Model-less Non-Technical Loss Detection Using Smart Meter Data

Abu Bakr Pengwah, *Student Member, IEEE*, Reza Razzaghi, *Senior Member, IEEE*,
Lachlan L. H. Andrew, *Senior Member, IEEE*

***Abstract*—This paper proposes a novel method for estimating non-technical losses (NTL), typically due to electricity theft caused by bypassed meters in low-voltage distribution networks. First, the voltage sensitivity coefficients are estimated by a weighted least squares based approach using the residuals obtained from an ordinary least squares-based approach. Then, the voltage measurements of smart meters and the estimated sensitivity coefficients are used to estimate the actual consumption of customers. The differences between the measured and the estimated consumption values are compared against a threshold and customers whose differences exceed the threshold are flagged as fraudulent customers. The performance of the algorithm is evaluated on the IEEE European Low Voltage (LV) network in various case studies, and on large randomly generated distribution networks. The results exhibit the capability of the proposed algorithm in correctly identifying fraudulent meters, together with their actual consumption values.**

***Index Terms*—Power distribution networks, non-technical losses, smart meters, electricity theft.**

I. Introduction

ELECTRICITY demand in distribution networks has been increasing in recent years due to the electrification of transport and buildings. With distribution network service providers (DNSP) having to supply this substantial demand, electricity loss becomes an increasingly important challenge. Electricity losses can be classified into two major categories: technical and non-technical losses. Technical losses, such as Ohmic losses, are inherent to electrical systems. On the other hand, any use of electricity that does not get reported to DNSPs and is not consequently billed is classified as non-technical losses (NTL) [1]. NTL is often theft, such as illegal connections of meters (including electricity theft by bypassed meters) and meters deliberately rendered faulty.

In this paper, the objective is to identify NTLs in the form of electricity theft caused by meters being bypassed at the point of connection.

A. Prior Art

Detection of NTL events has been extensively studied in the literature. A traditional method is in-person meter inspections [2], [3]. However, the substantial cost of the on-site inspections, especially for large networks, and the recurrence of human errors lead to such traditional methods being infeasible. With the advent of advanced metering infrastructure (AMI), such as smart meters, used for billing and operational procedures, the access to regular intervals of measurements opened up new avenues for NTL detection.

State estimation has been one of the approaches used in detecting NTL events. In [4], AMI measurements and the network model is provided to a distribution system estimator (DSE). The DSE outputs estimated measurements of customers that are compared to the measured ones. Based on an Analysis of Variance (ANOVA) statistic, customers with biased differences in the measurements are flagged as fraudulent. In [5], a load flow algorithm is presented in which the differences between the measured and estimated power measurements, using input voltage measurements and line impedance values, are compared to a derived threshold.

Due to the full network model not readily accessible to DNSPs, new studies have focused on data-driven approaches to estimate NTL events. In [6], phasor measurement units (PMU) and intelligent electronic device (IED) are installed in the distribution network to obtain an active representation of the network model. Variance analysis and path finding algorithms based on the differences between measured and estimated power measurements are used to identify the NTL location. In [7], a method based on the disturbances in the estimated line resistances is presented. The incorporation of temperature sensors allows accurate estimation of the resistances of transmission lines, which are then used to separate the technical losses from the total losses in the network. An intermediate monitoring device is used in conjunction with smart meter readings in [8] to detect differences between the upstream and downstream power flow. The main limitations of the aforementioned papers are the cost of the additional metering devices and determining their optimal location in distribution networks.

In the context of NTL detection using smart meter data, in [9] a residuals-based approach is presented. The distribution of the residuals between the estimated and measured active and reactive powers captures NTL events as outliers that can be readily identified, albeit requiring the noise distribution to be known. If the NTL event is consistent across a sampled period, the authors in [10] present a multiple linear regression approach that calculates customers' anomaly coefficients. These anomaly coefficients coupled with a t-statistic test are used for determining whether a particular customer is maliciously stealing electricity or whether its smart meter is faulty. Correlation analysis has also been employed in [11]. A master metering device measurements obtained from the distribution transformer is correlated with the smart meter readings, with any correlation below a specific threshold signifying possibility of electricity theft. While the correlation technique provides a means of identifying fraudulent meters,

it fails to provide any information about the quantity of power that was being misreported. A new technique that detects NTL through spectral properties of the consumption curve is presented in [12]. The downside of using spectral techniques is that any change in the theft pattern can lead to either a false positive or false negative.

Using certain patterns in electricity theft, machine learning algorithms have also been applied onto smart meter readings with the goal of identifying NTL events. While there are multiple studies presenting the applications of machine learning in NTL detection, with some being [13]–[15], these approaches require accurate and large training samples, which might not be readily accessible to DNSPs.

B. Contributions of Paper

Compared to the aforementioned limitations of requiring the full network model [4] (which includes the line impedance values and the topology), or additional metering devices installed in the network [8], or accurate training samples [16], this paper proposes an algorithm that

1) estimates the network parameters from smart meter data, which can be corrupted with electricity theft occurrences, using a weighted least squares method adapted from [17];
2) identifies the times during which the NTL events occur; and
3) estimates the true consumption of the corresponding customers based solely on the voltage measurements and the estimated network parameters.

C. Paper Organisation

The rest of this paper is detailed as follows; The optimization problem estimating the voltage sensitivity coefficients is formulated in Section II. Section III presents the proposed theft detection algorithm that estimates the true consumption of customers based on a novel iterative approach. Section IV demonstrates the performance of the proposed algorithm on the IEEE European LV test feeder and large networks. Finally, Section V provides the concluding remarks.

II. Voltage Sensitivity Coefficients

This section describes the proposed voltage sensitivity coefficients estimation process in the presence of NTL. First, a brief background on the voltage sensitivity coefficients is provided, followed by the ordinary least squares algorithm. Finally, the weighted least squares approach is described.

A. Background

Voltage sensitivity coefficients relate the changes in voltages of customers to changes in their imported/exported power from/to the grid. Voltage sensitivity coefficients have been used in various applications including, operation and maintenance procedures [18], overvoltage mitigation [19], state estimation [20], topology estimation [21]. These coefficients inherently capture the relationship between the consumption of customers and their measured voltages, i.e, they reflect the distribution network model. From [21], the sensitivity matrices, $\mathbf{S}^r$ and $\mathbf{S}^x$, are given as

$$\Delta\mathbf{V} = \mathbf{S}^r\Delta\mathbf{I}^{re} + \mathbf{S}^x\Delta\mathbf{I}^{im}\,, \tag{1}$$

where $\mathbf{V}$, $\mathbf{I}^{re}$ and $\mathbf{I}^{im}$ correspond to the synchronized matrices of customers' instantaneous voltages, real and imaginary parts of currents respectively and the Δ operator denotes differences between measurements at different times (typically consecutive). Since smart meters do not provide voltage and current phasor information, the real and imaginary parts of the load currents, $\mathbf{I}^{re}$ and $\mathbf{I}^{im}$, are defined based on the power factor angle, ϕ, whilst assuming the voltage angles are the nominal angle.

Corresponding to a particular voltage level, if these coefficients are known, the imported power from the grid can be estimated and subsequently compared to measured ones. The deviations between the measured and estimated quantities are useful tools in detecting electricity theft caused by bypassed meters [22]. However, the estimation of the sensitivity matrices is not trivial in the presence of NTL, as will be explained in Section II-C. As such, a novel application of the weighted least squares algorithm in [17] is proposed to improve the estimates of the voltage sensitivity coefficients, despite interference from electrical theft. In the rest of this section, the method of estimating the sensitivity matrices, $\mathbf{S}^r$ and $\mathbf{S}^x$, in the presence of NTLs is presented.

B. Ordinary Least Squares (OLS)

With no explicit details on the NTL and the model of the network, the first step is to obtain a crude estimate of the sensitivity matrices, $\mathbf{S}^r$ and $\mathbf{S}^x$. Given smart meter measurements of voltages, $\mathbf{V}$, currents, $\mathbf{I}$ and power factor angle, ϕ, (1) can be formulated with the sensitivity matrices as the main decision variables. For a distribution network with N customers, the measurement matrices are of sizes $N \times T$, where T is the number of sampled measurements (recorded time stamps by smart meters), while the sensitivity matrices are each of size $N \times N$. The approach of estimating the sensitivity matrices is taken from [21].

The method in [21] is summarized as the following. The linear problem in (1) is formulated as a least squares error function. Linear constraints derived from the physical properties of sensitivity coefficients and from the topological properties of distribution networks are imposed onto the least squares error function. The linear constraints force the estimated sensitivity matrices to be symmetric and the diagonal elements to be the largest in their respective rows. The constraints dealing with the topological properties of distribution networks involve estimating net positive values of distance matrices (impedance found in the path linking two customers), and obeying both the semi-definiteness rule and triangle inequality for distance matrices of radial trees.

After vertically concatenating the real and imaginary parts of currents, the resultant second order cone programming

problem is defined as

$$\min_{\mathbf{S}^r,\mathbf{S}^x} \quad \left\| \Delta\mathbf{V} - \begin{bmatrix}\mathbf{S}^r & \mathbf{S}^x\end{bmatrix} \begin{bmatrix}\Delta\mathbf{I}^{re} \\ \Delta\mathbf{I}^{im}\end{bmatrix} \right\|_F^2 \qquad (2)$$
$$\text{subject to:} \quad (3)\text{ - }(7),\ (9)\text{ in [21]}.$$

The first estimates of the sensitivity matrices obtained from (2) are denoted as $\hat{\mathbf{S}}^r$ and $\hat{\mathbf{S}}^x$.

C. Weighted Least Squares

When electricity theft occurs, the values of $\mathbf{I}$ above are not accurate. It is useful to break (1) into two parts

$$\Delta\mathbf{V} = \mathbf{S}^r\Delta\mathbf{I}^{re} + \mathbf{S}^x\Delta\mathbf{I}^{im} + \mathbf{S}^r\Delta\mathbf{I}^{re,ntl} + \mathbf{S}^x\Delta\mathbf{I}^{im,ntl}, \quad (3)$$

where $\Delta\mathbf{I}^{re,ntl}$ and $\Delta\mathbf{I}^{im,ntl}$ correspond to the real and imaginary current differences being consumed by the bypassed meters. As such, the estimates of $\hat{\mathbf{S}}^r$ and $\hat{\mathbf{S}}^x$ (2) have high errors. If these errors are ignored, this can result in significant deviations in the estimated consumption of the customers leading to erroneous estimates of NTLs as documented in [5]. Note that, the errors due to the bypassed meters are encapsulated in the voltage residuals of the fitted equation (1). If $\Delta\mathbf{I}^{re,ntl} \neq \mathbf{0}$ or $\Delta\mathbf{I}^{im,ntl} \neq \mathbf{0}$, then $\Delta\mathbf{V}$ in (1) using $\hat{\mathbf{S}}^r$ and $\hat{\mathbf{S}}^x$ will not match the measured $\Delta\mathbf{V}$. This will cause a dramatic shift in the residuals, which signifies the period during which the bypassed meters are active. These residuals have been used in previous works with the objective of identifying the bypassed meters [23]. In this paper, the relationship between shifts in residuals and electricity theft occurrences is exploited to reduce the errors in the estimates of $\mathbf{S}^r$ and $\mathbf{S}^x$. A weighted least squares (WLS) scheme based on the residuals of the OLS approach is proposed to improve the estimate of the sensitivity matrices when the dataset is corrupted with theft occurrences.

WLS is a modification to OLS that incorporates a weighting matrix, $\mathbf{W}$, that can indicate which samples $t \in T$ contain more useful information and less noise [17]. Ideally, the weights should be chosen based on the variances of the different components, but in this application those variances are not known. Instead, the residual matrix, $\mathbf{r}$, corresponding to $\hat{\mathbf{S}}^r$ and $\hat{\mathbf{S}}^x$ is first calculated as

$$r_{n,t} = \left| V_t^{tr} - V_{n,t} - \begin{bmatrix}\hat{\mathbf{S}}^r & \hat{\mathbf{S}}^x\end{bmatrix} \begin{bmatrix}\mathbf{I}_t^{re} \\ \mathbf{I}_t^{im}\end{bmatrix} \right|, \qquad (4)$$

where V_t^{tr} is the measured voltage at the secondary side of the distribution transformer at time t.

Instead of weighting each time step, the region where the residuals are high is identified. This ensures that the period where the bypassed meter starts and stops consuming is under weighted. A simple moving average process with a window size of 12, which corresponds to a 1-hr sampling period for a smart meter recording at $5-$mins intervals, is applied on $\mathbf{r}$. In this paper, smart meters recording at 5-mins intervals were used, but the proposed approach can be applied to smart meters with coarser resolutions. The mean residual taken over all customers, $\bar{\mathbf{r}}_t$, is calculated at each time t and the weighting matrix, $\mathbf{W}$, is defined as the reciprocal of the mean residual,

$$\mathbf{W} = \begin{bmatrix} \bar{r}_{t_1} & 0 & \dots & 0 \\ 0 & \bar{r}_{t_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{r}_{T-1} \end{bmatrix}^{-1}. \qquad (5)$$

Finally, the weighted least squares error problem is formulated as

$$\min_{\mathbf{S}^r,\mathbf{S}^x} \quad \left\| \mathbf{W}\left(\Delta\mathbf{V} - \begin{bmatrix}\mathbf{S}^r & \mathbf{S}^x\end{bmatrix} \begin{bmatrix}\Delta\mathbf{I}^{re} \\ \Delta\mathbf{I}^{im}\end{bmatrix} \right) \right\|_F^2 \qquad (6)$$
$$\text{subject to:} \quad (3)\text{ - }(7),\ (9)\text{ in [21]}$$

with the same constraints as in (2). The process is iterated until the maximum difference between the new and old weighting matrices is below a tolerance of 10^{-4}.

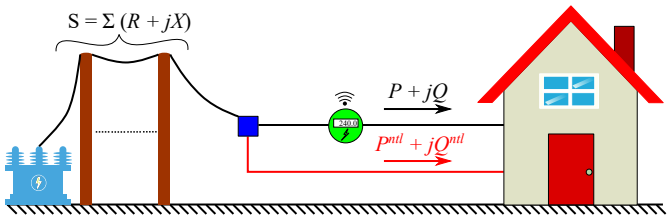

Fig. 1. NTL in the form of bypassed smart meter which measures $P + jQ$, while the unmetered load consumes $P^{ntl} + jQ^{ntl}$. The overall model of the distribution network is encapsulated in the sensitivity estimates, $\mathbf{S}$.

III. NTL Algorithm

Once an estimate of the sensitivity model of the network is obtained, the actual consumption of customers can be estimated by solving the voltage sensitivity equation (1), with the load current being the main decision variables. This section first presents a description of the NTL that is caused by the electricity theft through a bypassed meter. Then, the algorithm that identifies which customer is misreporting their energy usage, together with their actual usage, is detailed.

A. Electricity Theft

In this paper, the objective is to identify customers that bypass their smart meters. This type of NTL is shown in Fig. 1. A bypassed meter is present when an unmetered load is connected at the point of connection [24]. As such, the true consumption value does not get reported to the DNSP. However, assuming negligible electrical distance between the bypassed meter and the unmetered load, not only can the voltage of the unmetered load be considered equal to that of the bypassed meter but the estimated $\mathbf{S}^r$ and $\mathbf{S}^x$ matrices up to the point of theft still serve as a viable sensitivity model of the distribution network. The remainder of this section introduces an algorithm that identifies the bypassed meters based on the measured voltages and the estimated sensitivity model.

The complete equation relating the estimated sensitivity model, the measured voltages, $\mathbf{V}$, and currents, $\mathbf{I}$, is

$$V_t^{tr} - \mathbf{V}\angle\boldsymbol{\theta}^v = \mathbf{SI}\angle\boldsymbol{\theta}^i, \qquad (7)$$

where $\mathbf{S} = \mathbf{S^r} + j\mathbf{S^x}$, $\boldsymbol{\theta}^v$ and $\boldsymbol{\theta}^i$ correspond to the voltage phasor angle and current phasor angle, respectively. In the presence of a bypassed meter, (7) is expanded with $\mathbf{I}$ consisting of both the measured current $\mathbf{I}^{mea}$ and the bypassed current $\mathbf{I}^{ntl}$ as

$$V^{tr} - \mathbf{V}\angle\boldsymbol{\theta}^v = \mathbf{S}\left(\mathbf{I}^{mea}\angle\boldsymbol{\theta}^{i,mea} + \mathbf{I}^{ntl}\angle\boldsymbol{\theta}^{i,ntl}\right). \quad (8)$$

If V^{tr}, $\mathbf{V}$, $\boldsymbol{\theta}^v$, $\mathbf{S}$, $\mathbf{I}^{mea}$ and $\boldsymbol{\theta}^{i,mea}$ are known, solving (8) will yield the actual current being drawn by the bypassed meter, $\mathbf{I}^{ntl}\angle\boldsymbol{\theta}^{i,ntl}$. However, smart meters do not provide the voltage and current phasor angles. Therefore, (8) is simplified to a case for which a solution can be obtained by a novel iterative approach.

In distribution networks, the power factor of all the loads is close to unity, i.e, the differences between power factor angles are negligible [25]. This leads to the simplification, $\boldsymbol{\phi}^{mea} \approx \boldsymbol{\phi}^{ntl}$. If ϕ is defined as $\theta^v - \theta^i$, this simplifies (8) to

$$V^{tr} - \mathbf{V}\angle\boldsymbol{\theta}^v = \mathbf{S}\left(\mathbf{I}^{mea} + \mathbf{I}^{ntl}\right)\angle(\boldsymbol{\theta}^v - \boldsymbol{\phi}^{mea}). \quad (9)$$

The simplification of the bypassed meter having the same power factor angle as the measured load will be shown numerically to have minimal impact on the performance of the algorithm, even in extreme scenarios. The unknowns in (9) are the bypassed meters' currents $\mathbf{I}^{ntl}$ and the voltage phasor angles θ^v. To find these, an iterative algorithm that estimates the voltage phasor angle together with the actual load currents is proposed.

B. Estimating the load currents and voltage phasor angles

Starting with the measured voltages, currents and power factor angles, the goal is to obtain the actual current and voltage phasor angles that fit (9). First, the voltage phasor angles, $\boldsymbol{\theta}^v$, are initialized as zero. The load currents, in the kth iteration, corresponding to the measured voltages, estimated network model, power factor angles and voltage phasor angles in the $k-1$st iteration is calculated as

$$\mathbf{I}_k = \mathbf{S}^\dagger \frac{V^{tr} - \mathbf{V}\angle\boldsymbol{\theta}^v_{k-1}}{\angle(\boldsymbol{\theta}^v_{k-1} - \boldsymbol{\phi}^{mea})}, \quad (10)$$

where $\mathbf{S}^\dagger$ is the pseudo-inverse of the complex sensitivity matrix. A temporary variable, $\boldsymbol{\beta}_k$, is calculated from the currents estimated in the kth iteration as

$$\boldsymbol{\beta}_k = \frac{V^{tr} - \mathbf{S}\mathbf{I}_k\angle(\boldsymbol{\theta}^v_{k-1} - \boldsymbol{\phi}^{mea})}{\mathbf{V}}. \quad (11)$$

At the last step, the voltage phasor angle is adjusted given the calculated temporary variable, and is calculated as

$$\angle\boldsymbol{\theta}^v_k = \frac{\boldsymbol{\beta}_k}{|\boldsymbol{\beta}_k|}. \quad (12)$$

The overall process continues until the maximum difference in $\mathbf{I}$ between the kth and $k-1$st iteration is below 10^{-4} A. The region of convergence of the proposed iterative approach was tested in 1000 random instances of topologies with several values of $\mathbf{V}$, $\mathbf{I}$, and $\boldsymbol{\phi}$ and the algorithm always converged for average power factors higher than 0.25.

C. NTL Detection

Once the estimates of load currents, $\mathbf{I}$, and the voltage phasor angles, $\boldsymbol{\theta}^v$, have been obtained from the iterative process, the bypassed meter can be detected by observing $\mathbf{P}^{mis} = \hat{\mathbf{P}} - \mathbf{P}$, the difference between the estimated and measured active powers of the customers, respectively. The mismatch matrix, $\mathbf{P}^{mis}$, is the stolen active power.

The mismatch matrix is compared to a threshold, P_τ, called the minimum detectable power (MDP) [5]. In [5], MDP is calculated using the full network model, which includes knowledge of the line impedance values and topology of the network. However, the full network model is seldomly accessible to DNSPs. To this end, a method of estimating the MDP parameters using the voltage sensitivity coefficients is proposed. The threshold must reflect several parameters, such as the measurement uncertainty, e^v, and the network sensitivities, but there is no specific formula that must be satisfied. For concreteness, the MDP threshold, P_τ, is calculated by considering the maximum voltage error at all customers in the network and is given as

$$P_\tau = \frac{1}{N}\Re\left(V_1^{tr} e^v \mathbf{J}_N^T \mathbf{S}^\dagger \mathbf{J}_N\right)^*, \quad (13)$$

where the * operator corresponds to the conjugate function, $\Re$ corresponds to the real part operator,and $\mathbf{J}_N$ is an $N \times 1$ matrix of ones. The mean is considered since the spread in the values of P_τ is wide. This is due to upstream customers closer to the distribution transformer having a net smaller impedance compared to downstream customers which result in a broad spectrum of P_τ. The definition of e^v is based on the type of smart meters present in a network. Smart meters are categorized in different classes with the most notable one being the 0.2S model, in which the maximum voltage error can be 0.2% [26]. For a LV distribution transformer with a nominal voltage of 240 V, this corresponds to a maximum error of $e^v = 0.48$ V.

Measurement noise affects this estimation in two ways: it causes inaccuracy in the sensitivity estimate $\mathbf{S}$, and errors in the individual estimates of the power mismatch. In principle, the accuracy of $\mathbf{S}$ can be restored by sampling over a longer window, proportional to the noise variance or the square of the meter tolerance. Errors in the power mismatch estimates can also be addressed by smoothing, as follows.

First, a simple moving average process with a window size of 12 samples is applied on $\mathbf{P}^{mis}$ to mitigate the short-term fluctuations due to e^v. If the noise samples are independent (like thermal noise and not, for example, like noise dominated by limited resolution of the meter) the probability of a consecutive sequence of samples of $\mathbf{P}^{mis}$ being false positives drops exponentially in the length of the sequence, whereas theft is highly correlated. Then, to identify the bypassed meters, the entries of the mismatch matrix are compared against P_τ. The matrix $\boldsymbol{\mathcal{E}}$ of size $N \times T$ is defined as

$$\mathcal{E}_{n,t} = \begin{cases} 1 & \text{if } \min_{i=0,\ldots k-1} P^{mis}_{n,t+i} \geq P_\tau, \\ 0 & \text{otherwise}, \end{cases} \quad (14)$$

where $k = 6$ samples correspond to 30 minutes at 5-minute intervals. For all cases in which $\boldsymbol{\mathcal{E}} = 1$, the estimated

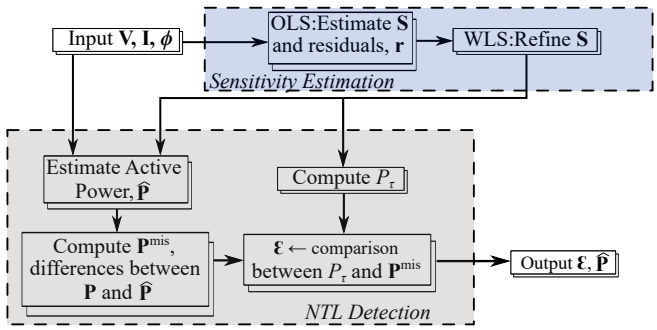

Fig. 2. Flowchart for proposed algorithms.

consumption value, $\hat{\mathbf{P}}$ of the corresponding customers is also output by the algorithm. The overall algorithm is summarized by the flowchart shown in Fig. 2.

IV. Performance Evaluation

The performance of the proposed algorithms is evaluated by simulations on the IEEE Low Voltage European Test Feeder network [27]. The IEEE network, shown in Fig. 3, is implemented in Matlab [28] and various case studies are presented: i) base scenario showing improvement in estimated sensitivity matrices using WLS; ii) impact of random theft occurrences; iii) impact of different theft levels ranging from 300 W to 5 kW; iv) impact of dissimilar power factor angles; v) impact of multiple NTLs; and vi) impact of network topology.

The algorithm has two major outputs: i) status of electricity theft (active/inactive) at customer i at time t; ii) the estimated consumption of customer i at time t, if theft occurs. The performance of the algorithm in terms of the two outputs is assessed using the three statistics, namely sensitivity (true positive rate), specificity (true negative rate) and accuracy (Acc) [29]. Sensitivity is defined as the fraction of correctly classified theft occurrences penalized by the number of false negatives (FN), while specificity is defined as the fraction of correctly classified non-theft occurrences penalized by the number of false positives (FP). Finally, accuracy is defined as

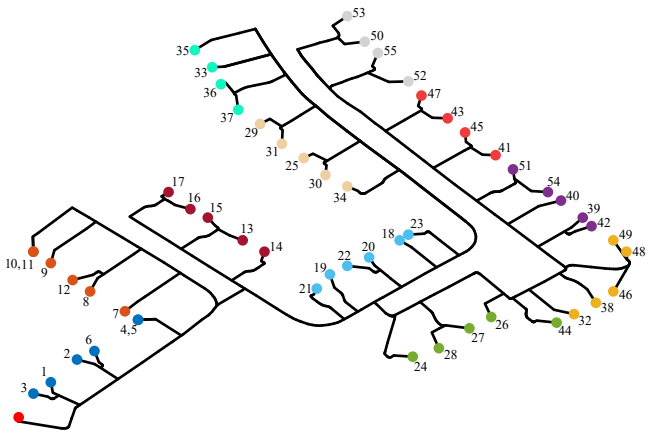

Fig. 3. IEEE European LV Test Feeder [27] with the customers numbered from 1 to 55. The customers are grouped into 11 different clusters shown by the various colors. The clusters are formed based on the approximate distances between the customers and the LV transformer.

TABLE I
Metrics Definition

NTL Status	Predicted Active	Predicted Inactive
Active	TP (Predicted Active NTL as Active)	FN (Predicted Active NTL as Inactive)
Inactive	FP (Predicted Inactive NTL as Active)	TN (Predicted Inactive NTL as Inactive)

the ratio of the sum of true positives (TP) and true negatives (TN) over the total number of samples. The basic metrics are summarized in Table I and the statistics are calculated as

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \tag{15}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}, \tag{16}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{17}$$

The customers have been grouped into 11 clusters based on their approximate distances to the LV transformer, represented by various colours. It consists of an LV distribution transformer at 0.416 kV and 55 customers. To simulate Australian distribution networks which are multiple-earthed-neutral (MEN) networks [30], ground rods are attached to the neutral lines at each customer's premises.

The active and reactive load profiles, corresponding to a day, are taken from the PES database [27]. To increase the size of the data-set to 1 month, power factor samples are drawn from independent uniform random distributions with minimum and maximum of 0.9 and 1, following which additional reactive power consumption values of customers are calculated given the active power load profiles. The active and reactive power samples are input to the algorithm.

To simulate the electricity theft, a random period in the active power samples corresponding to a particular customer is chosen and increased by random values drawn from an independent Gaussian distribution with mean μ and standard deviation σ. The reactive power of the bypassed meter is adjusted to keep the overall power factor similar to the measured load, except in subsection IV-D. In all subsequent studies, only NTLs in the aforementioned form are simulated; in particular, none are located between metered loads. The load-flow analysis is performed to calculate the customers' voltages, which are then given as the last input to the algorithm.

A. Base scenario

In the base case scenario, 55 cases are simulated, each with theft occurring at a different one of the 55 nodes. In each case, $\mu = 3$ kW and $\sigma = 500$ W . The set of measured voltages, active and reactive power obtained from the load-flow analysis are given as input to the voltage sensitivity algorithm. The 2-norm error (matrix norm induced by the vector 2-norm) between the true and the complex sensitivity matrices ($\mathbf{S}^r + j\mathbf{S}^x$) estimated using the two different approaches, namely OLS [21] and the proposed weighted least squares, is plotted in Fig. 4. The proposed weighted least squares scheme clearly outperforms [21], with more accurate sensitivity coefficients returned across all cases as the location of the electricity theft changes.

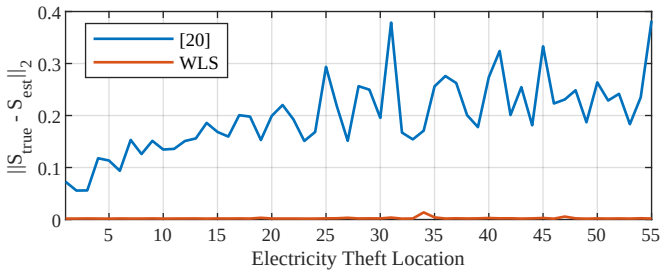

Fig. 4. 2-norm (matrix norm induced by the vector 2-norm) error between true and estimated $\mathbf{S}$ for an NTL with $\mu = 3$ kW and $\sigma = 500$ W at different customer in the distribution network.

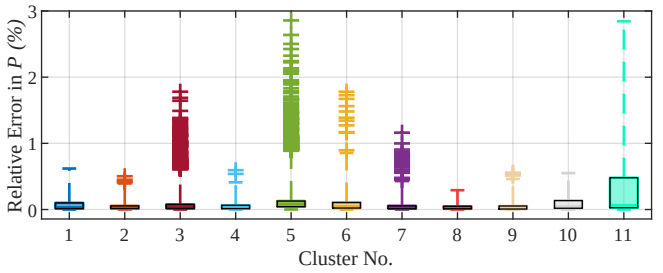

Fig. 5. Base case with $\mu = 3$ kW and $\sigma = 500$ W theft at different customers. Boxplot of relative error between true active power consumption of customer stealing electricity and its estimated value. The 11 clusters are color coded in a similar fashion to Fig. 3.

The estimated sensitivity coefficients from the weighted least squares are then given as input to the NTL detection algorithm. The NTL algorithm has two outputs: 1) whether a particular customer is stealing electricity at time t and 2) the actual consumption of the customer stealing electricity. The proposed NTL detection algorithm correctly detects the NTL location across all instances and times, t, i.e., accuracy $= 1$, sensitivity $= 1$ and specificity $= 1$. The distribution of the relative error, as a percentage, between the true and the estimated active power profile of the customer stealing electricity is plotted in Fig. 5. The 11 boxplots correspond to the 11 clusters represented by similar colors in Fig. 3. The maximum relative error is consistently below 3% ($\approx$ 60W). This shows that the iterative approach accurately converges to the true consumption of the customer stealing electricity.

B. Impact of random theft occurrences

1) Fraction of Time Theft Occurs: To study the impact of the amount of theft occurrences on the dataset, random theft occurrences corresponding to different portions of the data-set are simulated. One month of samples is used and in this case study, different lengths of theft occurrences are simulated. The fraction of time theft occurs is defined as the ratio between the number of samples during which theft occurs and the total number of samples. With a fraction value of 1% indicating that across the 8640 samples (corresponding to 1 month), theft occurs on 864 independently chosen samples, the model is simulated with various fraction of times ranging from as low as 1% to 100% (electricity theft occurs across all samples). Across all instances, the characteristics of the electricity theft are $\mu = 3$ kW and $\sigma = 500$ W and its location is swept across all customers. The accuracy, sensitivity and specificity with respect to each fraction of time is calculated.

TABLE II
PERFORMANCE EVALUATION FOR VARIOUS ACTIVE PERIODS OF THEFT

Fraction of Time Theft occurs (%)	Accuracy	Specificity	Mean Relative error in $P(\%)$
1	1	1	0.0111
5	1	1	0.0129
10	1	1	0.0165
25	1	1	0.0374
50	1	1	0.142
75	1	1	0.450
90	1	1	1.195
100	0.999	0.9991	12.105

The results are summarized in Table II. The theft detection algorithm correctly identifies the instances in which the status of the NTL is active, i.e., the sensitivity is 1 across all simulated theft occurrences. The accuracy and specificity is 1 across all fractions except for the case where the NTL is active across the whole sampled period (worst-case scenario). In the worst-case scenario, the data-set is significantly contaminated with theft such that the proposed weighted least squares scheme estimates higher errors in the voltage sensitivity matrices. This is shown in Fig. 6. The 2-norm error can be seen to dramatically increase for the case in which the NTL is always active, because if WLS applies low weight to all of the corrupted measurements then it reduces to OLS. This error gets carried forward to the theft detection algorithm in which, at the same fraction, the mean relative error between the true and estimated active power consumption values of the bypassed meter increases to around 12% (≈ 360 W). However, this error is still much less than the 3 kW of mean stolen active power.

2) Random step changes in stolen active power: In the previous section, the impact of the fraction of time the NTL is active was studied. In this section, the ability of the algorithm to track changes in the active power theft profile is studied. In this scenario, the stolen active power starts at exactly 3 kW. Within the period in which the NTL is active, random step changes ranging from ± 100 W to ± 500 W are added to the theft profile at every 10-mins intervals. The fraction of time theft occurs is kept to 50%. The experiment is repeated with the location of the NTL swept across all customers and the accuracy, sensitivity and specificity is calculated.

The accuracy, sensitivity and specificity are all 1. A specific theft period is plotted in Fig. 7. The measured active

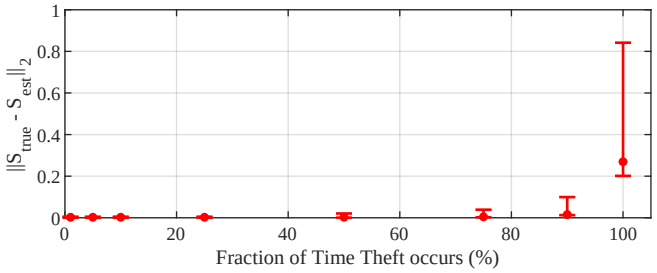

Fig. 6. 2-norm (matrix norm induced by the vector 2-norm) error between true and estimated $\mathbf{S}$ for electricity theft with $\mu = 3$ kW and $\sigma = 500$ W across various active periods of theft.

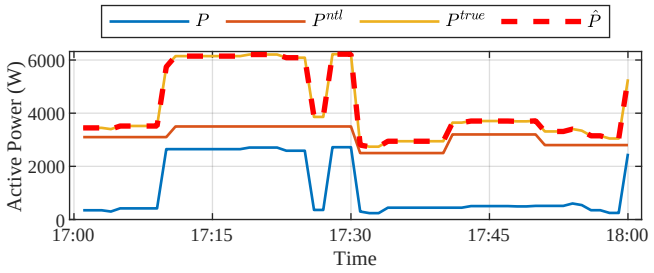

Fig. 7. Active power profiles within a 1-hr period during which NTL is active. The blue, orange, yellow and dashed-red lines correspond to the measured active power, stolen active power, true active power consumption and estimated active power consumption of the customer 55, respectively.

power (blue), active power theft profile (orange), true active power consumption (yellow) and the estimated active power consumption (dashed red) are shown for a 1-hr period in which the NTL is active at customer 55. The plot exhibits successful tracking of the true consumption of the customer as the theft profile changes dynamically.

C. Impact of different theft levels

To assess the proposed MDP threshold, the algorithm is applied to scenarios with different values of active power being stolen by the customers. The bypassed meter's active power consumption is varied from 500 W to 5 kW, with its reactive power adjusted such that the power factor of the unmetered load matches the measured load's power factor. The fraction of time theft occurs is kept fixed at 50% across the sampled period and the times at which theft occurs are randomly determined. The threshold P_τ is 270 W for the test network.

The results are shown in Table III. For any stolen active power exceeding P_τ, the proposed algorithm not only correctly identifies the instances when the bypassed meter is active but also estimates the actual consumption of the customer stealing electricity with minimal errors (0.1% on average). While the specificity remains constant, the sensitivity drops below 1 for mean stolen active power below 400 W. This is due to the randomness in the stolen active power which leads to some values of stolen active power being below P_τ. It is worth noting that having different thresholds for different customers could entail a more sensitive algorithm in which, based on the location, some bypassed meters might not be easily identified.

TABLE III
PERFORMANCE EVALUATION FOR VARIOUS LEVELS OF THEFTS

Stolen Active Power.	Accuracy	Specificity	Sensitivity	Mean Relative error in $P(\%)$
300	0.999	1	0.867	0.0878
400	1.0	1	0.999	0.115
500	1	1	1	0.0847
1000	1	1	1	0.0900
2000	1	1	1	0.0728
3000	1	1	1	0.1439
5000	1	1	1	0.0686

TABLE IV
PERFORMANCE EVALUATION FOR DISSIMILARITIES IN POWER FACTOR OF MEASURED AND BYPASSED METERS

Power Factor of bypassed meter	Accuracy	Specificity	Mean Relative error in $P(\%)$
0.5	0.9995	0.9995	15.0
0.6	0.9999	0.9999	10.8
0.7	1	1	7.4
0.8	1	1	4.51
0.9	1	1	1.66
0.95	1	1	0.119
1.0	1	1	2.05

D. Impact of dissimilar power factor angles

In this sub-section, the impact of the assumption of the bypassed meter having a similar power factor as the measured load is investigated. In previous studies, the bypassed meter had the same power factor as the measured load. In this case study, the bypassed meter's power factor is varied from 0.5 to unity irrespective of the measured power factor, which is on average 0.95. The active power distribution of the bypassed meter's has the parameters $\mu = 3$ kW and $\sigma = 500$ W across all instances and all locations. The model is simulated with the particular power factor and the results are tabulated in Table IV.

The sensitivity is 1 across all values of the power factor, indicating the ability of the proposed algorithm in correctly detecting the times during which the NTL is active. In terms of accuracy, for minute differences between the power factor of the measured load and the bypassed meter, the proposed algorithm is successful in identifying the instances in which the bypassed meter is active. The mean relative error also increases with increasing dissimilarities between the measured load's and the bypassed meter's power factors. When the power factor of the bypassed meter drops below 0.7, the latter's phase angle becomes significant. The proposed iterative approach does not converge to the true solution but instead has some FPs, i.e., specificity < 1. The mean relative error increases to approximately 15% in the worst case scenario. While not insignificant, this worst case scenario of 15% only amounts to approximately 450 W of active power, which is low compared to the mean stolen active power of 3 kW. It is worth mentioning that it is unlikely that the bypassed meter's power factor would be this far from unity.

E. Impact of measurement errors

1) Impact of uniform smart meter class: To study the impact of measurement noise on the performance of the proposed algorithm, three accurate smart meter classes corresponding to 0.1S, 0.2S and 0.5S and two less accurate smart meter classes 1, and 2 are considered [26], [31]. A class x meter is defined as having a relative error up to $x\%$ on the voltage readings. Similar to [31], current measurements can have larger errors compared to voltage measurements, and to model this, errors up to 0.1%, 0.5%, and 1% are added to the current and power factor measurements from smart meter classes 0.1S, 0.2S, and 0.5S, respectively while for the less accurate smart meter classes 1 and 2, errors up to 2% are added. Truncated zero mean normal distributions are taken as

TABLE V
PERFORMANCE EVALUATION FOR SEVERAL SMART METER CLASSES. OUTPUTS MARKED WITH '–' SIGNIFIES INACCURATE ESTIMATES.

	Smart Meter Class	0.1S	0.2S	0.5S	1	2
	$\|\|\mathbf{S}_{\text{true}} - \mathbf{S}_{\text{est}}\|\|_2$	0.0312	0.0599	0.146	0.295	0.609
Accuracy	$\mathbf{S}_{\text{true}}$	0.963	0.870	0.731	0.664	0.623
	$\mathbf{S}_{\text{est}}$	0.967	0.805	0.641	0.648	0.679
Specificity	$\mathbf{S}_{\text{true}}$	0.963	0.870	0.730	0.664	0.623
	$\mathbf{S}_{\text{est}}$	0.967	0.805	0.642	0.649	0.680
Sensitivity	$\mathbf{S}_{\text{true}}$	1	0.979	0.870	0.704	0.582
	$\mathbf{S}_{\text{est}}$	0.998	0.753	0.305	0.300	0.296
Mean Relative error in $P(\%)$	$\mathbf{S}_{\text{true}}$	34.0	67.8	176	338	–
	$\mathbf{S}_{\text{est}}$	34.5	155	–	–	–

the noise distributions with the truncated values depending on the smart meter classes. Across all smart meter classes, the characteristics of electricity theft are $\mu = 3$ kW and $\sigma = 0$ W, and its location is swept across all customers.

Table V shows that the norm error between the true and estimated $\mathbf{S}$ matrices increases proportionally to the meter error. This error is carried forward to the NTL detection algorithm. Using the estimated sensitivity model, $\mathbf{S}_{\text{est}}$, the smart meter class 0.1S is the best performer with accuracy, specificity, and sensitivity all above 0.9. A drop in performance is seen going from the 0.1S class to the class 2 smart meter. However, the accuracy does not drop linearly to zero but rather flattens around 0.6. The relative error in the estimated $\mathbf{P}$ increases as the meter error increases. Using $\mathbf{S}_{\text{est}}$, for smart meter classes 0.5S and worse, the estimated consumption of consumers cannot be accurately estimated and is marked by '–'. The impact of measurement errors on the estimated $\mathbf{S}$ depends on the number of available measurements. The sample size can be increased to decrease the errors in the estimate of $\mathbf{S}$ for a smart meter class 2 compared to the 0.1S, without changing the number of samples of real-time data required for the NTL detection step.

The performance metrics improve as errors in $\mathbf{S}$ decrease. This is shown in the rows of $\mathbf{S}_{\text{true}}$, in which the true value of the sensitivity model is given as input to the NTL detection algorithm. With $\mathbf{S}_{\text{true}}$, the sensitivity improves significantly to 0.582, from 0.296 using $\mathbf{S}_{\text{est}}$, for the class 2 smart meters. The accuracy for the class 2 smart meters using $\mathbf{S}_{\text{est}}$ is higher than that of $\mathbf{S}_{\text{true}}$ since higher errors in $\mathbf{S}_{\text{est}}$ leads to a higher threshold P_τ, following which it is less likely that the higher threshold will be surpassed for $k = 6$ consecutive samples for a meter not associated with electricity theft. The reliability of indicating theft is happening is important in terms of cost to network operators. As shown by the results, as the smart meter class worsens, the performance of the proposed algorithm degrades, albeit remaining reliable (sensitivity above 0.7) in detecting theft using $\mathbf{S}_{\text{true}}$ for class 1 smart meters.

2) Fraction of class 1 smart meters: In this case study, a scenario is studied in which the network consists of different classes of smart meters. Two smart meter classes, namely 0.2S and 1, are considered. The penetration level of class 1 smart meters, defined as the percentage of class 1 smart meters divided by the total number of smart meters (both 0.2S and 1), is varied from 0% to 100%. The customers associated with each smart meter class are chosen at random. The errors are drawn from the aforementioned truncated normal distributions and the characteristics of electricity theft are $\mu = 3$ kW and $\sigma = 0$ W, and its location is swept across all customers. The accuracy, specificity, and sensitivity are tabulated in Table VI.

TABLE VI
PERFORMANCE EVALUATION FOR VARIOUS PENETRATION LEVELS OF CLASS 1 SMART METERS

		Accuracy		Specificity		Sensitivity	
Penetration Level (%)	$\|\|\mathbf{S}_{\text{true}} - \mathbf{S}_{\text{est}}\|\|_2$	$\mathbf{S}_{\text{true}}$	$\mathbf{S}_{\text{est}}$	$\mathbf{S}_{\text{true}}$	$\mathbf{S}_{\text{est}}$	$\mathbf{S}_{\text{true}}$	$\mathbf{S}_{\text{est}}$
0	0.0599	0.870	0.805	0.870	0.805	0.979	0.753
20	0.150	0.804	0.645	0.804	0.646	0.874	0.441
40	0.198	0.750	0.647	0.750	0.647	0.840	0.488
60	0.233	0.714	0.664	0.714	0.664	0.722	0.351
80	0.266	0.686	0.654	0.686	0.654	0.744	0.316
100	0.295	0.664	0.648	0.664	0.649	0.704	0.300

As the fraction of class 1 smart meters increases, the 2-norm error between the true and estimated $\mathbf{S}$ matrices also increases with the sharpest increase going from a penetration level of 0% to 20%. Accuracy, specificity, and sensitivity also decrease as the penetration level increases. With the inclusion of class 1 smart meters, the consumption of the customers cannot be accurately estimated. In terms of reliability, using $\mathbf{S}_{\text{true}}$ the sensitivity metric has a drop in performance going from a penetration level of 40% to 60% but remains above 0.7 in all cases.

F. Impact of multiple NTLs

The performance evaluation of the algorithm is extended to scenarios where more than 1 bypassed meter is present in the network. The occurrence level of bypassed meters, defined as the number of bypassed meters divided by the number of customers, is varied from 5% to 100% (every meter is bypassed). The mean value of the active power being stolen by the bypass is chosen randomly from the range $\mu = 2$ kW to $\mu = 5$ kW, with $\sigma = 500$ W. The level of theft is chosen uniformly at random between 2 kW and 5 kW in order to evaluate the effectiveness of the proposed algorithm in detecting both high and low level of thefts. The fraction of time theft occurs at each bypassed meter is uniformly distributed between 10% and 95%, chosen independently.

First, the performance of the proposed weighted least squares method is shown in Fig. 8. As the occurrence level of bypassed meters increases, the 2-norm error in the sensitivity estimates also increases. This is due to more time samples being down-weighted since more customers steal at different times (excluding the overlap periods). A significant increase is seen going from an occurrence level of 25% to 50%. The sensitivity estimation process can be improved by using historical data as this can provide more measurement periods during which electricity theft does not occur.

This error gets carried forward to the theft detection algorithm as shown in Table VII. While the accuracy, sensitivity and specificity remain close to 1 as the occurrence level increases from 5% to 10%, the mean relative error between the estimated and true consumption at the bypassed meters increases from 0.4% to 3.36%. Beyond the 10% occurrence

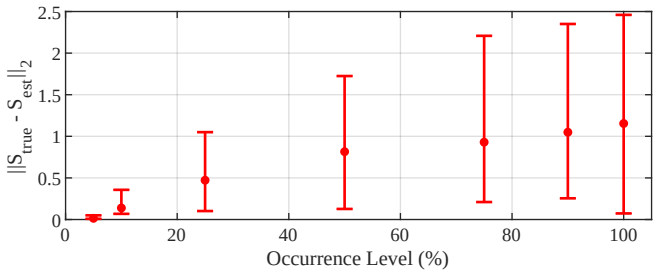

Fig. 8. 2-norm (matrix norm induced by the vector 2-norm) error between true and estimated $\mathbf{S}$ against various occurrences level of bypassed meters in the network.

TABLE VII
PERFORMANCE EVALUATION FOR VARIOUS OCCURRENCE LEVEL OF BYPASSED METERS

Occurrence Level (%)	Accuracy	Specificity	Sensitivity	Mean Relative error in P(%)
5	1	1	1	0.376
10	0.9895	0.9889	1	3.36
25	0.9860	0.9843	1	19.36
50	0.9663	0.9523	9995	9.21
75	0.9088	0.8544	0.9933	12.1
90	0.8722	0.7749	0.9721	137
100	0.5620	0.5982	0.5284	1400

level, the accuracy starts decreasing, reaching 0.59 for ubiquitous theft. This is attributed to the specificity which implies that the number of FN increases, i.e., for some instances in which the NTL is inactive, some customers are estimated to be connected to active bypassed meters. The mean relative error between the estimated and true consumption of the bypassed meters also increases beyond 100% in the high occurrence cases, which are highly unlikely to occur.

G. Impact of distributed energy resources (DERs)

Distribution networks increasingly host more distributed energy resources (DERs). To study the impact of DERs, customers with photovoltaic (PV) systems are simulated, with generation profiles generated from the PV generation profiles in [32]. For customers with PVs, voltage, net current, and power factor of the aggregated load are measured by the smart meter. The customers with DERs are chosen as the ones stealing electricity from the grid and the theft instances are randomly determined. The theft characteristics are $\mu = 3$ kW and $\sigma = 500$ W. The PV penetration level, defined as the percentage of customers with PVs, is varied from 25% to 100%. The location of the DERs (and theft) is swept across all customers in the network. To focus on the impact of DERs, these simulations had no measurement errors.

The accuracy and specificity are tabulated in Table VIII. Across all penetration levels, the sensitivity is 1. The accuracy and specificity are slightly below 1 but always above 99%. The negligible decrease in performance is attributed to an increase in the 2-norm error of the estimated sensitivity matrix, from 0.0018 for no PV penetration to 0.006 on average across all penetration levels. The mean relative error between the actual and the estimated active power profile increases to 4.92% for

TABLE VIII
PERFORMANCE EVALUATION FOR VARIOUS PV PENETRATION LEVEL

PV Penetration Level (%)	Accuracy	Specificity	Mean Relative error in P(%)
0	1	1	0.1
25	0.9974	0.9974	2.25
50	0.9997	0.9971	2.86
75	0.9966	0.9966	3.52
100	0.9958	0.9958	4.92

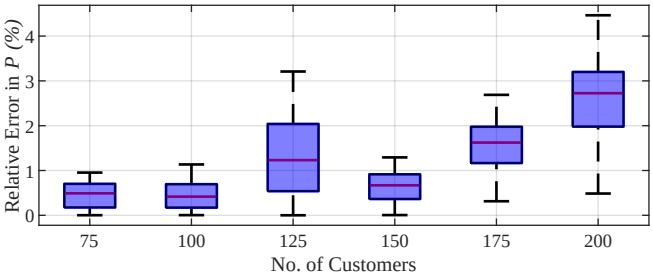

Fig. 9. Boxplot of relative error in the estimated active power profile of customer stealing electricity.

the case with 100% PV penetration, compared to 0.1% without DERs, all due to the error in the estimated sensitivity matrix. This shows that the proposed methodology remains robust on active distribution networks with high DER penetration.

H. Impact of Network Size

To study the impact of the size of the networks, networks with 75, 100, 125, 150, 175 and 200 customers, with 50 different topologies per size, are randomly generated using constrained Prüfer sequences [33]. Realistic distribution networks are generated by constraining the maximum number of branches of buses to less than 5. The line impedance values are sampled from independent uniform random distributions with the bounds similar to the minimum and maximum impedance values of the lines in the IEEE network. The load profiles are sampled from the aforementioned independent distributions. The network is simulated with the distribution of the NTL having parameters $\mu = 3$ kW and $\sigma = 500$ W. Across all network sizes, accuracy, sensitivity and specificity are 1. The proposed approach correctly identifies the instances in which the bypassed meter is active and its location, irrespective of the size of the network. The distribution of the relative error between the estimated and actual active power of the bypassed meter is plotted in Fig. 9.

The relative error increases slightly from 0.4% to 1.7%, on average, as the size of the network increases from 75 to 200 customers. This can be attributed to the mean 2-norm error in the estimated sensitivity matrices which also increases from 0.0015 for a network with 75 customers to 0.004 for another with 200 customers. However, the error in the estimated active power profile is still much lower than the true consumption at the bypassed meter.

I. Computation Time

The proposed method has two main steps: voltage sensitivity estimation and NTL detection algorithm. The estimation of the

voltage sensitivity coefficients via the weighted least squares algorithm has computational complexity $\mathcal{O}(N^3 + N^2T + T^2N)$ [34]. For a similar T, the NTL detection algorithm comprises of matrix multiplication and inversion and has computational complexity $\mathcal{O}(K(N^3 + N^2T))$, where K is the number of iterations taken to reach convergence [35]. Thus, the overall complexity of the algorithm is $\mathcal{O}(N(KN^2 + N^2 + KNT + NT + T^2))$. As N increases, K also increases such that $N^3 = \mathcal{O}(KN^3)$. Thus the overall computational complexity is simplified to $\mathcal{O}(K(N^3 + N^2T) + T^2N)$. This shows that N has a clear impact on computing time.

With T fixed to 1 month of samples, the time cost of the proposed method is measured for networks of several sizes ranging from 10 customers to 100 customers. The simulation was carried out in MATLAB 2022a with MOSEK as the solver. The average computation time required for the entire process increases quadratically from 2.4 seconds for 10 customers to 3 minutes 100 customers, with most of the time taken by the voltage sensitivity estimation algorithm. While the voltage sensitivity estimation algorithm is expensive to perform, due to dealing with large data matrices, it can be performed infrequently, such as once per month. However, the NTL algorithm needs to be applied to incoming smart meter samples. The average computation time for the NTL algorithm increases from 0.14 seconds to 1.8 seconds for 10 and 100 customers, respectively (around 0.02 seconds per customer). This shows that the process is scalable to a large number of customers.

J. Comparison with the state-of-the-art

The proposed approach is compared against state-of-the-art methods [5], [9], [11] in terms of their required inputs, methodology, and their outputs. Table IX summarizes the differences between the approaches. While [11] and the proposed method do not require the prior network model (including line parameters), they are prerequisites for [5] and [9]. All approaches make use of smart meter readings, with the caveat in [11] being an additional master meter placed at the LV transformer. Methodology-wise, [5] and [9] employ load flow and state estimation analyses which are comparable to the weighted least squares approach in the proposed method. The Pearson correlation approach in [11] is the simplest out of all. In terms of the outputs, all approaches estimate the location of the NTL. However, only the proposed method and the algorithms in [5] and [9] can output the actual stolen power associated with the NTL.

TABLE IX
COMPARISON WITH THE STATE-OF-THE-ART

		[5]	[9]	[11]	Proposed Method
Inputs	Model-Less	✗	✗	✓	✓
	Smart Meter	✓	✓	✓* Master Meter	✓
Methodology	Algorithm	Load-flow	State Estimation	Pearson Correlation	Weighted Least Squares
Outputs	NTL Location	✓	✓	✓	✓
	Theft Amount	✓	✓	✗	✓

V. CONCLUSION

In this paper, a model-less method for estimating electricity theft using smart meters in distribution networks has been presented. The proposed approach of estimating the sensitivity coefficients has been shown to provide improved estimates compared to an ordinary regression technique. The key idea in the proposed method is incorporating a weighted least squares scheme that can accurately estimate these coefficients despite interference from electricity theft. These coefficients are given as input to a novel NTL detection algorithm, which both identifies the customers associated with bypassed meters and their actual consumption data. The extensive simulation results, conducted on the IEEE network, exhibit the ability of the proposed algorithms to correctly identify the customer with a bypassed meter, together with its actual consumption value with minimal errors. A promising direction of future research would be extending the algorithm to scenarios with missing measurements.

REFERENCES

[1] L. T. Faria, J. D. Melo, and A. Padilha-Feltrin, "Spatial-temporal estimation for nontechnical losses," *IEEE Transactions on Power Delivery*, vol. 31, no. 1, pp. 362–369, 2016.

[2] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2661–2670, 2019.

[3] J. I. Guerrero, I. Monedero, F. Biscarri, J. Biscarri, R. Millan, and C. Leon, "Non-technical losses reduction by improving the inspections accuracy in a power utility," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1209–1218, 2017.

[4] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, 2013.

[5] T. S. D. Ferreira, F. C. L. Trindade, and J. C. M. Vieira, "Load flow-based method for nontechnical electrical loss detection and location in distribution systems using smart meters," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3671–3681, 2020.

[6] J. B. Leite and J. R. S. Mantovani, "Detecting and locating non-technical losses in modern distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1023–1032, 2016.

[7] M. Tariq and H. V. Poor, "Electricity theft detection and localization in grid-tied microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1920–1929, 2016.

[8] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, "Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid," *IEEE Access*, vol. 7, pp. 129 043–129 053, 2019.

[9] L. M. Raggi, F. C. Trindade, V. C. Cunha, and W. Freitas, "Non-technical loss identification by using data analytics and customer smart meters," *IEEE Transactions on Power Delivery*, vol. 35, no. 6, pp. 2700–2710, 2020.

[10] 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," *International Journal of Electrical Power & Energy Systems*, vol. 91, pp. 230–240, 2017.

[11] P. P. Biswas, H. Cai, B. Zhou, B. Chen, D. Mashima, and V. W. Zheng, "Electricity theft pinpointing through correlation analysis of master and individual meter readings," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3031–3042, 2019.

[12] V. Botev, M. Almgren, V. Gulisano, O. Landsiedel, M. Papatriantafilou, and J. van Rooij, "Detecting non-technical energy losses through structural periodic patterns in ami data," in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 3121–3130.

[13] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2436–2442, 2011.

[14] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, 2010.

[15] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving svm-based nontechnical loss detection in power utility using the fuzzy inference system," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1284–1285, 2011.

[16] P. Massaferro, J. M. D. Martino, and A. Fernández, "Fraud detection on power grids while transitioning to smart meters by leveraging multi-resolution consumption data," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2381–2389, 2022.

[17] Y. Wu, Y. Xiao, F. Hohn, L. Nordström, J. Wang, and W. Zhao, "Bad data detection using linear wls and sampled values in digital substations," *IEEE Transactions on Power Delivery*, vol. 33, no. 1, pp. 150–157, 2018.

[18] B. Liu, K. Meng, Z. Y. Dong, P. K. C. Wong, and X. Li, "Load balancing in low-voltage distribution network via phase reconfiguration: An efficient sensitivity-based approach," *IEEE Transactions on Power Delivery*, vol. 36, no. 4, pp. 2174–2185, 2021.

[19] E. L. da Silva, A. M. N. Lima, M. B. de Rossiter Corrêa, M. A. Vitorino, and L. T. Barbosa, "Data-driven sensitivity coefficients estimation for cooperative control of pv inverters," *IEEE Transactions on Power Delivery*, vol. 35, no. 1, pp. 278–287, 2020.

[20] Q. Liu, Y. Wang, S. Wang, D. Liang, Q. Zhao, and X. Zhao, "Voltage regulation strategy for dc distribution networks based on coordination of centralized control and adaptive droop control," *IEEE Transactions on Power Delivery*, vol. 37, no. 5, pp. 3730–3739, 2022.

[21] A. B. Pengwah, L. Fang, R. Razzaghi, and L. L. H. Andrew, "Topology identification of radial distribution networks using smart meter data," *IEEE Systems Journal*, pp. 1–12, 2021.

[22] X. Cui, S. Liu, Z. Lin, J. Ma, F. Wen, Y. Ding, L. Yang, W. Guo, and X. Feng, "Two-step electricity theft detection strategy considering economic return based on convolutional autoencoder and improved regression algorithm," *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2346–2359, 2022.

[23] Y.-L. Lo, S.-C. Huang, and C.-N. Lu, "Non-technical loss detection using smart distribution network measurement data," in *IEEE PES Innovative Smart Grid Technologies*. IEEE, 2012, pp. 1–5.

[24] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1319–1330, 2013.

[25] D. Flynn, A. B. Pengwah, R. Razzaghi, and L. L. H. Andrew, "An improved algorithm for topology identification of distribution networks using smart meter data and its application for fault detection," *IEEE Transactions on Smart Grid*, doi: 10.1109/TSG.2023.3239650.

[26] *American National Standard for Electricity Meters - 0.2 and 0.5 Accuracy Classes*, ANSI standard C12,20-2010.

[27] *IEEE PES Distribution Systems Analysis Subcommittee's Radial Test Feeders, The IEEE European Low Voltage Test Feeder*. [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources/

[28] Y. Z. Gerdroodbari, R. Razzaghi, and F. Shahnia, "Decentralized control strategy to improve fairness in active power curtailment of pv inverters in low-voltage distribution networks," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2282–2292, 2021.

[29] Q. Li, Y. Liu, Z. Liu, P. Zhang, and C. Pang, "Efficient forwarding anomaly detection in software-defined networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 11, pp. 2676–2690, 2021.

[30] Ausgrid, "Network standard nw000-s0051 ns116 design standards for distribution equipment earthing," ausgrid.com.au, Accessed Jun. 22, 2016. [Online]. Available: https://www.ausgrid.com.au/ASPs-and-Contractors/Technical-documentation/Network-Standards

[31] H. Carstens, X. Xia, and S. Yadavalli, "Measurement uncertainty in energy monitoring: Present state of the art," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2791–2805, 2018.

[32] P. Mohammadi and S. Mehraeen, "Challenges of PV integration in low-voltage secondary networks," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 525–535, 2017.

[33] B. Y. Wu and K.-M. Chao, *Spanning trees and optimization problems*. CRC Press, 2004.

[34] M. Radulović, Ž. Zečević, and B. Krstajić, "Dynamic phasor estimation by symmetric taylor weighted least square filter," *IEEE Transactions on Power Delivery*, vol. 35, no. 2, pp. 828–836, 2019.

[35] J. F. Traub, "Computational complexity of iterative processes," *SIAM Journal on Computing*, vol. 1, no. 2, pp. 167–179, 1972.

Abu Bakr Pengwah (Student Member, IEEE) received the B.E. degree in electrical engineering from Monash University, Victoria, Australia, in 2019. He is currently working toward the Ph.D. degree at Monash University.

His research interests include modelling and identification of distribution networks, smart grids and data analysis.

Reza Razzaghi (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology of Lausanne, Lausanne, Switzerland, in 2016.

In 2017, he joined Monash University, Melbourne, Australia, where he is currently a Senior Lecturer with the Department of Electrical and Computer Systems Engineering. His research interests include distributed energy resources, microgrids, power system protection, dynamics, and transients.

Dr. Razzaghi was the recipient of the 2019 Best Paper Award of the IEEE Transactions on Electromagnetic Compatibility and the 2013 Basil Papadias Best Paper Award from the IEEE PowerTech Conference.

Lachlan L. H. Andrew (Senior Member, IEEE) received the B.Sc. degree in computer science, and B.E. and Ph.D. degrees in electrical engineering from the University of Melbourne, Melbourne, VIC, Australia, in 1992, 1993, and 1997, respectively.

Since 2019, he has been with the University of Melbourne, Australia. From 2014 to 2018, he was with Monash University, Melbourne, VIC, Australia, and from 2008 to 2014, he was with the Swinburne University of Technology, Melbourne, VIC, Australia. From 2010 to 2014, he was an ARC Future Fellow. He was a Senior Research Fellow with the University of Melbourne and a Lecturer with Royal Melbourne Institute of Technology, Melbourne, VIC, Australia. From 2005 to 2008, he was a Senior Research Engineer with the Department of Computer Science, Caltech, Pasadena, CA, USA. His research interests include algorithmic aspects of energy management, and performance analysis of network resource allocation algorithms.

Dr. Andrew was the co-recipient of the ACM SIGMETRICS Test of Time Award in 2021, the IEEE William R. Bennett Prize in 2014, and the Best Paper Award at International Green Construction Code (IGCC) 2012, IEEE International Conference on Computer Communications (INFOCOM) 2011, and IEEE Mobile Adhoc and Sensor Systems (MASS) 2007. He is a member of the ACM.