2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) | 979-8-3503-5866-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICT4SD59951.2023.10303520

Data Driven Energy Theft Localization in a Distribution Network

A S M Jahid Hasan
Dept. of ECE
North South University
jahid.hasan12@northsouth.edu

Md Saydur Rahman
Dept. of ECE
UC Riverside
mrahm054@ucr.edu

Md Shazid Islam
Dept. of ECE
UC Riverside
misla048@ucr.edu

Jubair Yusuf
Member, IEEE
jyusuf177@gmail.com

***Abstract*—Energy theft unsettles the regular operation of utility planners and can make the grid more vulnerable. The customers behind the meter can also undergo both financial and security risks for any energy theft happening in the electrical network. This paper presents a straightforward approach to detect the electricity theft event at nodes in the distribution network. The proposed methodology utilizes 15-minute voltage magnitude profiles at all the nodes in the network and investigates the electricity theft node by using the statistical properties of the available data. A test network with 59 nodes is employed to test this algorithm. Results show that this approach can provide a reasonable accuracy to localize the node of the electricity theft event.**

***Keywords*—Energy theft, Fault, Data, Nodes, Network**

I. Introduction

Electricity theft as known as energy theft poses significant financial risks to power grids worldwide. It can occur through physical attacks, such as bypassing meters, as well as cyber-attacks, including session hijacking and firmware manipulation. Detecting and preventing electricity theft is crucial for the proper functioning and economic development of any country. According to a study conducted by the Northeast Group in 2017, it was found that utilities globally suffer an estimated loss of $96 billion annually due to electricity theft and other non-technical losses (NTLs) [12].

Traditionally, there are two main approaches for electricity theft detection: hardware-based and data-driven methods. Hardware-based methods utilize specialized equipment with anti-tampering features to identify fraudulent users. On the other hand, data-driven methods, including game theory, system state analysis, and machine learning, analyze power load profiles to detect suspicious consumption patterns.

While smart meters in advanced metering infrastructure (AMI) bring numerous benefits to power grids, they also introduce security vulnerabilities [13]. Malicious customers can manipulate smart meter readings to reduce their electricity bills, posing a significant challenge for utilities worldwide. Developing nations may face challenges in implementing robust ICT infrastructure for smart grids. However, the widespread adoption of smart meters enables utilities to collect vast amounts of high-frequency consumption data, which can be utilized for advanced services such as load forecasting, demand response, phase identification etc.

Addressing electricity theft to ensure smart grid security is paramount. Efforts are underway to enhance detection methods and strengthen cybersecurity measures, aiming to mitigate financial losses caused by theft. Data driven techniques offer a promising approach to detect normal and abnormal consumption patterns using data from IoT-based smart meters. These techniques leverage existing data, making them a cost-effective alternative to manual electricity theft detection approaches [5]- [10]. Extensive analysis of AMI data allows the identification of branches with faults or anomalies, enabling predictive capabilities for energy theft forecasting. Detecting the specific branch involved is particularly valuable for smart grid systems, facilitating appropriate actions such as revenue compensation and billing adjustments. Moreover, integrating ML and DL techniques in the energy sector holds potential for improving grid efficiency, reducing losses, and enhancing overall performance. Advanced analytics and predictive modeling empower utility companies to mitigate the impact of energy theft, optimize operations, and establish a more secure and sustainable energy infrastructure.

This paper presents a statistical approach to detect the electricity theft event for any node in a distribution network. This approach is developed on the notion of the statistical difference between time-series voltage profiles of the customers in the network. The proposed method is very easy-to-implement and only leverages the 15-minute voltage magnitudes to provide an approximation for detecting the node of energy theft event. This algorithm can work as an option for the utility planners while only voltage data is available for the nodes in the network.

II. Related Works

Various techniques have been extensively studied and applied in the context of energy theft detection. Numerous research works have focused on developing algorithms and models that can effectively identify instances of energy theft from utility meters. Specially data driven methods are of prime interest as they have diverse application in power systems

ranging from estimation of renewable energy generation or state of charge (SOC) in battery energy storage systems (BESS), to predictive maintenance of transformers or utility-scale BESS [1]–[4]. In [5] authors introduce data-driven model for detecting electric energy theft in real-world scenarios. By analyzing multivariate time series data from industrial sites, their model can identify anomalies and accurately determine the extent of theft. Thorough extensive training on labeled input sequences, the presented DNN model could not achieve reliable theft detection and characterization. In [6] authors examine different methods of energy theft in energy meters and reviews existing solutions proposed by researchers. They propose K-SMOTE-PCA-XGBoost, which combines K-means clustering, SMOTE for dataset balancing, PCA for dimensionality reduction, and XGBoost for classification. To validate they use 5-fold Nested Cross-validation and compare it with other popular machine learning methods. But their results could not demonstrate promising outcomes in detecting electricity theft with high detection rates and low false positive rates. In [7] authors propose to detect and quantify electricity theft by utilizing statistical and machine learning techniques. A theft detection unit is employed to identify suspicious data flows and detect electricity theft cyber-attacks in the customer domain. Historical data on load consumption, area, time, and temperature are used to construct the theft detection unit using a fine tree regression model. But it did not offer an effective approach for detecting and quantifying electricity theft. In [8] authors presents an unsupervised model that establishes a relationship between kWh measurements and line-line voltage magnitude measurements obtained from smart meters in distribution secondaries. It can exploit behaviors of the linear model using both normal and abnormal smart meter data. Notably, their algorithm does not require training samples for theft cases or comprehensive network topology and parameter information. But it only relies on a customer-to-transformer association map, making it an inefficient approach for detecting theft. An Unsupervised method for detecting electricity theft also has been used in [9]. They use ratio profiles, enhanced wavelet-based feature extraction and fuzzy c-means clustering. They also introduce a new anomaly score based on cluster membership information. Moreover, Machine learning algorithms have been very useful in processing multi-modal data [11]. In [10] authors tested customer-specific and generalized benchmark detectors to assess their performance in detecting data poisoning attacks and electricity theft. The detectors included random forest, AdaBoost, ARIMA, SVM, deep feed forward, GRU, and AEA. They used ensemble averaging and sequential ensemble methods using AEA, GRUs, and feed forward layers. They found that sequential ensemble provides greater resilience against data poisoning attacks compared to ensemble averaging. Also, Genetic Algorithom has been used in a dynamic system with discrete time and spatial neighborhood structure [14].

III. System Architecture

A European test circuit is selected for our experiments. The network configuration is obtained from SimBench, an open source platform for power system network analysis and research developed under German Federal Government's Energy Research Program. The network information for the circuit is shown in the table below:

TABLE I
Attributes of Network Model

Attribute	Network-1
SimBench Code	1-LV-urban6–0-sw
Type	Distribution, Urban, Radial
Voltage Level (V)	400 L-L
Buses	59
Lines	57
Loads	111
Distributed Generation	5
Distributed Generation Type	PV

The following figure 1 shows the single-line diagram of the network:

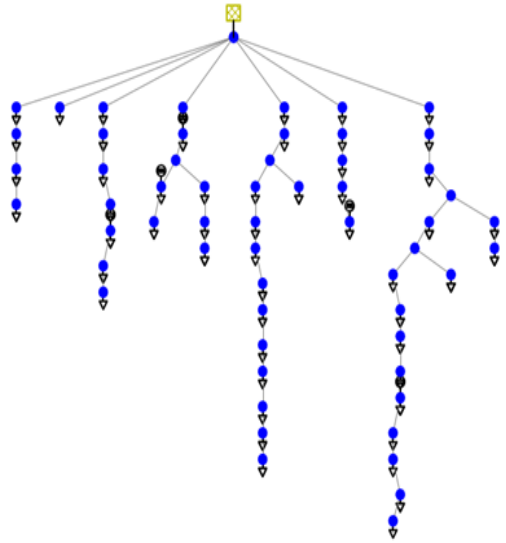

Fig. 1. Single line diagram of Network

For a network we denote a bus or a node by $n \epsilon \mathcal{N}$ where $\mathcal{N}$ is the set of all nodes. The loads in the network are represented by $l^n \epsilon \mathcal{L}$. Here l^n means load number l at bus number n. $\mathcal{L}$ represents the set of all loads. We introduce two functions $\delta(l^n)$ and $\pi(n, m)$. $\delta(l^n)$ gives us the status of load l^n, meaning if it is active or not. $\pi(n, m)$ informs us if n and m are two adjacent nodes, meaning there is a line directly connecting these two nodes.

$$\delta\left(l^n\right) = \begin{cases} 0, & \text{if } l^n \text{ inactive} \\ 1, & \text{if } l^n \text{ active} \end{cases} \tag{1}$$

$$\pi(n,m) = \begin{cases} 0, & \text{if } n \text{ and } m \text{ have no direct connection} \\ 1, & \text{if } n \text{ and } m \text{ have direct connection} \end{cases} \quad (2)$$

IV. Methodology

The implementation of the proposed algorithm consists of three sub-tasks. These are as follows: (i) generating energy theft dataset, (ii) threshold selection for the presence of theft in a branch of the network and (iii) theft node localization. The following three subsections describe each of these activities in detail.

A. Generating Energy Theft Dataset

When there is an energy theft event, this corresponds to one extra node being created within the network having active loads in that network. To imitate this scenario, we assume that loads of one of the nodes within the network are inactive. Let's assume this node is φ. This is our scenario when no theft occurs. Then we activate the loads within the node φ and this node becomes our point of energy theft. We repetitively do this for all nodes that results in that results in generating total $|\mathcal{N}|$ number of scenarios of theft. Algorithm-1 reprsents this task as shown here

Algorithm 1

Select the network
Get all network configuration information
Run load flow for the network
Get all bus voltage profiles with theft V^T
for $\forall n \in \mathcal{N}$ **do**
    Assign $\varphi \leftarrow$ n
    Assign $\delta\ (l^n) \leftarrow 0$
    Run load flow for the network
    Get all bus voltage profiles with no theft V^{NT}
    Save the voltage profile $\mathcal{V}_n^{NT}$ as the voltage profile without theft in scenario n
end for

B. Branch Detection and Threshold Selection

In this task we consider each scenario and try to detect the presence of energy theft in a branch using statistical analysis. We perform a Welch's t-test on the voltage profiles V^T and V_n^{NT} and get the p-values for each of the nodes. Now we compare the p-values with a predefined threshold. The nodes with lower p-values than the threshold indicate that there is a significant difference in bus voltage profiles in those nodes with and without theft cases. We find the threshold p^{Th} by using the following algorithm 2.

C. Locating Theft Node

After we have selected the threshold p^{Th}, we can now check if there is any possibility of theft in any branch by comparing all the p^n with p^{Th}. If we can find a p^n lower than our threshold value p^{Th}, then we can check which of the nodes has the lowest p^n. Intuitively, the closest node to theft should be the most affected node and thus it should have the lowest p-value. We can use the following algorithm to validate our argument.

Algorithm 2

Initialize p^{Th}
Set a tolerance value ϵ
Set iteration limit iter_lim
Set Δprecision $\leftarrow$ 1, $precision_{pre} \leftarrow$0 and iter $\leftarrow$ 0
while ($\epsilon \leq \Delta$precision and and iter$\leq$ iter_lim) **do**
    for $\forall n \in \mathcal{N}$ **do**
        Get p-value for node n p^n
        if ($p^n < p^{Th}$) **then**
            Check if n is from the same branch as theft node φ
        end if
    end for
    Get the true positive tp and false positive fp detections
    Calculate $precision_{post} = \frac{tp}{tp+fp}$
    Calculate Δprecision=$precision_{post} - precision_{pre}$
    Assign $precision_{pre} \leftarrow precision_{post}$
    Assign $p^{Th} = p^{Th}/10$
    iter $\leftarrow$ iter +1
end while
Finalize p^{th}

Algorithm 3

for $\forall n \in \mathcal{N}/\{\varphi\}$ **do**
    if ($p^n < p^{Th}$) **then**
        theft$\leftarrow$True
        break
    end if
end for
if (theft==True) **then**
    z $\leftarrow$ argmin p^n
    if ($\pi(\varphi, z) == 1$) **then**: validated
    end if
end if

V. Result Analysis

We apply our algorithms described in the methodology section on the network mentioned above. For our power flow simulation, we use python based Pandapower package. The load dataset used here is for a whole year with 15-minute intervals. An example case is presented here for better understanding of how the algorithms work. First, we run the load flow simulation of the network with all the loads active. Then, the we chose a node in the network randomly as our theft node. For our example we choose node 8 as our theft node. We deactivate the loads that are connected to this node and run the load flow simulation again. Thus, we can present the first load flow simulation as the case where theft is present (the active loads in node 8) and the second load flow simulation as the case where no energy theft is occurring.

Similarly, we chose each of the remaining nodes in the network as the theft node and perform the load flow simulation again with deactivated loads in that to generate our energy theft dataset for all possible cases. After generating the energy theft dataset, we select the threshold and detect the theft branch by using our algorithm 2. We try different threshold values to find the optimal threshold value. Figures 2 and 3 show the results for the network for two different threshold values, where a theft branch is detected perfectly and with error, respectively. The green colored nodes are the nodes with p-value lower than the test threshold value.

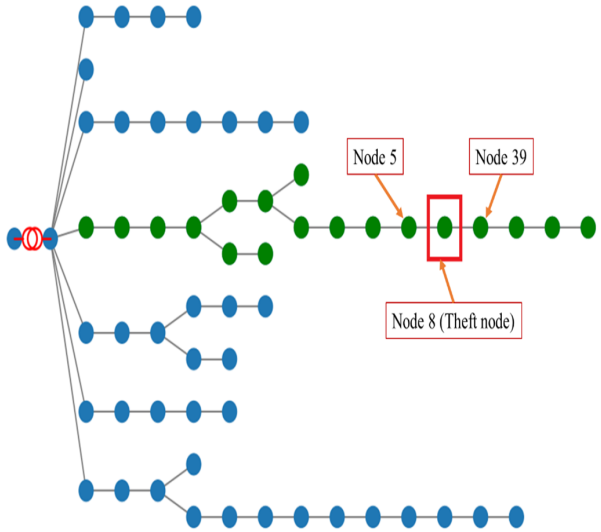

Fig. 2. Accurate branch detection using Algorithm 2 for network

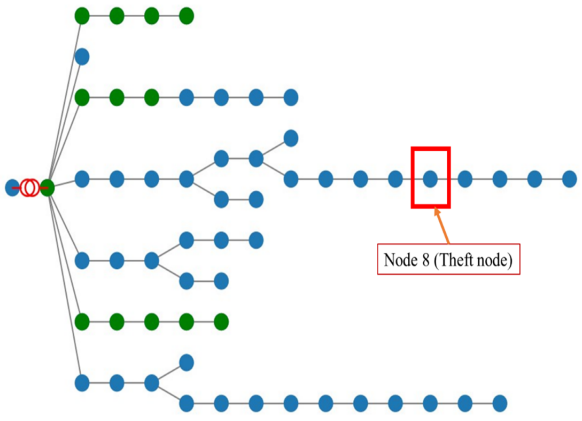

Fig. 3. Inaccurate branch detection using Algorithm 2 for network

After running our algorithm 2, we set the threshold for p-value as $p^T h = 10^{-4}$ for our network. A plot of the threshold versus precision for the network is given in figure 4

After setting our algorithm for branch detection we move towards the theft node localization. We select several test cases for different nodes and for each case we find that the lowest p-value node, z is always adjacent to the theft node φ. This validates our assumption and we can use the lowest p-value node to find the location of the theft, as the theft will be occurring within the directly connected to node z. We now move on to localization of the theft node by identifying a nearby node through use of algorithm 3. We can see from figure 2 that the adjacent nodes of the theft node 8 are node 5 and node 39, meaning both are directly connected with node 8. Table 2 presents the p-values within the nodes of this branch, sorted as nodes with p-values from lowest to highest.

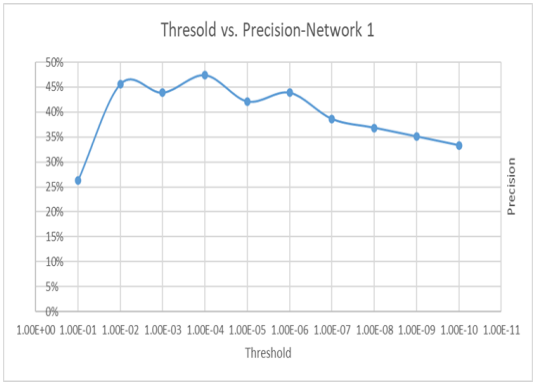

Fig. 4. Threshold vs. precision plot for network

Node	p-value	Node	p-value
8	1.16×10^{-36}	42	6.01×10^{-26}
39	2.53×10^{-36}	41	1.47×10^{-25}
22	1.24×10^{-35}	31	2.42×10^{-24}
33	3.54×10^{-35}	23	2.33×10^{-18}
18	4.03×10^{-35}	14	4.72×10^{-18}
5	1.51×10^{-33}	13	5.42×10^{-18}
19	6.22×10^{-33}	37	1.10×10^{-15}
24	4.60×10^{-30}	25	1.20×10^{-15}
28	2.67×10^{-29}	38	1.22×10^{-09}

TABLE II
P-VALUES IN THE NODES OF THE IDENTIFIED BRANCH

From table 2, we see that the lowest p-value in the branch can be found in node 8 or the theft node. In practice, this node is nonexistent in the network to utilities since it is generated as a result of energy theft. The next lowest p-values in the branch can be found in node 39. This is an adjacent node to theft node and thus we can use the lowest p-value for localization of theft node, that a theft is occurring nearby node 39. This validates our assumption and we can use the algorithm to find the location of the theft, as the theft will be occurring within the directly connected to node with the lowest p-value. Like the example shown here, where theft is occurring at node 8, we obtain similar results for other nodes as well by setting any other node as the theft node.

VI. Conclusion and Future Works

The development of energy theft detection techniques, particularly branch detection, holds great promise in combating the widespread issue of energy theft. Identifying and pinpointing instances of theft within the energy distribution network has been made possible through the application of advanced data analytics. Utilities can effectively detect and address energy theft at the point of occurrence by leveraging the branch detection approach that involves the identification of abnormal load patterns and deviations from expected energy consumption. A straightforward and easy-to-implement statistical approach is proposed in this work that can help the utilities and the customers behind the meter to detect energy theft successfully. The proposed algorithm provides a reasonable approximation to detect energy theft. This approach can be an alternative to complex machine-learning based energy theft detection algorithms and provide a reasonable head start before implementing any state-of-the-art data-driven algorithms over a large complex distribution network. Future work will should focus on creating scalable and adaptable frameworks capable of handling diverse energy generation and distribution scenarios. Exploring innovative technologies and leveraging existing infrastructure can contribute to reducing implementation costs and facilitating the wider adoption of data-driven algorithms for energy theft detection.

References

[1] F. Kabir, N. Yu, W. Yao, R. Yang and Y. Zhang, "Joint Estimation of Behind-the-Meter Solar Generation in a Community," in IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 682-694, Jan. 2021, doi: 10.1109/TSTE.2020.3016896.

[2] A. S. M. J. Hasan, J. Yusuf and R. B. Faruque, "Performance Comparison of Machine Learning Methods with Distinct Features to Estimate Battery SOC," 2019 IEEE Green Energy and Smart Systems Conference (IGESSC), Long Beach, CA, USA, 2019, pp. 1-5, doi: 10.1109/IGESSC47875.2019.9042399.

[3] F. Kabir, B. Foggo and N. Yu, "Data Driven Predictive Maintenance of Distribution Transformers," 2018 China International Conference on Electricity Distribution (CICED), Tianjin, China, 2018, pp. 312-316, doi: 10.1109/CICED.2018.8592417.

[4] A. S. M. Jahid Hasan, J. Yusuf, L. Enriquez–Contreras and S. Ula, "Bad Cell Identification of Utility-Scale Battery Energy Storage System through Statistical Analysis of Electrical and Thermal Properties," 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Espoo, Finland, 2021, pp. 01-05, doi: 10.1109/ISGTEurope52324.2021.9640168.

[5] A. Ceschini, A. Rosato, F. Succetti, R. Araneo and M. Panella, "Multivariate Time Series Analysis for Electrical Power Theft Detection in the Distribution Grid," 2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe), Prague, Czech Republic, 2022, pp. 1-5, doi: 10.1109/EEEIC/ICPSEurope54979.2022.9854628.

[6] A. I. Kawoosa and D. Prashar, "Application of XGBoost ensemble method for energy theft detection in Smart Energy Meters," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9965151.

[7] A. Ali, M. Mokhtar and M. F. Shaaban, "Theft Cyberattacks Detection in Smart Grids Based on Machine Learning," 2022 5th International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Cairo, Egypt, 2022, pp. 1-4, doi: 10.1109/ICCSPA55860.2022.10019036.

[8] Y. Gao, B. Foggo and N. Yu, "A Physically Inspired Data-Driven Model for Electricity Theft Detection With Smart Meter Data," in IEEE Transactions on Industrial Informatics, vol. 15, no. 9, pp. 5076-5088, Sept. 2019, doi: 10.1109/TII.2019.2898171.

[9] R. Qi, J. Zheng, Z. Luo and Q. Li, "A Novel Unsupervised Data-Driven Method for Electricity Theft Detection in AMI Using Observer Meters," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-10, 2022, Art no. 2513010, doi: 10.1109/TIM.2022.3189748.

[10] A. Takiddin, M. Ismail, U. Zafar and E. Serpedin, "Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids," in IEEE Transactions on Smart Grid, vol. 12, no. 3, pp. 2675-2684, May 2021, doi: 10.1109/TSG.2020.3047864.

[11] M. S. Islam, M. S. Rahman and M. A. Amin, "Beat Based Realistic Dance Video Generation using Deep Learning," 2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), Dhaka, Bangladesh, 2019, pp. 43-47, doi: 10.1109/RAAICON48939.2019.22.

[12] Zhu, Lipeng, et al. "Deep Active Learning-Enabled Cost-Effective Electricity Theft Detection in Smart Grids." IEEE Transactions on Industrial Informatics, 2023.

[13] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang and Y. Zhao, "Load profiling and its application to demand response: A review," in Tsinghua Science and Technology, vol. 20, no. 2, pp. 117-129, April 2015, doi: 10.1109/TST.2015.7085625.

[14] Tumpa, Farhana Akter, Md Saydur Rahman, and Md Shazid Islam. Utilizing Genetic Evolution to Enhance Cellular Automata for Accurate Image Edge Detection. No. 10279. EasyChair, 2023.