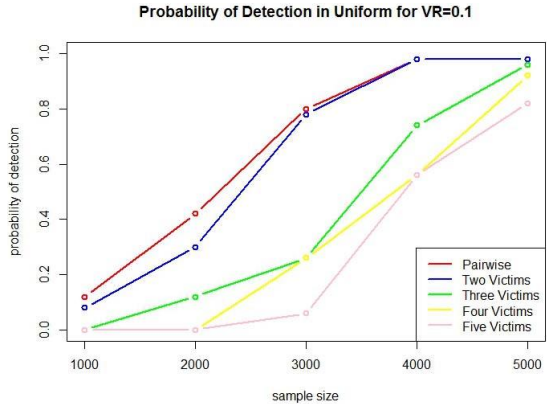


Fig. 4. Probability of detection for Uniformly(625.675) distributed uncorrelated data.

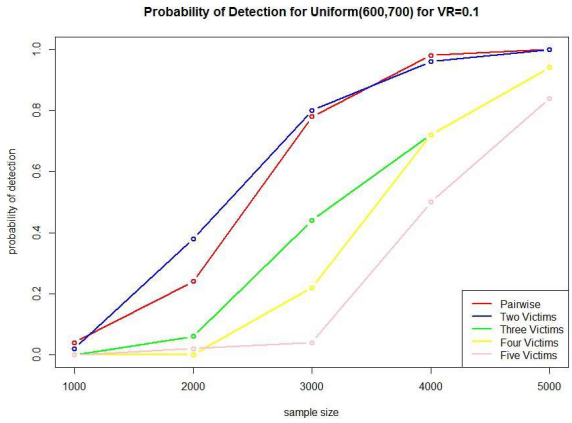


Fig. 5. Probability of detection for Uniformly(600.700) distributed uncorrelated data.

setting. In addition, the mean attack is set to $E(\alpha) = 80$, which is less than the range of the measurements. The same mechanism is employed for generating the attack variables and the same attack scenarios are considered. Figure 5 depicts the detection probability calculated based on 50 simulated data sets. Only a slight deterioration in the detection probability is observed, compared to the previous setting.

Next, we consider 50 data sets generated from a Gamma$(625, 675)$, while the corresponding attacks α are also Gamma distributed, so that the prespecified mean (130) and VR (0.10) requirements are met. Figure 6 shows the probability of detection for the Gamma distributed measurements. The presence of a longer right tail in the smart meter measurement leads to a higher sample size requirement for achieving the same detection probability to the Uniform case. Further, all the conclusions drawn from the Uniform distribution setting continue to hold.

Next, we examine a scenario involving multiple attacks of both pairwise and one attacker-many victims types, with measurements generated from different distributions. Specifically, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ come from a Uniform$(625, 675)$ distribution, and we consider 10 pairwise attacks, 5 One Attacker-Two Victims attacks and 3 One Attacker-Three Victims attacks from a Uniform$(106.3, 153.7)$ distribution, which results in a minimum $VR = 0.1$. Further, we set the sample size to $n = 5000$. Figure 7 shows the resulting heat map of the correlation matrix estimate.

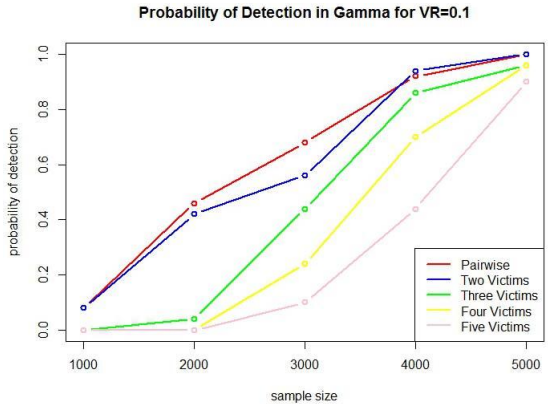


Fig. 6. Probability of detection Gamma distributed uncorrelated data.

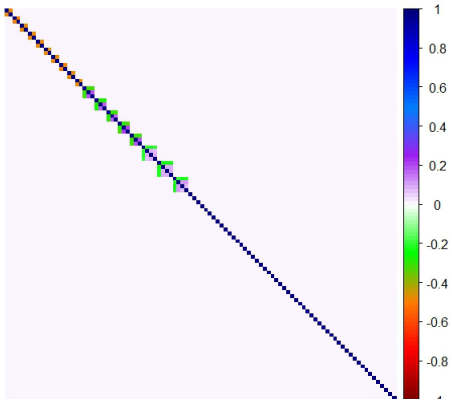


Fig. 7. Heat map of the correlation matrix in the independent Uniform "mixed attacks" case.

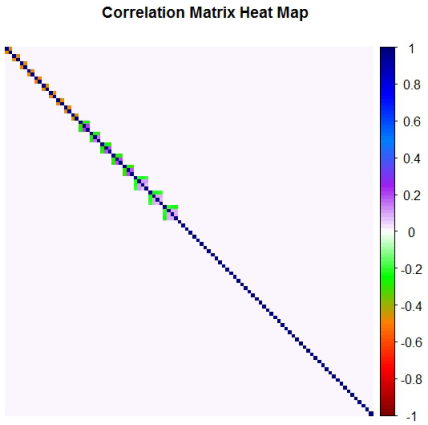


Fig. 8. Heat map of the correlation matrix in the independent Uniform "mixed attacks" case 2.

For $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from a Uniform$(600, 700)$ distribution, we follow the same experiment set-up and generate the attack variables from Uniform$(33, 127)$. Figure 8 illustrates the corresponding results.

For $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from a Gamma$(400, 1.5)$ distribution, we choose the same set-up, except for generating all the attack variables from a Gamma$(17.78, 6.75)$ distribution. Figure 9 depicts the results.

The following Table I contains detection results, based on 50 replicates of the attacks. This table shows the average number of successful detection for each type of attack, and the number in the bracket is the standard deviation of detected

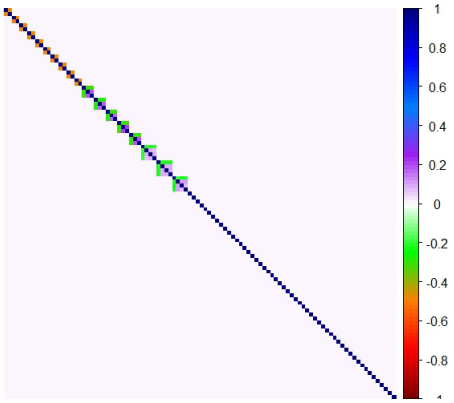


Fig. 9. Heat map of the correlation matrix in the independent Gamma "mixed attacks" case.

TABLE I

WE REPEAT TO LAUNCH 10 PAIRWISE, 5 2-VICTIMS AND 3 3-VICTIMS ATTACKS 50 TIMES, AND CALCULATE THE AVERAGE NUMBER OF SUCCESSFUL DETECTION FOR EACH TYPE OF ATTACK

	Pairwise	2 Victims	3 Victims
Uniform(625,675)	10(0.000)	5(0.000)	2.76(0.431)
Uniform(600,700)	10(0.000)	5(0.000)	2.82(0.388)
Gamma	9.98(0.141)	4.96(0.198)	2.84(0.370)

attack in 50 replications, which means that the bracketed number is the smaller the better. The same format of table holds for the rest of this paper.

It can be seen that for the pairwise setting, all 10 attacks are (essentially) always detected for both Uniform and Gamma distributed measurements. For the 3-victim setting, on average 2.76 (2.84) of the victims are successfully detected for Uniform (Gamma) measurements. Hence, even in the "mixed attack" scenario, our proposed detection algorithm nearly captures all the electricity theft activities according to Figure 7, Figure 9 and Table I for both Uniform and Gamma data generating mechanisms.

Remark: On sample size requirements. We conclude by commenting on the sample size needed for high detection and identification accuracy. Note that in our simulation scenarios we set VR=0.1, which as previously mentioned acts as a signal-to-noise measure for the problem at hand. This is a very stringent requirement which consequently requires a fairly large sample size for high detection accuracy. Our simulation results indicate that for measurements obtained every 2 minutes (see real data setting in Section III.B), an attack becomes detectable with high probability if it goes on for 4+ days. At lower sampling frequency (e.g. 15 minute intervals, which is the case for many utilities) the attack needs to last considerably longer.

However, note that [29] reports that for the supervised learning approach used, wherein ground truth data for attacks are needed, an attack becomes detectable if it is ongoing for more than a week. Reference [18] simulates various attacks leveraging an Irish data set comprising of 5,000 customers, whose electricity profiles were monitored twice hourly for 535 days. Various supervised learning techniques were assessed, assuming that the simulated attacks lasted for one week. In paper [26] that employs a similar neighborhood area network architecture, the signal-to-noise ratio used in

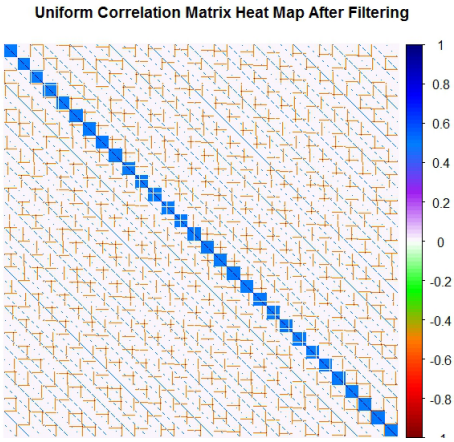


Fig. 10. Heat map of correlation matrix under dependent, no attack case, for data generated from a Uniform distribution.

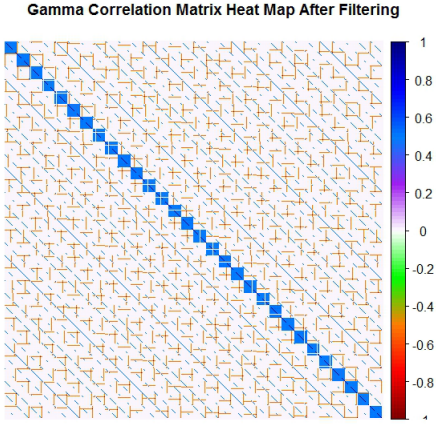


Fig. 11. Heat map of correlation matrix under dependent, no attack case, for data generated from a Gamma distribution.

their numerical work ranges between 1.1-9. For the proposed approach, when VR=0.2, the probability of detection is close to 1 for all attack scenarios and data generating distributions, for sample sizes around 2000. Finally, for VR=0.4, a sample size of 500 suffices for detecting with probability 0.95 a pairwise attack based on Uniform$(600, 700$ distributed independent data with $\mathbb{E}(\alpha) = 0.80$.

2) Dependent Case: For the general dependent (correlated) scenario, we set $P = 30$ and the significance level in the filtering technique to $q = 0.1$. Since we need to generate $\mathcal{O}(P^2)$ differencing variables, to ensure an adequate detection level, a large number of smart meter measurements needs to be generated. In our experiments, we set $n = 10,000$ measurements from $W \sim$ Uniform$(625, 675)$ and $W \sim$ Gamma$(400, 1.2)$ distributions, respectively. Further, we set $U_i \sim N(0, 100)$ and $\rho_i \sim$ Uniform$(0.80, 0.85)$.

Figure 10 and 11 show the heat maps of the correlation matrix estimates of $\mathbf{X}$, when there is no attack for different generating procedures.

For $W \sim$ Uniform$(625, 675)$, we separately launch different types of attacks. To make settings comparable to the independent case, we let $\mathbb{E}(\alpha) = 130$, i.e., approximately 20% of $\mathbb{E}(Y)$, and set *Variance Ratio* the same for each attack group. Analogously to the independent case, we generate α from different Uniform distributions, whose parameters

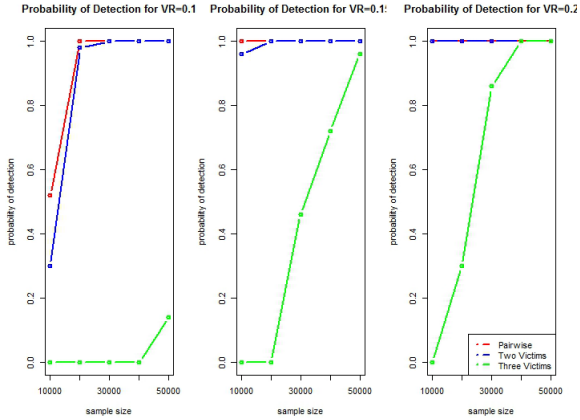


Fig. 12. Probability of detection for various attack scenarios for Uniformly distributed dependent data.

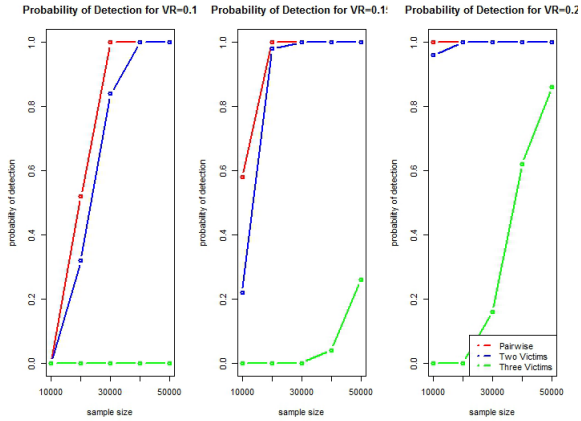


Fig. 13. Probability of detection for various attack scenarios for Gamma distributed dependent data.

are calculated based on the pre-specified mean and VR=0.1 (or 0.2) requirement. As before, we generate 50 data sets to calculate the detection probability. Figure 12 depicts the relationship between the probability of detection, sample size and VR for different types of attacks.

Based on these results we conclude that: (i) an increased sample size improves the probability of detection; (ii) a larger VR, significantly improves the probability of detection; (iii) the dependent case is considerably more challenging that the independent case, since detection is based on differences of the original measurements, that exhibit much higher variability (see for example, equation (II.7)).

For $W \sim$ Gamma$(400, 1.5)$, we let $\mathbb{E}(\alpha) = 120$ in order to be approximately 20% of $\mathbb{E}(Y)$. Figure 13 depicts the relationship between probability of detection, sample size and VR for different types of attacks for this setting. Due to the long right tailed nature of the Gamma distribution, the probability of detection is lower than that for the Uniform distribution.

The counterpart of Table I is shown next. It can be seen that performance deteriorates significantly for complex attack scenarios.

B. A Residential Buildings Example

Next, we use electricity consumption data from buildings at the University of Michigan, Ann Arbor. The form of

TABLE II

AVERAGE NUMBER OF SUCCESSFUL DETECTIONS (WITH STANDARD DEVIATION IN PARENTHESES) FOR 10 PAIRWISE, 5 TWO-VICTIM AND 3 THREE-VICTIMS ATTACKS, BASED ON 50 GENERATED DATA SETS EXHIBITING DEPENDENCE

	Pairwise	2 Victims	3 Victims
Uniform	4.68(1.096)	2.76(0.822)	0.92(0.274)
Gamma	2.60(1.979)	1.20(1.485)	0.40(0.495)

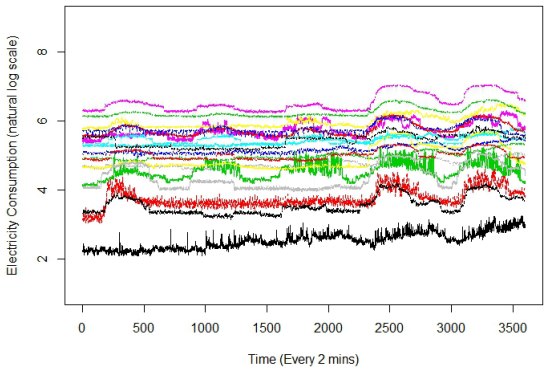


Fig. 14. Electricity consumption of 19 campus buildings (in natural log-scale).

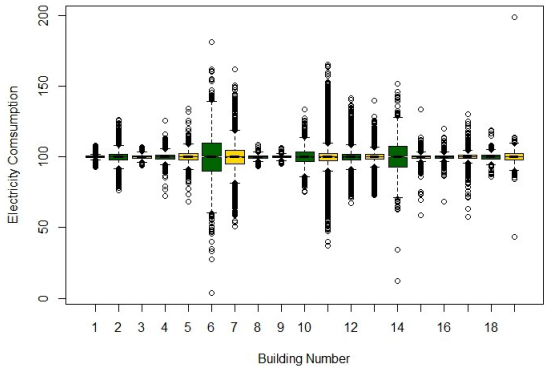


Fig. 15. Boxplots of electricity consumption for the 19 selected buildings.

data is time series based and recorded by the smart meters deployed in the buildings every 2 minutes. The data were preprocessed following the procedure presented in [30] that included exclusion of buildings with highly unusual behavior, such as missing recorded values, long intervals of constant values, abnormal spiky trends and so forth. The extracted data set comprises of $p = 19$ buildings and covers a duration of 5 days, for a total sample size of $n = 3600$. Figure 14 shows the electricity consumption patterns (in natural log-scale) recorded by the selected smart meters.

Further, boxplots of electricty consumption of the 19 buildings under consideration are depicted in Figure 15. It can be seen that the average consumption across all buildings is comparable and hence the emulated attacks launched (see below) are of the same magnitude for all buildings.

Figure 16 shows the heat map for the filtered correlation matrix estimate of the data. It is apparent that the correlation coefficients between different buildings are 0, which helps us to conclude that we fall under the independent scenario.

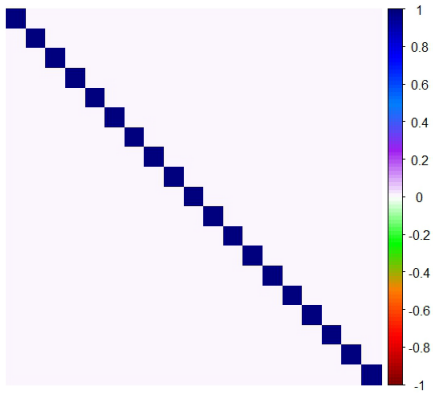


Fig. 16. Filtered heat map for correlation matrix of residential data.

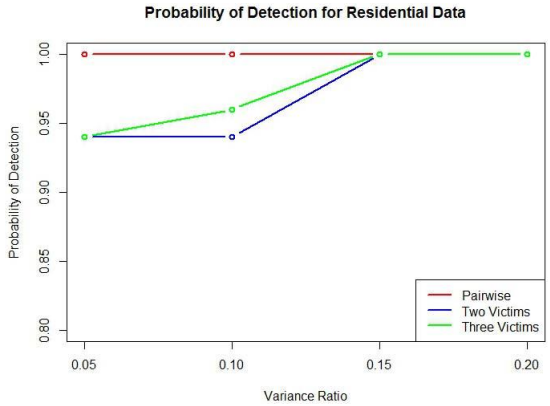


Fig. 17. Probability of detection for emulated attacks on the residential building data.

TABLE III

AVERAGE NUMBER OF SUCCESSFUL DETECTIONS (WITH STANDARD DEVIATION IN PARENTHESES) FOR 2 PAIRWISE, 1 TWO-VICTIM AND 1 THREE-VICTIMS ATTACKS, BASED ON 50 GENERATED ATTACKS BASED ON THE RESIDENTIAL BUILDINGS DATA

	Pairwise	2 Victims	3 Victims
$VR = 0.05$	1.92(0.274)	0.62(0.490)	0.70(0.463)
$VR = 0.10$	1.94(0.240)	0.98(0.141)	0.98(0.141)
$VR = 0.15$	1.94(0.240)	0.98(0.141)	0.98(0.141)
$VR = 0.20$	1.94(0.240)	0.98(0.141)	0.98(0.141)

As a result, the detection algorithm for the independent case is appropriate for this real data example.

To test the algorithm in the real data example, we generate the attack variable from different Uniform distributions such that $\mathbb{E}(\alpha)$ is 20% of the mean consumption level of the selected buildings, and set the *Variance Ratio* to the same value in the set $(0.05, 0.10, 0.15, 0.20)$ for each attack group. Figure 17 depicts the relationship between probability of detection rate and VR for different types of attacks, based on 50 generated attack data sets.

Moreover, we also launch different types of attacks simultaneously, i.e., mixed attacks, to test our detection method. Specifically, we generate 2 pairwise attacks, one 2-Victims attack and one 3-Victims attack. Figure 18 shows the resulting heat map of the correlation matrix estimate, when $VR = 0.15$.

Finally, the following detection Table III is given by replicating the mixed attacks scenario 50 times. It can be seen that once VR≥ 0.10 performance becomes very satisfactory.

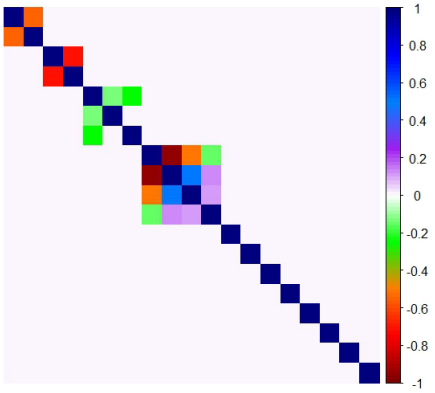


Fig. 18. Filtered heat map for Uniform mixed attacks in real data.

The obtained results amply demonstrate that the proposed detection methodology exhibits good performance and the key conclusions obtained from the synthetic data are applicable to real smart meter measurements.

Finally, we used time series techniques presented in [30] that were applied to the same data set. They included sparse vector autoregressive models and dynamic factor models. Note that the average magnitude of the simulated attacks is 20% of the mean consumption level of each building, which translates to a signal to noise ratio of 1.2 for these models. The detection probability for the simulated scenario presented above is 80%; namely, it corresponds to the proportion of "anomalies" detected in the 9 buildings involved in the attack scenario under consideration. However, such techniques are not able to go beyond the detection stage and thus identify attackers and victims, unlike the proposed approach. The latter constitutes a *key property* of the developed methodology that is particularly useful to utility companies.

C. Some Practical Guidelines to Aid the Proposed Detection Algorithms

Note that regularizing the sample correlation matrix may result in more complicated patterns than expected according to our theoretical results. Examples include setting more elements to zero, the presence of opposite signs within sub-blocks, etc. These issues will result in false positives/negatives for the detection algorithm in the independent case. To address them, we propose certain rules that can be applied as further post-processing of the correlation matrix obtained from the Universal Thresholding method.

Firstly, we denote two types of positions in the covariance matrix, one is the Attacker-Victim position, where all the elements should have negative sign theoretically; the other is the Victim-Victim position, where all the elements should have a positive sign.

1) Under the Pairwise Attack scenario, when dealing with the covariance matrix estimate, if the entry in the Attacker-Victim position has positive sign, we will consider it as noise signal and conclude that there is no attack between these two smart meters;
2) Under the One Attacker-Many Victims scenario, we might have more than one meter node that has negative covariance values with the rest after regularization, which will make it more difficult in identifying the Attacker and Victims. For example, in One Attacker-Three Victims case, the sub-covariance matrix block estimate for attack α is B. We could see that node a has negative covariance value with b and c. In the mean time, node d has negative covariance value with b. But based on the assumptions and detection method that we propose in this paper, there could be one and only one node that has negative covariance value with the rest theoretically. Therefore, it is hard to identify which node is the Attacker given B. To address this, we denote the cardinality of node a as $N(a) = |\{k : Cov(a,k) < 0\}|$. Then, for all the nodes in the sub-covariance matrix block, we let the node with the largest cardinality be the Attacker, and consequently the rest are the victims. In the example below, node a is the Attacker.

$$B = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} + & - & - & + \\ - & + & + & - \\ - & + & + & + \\ + & - & + & + \end{bmatrix} \end{array}$$

3) We could also have some missing entries in the covariance estimate after regularization. Considering the toy example above for the three victims case, note that if the covariance estimate B has the following form with some missing entries in the Victim-Victim positions, we still identify node a, b, c, d as being within the same attack group.

$$B = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} + & - & - & - \\ - & + & + & \\ - & + & + & + \\ - & & + & + \end{bmatrix} \end{array}$$

If the missing values are in the Attacker-Victim positions, we could again use the cardinality rule to address this issue. For example, if estimate B has the following form, by applying our rules, we still identify a, b, c and d being within the same attack group, and the node a is the Attacker.

$$B = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} + & - & & - \\ - & + & + & \\ & + & + & + \\ - & & + & + \end{bmatrix} \end{array}$$

By applying this rule, we could resolve this problem when dealing with real noisy data sets.

IV. Conclusion

In this paper, we have primarily focused on how to address coordinated power theft activities detection problem by considering independent and dependent smart meter data generating mechanisms. For each case, two scenarios, pairwise and one attacker-many victims, have been thoroughly investigated.

We have separately developed an easy-to-implement detection algorithm to detect attacks and identify attackers and

victim nodes. The implementation of the strategy leverages a regularized covariance estimator, followed by close examination of patterns in the resulting matrix. Extensive numerical results based on both synthetic and real data illustrate the superior performance of the proposed methodology.

Note that there is a plethora of machine learning approaches that addresses the detection problem. However, identifying "attackers" and their corresponding "victims" is a more challenging problem that few of these approaches can address. Hence, this constitutes an important feature of the proposed methodology.

There are some open problems that merit additional investigation, including scenarios involving multiple attackers and multiple victims. However, such coordinated attacks are more difficult to launch, since they require a higher level of sophistication from the attacker's perspective.

REFERENCES

[1] T. B. Smith, "Electricity theft: A comparative analysis," *Energy Policy*, vol. 32, no. 18, pp. 2067–2076, Dec. 2004.

[2] T. Ahmad, H. Chen, J. Wang, and Y. Guo, "Review of various modeling techniques for the detection of electricity theft in smart grid environment," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2916–2933, Feb. 2018.

[3] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar. 2009.

[4] H. Gharavi and R. Ghafurian, "Smart grid: The electric energy system of the future [scanning the issue]," *Proc. IEEE*, vol. 99, no. 6, pp. 917–921, Jun. 2011.

[5] G. B. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 107–128, Sep. 2013.

[6] V. Pappu, M. Carvalho, and P. M. Pardalos, *Optimization and Security Challenges in Smart Power Grids*. Heidelberg, Germany: Springer, 2013.

[7] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, Jul. 2017.

[8] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, 2009, pp. 21–32, doi: 10.1145/1653662.1653666.

[9] Z. H. Yu and W. L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[10] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1686–1696, Jul. 2015.

[11] M. G. Kallitsis, S. Bhattacharya, S. Stoev, and G. Michailidis, "Adaptive statistical detection of false data injection attacks in smart grids," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 826–830.

[12] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[13] A. Sanjab and W. Saad, "Smart grid data injection attacks: To defend or not?" in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2015, pp. 380–385.

[14] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Sci. Technol.*, vol. 19, no. 2, pp. 105–120, Apr. 2014.

[15] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1319–1330, Jul. 2013.

[16] A. A. Cárdenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, "A game theory model for electricity theft detection and privacy-aware control in ami systems," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2012, pp. 1830–1837.

[17] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and R. C. Green, "High performance computing for detection of electricity theft," *Int. J. Elect. Power Energy Syst.*, vol. 47, pp. 21–30, May 2013.

[18] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2015.

[19] M. G. Kallitsis, G. Michailidis, and S. Tout, "Correlative monitoring for detection of false data injection attacks in smart grids," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2015, pp. 386–391.

[20] G. M. Messinis and N. D. Hatziargyriou, "Review of non-technical loss detection methods," *Electr. Power Syst. Res.*, vol. 158, pp. 250–266, May 2018.

[21] S. Axelsson, "The base-rate fallacy and the difficulty of intrusion detection," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 3, pp. 186–205, Aug. 2000.

[22] F. Li *et al.*, "Smart transmission grid: Vision and framework," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 168–177, Sep. 2010.

[23] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 5–20, 1st Quart., 2013.

[24] S. Sahoo, D. Nikovski, T. Muso, and K. Tsuru, "Electricity theft detection using smart meter data," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2015, pp. 1–5.

[25] S.-C. Yip, K. Wong, W.-P. Hew, M.-T. Gan, R. C.-W. Phan, and S.-W. Tan, "Detection of energy theft and defective smart meters in smart grids using linear regression," *Int. J. Elect. Power Energy Syst.*, vol. 91, pp. 230–240, Oct. 2017.

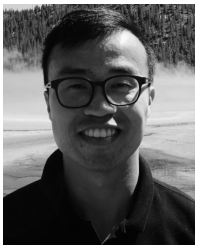
[26] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, "An anomaly detection framework for identifying energy theft and defective meters in smart grids," *Int. J. Elect. Power Energy Syst.*, vol. 101, pp. 189–203, Oct. 2018.

[27] J. L. Viegas, P. R. Esteves, and S. M. Vieira, "Clustering-based novelty detection for identification of non-technical losses," *Int. J. Elect. Power Energy Syst.*, vol. 101, pp. 301–310, Oct. 2018.

[28] N. Bailey, M. H. Pesaran, and L. V. Smith, "A multiple testing approach to the regularisation of large sample correlation matrices," *J. Econometrics*, vol. 208, no. 2, pp. 507–534, 2019.

[29] V. B. Krishna, K. Lee, G. A. Weaver, R. K. Iyer, and W. H. Sanders, "F-DETA: A framework for detecting electricity theft attacks in smart grids," in *Proc. 46th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, Jun./Jul. 2016, pp. 407–418.

[30] M. G. Kallitsis, S. Bhattacharya, and G. Michailidis, "Detection of false data injection attacks in smart grids based on forecasts," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2018, pp. 1–7.



Jin Tao was born in Liuzhou, China. He received the B.Sc. degree (Hons.) in mathematics and applied mathematics from the Renmin University of China in 2012, and the Ph.D. degree in statistics from the Department of Statistics, University of Florida, in August 2018, with the Top-Off Fellow Award at Gainesville, FL, USA. He started a new chapter at J. P. Morgan to pursue career goals in industry. His research interests include computational statistics, data mining, and inverse problems, with applications to engineering and finance problems.

George Michailidis received the B.S. degree in economics from the University of Athens, Greece, in 1987, and the M.A. degree in economics, the M.A. degree in mathematics, and the Ph.D. degree in mathematics from UCLA. After a postdoctoral position in operations research at Stanford University, he joined the Department of Statistics, University of Michigan in 1998, where he became a Full Professor in 2008. In 2015, he joined the University of Florida as the Founding Director of the Informatics Institute. His research interests include network analysis, queuing theory, stochastic control and optimization, applied probability, and machine learning. He is a fellow of the American Statistical Association and the Institute of Mathematical Statistics.