Indonesian Journal of Electrical Engineering and Computer Science
Vol. 15, No. 2, August 2019, pp. 758~768
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v15.i2.pp758-768

Electricity theft detection framework based on universal prediction algorithm

Abdulrahaman Okino Otuoze[1], Mohd Wazir Mustafa[2], Ibim Ebianga Sofimieari[3], Abdulhakeem Mohd Dobi[4], Aliyu Hamza Sule[5], Abiodun Emmanuel Abioye[6], Muhammad Salman Saeed[7]
[1,2,3,4,5,7]Department of Power Engineering, School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
[1]Department of Electrical and Electronics Engineering, Faculty of Engineering and Technology, University of Ilorin, Nigeria
[6]Department of Control and Mechatronics, School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
[3]Department of Electrical Engineering, University of Portharcourt, Nigeria
[4]Department of Electrical Engineering, Waziri Umaru Federal Polytechnic, Nigeria
[5]Department of Electrical Engineering, Hassan Usman Katsina Polytechnic, Nigeria
[6]Department of Electrical and Electronics Engineering, Akannu Ibiam Federal Polytechnic, Nigeria

Article Info

Article history:

Received Oct 12, 2018
Revised Feb 14, 2019
Accepted Mar 29, 2019

Keywords:

Anomaly detection
Botnets
Electricity theft
Lempel-ziv
Smart meters

ABSTRACT

Electricity theft has caused huge losses over the globe and the trend of its perpetuation constantly evolve even as smart technologies such as smart meters are being deployed. Although the smart meters have come under some attacks, they provide sufficient data which can be analysed by an intelligent strategy for effective monitoring and detection of compromised situations. So many techniques have been employed but satisfactory result is yet to be obtained for a real-time detection of this electrical fraud. This work suggests a framework based on Universal Anomaly Detection (UAD) utilizing Lempel-Ziv universal compression algorithm, aimed at achieving a real-time detection in a smart grid environment. A number of the network parameters can be monitored to detect anomalies, but this framework monitors the energy consumption data, rate of change of the energy consumption data, its date stamp and time signatures. To classify the data based on normal and abnormal behaviour, Lempel-Ziv algorithm is used to assign probability of occurrence to the compressed data of the monitored parameters. This framework can learn normal behaviours of smart meter data and give alerts during any detected anomaly based on deviation from this probability. A forced aggressivemeasure is also suggested in the framework as means of applying fines to fraudulent customers.

Corresponding Author:

Abdulrahaman Okino Otuoze,
Department of Power Engineering,
School of Electrical Engineering,
Universiti Teknologi Malaysia,
Jalan Ilmu, 81310 Skudai, Johor Bahru, Johor, Malaysia.
Email: ooabdulrahaman2@live.utm.my

1. INTRODUCTION

The operations of every of our infrastructure and activities such as health care delivery systems; water and electricity; various marketing structures and platforms; manufacturing and all industrial processes; security systems and operations; education and researches etc. are mostly computer-based and are being thought of to be integrated in planned smart cities whose operation solely depends on an intelligent grid known as smart grids (SGs). The introduction of smart grids (SGs) as the integration of digital computing and communication technology with the power infrastructure has improved the systems' intelligence and has modernized the operation of the conventional power delivery schemes for improved reliability, flexibility, sustainability, security, resiliency, and energy efficiency [1-3] but are subjected to the risk of cyber-attacks and its related threats, especially with internet of things (IoT) taking over [4]. These attacks can change their patterns, appearance and are constantly evolving, hence, requiring a real-time solution to keep track and bring the attacks under control [5] the sooner they occur. Electricity infrastructure, the most targeted by cyber-attacks [6, 7] are commonly subjected to electricity theft.

Electricity theft is a global menace whose trend of perpetuation constantly evolves even as smart technologies are being deployed [8-10]. It occurs sporadically and inflicts huge economic losses and also threatens power systems' sustainability [2]. World bank data reveal India loses about 25% of their generated power, Brazil faces about 16% loss while China and US reportedly lose 6% and 5%, respectively [1] with all countries worldwide having several bitter experiences. Northeast group LLC reported that worldwide, $89.3 billion are lost due to electricity theft on yearly basis [11]. So many approaches have been reported for energy theft detection using the data from conventional meters mainly by the application of artificial intelligence techniques [11-14] etc. but these solutions hardly focus on energy theft detection in real time which is a key aspect of the SGs. Nonetheless, the advent of smart electricity meters (SEMs) has helped greatly by mitigating the energy theft since customer's meters can easily be monitored and its consumption pattern can easily be analysed for inferential judgement using various techniques and strategies some of which are as applied to the data obtained from conventional meters [15-18]. Although, the vulnerability of SEMs as a cyber-physical system remains a crucial concern due to their being able to change their patterns, appearances and constantly evolve, hence, demands critical monitoring by a real-time solution to keep track and bring the attacks under control [5] the sooner they occur. The current trend of ICT with its associated cyber-attacks dictates that proactive steps be taken in leveraging the technology for increased security [11] as electricity theft remains a key issue to be addressed.

In this paper, a framework for energy theft detection based on 'anomaly detection technique' employing a universal prediction algorithm, known as Lempel-Ziv algorithm (LZA), is proposed. The customers' energy consumption data are to be classified by probability assignment of the LZA. Then, the probability assignment on the processed training data is carried out to build a statistical model which forms the basis for normal and abnormal behaviour classification to make decisions. A punishment for detected fraudulent customers to ease the on-site monitoring for utility officials is also suggested by a forced corrective technique. The next section of this paper explores smart meter data and electricity theft. Botnets and general intrusion detection systems are explained in Section 3 while anomaly detection and related works are presented in Sections 4 and 5 respectively. The universal prediction algorithm is explained in Section 6, Section 7 gives the details of the proposed framework and Section 8 concludes the study.

2. SMART METER DATA AND ELECTRICITY THEFT

The introduction of smart grids (SGs) as the integration of digital computing and communication technology with the power infrastructure has improved the systems' intelligence and has modernized the operation of the conventional power delivery schemes for improved reliability, flexibility, sustainability, security, resiliency, and energy efficiency [1-3]. Fundamental to SGs, is the implementation of advanced metering infrastructure (AMI) with SEMs as its key components, for the monitoring and control of systems' parameters [19-21]. This scheme helps to lessen energy theft risks but are vulnerable to cyber-attacks which could cause some adverse effects to customers and utilities. Such adverse effects include electricity service disruption, damage to infrastructure, electricity thefts etc. [1, 22].

SEMs are innovative energy meters deployed for the measurement of energy consumption data which in addition, provide real-time monitoring and basic information interchange between the customer and utility company. They securely communicate the stored data to advanced metering infrastructure (AMI) and are perfect replacements for the conventional analogue meters by the elimination of on-site monitoring and measurement reading by personnel [7, 18]. Additional functions of SEMs are as highlighted in Table 1 [7, 15, 21, 23-26]. Worldwide, SEMs are being gradually deployed with many nations setting landmark targets to completely replace their conventional meters [12, 27-29] own to these many functions and more.

The ability of SEMs to help in real-time collection of energy consumption data at regular intervals for advanced data analytics help energy theft detection at shorter periods which could have easily evade detection in longer periods (as applicable to conventional energy meters); and for further inference aimed at improving the power systems [2, 15, 30]. Its ability to detect fraud has helped in curbing theft but its vulnerability as applicable to any other cyber physical systems remain a major issue [1, 7]. A 24-hour test data obtained from a SEM is as shown in Figure 1.

Nonetheless, SEMs can be compromised by illegal direct connections to distribution lines; It could be via the utility service provider infrastructure by sniffing, large-scale meter take-over, emulation of own or foreign meter etc and; it could be via a web application e.g. for data injection, privacy breaches, billing errors [31]. Other ways by which energy theft is carried out include meter bypass, meter tampering, connivance with fraudulent staff, billing irregularities and any other physical or cyber engineered mechanism aimed at manipulating energy consumption to evade bill payment [11, 32, 33]. Unlike in the conventional mechanical meters which often suffer physical tampering, the digital SMs are exposed to alteration of metering data (locally or remotely) [22].

Various schemes have been proposed in the attempt to present a lasting solution to energy theft based on SEMs data analytics. Most of the schemes such as firmware security, key management security, source code development security etc. involve building a resilient defence mechanism by using these data [31]. Some of the methods employed in this analysis include statistical approach, development of some intelligent algorithms utilising machine learning classifiers e.g. support vector machines (SVM), Decision Trees, fuzzy classifications etc. [11, 34]. A fundamental issue related with machine learning classifiers as applicable to energy theft detection is the imbalance in the data resulting from the difference in normal and abnormal samples. This is because theft samples do not practically exist in wholesome as it is exposed to myriads of vulnerabilities. This vulnerability of SEMs has called for various dimensions of research input aimed at addressing the challenges. But these attacks, mainly for electricity thefts, are constantly evolving and in different modes, hence, the need for real time monitoring of the meters' normal behaviours to distinguish it from abnormal behaviours. This is best achieved by the provision of network security offering protection against miscellaneous attacks. Most of the attacks ravaging computer systems today are mainly caused by botnets [35-37].

Table 1. A Highlight of the Advantages of SEM

Advantages of Smart Electricity Meters
• Real-time monitoring and response of energy consumption usage and dynamic billings
• Enhanced load control and revenue generation
• Improved supply monitoring and management as outages and faults are easily detected and reported
• Local reading ability and remote control for disconnect, dynamic pricing and general information from the utility etc.
• Automated power restoration
• Improved metering by detection of tampering and energy losses especially resulting from fraudulent activities
• Device and energy status capture by supporting a non-intrusive load monitoring for home automation
• Support for Home Area Networks
• Interoperability within the SGs network by renewable energy integration support
• Frequent sampling intervals and provision of adequate data used in processing and classifications for pattern identification and consequent decision support

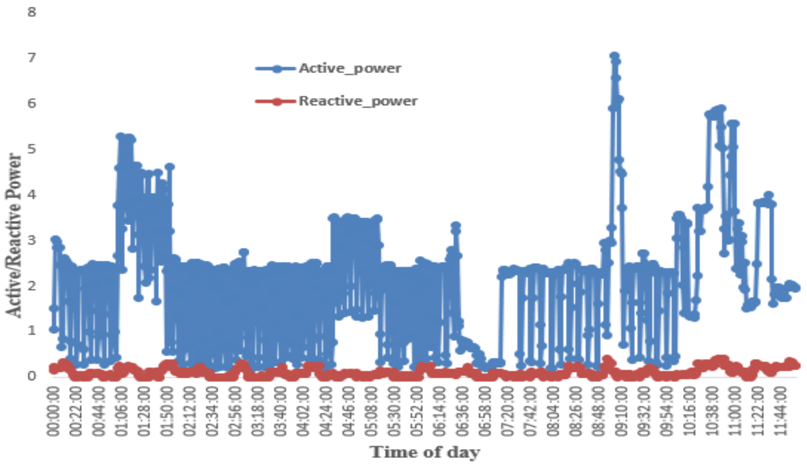

Figure 1. 24-hr sampled data of power consumption of a customer from a SM

3. BOTNETS AND INTRUSION DETECTION SYSTEMS

Botnets are compromised, and distributed software entities controlled by a server under the influence of a Bot-master using command and control (C&C) channel to infect machines while carrying out malicious activities such as stealing and manipulations of data, cyberattacks, Denial of Service (DoS) etc. While the server may be infected websites, email attachments, file sharing etc., the host (i.e. the compromised machine) could be either computers, mobile phones etc. [5, 38-40]. Botnets attacks escalate geometrically as criminals have found them a safe-haven to perpetuate attacks for several reasons. These attacks are either server-based or host-based [38]. The severity of the attacks is very high because every compromised machine (known as bot) could also become a server for launching further attacks. Botnets account for over 85% of spam mails and about 20% ad clicks with associated heavy financial losses in billions of dollars [38]. Intrusion prevention techniques such as authentication, authorization and privacy are applied to secure systems from bot attacks, but they are random and prone to breaches. Hence, an intrusion detection systems (IDSs) are required as another level of protection [41]. An efficient IDSs must first uncover the behaviours of the bots to aid the design, detection and blocking mechanism [38, 39]. This is done by exploring the communication patterns of the bots C&C channel which is its weakest link, since it is the only link the bot-master communicates with its bots, and block them before any serious harm is done [5, 39, 42, 43].

IDSs are sometimes considered as either host-based or network-based. The host-based monitors a single host activity such as log files and applications activities while the network-based scans for any suspicious activity from the network traffic of a given part of the network. The boundary of the different segments of the network is usually the best place for the network-based. IDS can be executed in offline or online mode. The offline mode involves manual periodic scanning to check for any possible intrusion since the previous scan. In this approach, scanning is carried out only at convenience but usually when suspicious activities are suspected. The online-based scanning is real time and able to run the scan at a pre-set interval of time [39, 44, 45]. IDS can also be classified based on other techniques depending on the nature of the attacks and with respect to its detection principle. Several techniques are employed in botnets detection, but a basic approach is to view them from three angles namely; the honeynets approach, the signature-based detection and anomaly detection [5, 40]. While the honeynets approach utilises set traps to collect, study and reveal information about bots, the signature-based (sometimes called misuse-based) relies on using learned information stored on the database to locate bots and requires constant updates of the database since new attacks are not learnt and cannot be detected; the anomaly detection technique is targeted at recognising the presence of irregularities in any of the systems' indicative parameters or in the network traffic [5, 45]. Figure 2 depicts a typical botnets and intrusion detection techniques as explained above. Undeniably, anomaly detection is more promising for the detection of botnets, since new attacks (also known as zero-day attacks) even without prior knowledge, can also be detected. Although the signature-based approach as discussed earlier works perfectly as a basis for detecting theft, the anomaly detection approach presents higher intelligence [22] since botnets are highly dynamic.

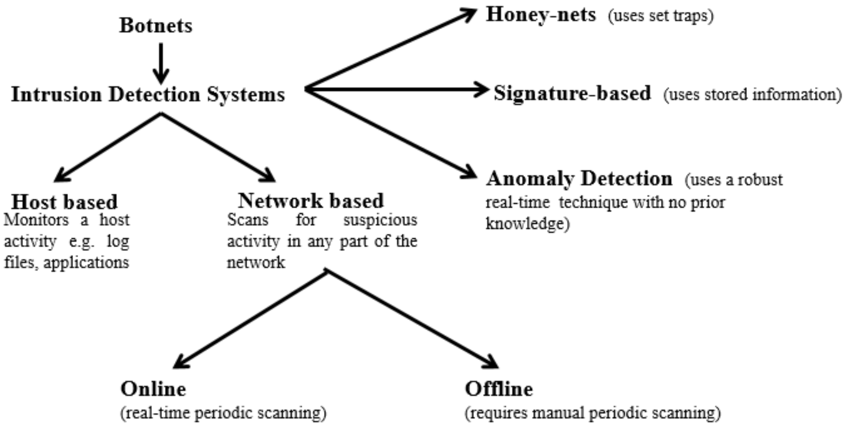
Botnets
Intrusion Detection Systems
Honey-nets (uses set traps)
Signature-based (uses stored information)
Anomaly Detection (uses a robust real-time technique with no prior knowledge)
Host based
Monitors a host activity e.g. log files, applications
Network based
Scans for suspicious activity in any part of the network
Online
(real-time periodic scanning)
Offline
(requires manual periodic scanning)

Figure 2. Botnets and intrusion detection techniques

4. ANOMALY DETECTIONS

Anomaly detections (ADs) involve the technique of finding outliers, intrusions or patterns in a given data set which do not conform to a defined normal or expected behaviour [45-47]. The anomalies (or outliers) are often induced in the data by adversaries for various purposes ranging from fraud, terrorism, cyber-attacks, war etc. and hence, ADs have found applications in theft and fraud detection, intrusion detection for cybersecurity, fault detection in critical systems etc. [7, 45]. Anomalies in any given dataset could cause severe losses (financial and technical), failure, and breakdown of any system in context. In power systems, huge losses (in revenues) are incurred mostly due to electricity theft resulting from data manipulations by cyber-attacks on SMs. Worldwide, huge losses (in billions of dollars) are being reported in various cases of electricity thefts. Hence, the need for intensified research efforts by developing more intelligent frameworks and algorithms for detecting anomalies in energy consumption data especially.

In ADs, traffic analysis is used on both packet and flow data since attacks are assumed different from normal pattern. Metrics such as response time, date and time stamps, rate, volume, range etc. are used in identifying anomalous data [42]. Figures 3 and 4 show samples anomaly of an attack on a SM by infected data on the time and power utilised, respectively. The anomaly in Figure 2 presents same time and values over different instants and it is usually aimed at maliciously paying for only an instant (the first time before successful attack) of the periods while Figure 3 presents an instance where the power consumption remains flat at some very low value over a considerable long period of time. These types of attacks may last for days, months or even years without an intelligent algorithm for real-time detection. ADs are faced with possible high false alarm rate and so, it utilises prior knowledge based on statistical evaluations (e.g. ARMA or Markov model) on the normal data to achieve higher performance [42].

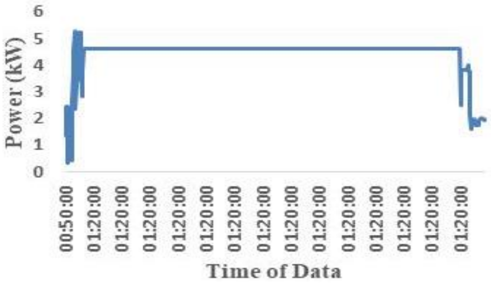

Figure 3. Time-infected data of a smart meter reading

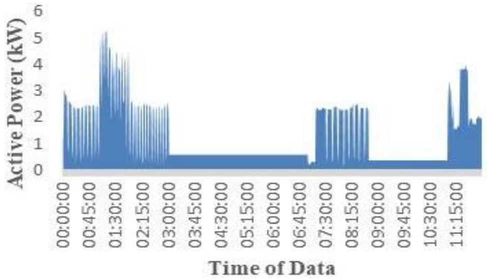

Figure 4. Power consumption infected data

5. RELATED WORKS

The advent of smart meters with its steady provision of real-time customer consumption data has necessitated the applications of big data analytics and machine learning in the study of electricity theft detection. Machine Learning applications such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Self-organising Maps (SOM) etc. are often employed for data pre-processing, feature extraction and classification [44]. Some works have been reported by training and modelling datasets using ANN, fuzzy classification, SVM [12, 13, 22, 48-50] and other classification strategies.

An anomaly-based general detection framework clearly independent of botnets relying on the technique of deep packet inspection although could not establish a result standing the test of ever expanding data, waspresented by Gu, et al. [51]. Spirić, et al. [52], using rough set theory, analysed customers' invoice and some other registered data, and predicted a list of fraudulent customers based on the consumption pattern of suspected customers and then hinted the need for on-field corrective and penalty measures where necessary. In another work, a temperature based predictive model was proposed for different types of circuit approximations to estimate technical losses which was tested on distribution feeders and linear circuits to detect power theft [18]. Depuru, et al. [32] introduced a high-performance computing by data encoding for speedy detection of energy theft. Using SVM and Rule-based algorithms, the encoded data were then classified to uncover electricity theft.

A centralized energy detection scheme was proposed utilizing the Kalman filter by Salinas and Li [2]. They proposed a privacy-preserving energy detection-based algorithm to identify fraudulent customers while Soniya and Wilscy [39] proposed an intrusion detection technique capable of early and randomised detection utilising traffic analysis of an end-point host in identifying bot's C&C communication. These algorithms depend on the systems' parameters' fore-knowledge. Universal anomaly detection algorithms able to learn and distinguish normal from abnormal pattern in a network traffic [5]. This algorithm has no fore knowledge of the system model or the characteristics of the threat or attack.

Despite the various efforts, issues of high false positive rates, the need for field inspection after determining suspected fraudulent customers, complicated data collection procedure, lack of real-time monitoring remain a major concern. These concerns can only be addressed by an effective anomaly detection technique as evident in the study presented by Siboni and Cohen [5]. In their work, Lempel-Ziv universal compression algorithm was utilised to assign optimally, probability assignments for normal behaviour (during learning) and able to estimate the likelihood of new data (during operation) and classify it accordingly. They tested a time-series data, enabling the network to be both protocol and encryption independent. They were able to suggest a system able to detect every hiding technique and concluded from their report that, LZA is applicable to any sequence of behaviours and not just the timing data and can be applied to detect anomalous behaviour in any given data set.

6. UNIVERSAL PREDICTION ALGORITHM

Provide a statement that what is expected, as stated in the "Introduction" chapter can ultimately result in "Results and Discussion" chapter, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

LZA, a lossless data compression algorithm is applied by a means of probability assignment to compress the data. It is characterised by a vanishing redundancy [5]. It is sometimes, considered a universal prediction algorithm and often referred to as an optimal universal compression algorithm [5, 42]. LZA is a parsing algorithm used to partition block of variables (or phrases) of sequence data such that a newly parsed block is the shortest variable not seen previously [53, 54]. Each phrase in this dictionary or block is usually represented by a rooted-tree defining paths to internal nodes with each phase containing suffixes of leaf-nodes added to the tree and can be used to define a statistical model for a given sequence [5, 42, 53]. Each sensor nodes is used to assign labels to the sensed data for instance, '0' can be used to denote normal data while '1' for the anomaly [55]. Probability is then randomly assigned with the anomalies defined to be of low probability. The anomalies deviate from the expected behaviour forecasted by the statistical model developed to classify and detect by comparing with probabilistic model of normal events [5, 41].

These parsing algorithms have found applications in numerous areas such as universal data compression scheme, entropy estimations, anomaly detections, data randomness test, statistical model estimations of given sequence, pattern matching [53, 54, 56, 57] etc. A good LZ dictionary or the sensed data contains the most useful aspect of the data and the compressed data based on the codewords of the phrases in the dictionary [58]. Using the principle of entropy in information technology, the most useful part of the sensed data is captured by extracting all the phrases in the file with the most frequent data stored into the dictionary [58]. Consider a phrase of length M symbols occurring n times in a file of length N symbols, its average entropy, H_M is as given in (1) [58]. This means H_M represents the components of the phrases with the higher length or more frequent in the file.

$$H_M = \frac{1}{M}\log_2 \frac{N}{nM} \tag{1}$$

Let the sensed data of the four parameters selected for monitoring by this framework, energy consumption data, its rate of change, its date stamp and time signatures be as represented by (2) to (5).

$$E = E_1, E_2, E_3, E_4, E_5, E_6, \dots, E_i \tag{2}$$

$$e = e_1, e_2, e_3, e_4, e_5, e_6, \dots, e_i \tag{3}$$

$$D(h, m, s) = D_1, D_2, D_3, D_4, D_5, D_6, \ldots, D_i \quad (4)$$

$$T = t_1, t_2, t_3, t_4, t_5, t_6, \ldots, t_i \quad (5)$$

Assuming the assigned label of any of the parameters at any instance, i is $S = 10010010011010101$, then, the parsed sequence is $S = 1|0|01|00|10|011|010|101|$ and Table 2 shows the code word and probability assignments for each of the parsed phrase contents in the sensed. The statistical model is depicted in Figure 5.

Table 2. LZ Probability Assignment and Code Word for 10010010011010101

Dictionary Location	Content (C)	Codeword	Probability Assignment (Pr(C/S))
001	1	0001	1/4
010	0	0000	3/4
011	01	0101	1/2
100	00	0100	1/4
101	10	0010	1/4
110	011	0111	1/4
111	010	0110	1/4
	101	1011	1/4

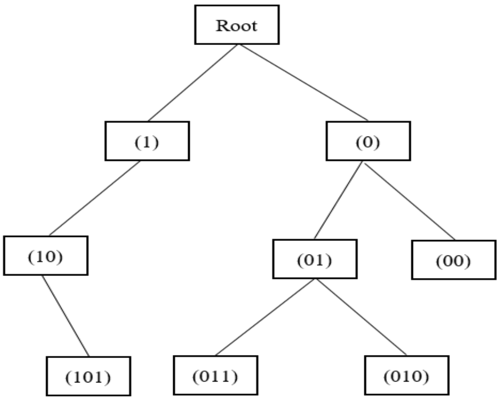

Figure 5. LZA statistical model for 10010010011010101

7. THE UNIVERSAL ANOMALY DETECTION FRAMEWORK

In the study of electricity theft detection using smart meter data, customers' energy consumption data are usually considered from an instant of time, say, t_i and at every interval of usually 1 minute but for analysis sake, the interval of say, t_{i+j} can be considered where j denotes an arbitrary interval of interest. This gives large volume of data for analysis because the training data usually require a minimum of one moth of one-minute sample rate. Both training and test data are labelled by a suitable classification of data. As presented in the framework of Figure 6, the training and testing stage based on LZA is carried out and a statistical model is developed based on the set of discrete data. Training is done only on the normal set of discrete sequences and the testing sequences are separately quantized using similar quantization and approach though same set of centroids.

Based on the statistical model, the probability of the dataset is assigned, and classification carried out using a set threshold while sequential probability assignment is used in the probability assignment of each testing sequence. The anomalies are those testing sequence for which the probability is lower than the set threshold. The customers within this category are then listed as suspected fraudulent customers. To eliminate the need for on-site inspection of those customers, a forced aggressive measure is suggested in the framework. This extra measure assists the utility punish energy fraudsters to reduce the hectic tasks of the on-field measures. Any customer who feels cheated or billed wrongly can appeal whereas confirmed fraudulent customers are automatically fined or recommend for court hearing for further punishment depending on the regulations of the utility. With low false positive rate (FPR), the scheme is just a key model to mitigating energy thefts.

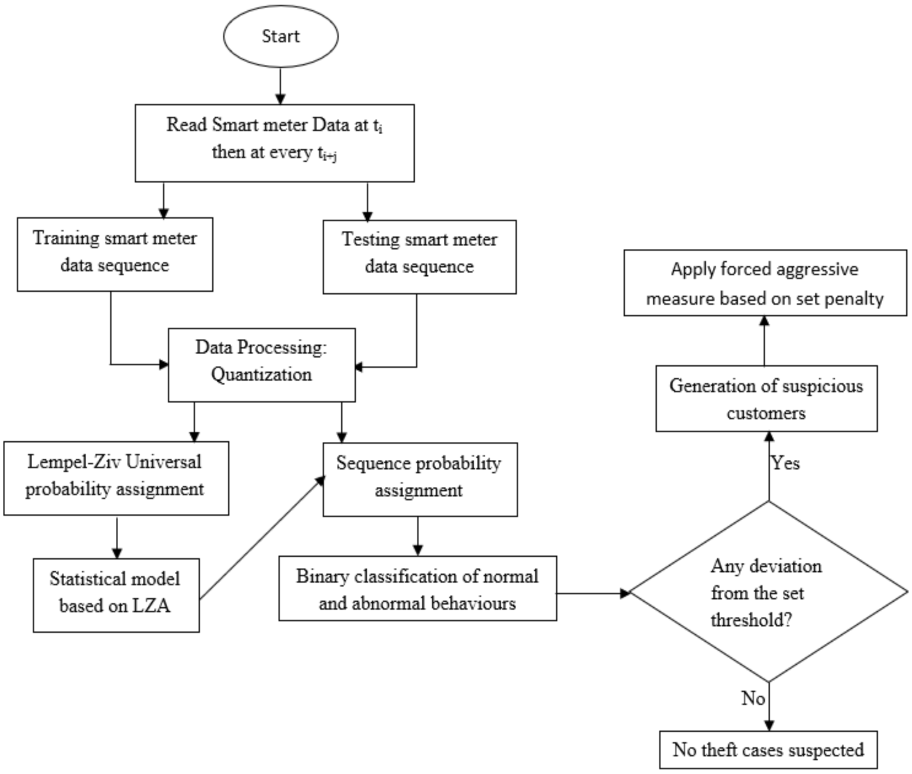
Start
Read Smart meter Data at t_i then at every t_{i+j}
Training smart meter data sequence
Testing smart meter data sequence
Data Processing: Quantization
Lempel-Ziv Universal probability assignment
Sequence probability assignment
Statistical model based on LZA
Binary classification of normal and abnormal behaviours
Any deviation from the set threshold?
Yes
No
Generation of suspicious customers
Apply forced aggressive measure based on set penalty
No theft cases suspected

Figure 6. Proposed framework for electricity theft detection based on universal prediction algorithm

8. CONCLUSION

A framework for electricity theft detection utilising Lempel-Ziv probability assignment algorithm has been presented. This framework suggests a real-time monitoring of smart meter data at some pre-set interval. This eliminates the possibility of customers evading detection. Any abnormal behaviour or pattern in the Date, Time stamps, rate and the energy consumption data will be detected. A forced corrective/penalty measure is also suggested in the framework. When implemented, the utility staff would have been saved the stress of on-the-field monitoring and confirmation of fraudulent customers. The implementation of the framework gives a positive step in the fight against energy theft. The framework is however a maturity model and therefore, requires constant implementation and modification using real-time data to suit the efficient detection of ever-dynamic nature of electricity thefts.

ACKNOWLEDGEMENTS

My supervisor, Prof. M. W. Mustafa is highly acknowledged for his tutelage and UTM administration, for constantly subscribing to credible database.

REFERENCES

[1] V. B. Krishna, K. Lee, G. A. Weaver, R. K. Iyer, and W. H. Sanders, "F-DETA: A framework for detecting electricity theft attacks in smart grids," in *Dependable Systems and Networks (DSN), 2016 46th Annual IEEE/IFIP International Conference on*, 2016, pp. 407-418: IEEE.
[2] S. A. Salinas and P. Li, "Privacy-preserving energy theft detection in microgrids: A state estimation approach," *IEEE Transactions on Power Systems,* vol. 31, no. 2, pp. 883-894, 2016.
[3] P. Siano, "Demand response and smart grids—A survey," *Renewable and sustainable energy reviews,* vol. 30, pp. 461-478, 2014.
[4] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer,* vol. 50, no. 7, pp. 80-84, 2017.
[5] S. Siboni and A. Cohen, "Universal Anomaly Detection: Algorithms and Applications," *arXiv preprint arXiv:1508.03687,* 2015.
[6] M. Erol-Kantarci and H. T. Mouftah, "Smart grid forensic science: applications, challenges, and open issues," *IEEE Communications Magazine,* vol. 51, no. 1, pp. 68-74, 2013.

[7] A. O. Otuoze, M. W. Mustafa, and R. M. Larik, "Smart grids security challenges: Classification by sources of threats," *Journal of Electrical Systems and Information Technology,* 2018.

[8] S. Saini, "Social and behavioral aspects of electricity theft: An explorative review," *International Journal of Research in Economics and Social Sciences,* vol. 7, no. 6, pp. 26-37, 2017.

[9] 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, 2018.

[10] S. Weerakkody and B. Sinopoli, "Challenges and Opportunities: Cyber-Physical Security in the Smart Grid," in *Smart Grid Control*: Springer, 2019, pp. 257-273.

[11] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics,* vol. 12, no. 3, pp. 1005-1016, 2016.

[12] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*, 2011, pp. 1-8: IEEE.

[13] J. Nagi, K. Yap, S. Tiong, S. Ahmed, and A. Mohammad, "Detection of abnormalities and electricity theft using genetic support vector machines," in *TENCON 2008-2008 IEEE Region 10 Conference*, 2008, pp. 1-6: IEEE.

[14] A. Nizar and Z. Dong, "Identification and detection of electricity customer behaviour irregularities," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, 2009, pp. 1-10: IEEE.

[15] D. N. Nikovski *et al.*, "Smart meter data analysis for power theft detection," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 2013, pp. 379-389: Springer.

[16] M. Anas, N. Javaid, A. Mahmood, S. Raza, U. Qasim, and Z. A. Khan, "Minimizing electricity theft using smart meters in AMI," in *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2012 Seventh International Conference on*, 2012, pp. 176-182: IEEE.

[17] S.-C. Yip, C. Tan, W.-N. Tan, M.-T. Gan, K. Wong, and R. C.-W. Phan, "Detection of Energy Theft and Metering Defects in Advanced Metering Infrastructure Using Analytics," in *2018 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE)*, 2018, pp. 15-22: IEEE.

[18] S. Sahoo, D. Nikovski, T. Muso, and K. Tsuru, "Electricity theft detection using smart meter data," in *Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society*, 2015, pp. 1-5: IEEE.

[19] D. Bian, M. Kuzlu, M. Pipattanasomporn, and S. Rahman, "Analysis of communication schemes for Advanced Metering Infrastructure (AMI)," in *PES General Meeting| Conference & Exposition, 2014 IEEE*, 2014, pp. 1-5: IEEE.

[20] N. M. G. Strategy, "Advanced metering infrastructure," *US Department of Energy Office of Electricity and Energy Reliability,* 2008.

[21] A. Cooper, "Electric company smart meter deployments: foundation for a smart grid," *Institute for Electric Innovation Report,* no. 6046, pp. accessed via http://www.edisonfoundation.net/iei/publications/documents/final%20electric%20company%20smart%20meter%20deployments-%20foundation%20for%20a%20smart%20energy%20grid.pdf on 7th February, 2019, 2016.

[22] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns," *IEEE Trans. Smart Grid,* vol. 7, no. 1, pp. 216-226, 2016.

[23] D. Mashima and A. A. Cárdenas, "Evaluating electricity theft detectors in smart grid networks," in *International Workshop on Recent Advances in Intrusion Detection*, 2012, pp. 210-229: Springer.

[24] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid—Challenges, issues, advantages and status," in *2011 IEEE/PES Power Systems Conference and Exposition*, 2011, pp. 1-7: IEEE.

[25] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *Green Technologies Conference, 2013 IEEE*, 2013, pp. 57-64: IEEE.

[26] V. C. Gungor *et al.*, "Smart grid technologies: Communication technologies and standards," *IEEE transactions on Industrial informatics,* vol. 7, no. 4, pp. 529-539, 2011.

[27] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *IEEE Transactions on Industrial Informatics,* vol. 12, no. 1, pp. 425-436, 2016.

[28] J.-S. Chou and I. G. A. N. Yutami, "Smart meter adoption and deployment strategy for residential buildings in Indonesia," *Applied Energy,* vol. 128, pp. 336-349, 2014.

[29] N. Uribe-Pérez, L. Hernández, D. de la Vega, and I. Angulo, "State of the art and trends review of smart metering in electricity grids," *Applied Sciences,* vol. 6, no. 3, p. 68, 2016.

[30] D. De Silva, X. Yu, D. Alahakoon, and G. Holmes, "A data mining framework for electricity consumption analysis from meter data," *IEEE Transactions on Industrial Informatics,* vol. 7, no. 3, pp. 399-407, 2011.

[31] S. Goel and Y. Hong, "Security Challenges in Smart Grid Implementation," in *Smart Grid Security*: Springer, 2015, pp. 1-39.

[32] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and R. C. Green, "High performance computing for detection of electricity theft," *International Journal of Electrical Power & Energy Systems,* vol. 47, pp. 21-30, 2013.

[33] 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.

[34] J. Nagi, A. Mohammad, K. Yap, S. Tiong, and S. Ahmed, "Non-technical loss analysis for detection of electricity theft using support vector machines," in *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, 2008, pp. 907-912: IEEE.

[35] S. S. Silva, R. M. Silva, R. C. Pinto, and R. M. Salles, "Botnets: A survey," *Computer Networks,* vol. 57, no. 2, pp. 378-403, 2013.

[36] S. Soltani, S. A. H. Seno, M. Nezhadkamali, and R. Budiarto, "A survey on real world botnets and detection mechanisms," *International Journal of Information and Network Security,* vol. 3, no. 2, p. 116, 2014.
[37] L. Liu, O. De Vel, Q.-L. Han, J. Zhang, and Y. Xiang, "Detecting and Preventing Cyber Insider Threats: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 20, no. 2, pp. 1397-1417, 2018.
[38] B. Soniya and M. Wilscy, "Using entropy of traffic features to identify bot infected hosts," in *Intelligent Computational Systems (RAICS), 2013 IEEE Recent Advances in*, 2013, pp. 13-18: IEEE.
[39] B. Soniya and M. Wilscy, "Detection of randomized bot command and control traffic on an end-point host," *Alexandria Engineering Journal,* vol. 55, no. 3, pp. 2771-2781, 2016.
[40] C.-T. Yen, S. Lugani, S. Mukhopadhyay, and K. Daftary, "Detecting botnets," ed: Google Patents, 2014.
[41] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," *ACM transactions on Information and system security (TiSSEC),* vol. 3, no. 4, pp. 227-261, 2000.
[42] S. Siboni and A. Cohen, "Botnet identification via universal anomaly detection," in *Information Forensics and Security (WIFS), 2014 IEEE International Workshop on*, 2014, pp. 101-106: IEEE.
[43] C. J. Dietrich, C. Rossow, and N. Pohlmann, "CoCoSpot: Clustering and recognizing botnet command and control channels using traffic analysis," *Computer Networks,* vol. 57, no. 2, pp. 475-486, 2013.
[44] A. Juvonen, T. Sipola, and T. Hämäläinen, "Online anomaly detection using dimensionality reduction techniques for HTTP log analysis," *Computer Networks,* vol. 91, pp. 46-56, 2015.
[45] N. R. Al-Dhubhani and F. Saeed, "A Prototype for Network Intrusion Detection System using Danger Theory," *Jurnal Teknologi,* vol. 73, p. 8, 2015.
[46] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR),* vol. 41, no. 3, p. 15, 2009.
[47] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE transactions on pattern analysis and machine intelligence,* vol. 36, no. 1, pp. 18-32, 2014.
[48] B. C. Costa, B. L. Alberto, A. M. Portela, W. Maduro, and E. O. Eler, "Fraud detection in electric power distribution networks using an ANN-based knowledge-discovery process," *International Journal of Artificial Intelligence & Applications,* vol. 4, no. 6, p. 17, 2013.
[49] C. Muniz, M. M. B. R. Vellasco, R. Tanscheit, and K. Figueiredo, "A Neuro-fuzzy System for Fraud Detection in Electricity Distribution," in *IFSA/EUSFLAT Conf.*, 2009, pp. 1096-1101: Citeseer.
[50] 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.
[51] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," 2008.
[52] J. V. Spirić, S. S. Stanković, M. B. Dočić, and T. D. Popović, "Using the rough set theory to detect fraud committed by electricity customers," *International Journal of Electrical Power & Energy Systems,* vol. 62, pp. 727-734, 2014.
[53] P. Jacquet and W. Szpankowski, "Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees," *Theoretical Computer Science,* vol. 144, no. 1-2, pp. 161-197, 1995.
[54] P. Jacquet, W. Szpankowski, and J. Tang, "Average profile of the Lempel-Ziv parsing scheme for a Markovian source," *Algorithmica,* vol. 31, no. 3, pp. 318-360, 2001.
[55] J. Uthayakumar, T. Vengattaraman, and J. Amudhavel, "A simple data compression algorithm for anomaly detection in Wireless Sensor Networks," *International Journal of Pure and Applied Mathematics,* vol. 117, no. 19, pp. 403-410, 2017.
[56] G. Louchard and W. Szpankowski, "Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm," *IEEE Transactions on Information Theory,* vol. 41, no. 2, pp. 478-488, 1995.
[57] D. Belazzougui and S. J. Puglisi, "Range predecessor and Lempel-Ziv parsing," in *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, 2016, pp. 2053-2071: Society for Industrial and Applied Mathematics.
[58] S. Kwong and Y. F. Ho, "A statistical Lempel-Ziv compression algorithm for personal digital assistant (PDA)," *IEEE Transactions on Consumer Electronics,* vol. 47, no. 1, pp. 154-162, 2001.

BIOGRAPHIES OF AUTHORS

Abdulrahaman Okino Otuoze received his B. Eng. Degree (2008) in Electrical Engineering, University of Ilorin, Ilorin, Nigeria, M. Eng. in Power and Machines (2012), University of Benin, Benin City, Nigeria and currently a PhD candidate, School of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru. He is also a Lecturer at the University of Ilorin, Ilorin, Nigeria and a registered Engineer by the council for the regulation of Engineering in Nigeria (COREN). He is a member of the Nigeria Society of Engineers (NSE) and Institute of Electrical and Electronics Engineers (IEEE). His research interest includes smart grids security, power systems distribution automation, electrical machines (design and performance analysis) and renewable energy systems.

Mohd Wazir Mustafa received his B. Eng. Degree (1988), M. Sc. (1993) and PhD (1997) from University of Strathclyde, Scotland, UK. He is currently a Professor and the Chair of the School of Electrical Engineering, Universiti Teknologi Malaysia. He is a member of Institution of Engineers, Malaysia (IEM) and a member of IEEE. His research interest includes power system stability, FACTS, wireless power transmission and power system distribution automation.

Ibim Sofimieari is a PhD candidate in the department of electrical power engineering, Universiti Teknologi Malaysia. He had his B. Sc and M. Sc (power systems and networks) in electrical engineering from Vinnytsia State Technical University, Ukraine. He is a member of IEEE and a member of Nigerian Society of Engineers (NSE). He worked briefly in the oil industry in Nigeria before proceeding to lecture in the University of Port Harcourt, Nigeria. He is interested in renewable energy integration, hybrid microgrids and power systems operations.

AbdulHakeem Mohd Dobi received a B.Eng. degree and M.Eng. degrees in electrical engineering from Bayero University, Kano (BUK) Nigeria, in 2005 and 2012 respectively. Currently, he is working toward a Ph.D. degree in power electronics at the Universiti Teknologi Malaysia (UTM) Johor Bahru. His research interests include power electronics, soft switching, resonance DC-DC converters, energy conversion and control.

Aliyu Hamza Sule is a PhD candidate, School of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru and a lecturer at Hassan Usman Katsina Polytechnic, Nigeria. His research interests incoudes Control systems

Abioye Abiodun Emmanuel received his B.Eng. degree in Electrical Engineering and M.Eng. in Electronics and Communication Engineering from University of Ilorin, Ilorin, Nigeria and Michael Okpara University of Agriculture, Umudike, Nigeria, respectively. Currently, he is pursuing his Doctor of Philosophy in Electrical Engineering at Universiti Teknologi Malaysia (UTM). His research interests include Internet of Things, precision agriculture, artificial intelligence and multivariable/process control system.

Muhammad Salman Saeed received his B. Eng. degree in Electrical Engineering (2011) from Bahauddin Zakariya University (BZU) Pakistan, M. Eng. in Electrical Power Enginnering (2013), the Islamia University of Bahawalpur, Pakistan and is currently a Ph.D candidate in the School of Electrical Engineering at Universiti of Teknologi Malaysia (UTM) Johor Bahru. His research interests include Non-technical loss reduction and smart grid security.