Detecting and Locating Non-Technical Losses in Modern Distribution Networks

Jônatas Boás Leite, *Member, IEEE*, and José Roberto Sanches Mantovani, *Member, IEEE*

***Abstract*—The recent addition of information and communication technologies in electric power distribution systems has introduced a new class of electricity theft or nontechnical loss. Energy consumption data can be hidden and altered through cyber-attacks that are characterized by the unauthorized access to the application database and digital tampering of smart meters. The development of cost-efficient algorithms to address these types of nontechnical losses also targets the reduction of commercial losses because the full protection of an information system is very expensive. Thus, this paper proposes a strategy to detect nontechnical losses using a multivariate control chart that establishes a reliable region for monitoring the measured variance. After the detection of nontechnical losses, a pathfinding procedure based on the A-Star algorithm is able to locate the consumption point with the non-technical loss. Moreover, a geographical information system application displays the consumption point that is the target of the cyber-attack. The numerical results demonstrate the selectivity and efficiency of the proposed methodology applied for monitoring a real distribution network.**

***Index Terms*—Commercial losses, cyber-attacks, multivariate procedure of monitoring and control, smart metering system, A-star algorithm.**

I. Introduction

ELECTRICAL power loss represents the difference between the quantity of energy injected into an electric distribution system and the quantity of energy that is billed. There are two types of electrical power losses: technical and non-technical. Technical losses comprise the power dissipation in the electrical system components (distribution lines and transformers), whereas non-technical losses are caused by unpredicted external actions against the electrical power system. Non-technical losses are the major source of commercial loss because of the difficulty of measuring them. The most probable causes of non-technical losses are related to frauds, such as the alteration of meter accuracy, consumption of unbilled energy bypassing utility meters, and tapping low-voltage lines.

Manuscript received August 27, 2015; revised December 29, 2015 and April 9, 2016; accepted May 26, 2016. Date of publication June 1, 2016; date of current version February 16, 2018. This work was supported in part by the São Paulo Research foundation—FAPESP under Grant 2014/22377-1 and Grant 2013/23590-8, and in part by CAPES and CNPq under Grant 305371/2012-6. Paper no. TSG-01007-2015.

The authors are with the Electrical Engineering Department, UNESP/FEIS, Ilha Solteira 15385-000, Brazil (e-mail: jonatasboasleite@gmail.com; mant@dee.feis.unesp.br).

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

Digital Object Identifier 10.1109/TSG.2016.2574714

The problem of non-technical losses is typical for utility companies around the world. For example, [1] enumerates and explains the main types of power system losses and indicates some ways to prevent these losses. The installation of pre-paid meters, antifraud conductors and utility information systems can aid in reducing commercial losses. In [2], the indication of solutions that reduce non-technical losses is followed by the proposition of a methodology based on knowledge discovery in databases able to identify suspect energy profiles.

In addition to the problem of non-technical losses, the utility industry has been attempting to address other important challenges, such as generation diversification, demand response and energy conservation, which cannot be addressed using the traditional view of the electricity grid. Smart grids are the next-generation of electricity grids with the capacity to overcome these great challenges [3]. In smart grid architecture, there are three domains: the service provider, the grid and the customer, which are interconnected using a communication network [4]. This architecture guarantees the full visibility and pervasive control over all assets and services of the utility company.

Smart grid characteristics change the nature of electricity theft. Attacks range from crude physical system manipulation to the remote penetration and control of complex computational systems [5]. In [6]–[8], new vulnerabilities of the smart grid infrastructure such as different types of cyber-attacks are identified. Cyber-attacks require multiple defense mechanisms that have high cost for protecting all vulnerable loads in large power systems. Cost-efficient load protection strategies should minimize the cost and prevent damages in the power grid. In this way, [9] assumes the feasibility of cybertampering on electronic meters [10] and proposes a framework to perform online data detection of irregularity in the measurements. The distribution network is divided into subsystems limited by feeder remote terminal units (FRTUs). Each subsystem is checked using the distribution power flow module. The non-technical losses are detected when the mismatch ratio is frequently greater than the predefined threshold. The calculation of the mismatch ratio depends on the average three-phase power consumption, power losses and power measurements for each subsystem. The use of average values requires additional stages to recognize consumption patterns based on historical load profiles.

Although the analysis of real power flows for detecting energy theft is common [11], [12], there are also methodologies based on state estimation [13]–[15]. In [13], for example, a guided search procedure of potential irregularities in the electricity consumption employs the weighted least squares

1949-3053 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

technique for the distribution state estimation. The demand measurements of distribution transformers are collected using the advanced metering infrastructure (AMI). The abnormal use of energy is detected whenever the calculated error is greater than the specified precision. In general, methodologies for detecting real-time malicious data injection in a smart grid demonstrate efficient defense mechanisms against cyber-attacks [14].

The main purpose of this work comprises the development of a cost-efficient methodology able to detect and locate non-technical losses caused by different types of cyber-attacks. Regarding the article structure, Section II describes the possible ways for defrauding the power system and shows a strategy for detecting and locating non-technical losses by exposing essential parts of the proposed strategy, such as the multivariate procedure of monitoring and control, and the derivative algorithm A*. The discussion of the testing results is presented in Section III, which precedes the conclusions in Section IV.

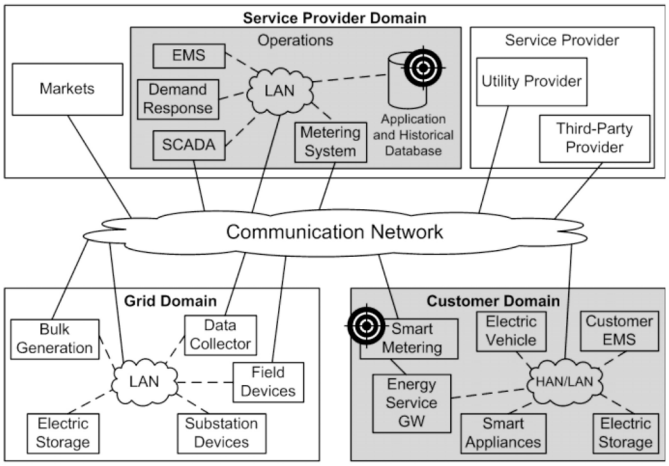

Fig. 1. A simplified information architecture for a smart grid [4] with the probable targets of cyber-attacks.

II. Detecting and Locating Non-Technical Losses

In the past, the recording of consumption data and billing were manual. Technicians visited consumers monthly and frequently uncovered electricity theft during the visual inspection of meters [16]. In the modern power system, the energy billing process is automatic with remote meter recording that eliminates the monthly visit of technicians. Fig. 1 shows a simplified information architecture diagram of the smart grid where critical components for electricity fraud are identified as follows:

1) *Database Attack:* In the operation sub-domain, the local area network (LAN) is a corporate network highly protected by firewalls and other defense mechanisms that make outside cyber intrusion difficult [8]. In this scenario, a significant number of security violations come from company insiders [17]. A utility employee knows how to access corporate computer systems to cause damage or gather information motivated by the prospect of financial gain [18]. An unauthorized employee can invade the application and historical database and delete the data of one particular consumer that becomes unregistered and hidden to the utility company;
2) *Smart Meter Attack:* The operation sub-domain and customer domain are linked by the AMI, which facilitates the bidirectional communication for the transference of control and power consumption data. The core components of the AMI are the smart meters, which are sources of measurement data and other energy-related information [19]. There is absolutely no way for a smart meter that is stuck to the wall and outside the customer's dependencies to remain secure from a physical attack, i.e., the smart meter can be ripped off the wall or smashed. In addition, the computer chips of the smart meter can be breached, the contents exchanged, or new data added [20]. Due to smart meter security constraints, an attack to the smart metering system should consist of the intentional and digital tampering of smart meters to corrupt the measured values [21].

Successful cyber-attacks against the essential components of the automatic billing process, such as the application and historical database and smart metering, must result in non-technical losses whenever the consumption data are excluded or corrupted. Because the full protection of the information system is costly, all available resources of the smart grid must be used for the information protection in a cost-efficient way. Thus, the detection of non-technical losses using data from the grid and customer domains can be developed as an advanced application of the energy management system (EMS). If the EMS has a geographical information system (GIS) application, the detector of non-technical losses can also locate the consumption point where the electricity theft is occurring.

Many error detection schemes compare the received information with reliable information provided by a suitable hash function or checksum algorithm. In information theory concepts, error detection schemes use different methods, such as parity bits, checksums, cyclic redundancy checks, cryptographic hash functions and error-correcting codes, that always compare the received or derivative information with the reliable information. The data from the grid domain are quite reliable because the communication network among field devices utilizes specific standards of the power system (DNP3 and IEC61850) implemented in accordance with the security policy. Moreover, several physical dependencies of the grid domain are protected using a building security system that includes services with alarm monitoring, video control, managed access, and security guards.

Fig. 2 shows the block diagram of the proposed procedure for detecting and locating non-technical losses in distribution networks. The input data of the non-technical loss detector come from the grid and customer domains. Data from the grid domain are reliable states measured by field devices, such as the phasor measurement unit (PMU) and intelligent electronic device (IED). Field devices measure states at terminals of distribution transformers or automatic switches. Reliable states are compared with states calculated by a state estimator that utilizes data from the smart metering system. An algorithm

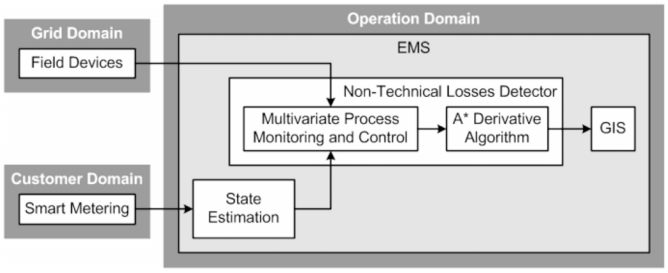

Fig. 2. Block diagram of the detection procedure of non-technical losses.

based on the multivariate procedure of monitoring and control employs the input data to detect a possible power loss at distribution transformer terminals. If the abnormal consumption of energy is detected, an A-star (A*) derivative algorithm searches the consumption point with the power loss. Next, the GIS application receives the identification of the fraudulent consumption point and determines its coordinates in the geographical map.

The state estimation method essentially estimates the set of unknown states using a set of measurements [22]. The AMI in the smart grid can provide the set of measurements using smart meters that measure different types of distribution network parameters, e.g., magnitude and phase-angle of the voltage, current and frequency [23], with a given timestamp [24]. Smart meters are typically placed next to the consumption point in the low-voltage network. In this way, a state estimation method, as provided by [25] and [26], which employs a suitable distribution transformer model for incorporating the state estimation of low-voltage into medium-voltage networks, is required by the proposed detection procedure because the reliable data of comparison come from field devices placed in the medium-voltage networks.

A. Detecting Non-Technical Losses

The comparison of network states in the power system has a selectivity problem because large magnitudes are compared to detect small errors. Voltage and current measurements have large magnitudes, whereas small errors result from voltage and current differences. The utilization of the multivariate procedure of monitoring and control overcomes the selectivity problem in the power system. This procedure is applicable when there are two or more related process variables of interest. In the industry, an automatic inspection procedure simultaneously monitors several parameters on each manufacturing unit to control the final product quality [27].

In univariate statistical quality control, the normal distribution describes the behavior of only one continuous quality characteristic or output variable. This same approach can be used in the case with q output variables where the behavior is described by the multivariate normal distribution. The proposed non-technical loss detector makes use of two variables, i.e., the differences in the voltage and current. Hence, $q = 2$ and a bivariate normal distribution describe the sample space of the measured variables.

Fig. 3 shows the algorithm for detecting non-technical losses. The control chart monitors the sample-generalized

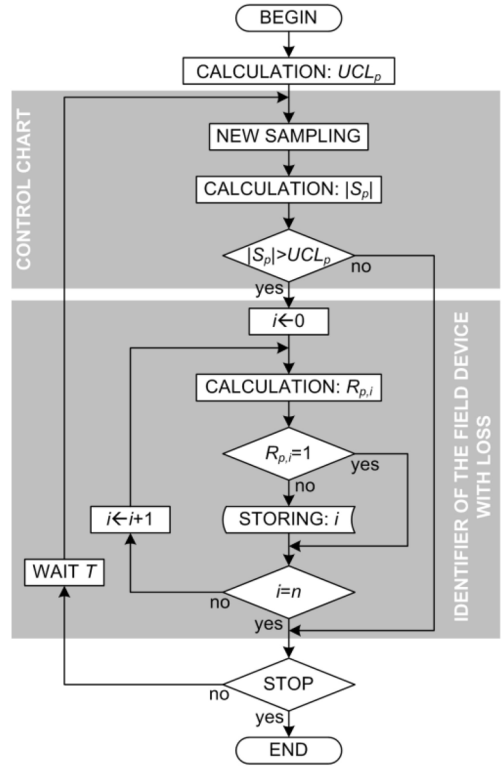

Fig. 3. Algorithm based on the multivariate procedure of monitoring and control for detecting non-technical losses.

variance of voltage and current differences. If the variance is greater than the upper control limit, there is some non-technical loss in the sample space, and most likely, the inspected feeder should have more than one loss. Then, an identification routine is executed to determine field devices that revealed these losses in the distribution feeder. The sample space is randomly built using the acquired information from field devices at the grid domain, which makes the calculations unpredictable for cyber-attacks. Routines involved in the detection procedure are depicted as follows.

1) Sampling: The sampling procedure calculates an error value that represents the voltage and current difference of each field device during a time period, T, as given by (1) and (2).

$$d_{p,1,i} = -\log_{10}\left|\dot{V}_{p,i}^{FD} - \dot{V}_{p,i}^{SE}\right| \quad p = 1, 2 \ or \ 3 \ i = 1, 2, \ldots N \tag{1}$$

$$d_{p,2,i} = -\log_{10}\left|\dot{I}_{p,i}^{FD} - \dot{I}_{p,i}^{SE}\right| \quad p = 1, 2 \ or \ 3 \ i = 1, 2, \ldots N \tag{2}$$

where

$d_{p,1,i}$ voltage difference in the ith field device for phase p;
$\dot{V}_{p,i}^{FD}$ voltage magnitude measured by the ith field device for phase p;
$\dot{V}_{p,i}^{SE}$ voltage profile estimated in the bus of the ith field device for phase p;
$d_{p,2,i}$ current difference in the ith field device for phase p;
$\dot{I}_{p,i}^{FD}$ current magnitude measured by the ith field device for phase p;

$\dot{I}_{p,i}^{SE}$ current flow estimated in the branch of the ith field device for phase p;
N sample space size.

The number of comparisons defines the sample space size of N. The elements of the sample space are randomly chosen to preserve the sampling diversity; furthermore, N is less than or equal to the total number of field devices with measurement units at the grid domain. In this way, the detection algorithm can rapidly and safely check large systems with many distribution feeders once the sampling routine is not performed for all field devices present in the inspected feeder.

2) Upper Control Limit (UCL): The detection algorithm continuously performs two main routines: monitoring of the control chart and identification of the field devices, which are dependent on the preliminary calculation of the upper control limit as given by (3).

$$UCL_p = \left|\Sigma_p\right|\left(b_1 + a_{p,0}\sqrt{b_2}\right),\ p = 1, 2, or3 \tag{3}$$

$$\left|\Sigma_p\right| = \begin{vmatrix} \sigma_{p,1}^2 & \sigma_{p,12} \\ \sigma_{p,21} & \sigma_{p,2}^2 \end{vmatrix} \tag{4}$$

$$b_1 = \frac{1}{(N-1)^q}\prod_{i=1}^{q}(N-i) \tag{5}$$

$$b_2 = \frac{1}{(N-1)^{2q}}\prod_{i=1}^{q}(N-i)\left(\prod_{j=1}^{q}(N-j+2) - \prod_{j=1}^{q}(N-j)\right) \tag{6}$$

$$a_{p,i} = \left\lfloor \frac{1}{q-\mathrm{sgn}(i)(q-1)} \sum_{j=i-(\mathrm{sgn}(i)-1)}^{i-(\mathrm{sgn}(i)-1)q} \times \left(\frac{\mu_{p,j} - \min\limits_{1\le k\le N}\left\{d_{p,j,k}^{PRE}\right\}}{\sigma_{p,j}}\right)\right\rfloor + 1 \tag{7}$$

In the above equations,

UCL_p upper control limit for phase p;
Σ_p covariance matrix of differences for phase p;
$\sigma_{p,1}^2$ variance of voltage differences for phase p;
$\sigma_{p,2}^2$ variance of current differences for phase p;
$\sigma_{p,12}$ covariance between voltage and current differences for phase p;
$\sigma_{p,21}$ same as $\sigma_{p,12}$;
b_1 product of an asymptotic normal approximation used to build the control chart with N samples and q output variables [27];
b_2 same as b_1;
$a_{p,i}$ adjustment coefficient of reliable regions for phase p. If $i=0$, it is the average confidence interval, if $i=1$, it is the confidence interval of voltage differences, and if $i=2$, it is the confidence interval of current differences;
$\mu_{p,j}$ mean of differences for phase p. If $j=1$, it is the mean of voltage differences and if $j=2$, it is the mean of current differences;
$\sigma_{p,j}$ standard deviation of differences for phase p. If $j=1$, it is the standard deviation of voltage differences, and if $j=2$, it is the standard deviation of current differences.

The calculation of UCL_p depends on the preliminary sampled values, $d_{p,j,k}^{PRE}$, used to obtain the covariance matrix Σ_p, which characterizes the bivariate dispersion for a distribution network without any non-technical loss. The calculation of the covariance matrix can be obtained using the Monte Carlo simulation method, which is able to reproduce the required distribution network behavior.

3) Monitoring of the Control Chart: The monitoring routine of the control chart begins with new sampling. In (8)-(11), the statistical parameters of the new sample space are calculated along with the sample generalized variance required to check the feeder.

$$\left|S_p\right| = \begin{vmatrix} s_{p,1}^2 & s_{p,12} \\ s_{p,21} & s_{p,2}^2 \end{vmatrix} \tag{8}$$

$$s_{p,i}^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(d_{p,i,j} - \bar{d}_{p,i}\right)^2 \quad i = 1\ and\ 2 \tag{9}$$

$$s_{p,21} = s_{p,12} = \frac{1}{N-1}\sum_{i=1}^{N}\left(d_{p,1,i} - \bar{d}_{p,1}\right)\left(d_{p,2,i} - \bar{d}_{p,2}\right) \tag{10}$$

$$\bar{d}_{p,i} = \frac{1}{N}\sum_{j=1}^{N} d_{p,i,j} \quad i = 1\ and\ 2 \tag{11}$$

where

$\left|S_p\right|$ sample generalized variance for phase p;
$s_{p,1}^2$ sample variance of voltage differences for phase p;
$s_{p,2}^2$ sample variance of current differences for phase p;
$s_{p,12}$ sample covariance between voltage and current differences for phase p;
$s_{p,21}$ same as $s_{p,12}$;
$\bar{d}_{p,i}$ sample mean of differences for phase p. If $j=1$, it is the sample mean of voltage differences, and if $j=2$, it is the sample mean of current differences.

The control chart is a statistical procedure aimed at the reduction of the measured variability through the monitoring of the sample generalized variance behavior. If the sample generalized variance is inside the reliable region, i.e., the value of $\left|S_p\right|$ is less than or equal to the value of UCL_p, the measured variability is under control. Otherwise, the measured variability is out-of-control, indicating the existence of some non-technical loss in the checked feeder.

4) Identification of Field Devices: The first step of the corrective procedure comprises the identification of field devices with abnormal consumption of energy detected by the previous control chart routine. The identification routine is based on the dispersion diagram of difference values of the current and voltage, as shown by Fig. 4.

The reliable region is defined as the projection of the bivariate dispersion of Σ_p, which is limited by the confidence intervals of the current and voltage differences. Confidence intervals have one lower limit and no upper limit because large

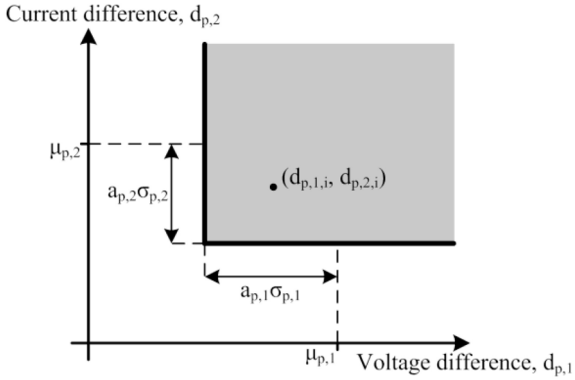

Fig. 4. Emphasis of the reliable region in the dispersion diagram.

dissimilarities produce low values of $d_{p,j,i}$, whereas similarities produce high values, according to (1) and (2). The position identification of the point $i = (d_{p,1,i}, d_{p,2,i})$ in the dispersion diagram, i.e., inside or outside of the bounds of the confidence intervals, is achieved by the calculation of the discrete range, $R_{p,i}$, according to (12).

$$R_{p,i} = \frac{1}{q}\sum_{j=1}^{q} \operatorname{sgn}\left(d_{p,j,i} - \mu_{p,j} + a_{p,j}\sigma_{p,j}\right), \quad i = 1, 2, \ldots n \tag{12}$$

In the identification routine, all of the n field devices of the inspected feeder are checked through the calculation of $R_{p,i}$. If the value of $R_{p,i}$ is equal to one, the point i is inside of the reliable region. Otherwise, the point i is outside of the reliable region, where there are points produced by field devices with abnormal use of energy. The identification procedure stores the index i of the field device with $R_{p,i} \neq 1$. This index is subsequently utilized by the A* derivative algorithm for locating the consumption point with the detected power loss.

B. Locating Non-Technical Losses

The location of the consumption point with the detected power loss requires the use of a search procedure. The A* algorithm is widely used as a pathfinding procedure and employs the best-first search for finding the least-cost path from an initial node to a target node [28]. Pathfinding algorithms are applied in a wide variety of areas, including telephone call routing in communication networks and vehicle routing in road networks [29]. Hart *et al.* [30] originally presented the A* algorithm as a technique that prescribes how the information from a problem domain can be incorporated in a formal mathematical approach for a graph analysis problem. Thus, the least-cost path is found using an evaluation function to determine which nodes the search procedure must visit in the graph tree. The evaluation function is given by the actual cost from the initial node to the actual node plus the estimated cost from the actual node to the target node [31].

In the present location problem, the initial node is the field device identified by the multivariate procedure of monitoring and control. However, the target node is not known because it corresponds to the consumption point with the non-technical loss. This characteristic differentiates the present location problem from the typical routing problem that is solved using the A* algorithm. Furthermore, a layered graph tree can represent the topology of the distribution network section with the non-technical loss. If the root layer has only the initial node, then the target node is located in any one of the downstream layers. In this way, the A* derivative algorithm employs an evaluation function, $L_{p,i,i+1}$, for the layer progression, i.e., to leave the i layer and enter into the $i+1$ downstream layer.

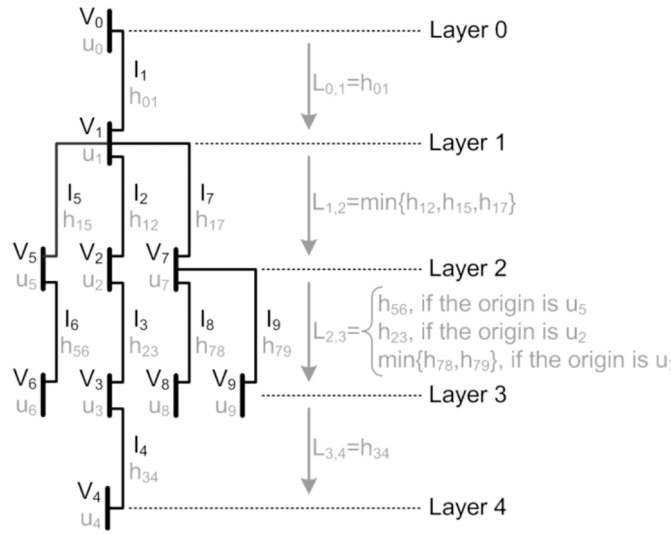

Fig. 5. Section of the distribution network with the evaluation function of layer progression.

Fig. 5 shows the topology of a distribution network section with bus voltages and branch currents. From the viewpoint of graph representation, buses are nodes, branches are edges, and the layer progression goes from the root node to more peripheral nodes. The value of $L_{p,i,i+1}$ is calculated for each layer and reduced during the progression because the value of $L_{p,i,i+1}$ is the minimum estimated electrical distance to the node with the power loss.

$$L_{p,i,i+1} = \begin{cases} \min\limits_{1\le j\le NB_1}\{h_{p,1j}\} & \textit{if the actual node is } u_1 \\ \min\limits_{1\le j\le NB_2}\{h_{p,2j}\} & \textit{if the actual node is } u_2 \\ \vdots & \vdots \\ \min\limits_{1\le j\le NB_k}\{h_{p,kj}\} & \textit{if the actual node is } u_k \end{cases} \tag{13}$$

where

$h_{p,kj}$ estimated electrical distance from node u_j to the target node when the actual node is u_k for phase p;

NB_k number of downstream and adjacent edges to node u_k.

In (13), the calculation of $L_{p,i,i+1}$ minimizes $h_{p,kj}$ and determines the edge with the least estimated electrical distance that represents the best path for the layer progression. After the layer progression, the new actual node is the destination node of the edge with the least estimated electrical distance at the anterior layer. In this way, the search procedure progresses to the next layer until reaching a value near zero, i.e., reaching the target node. If the power loss is added to the load at the

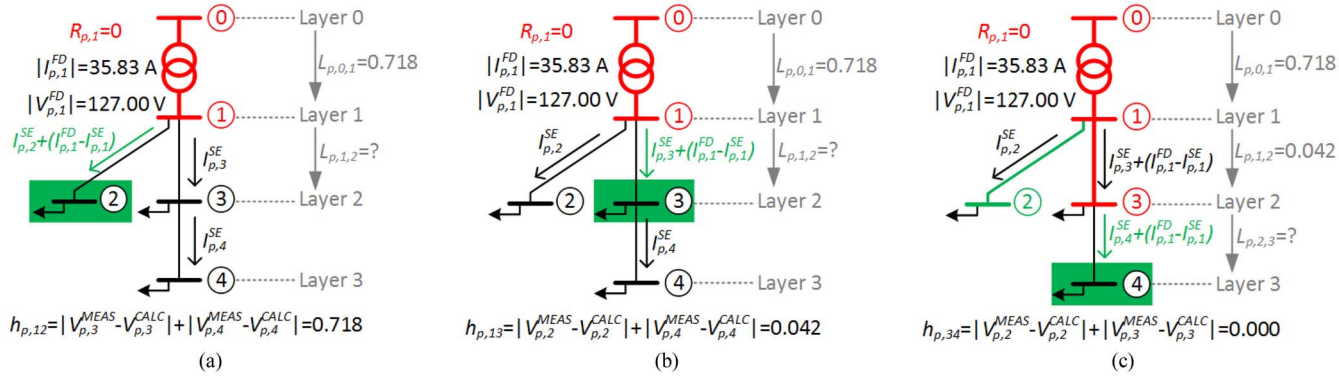

Fig. 6. Schemes for calculating the electrical distances: (a) via bus 2; (b) via bus 3; and (c) via bus 4.

target node and voltage profiles are again calculated, the difference between the calculated and measured voltage profiles must also be near zero. This strategy estimates the electrical distance, $h_{p,kj}$, which is equal to the summation of the absolute difference of the voltage profiles, as given by (14).

$$h_{p,kj} = \sum_{l=1,\, l\neq j}^{N_{SM}} \left| \dot{V}^{MEAS}_{p,l} - \dot{V}^{CALC}_{p,l} \right| \tag{14}$$

where

$\dot{V}^{MEAS}_{p,l}$ voltage profile measured by the lth smart meter for phase p;

$\dot{V}^{CALC}_{p,l}$ voltage profile in the lth smart meter calculated by the forward sweep algorithm for phase p;

N_{SM} number of smart meters minus the meter in the jth bus with the addition of the power loss.

The forward sweep algorithm calculates voltage profiles from the root bus toward more peripheral buses, correcting current flows of branches between the jth bus and root bus, as given by (15) and (16).

$$\left[\dot{V}^{CALC}_{l}\right]_{3x1} = \left[\dot{V}^{CALC}_{U,l}\right]_{3x1} - \left[\dot{K}_l\right]_{3x3}\left[\bar{Z}_l\right]_{3x3}\left[\dot{I}^{SE}_{1}\right]_{3x1} \tag{15}$$

$$\left[\dot{K}_l\right]_{3x3} = \begin{cases} 0 & \textit{if off} - \textit{diagonal} \\ \dfrac{\dot{I}^{SE}_{p,l} + \dot{I}^{FD}_{p,1} - \dot{I}^{SE}_{p,1}}{\dot{I}^{SE}_{p,1}} & \textit{if the branch l is between the } j^{th} \textit{ bus and root bus} \\ \dfrac{\dot{I}^{SE}_{p,l}}{\dot{I}^{SE}_{p,1}} & \textit{otherwise} \end{cases} \tag{16}$$

where

$\left[\dot{V}^{CALC}_{l}\right]_{3x1}$ voltage profile in the lth bus;

$\left[\dot{V}^{CALC}_{U,l}\right]_{3x1}$ upstream voltage profile of the lth bus;

$\left[\bar{Z}_l\right]_{3x1}$ impedance matrix of the lth branch;

$\left[\dot{K}_l\right]_{3x3}$ normalized current matrix in the lth branch;

$\left[\dot{I}^{SE}_{1}\right]_{3x1}$ current flow estimated in the branch of the identified field device;

$\left[\dot{I}^{FD}_{1}\right]_{3x1}$ current magnitude measured by the identified field device;

In the forward sweep algorithm, the voltage profile in the root bus, $[\dot{V}^{CALC}_{0}]_{3x1}$, is a fixed voltage reference equal to the voltage profile measured by the field device, $[\dot{V}^{FD}_{i}]_{3x1}$, which revealed the power loss. The algorithm initiates from the root layer and computes the downstream voltages using the updated upstream adjacent voltage and the normalized current matrix. Equation (16) expresses the way to build the normalized current matrix where the current flow error, $(\dot{I}^{FD}_{p,1} - \dot{I}^{SE}_{p,1})$, is added to the normalized current flow of branches between the jth bus and the root bus to correct the current flows that were wrongly estimated due to the non-technical loss. Indeed, the value of $L_{p,i,i+1}$ should be very close to zero when the jth bus has the detected non-technical loss.

TABLE I
MEASURED AND CALCULATED VOLTAGE MAGNITUDES

l	$\lvert\dot{V}^{MEAS}_{p,l}\rvert$	$\lvert\dot{I}^{SE}_{p,l}\rvert$	$\lvert\dot{V}^{CALC}_{p,l}\rvert$		
			$h_{p,12}$	$h_{p,13}$	$h_{p,34}$
1	-	29.82 A	-	-	-
2	126.25 V	11.87 A	125.87 V	126.25 V	126.25 V
3	125.49 V	17.95 A	125.87 V	125.49 V	125.49 V
4	125.15 V	6.01 A	125.49 V	125.11 V	124.73 V

1) Case Study (Layer Progression): This example case study illustrates the progression through two layers in a segment of the distribution network for locating the consumption point with non-technical loss, as given by Fig. 6. This case starts from the identified distribution transformer with $R_{p,1} = 0$, i.e., the estimated voltage matches with the measured magnitude, but the estimated current does not. The first evaluation of the layer progression through the distribution transformer is $L_{p,0,1} = 0.718$; hence, the next step comprises the calculation of $L_{p,1,2}$, which determines the path for progressing from layer 1 to layer 2.

There are two possible paths, i.e., via bus 2 or via bus 3, which are compared using the electrical distance, $h_{p,kj}$. Fig. 6 (a) shows the scheme for calculating the electrical distance via bus 2 where the detected current error, $(\dot{I}^{FD}_{p,1} - \dot{I}^{SE}_{p,1})$, also flows through the branch between buses 1 and 2. The fourth column of Table I presents the magnitudes of the calculated voltages to this supposition used in the calculation of $h_{p,12}$.

Fig. 6 (b) shows an analogous scheme for calculating the value of $h_{p,13}$. Because the value of $h_{p,13}$ is smaller than that of $h_{p,12}$, the layer progression follows bus 3, i.e., $L_{p,1,2} = h_{p,13}$. Then, a new step of layer progression is performed, as shown

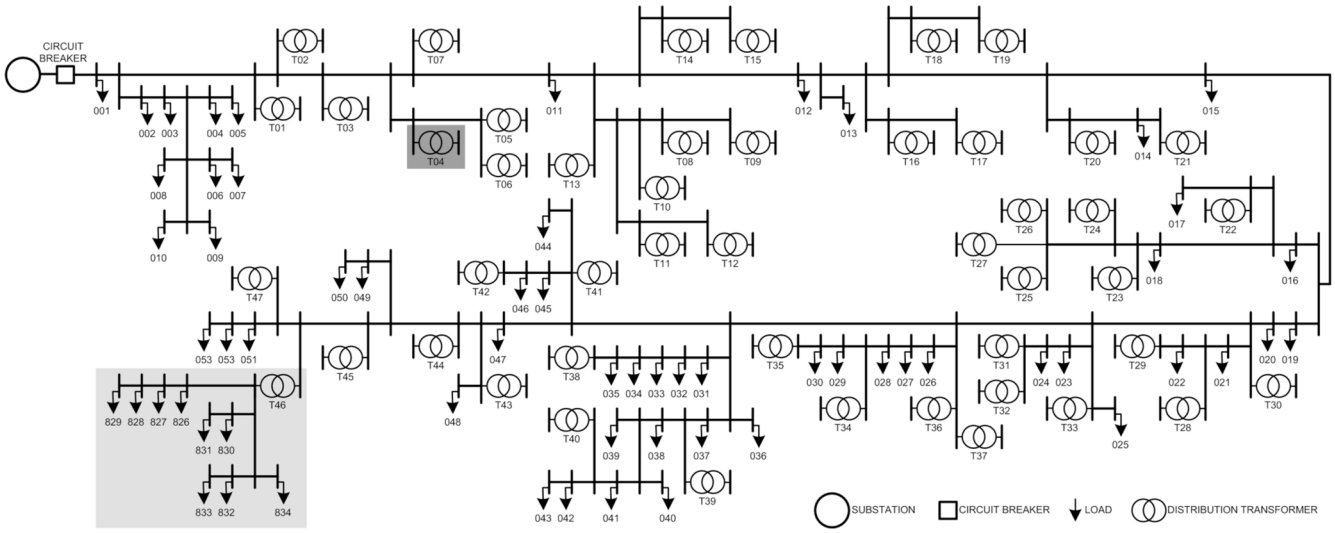

Fig. 7. Topologic diagram of a real distribution network.

by Fig. 6 (c), when the electrical distance is zero ($h_{p,34} = 0$), indicating localization of the non-technical loss at bus 4.

III. Simulation Results

The proposed methodology is evaluated under a real distribution network with the topological diagram shown in Fig. 7. The testing distribution network provides energy to 834 monitored loads, including 53 loads at the medium-voltage (MV) network. There are 47 distribution transformers with a suitable field device for monitoring. The topological diagram illustrates all distribution transformers and loads of the MV network, but it only shows an LV network with numbered loads from #826 to #834. The real distribution network behavior is achieved using a smart grid simulation environment [32].

A. Algorithm Adjustments

The main parameter of adjustment is the sample space size that influences the efficiency of the detection algorithm. Fig. 8 illustrates the upper control limit based on the sample space size. The figure shows a decrease in the *UCL* value caused by an increase in the *N* value. Therefore, the sample space size should also affect the selectivity of the detection algorithm because it determines the reliable region.

Fig. 8 highlights three points ($N = 10$, 20 and 30) that are used to evaluate the impact of the sample space size on the efficiency and selectivity of the detection algorithm. In addition, one unregistered load is inserted in the LV network for emulating the non-technical loss. The unregistered load is randomly varied from 0 to 10 kVA in a total of 350 power changes for each value of *N*. The random power values are divided into two groups: one group from 0 to 1 VA, and the other group from 1 VA to 10 kVA.

The algorithm behavior is shown through a bar chart of the successful rate by the apparent power groups and sample space sizes. The successful rate is obtained by the relation among the amount of detected instances and the total number of simulated instances where each instance is a power change.

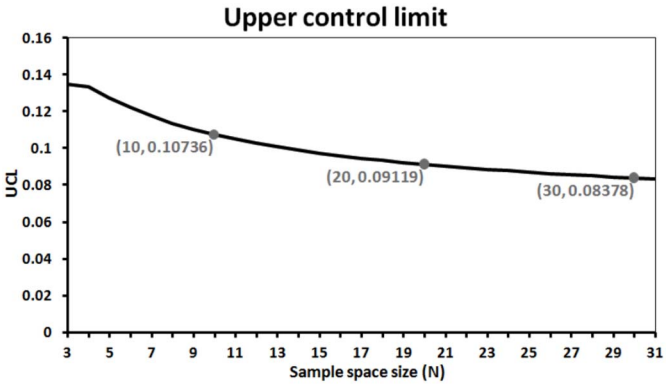

Fig. 8. Characteristic of upper control limit by the sample space size.

Fig. 9 presents a bar chart of algorithm behavior that reveals the highest success rate for $N = 30$ and the apparent power group from 1 VA to 10 kVA. In addition, the apparent power group from 0 to 1 VA is successfully approximately 10% of the time, regardless of the sample space size. These results indicate that the algorithm behavior for $N = 20$ and 30 is very similar.

Another way of analyzing the algorithm behavior consists of comparing the measured success rate with the expected success rate. The division of the sample space size by the population size is the expected success rate because there is only one load with power loss in the distribution network. The measured rate is obtained from the bar chart. Thus, the comparison can be performed using the percentage error.

Table II presents the percentage error that is computed by subtracting the measured rate from the expected rate and then dividing the difference by the expected success rate. In the power group from 0 to 1 VA, the percentage error increases from 61.84% to 83.32% with increasing sample space size. Despite the high error percentage, these results indicate that the selectivity of the detection algorithm is minor compared to the large sample space size.

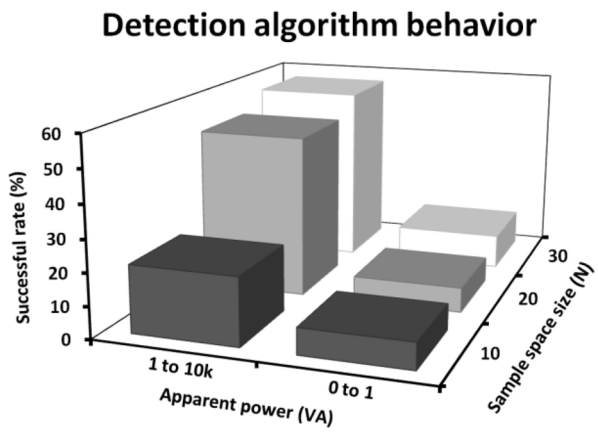

Fig. 9. Successful rate by the apparent power and sample space size.

TABLE II
SUCCESSFUL RATE OF DETECTION ALGORITHM

Power group (VA)	N	Expected (%)	Measured (%)	Error (%)
0 to 1	10	21.28	8.12	61.84
	20	42.55	7.69	81.93
	30	63.83	10.65	83.32
1 to 10k	10	21.28	20.91	1.74
	20	42.55	50.32	-18.26
	30	63.83	56.21	11,94

In the power group from 1 VA to 10 kVA, the error is notably reduced, indicating the enhancement in the efficiency of the detection algorithm. For example, the percentage error for $N = 10$ is equal to 1.74%. Moreover, the percentage error for $N = 20$ is negative because the measured success rate is greater than the expected rate. The low success rate for $N = 10$ inhibits the adjustment with a small sample space size, whereas the similar behavior for $N = 20$ and 30 constrains the adjustment with very large sample space size. Hence, the best adjustment can be achieved with $N = 20$, which also guarantees the diversity of the sample space.

B. Database and Smart Meter Attacks

The assessment of the strategy for locating non-technical losses comprises the simulation of successful cyber-attacks against the database and AMI. In the first type of cyber-attack, all registers of the target customer are deleted from the application database, thereby hiding the energy consumption data once the measurement data are not collected by the EMS. In the second type of cyber-attack, the EMS can collect metering data from the AMI, but the collected data are corrupted due to the intentional cybertampering of the smart meter accuracy.

Table III presents some characteristics of simulated non-technical losses. The second column shows the type of cyber-attack, and the third column establishes the accuracy of the measurement data provided by smart meters. The load column provides the value of the installed power in the customer with abnormal use of energy, and the last three columns characterize the testing performed in the assessment of the proposed methodology. For example, NTL #1 is simulated in the #T1, #T2 and #T3 testing, whereas NTL #7 is only simulated in the #T3 testing.

TABLE III
DESCRIPTION OF THE SIMULATED NON-TECHNICAL LOSSES (NTL)

NTL	Cyber-Attack	Accuracy	Load (kVA)	#T1	#T2	#T3
#1	Database	-	12	X	X	X
#2	Smart Meter	0.5	8	-	X	X
#3	Database	-	8	-	X	X
#4	Smart Meter	0.6	4	-	X	X
#5	Database	-	6	-	X	X
#6	Smart Meter	0.7	7	-	-	X
#7	Database	-	11	-	-	X
#8	Smart Meter	0.8	9	-	-	X
#9	Database	-	9	-	-	X
#10	Smart Meter	0.9	50	-	-	X

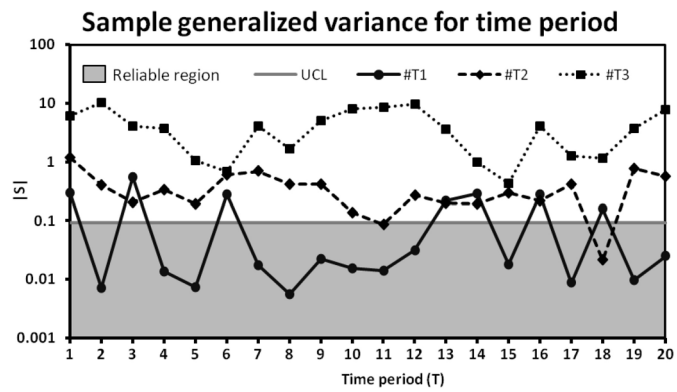

Fig. 10. Monitoring of the sample-generalized variance along time.

In the performed tests, the sample space size is adjusted to $N = 20$, and the sample-generalized variance is monitored along with time, as shown by Fig. 10. #T1 with one NTL has the maximum value of $|S|$ equal to 0.5565 and a success rate of 35%, i.e., the value of $|S|$ is greater than UCL for seven of the 20 monitored values. #T2, with five NTLs, exhibited an increased success rate (90%), which reached 100% in #T3, with 10 NTLs, when the minimum value of $|S|$ is equal to 0.4333. The obtained results are consistent with the expectations because the variance of the sample space increases proportionally to the number of non-technical losses, suggesting the reduction of the sample space size whenever the inspected distribution network has the suspicion of many customers with abnormal use of energy.

In addition to the monitoring of the sample-generalized variance, the computational cost is measured in #T1 using an Intel Core 2 Duo CPU running at 2.0 GHz, with 2.0 GB of RAM. Fig. 11 reveals that the average processing time increases from 27 ms to 85 ms whenever the value of $|S|$ is greater than that of UCL. The increased processing time comes from the identification procedure that checks all field devices in searching for abnormal consumption of energy.

After detection and identification procedures, the proposed method performs a search procedure for locating the NTL. Fig. 12 presents the topological diagram of a section of the distribution network where the path from the root node to

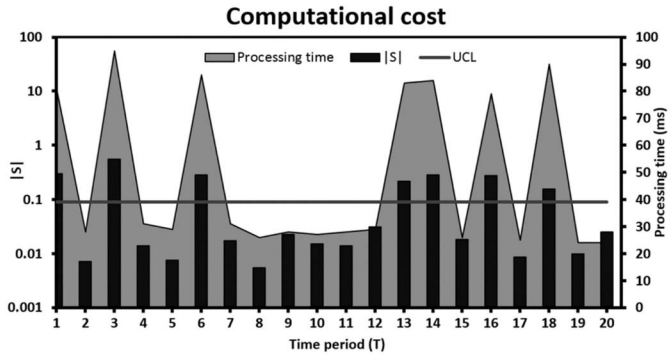

Fig. 11. Computational cost required by the #T1 testing.

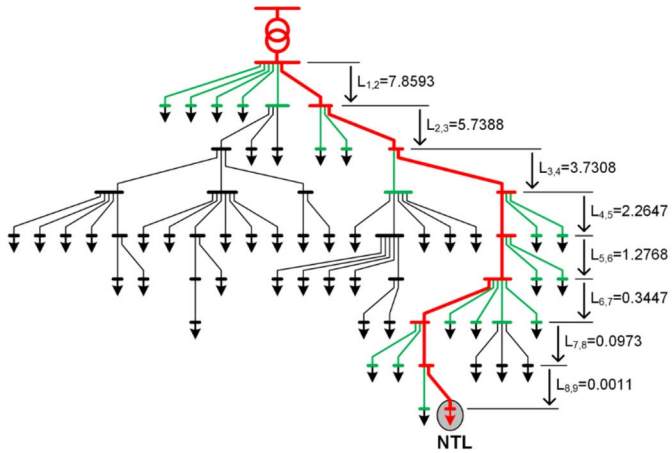

Fig. 12. Emphasis of the path built by the search procedure in a low-voltage network.

the node with NTL is highlighted using red color. This path crosses nine layers and is built using the evaluation function of layer progression, which provides the estimated electrical distance to the NTL. Green branches are evaluated edges with greater electrical distance than red branches. The value of the evaluation function decreases as much as the search procedure approximates of the NTL. The progression from the first layer has a maximum value, $L_{1,2} = 7.8593$, whereas the progression to the layer with NTL has the minimum value, $L_{8,9} = 0.0011$, during the construction of the path. These results illustrate the most severe case with the NTL at the eighth layer, but there are other cases with the simulation of NTL in the first, third and sixth layers. In all simulated cases, the search procedure was successful in locating the non-technical loss.

C. GIS Application

The algorithm for detecting and locating non-technical losses is a tool of the EMS and requires the use of a supervisory interface able to display the electric power system behavior to the system operator. If the supervisory interface is integrated with a GIS application, the locations of the non-technical losses are obtained using the street map in the background of the georeferenced topological diagram of the distribution network [33].

Fig. 13 presents a screen detail of the supervisory interface with non-technical losses highlighted by a red flag. When an end point of the distribution network is highlighted, the non-technical loss is caused by a cyber-attack against the database because there is not a customer connected to this end point. In an attack against the smart meter, the corrupted measurements of the customer are collected; hence, the red flag is placed above the customer symbol. In this way, the system operator easily performs the diagnosis and rapidly sends a maintenance team to the fraud address whenever necessary.

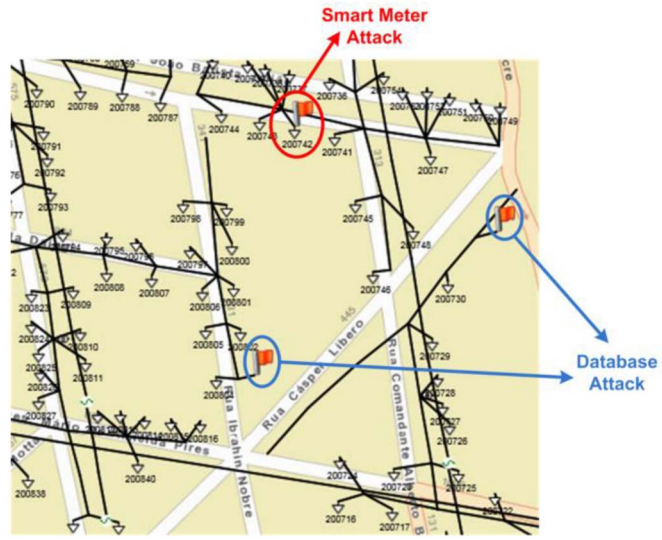

Fig. 13. Screen portion of the supervisory interface with the GIS application.

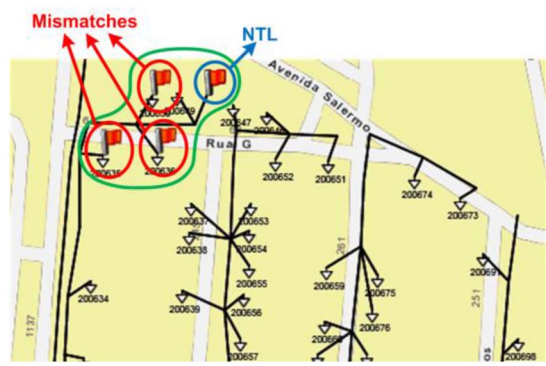

Fig. 14. Imprecise location of the NTL in the power group from 0 to 1 VA.

In mismatch cases, the pathfinding algorithm cannot complete the path toward the non-technical loss. Fig 14 shows an example of mismatches that frequently occur in the power group of 0 to 1 VA. As the power loss is small, the evaluation function starts with small values. Before the pathfinding algorithm reaches the node with non-technical loss, the value of the evaluation function is already close to zero, terminating the search process. Fortunately, mismatches are displayed in a small area surrounding the real non-technical loss.

IV. Conclusion

The intense utilization of information technologies makes the power grid vulnerable to cyber-attacks. Due to the high cost for protecting the whole power system, an alternative is the development of cost-efficient methodologies able to

reduce commercial losses caused by cyber-attacks that corrupt energy metering data. The proposed methodology employs resources of metering, communication and information from the smart grid, and mathematical tools that are typically used in statistical quality control.

The proposed methodology does not compare average power flows. The detection algorithm compares reliable measurements with magnitudes produced by a state estimator in real-time. Hence, the proposed strategy does not require an additional stage to recognize consumption patterns. Furthermore, the use of the pathfinding algorithm together with GIS application increases the diagnostic capability of the proposed methodology because it permits identifying the type of cyber-attack and provides the address of the customer with abnormal energy use.

REFERENCES

[1] J. R. Agüero, "Improving the efficiency of power distribution systems through technical and non-technical losses reduction," in *Proc. IEEE PES T&D*, Orlando, FL, USA, 2012, pp. 1–8.

[2] 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 Trans. Power Del.*, vol. 26, no. 4, pp. 2436–2442, Oct. 2011.

[3] J. D. Glover, M. S. Sarma, and T. J. Overbye, "Power system analysis and design," in *Power Distribution*, vol. 1, 5th ed. Stamford, CT, USA: Cengage Learning, 2010, pp. 757–769.

[4] G. M. Lee and D. H. Su, "Standarization of smart grid in ITU-T," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 90–97, Jan. 2013.

[5] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy*, vol. 7, no. 3, pp. 75–77, May/Jun. 2009.

[6] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Distributed Internet-based load altering attacks against smart power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 667–674, Dec. 2011.

[7] S. M. Amin, "Smart grid security, privacy, and resilient architectures: Opportunities and challenges," in *Proc. IEEE PES Gen. Meeting*, San Diego, CA, USA, 2012, pp. 1–2.

[8] D. Wei, Y. Lu, M. Jafari, P. M. Skare, and K. Rohde, "Protecting smart grid automation systems against cyberattacks," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 782–795, Dec. 2011.

[9] Y. Guo, C.-W. Ten, and P. Jirutitijaroen, "Online data validation for distribution operations against cybertampering," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 550–560, Mar. 2014.

[10] Y. H. Chang, P. Jirutitijaroen, and C.-W. Ten, "A simulation model of cyber threats for energy metering devices in a secondary distribution network," in *Proc. 5th Int. Conf. Crit. Infrastruct. (CRIS)*, Beijing, China, Sep. 2010, pp. 1–7.

[11] Z. Xiao, Y. Xiao, and D. H.-C. Du, "Exploring malicious meter inspection in neighborhood area smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 214–226, Mar. 2013.

[12] J. Valenzuela, J. Wang, and N. Bissinger, "Real-time intrusion detection in power system operations," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[13] Y.-L. Lo, S.-C. Huang, and C.-N. Lu, "Non-technical loss detection using smart distribution network measurement data," in *Proc. IEEE PES ISGT*, Tianjin, China, 2012, pp. 1–5.

[14] Y. Huang *et al.*, "Bad data injection in smart grid: Attack and defense mechanisms," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 27–33, Jan. 2013.

[15] C.-H. Lo and N. Ansari, "CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 33–44, Jun. 2013.

[16] 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.

[17] J. A. Momoh, *Electric Power Distribution, Automation, Protection and Control*, 1st ed. Boca Raton, FL, USA: CRC Press, 2007.

[18] J. D. McDonald, *Electric Power Substations Engineering*, 1st ed. Boca Raton, FL, USA: CRC Press, 2003.

[19] H. S. Cho, T. Yamazaki, and M. Hahn, "Determining location of appliances from multi-hop tree structures of power strip type smart meters," *IEEE Trans. Consum. Electron.*, vol. 55, no. 4, pp. 2314–2322, Nov. 2009.

[20] F. M. Cleveland, "Cyber security issues for advanced metering infrastructure (AMI)," in *Proc. IEEE PES Gen. Meeting*, Pittsburgh, PA, USA, 2008, pp. 1–5.

[21] G. N. Ericsson, "Cyber security and power system communication—Essential parts of a smart grid infrastructure," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1501–1507, Jul. 2010.

[22] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*. New York, NY, USA: Springer, 1999.

[23] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status," *Renew. Sustain. Energy Rev.*, vol. 15, no. 6, pp. 2736–2742, Aug. 2011.

[24] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *Proc. IEEE Green Technol. Conf.*, Denver, CO, USA, 2013, pp. 57–64.

[25] J. B. Leite and J. R. S. Mantovani, "State estimation of distribution networks through the real-time measurements of the smart meters," in *Proc. IEEE PES Powertech*, Grenoble, France, 2013, pp. 1–6.

[26] J. B. Leite and J. R. S. Mantovani, "Distribution system state estimation using the Hamiltonian cycle theory," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 366–375, Jan. 2016.

[27] D. C. Montgomery, *Multivariate Process Monitoring and Control*, 6th ed. Hoboken, NJ, USA: Wiley, 2009.

[28] D. Delling, P. Sanders, D. Schultes, and D. Wagner, "Engineering route planning algorithms," in *Algorithmics of Large and Complex Networks: Design, Analysis and Simulation*. Heidelberg, Germany: Springer, 2009, pp. 117–139.

[29] W. Zeng and R. L. Church, "Finding shortest paths on real road networks: The case for A*," *Int. J. Geogr. Inf. Sci.*, vol. 23, no. 4, pp. 531–543, Apr. 2009.

[30] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 2, pp. 100–107, Jul. 1968.

[31] K. M. Passino and P. J. Antsaklis, "A metric space approach to the specification of the heuristic function for the A* algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 1, pp. 159–166, Jan. 1994.

[32] J. B. Leite and J. R. S. Mantovani, "Development of a smart grid simulation environment, Part I: Project of the electrical devices simulator," *J. Control Autom. Elect. Syst.*, vol. 26, no. 1, pp. 80–95, Feb. 2015.

[33] J. B. Leite and J. R. S. Mantovani, "Development of a smart grid simulation environment, Part II: Implementation of the advanced distribution management system," *J. Control Autom. Elect. Syst.*, vol. 26, no. 1, pp. 96–104, 2015.

Jônatas Boás Leite (S'10–M'15) received the B.Sc. and Ph.D. degrees in electrical engineering from UNESP/Ilha Solteira, SP, Brazil, in 2010 and 2015, respectively. He is currently a Post-Doctoral Researcher in the Electrical Engineering Post-Graduate Program with UNESP and the Electrical and Computer Engineering Department, Texas A&M University/College Station, TX, USA. His research areas are planning and control of electric power systems.

José Roberto Sanches Mantovani (M'06) received the B.Sc. degree from UNESP/Ilha Solteira, SP, Brazil, in 1981, and the M.S. and Ph.D. degrees in electrical engineering from UNICAMP/Campinas, SP, Brazil, in 1987 and 1995, respectively. He is currently a Professor with the Electrical Engineering Department, UNESP. His research areas are planning and control of electric power systems.