ISSN 2278-3091

Volume 9, No.5, September - October 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse96952020.pdf

https://doi.org/10.30534/ijatcse/2020/96952020

Identification of Data Falsification in Smart Meter Consumption using an Integrated Statistical Technique

Siti Hawa Binti Mokhtar[1], Fiza Abdul Rahim[2], Zul-Azri Ibrahim[3]

[1]College of Graduate Studies, Universiti Tenaga Nasional, Malaysia, ct.hawa10@yahoo.com

[2] Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Malaysia, fiza@uniten.edu.my

[3]College of Computing and Informatics, Universiti Tenaga Nasional, Malaysia, zulazri@uniten.edu.my

ABSTRACT

False smart meter consumption data injected from compromised smart meters in Advanced Metering Infrastructure (AMI) is a threat that affects both utilities and consumers. With the increasing amount of smart meters' deployment, this will cause difficulties in the identification of fraud and malicious attempts. Due to the large scale of potential evidence, it is regarded as a grand challenge for forensic investigators in identifying relevant patterns of events. Furthermore, most of the existing works only deal with electricity theft from customers. Derived from these motivations, this study is focusing on the identification of data falsification in smart meter consumption and propose an integrated statistical technique combining interquartile range (IQR) and K-means. It is assumed that solving this challenge would help in structuring investigation findings, which able to aid investigator of law enforcement agencies and other stakeholders in reasoning and identifying compromised smart meters.

Key words: falsification attack, smart meter, forensics, statistical, energy.

1. INTRODUCTION

Societies around the world critically depend on the proper functioning of their critical infrastructures (CI) services such as energy supply, telecommunications, financial systems, water, and governmental services. The trend in designing and managing CI is to use complex information technology (IT) infrastructures interconnected through networks, which known as critical information infrastructure (CII).

In Malaysia, Critical National Information Infrastructure (CNII) is defined as those assets (real and virtual), systems and functions that are vital to the nations that their incapacity or destruction would have a devastating impact on national economic strength, national image, national defense and security, government capability to functions and public health and safety [1]. CNII sectors in Malaysia include national defense and security, banking and finance, information and communications, energy, transportation, water, health services, government, emergency services, food, and agriculture.

In the energy sector, the use of electric meters was initially applied to industrial and commercial customers due to the need for more sophisticated rates and more granular billing data requirements [2]. The usage was gradually expanded to all customer classes to accommodate a large number of customers. Automated meter reading (AMR) has been used to collect meter data by utilizing one-way communication.

Along with the transition of the traditional electrical grid to the growing development of a smart grid, advanced metering infrastructure (AMI) has been intensely developed during recent years. AMI is responsible for collecting, measuring, and analyzing energy consumption data, transmitting this information from a smart meter to a data concentrator, and then to a head-end system in the utility side.

The infrastructure includes smart meters, home network systems, communication networks from the meters to local data concentrators, back-haul communications networks to corporate data centers, meter data management systems (MDMS) and, finally, data integration into existing and new software application platforms [3]. As one of the most crucial components in the smart grid, it is important to ensure the security of overall AMI infrastructure [4], [5].

The employment of IT in the CI permits huge savings by reducing costs by switching to networking technologies while allowing large scale deployments based on off-the-shelf computing equipment [6]. With the benefits of two-way communication through a very complex combination of devices, services, protocols, and standards, along with the huge accessibility to technologies and methods, CII can be easily exposed to cyber-attacks. This makes AMI vulnerable to increased security threats at both the physical and logical layer [5], [7].

Furthermore, because of its bidirectional, interoperable, and software-oriented nature, AMI is very prone to cyber-attacks

[5]. If proper security measures are not taken, a cyber-attack on AMI can potentially bring a huge impact in the whole smart grid infrastructure, which might affect the society. For example, in mid-2010, the Stuxnet worm was engineered to specifically target infrastructure-monitoring computer systems built by Siemens and successfully gained control of key computer systems in Iran's nuclear facilities [8]. It was reported that the worm then spread to another similar computer systems in other countries [9], [10].

In late 2012, the smart grid software company Telvant reported a breach of their network by a group of Chinese hackers, resulting in the stolen project information, and the spread of malicious software across their network. Despite the disruption, the Telvant hack was relatively minor, a breach on a larger scale could result in devastating prolonged electric blackouts, and complete disruption of a utility communication systems [11].

Numerous works of literature also reported on potential cyber-attacks against the smart grid and its infrastructure [12]–[16]. Moreover, it is not only business applications but also communication links and underlying control systems that are susceptible to cyber-attacks.

In AMI specifically, the smart meter possesses huge cyber threat vulnerabilities that can cause catastrophic failure not only to the consumer but also to the utility provider, and the impact is nationwide. The smart meter is exposed to cyber-attacks because of its nature that allows communication through open space that supposed to enable consumer and utility provider to have the ability to monitor the smart meter from remote using a wireless connection and a broadband public network [17].

Based on [18], the vulnerabilities of a smart meter can be separated into three major areas that can be exploited by the cyber-criminal; physical hardware, the smart meter network topology and the software that used by the consumer and utility provider to monitor the smart meter condition.

For physical hardware, the adversary can replace the smart meter with a replication device to false the information in the smart meter or will physically breach the smart meter appliance to gain access. This attack may attain a high percentage of success due to direct access from the adversary [12]. Once the adversary manages to gain access in the smart meter, they can tamper the smart meter information, which can lead to wrong billing information.

For network topology, several attacks can be deployed in this layer such as Near-me Area Network (NAN) sniffing [19], signal jamming and denial of service attack [20], man-in-the-middle (MITM) attack [21], [22], and worm propagation [23], [24].

NAN sniffing attacker able to get information about future price information, control structure, and power consumption once they successfully decrypt the network encryption used by the smart meter communication [19]. In comparison, a jamming attack in AMI targeted to flood the smart meter wireless media in order to prevent it from communicating with the utility provider. Once the node already been compromised, the adversary will take control of the node and can start sending a random packet to compromise the network [20].

The MITM attack can be categorized as a combination of eavesdropping, injection and spoofing attack. By doing a MITM attack, the adversary can distribute and false encryption to other nodes in the network [21], [22]. In AMI environment, a computer worm can be spread in various ways; it may convey a payload that plays out a distinction instruction, which kills the inner meter switch associated with user home. The disconnect instruction in a synchronized way may cause sudden load drop in the system, which may lead to generators stumbling and power failure.

Whereas, the vulnerability of smart meter in software involves attacks on the smart meter web application level [19]. The unauthorized aggregating and correlating smart reading produce interesting information which later can be used by the attacker for various purposes.

Notably, due to the diverse angle of cyber-attacks that may compromise AMI, it is essential to categorize and secure where smart metering data is collected, stored, transported, analyzed. By adequate categorizing through a communication network and security, possible relationships between source and criminals can be reduced.

Concerning the time granularity, the new generation of smart meter technology can typically record and transmit at intervals of about 2 seconds or less [25]. For example, the smart meter data of a large utility company may generate millions or even billions of events per second [26]. According to a study done by ABI Research, there will be about 780 million smart meters that will be installed worldwide by 2020 [27].

The growing numbers of smart meter installation later contribute to generating a large amount of data. The data is needed to provide utility providers with capabilities for forecasting demand, shaping customer usage patterns, preventing outages, optimizing unit commitment and more [28]. On top of that, customer's personal data also can be derived from meter logs and the energy consumptions patterns [29].

The massive growth in the size of available digital data plays a major role in the increase in the size of digital evidence [30]. According to a summary report of CyberSecurity Malaysia, their Digital Forensics Department faced the increasing number of cases, along with the growing number of exhibits and size of the media which they need to tackle [31]. It is also challenging to handle incidents that arise during the unavailability of technical experts [32].

Particularly in AMI environment, a large number of IoT devices installed in one consolidated environment producing very big datasets of activity logs. Hence, fraudulent and malicious activities/users ae harder to spot using traditional log correlation or visualization techniques.

Whilst there is variety data being transferred to and from AMI environment, this study will focus on data falsification of consumption data. Consumption data can be falsified or modified due to calibration inaccuracies, faulty meter, false readings by meter reader either consciously or not, flaws in head-end-system, temper attempt either by meter owner himself or outsider attacker.

False consumption data from smart meter could be either less than or even more than the actual usage. While higher consumption data readings may seem like profit for electricity company when translated to monthly bills, it causes loss of trust between consumers towards the company that could result in higher unpaid bill numbers [33]. Billing more would also cause unused generated energy that was calculated by demand from billing data.

Smaller reading of actual consumption did not also solely impact annual revenues; these losses also indirectly mess with another power system such as power surge, where the system is expecting smaller usage but the demand is a lot more than expected [33]. The severe consequences of that situation would be a nationwide blackout.

In order to make sure that what is billed to the customer is what he only use for that month, necessary prevention and detection mechanism must be there. However, a deeper investigation like physical inspection of meters are time-consuming and require high-cost labor on every meter on a daily basis, and more complex analytics needs more data that is challenging to obtain due to a multitude of privacy and security concerns.

Hence, this study will focus on identifying data falsification in smart meter consumption data which is able to identify compromised meters from non-compromised meters over margins of false data on a large scale.

2.ADVANCED METERING INFRASTRUCTURE (AMI)

2.1 Overview of AMI

In the energy sector, the use of electric meters was initially applied to industrial and commercial customers due to the need for more sophisticated rates and more granular billing data requirements [2]. The usage was gradually expanded to all customer classes to accommodate a large number of customers. AMR has been used to collect meter data by utilizing one-way communication.

Along with the transition of the traditional electrical grid to the growing development of the smart grid, AMI is one of the components of the electrical network combining the energy and telecommunications infrastructure. The main actor in this system is a new type of energy metering device called smart meter [34]. AMI is responsible for collecting, measuring, and analyzing energy consumption data. It transmits information from a smart meter to a data concentrator, and then to a head-end system in the utility side. AMI works in two-way communication, starting with a request by the energy provider or pre-programmed microcontroller-based system to continuously and automatically records the readings and send information back for a pre-determined time [35]. One of the critical elements to implement AMI and smart meters is to use this functionality to manage electricity generation and distribution [36].

In AMI, a nearly real-time monitoring display of consumer’s energy consumption data has made it easier for utility companies to monitor anomalies in the network [37], [38]. However, a large number of users and smart meters give rise to the need to deliver better maintenance and monitoring more efficiently while keeping consumers informed on their own consumption habits [39]. As one of the most crucial components in the smart grid, it is vital to ensure the security of overall AMI infrastructure [4], [5].

2.2 Smart Meter

Smart meters support instant power consumption reading, remote connection and disconnection of supply, remote system update and downgrade, event reporting functions and steal detection functions [40]. Some smart meter is also able to inform users about the quantities measured [39] through communication with display device interference called in-home monitor that displays info such as real-time price.

Smart meters include various communication networks, depending on the geographical range and purpose. It sends information about the meter’s environment, status, as well as meter readings to Data Concentrator Unit (DCU) using Local Area Network (LAN) or Neighbourhood Area Network (NAN) [41]. After that, DCU which function similar to routers, acts as a gateway between smart meters and Head-End-System(HES) [41]. It collects, stores and transmits data/messages to and from smart meters and HES using WAN [41] which is usually the backhaul network that uses long-range and high-bandwidth communication technologies, such as long-range wireless (e.g., WiMAX), cellular (e.g., 3G, EVDO, EDGE, GPRS, or CDMA), satellite, or Power Line Communication (PLC) [42]. The data have to travel through various communication protocols [43] before collected and cleansed to use in other application for billing and analyzing purpose.

As it gets smarter day by day, the smart meter also able to tackle energy efficiency issues by having demand response and smart pricing features. Consumption patterns of users are utilized to forecast generation of electricity supply align with demand [44]. Tariff of Use (TOU) can also be set accordingly and give the consumer the freedom to pay peruse. Hence, many utility companies are leveraging smart meter in producing more cost-effective and reliable energy management.

However, smart meter consumption data include large volumes of time-series measurements that are subject to noise and may not be effective when used in the raw form [45]. Moreover, it has been widely acknowledged that given the important role and interconnected nature of AMI, it could potentially be the target of powerful and organized adversaries [46].

Bugden and Stedman [36] suggested a future research should investigate how utilities and policymakers can protect vulnerable groups from risks imposed by smart meter-enabled changes.

2.3 AMI Threats and Security Features

From attack resilience AMI model created by [41], we found that threats in AMI can be classified into three categories; ‘physical’ level, ‘manufacturing’ levels, and ‘network’ level. Physical level attacks are various such explained in [47] [48]. Some of the attacks include ‘Direct hooking from the line’, ‘Bypassing the energy meter’, ‘Injecting foreign element into the energy meter’, ‘Disc Physical Obstruction’, and ‘ESD attack’. Some of these attacks are indeed preventable by a smart meter, but some others are not. Besides, smart meter technology also comes along with a bundle of newer threats such as network attacks [49] and other attacks related to the Internet of Things (IoT) [43].

On the manufacturing level, threats faced are more severe than at other levels. This is because its consequences have mass and immediate effect. Imagine possible incidents that could happen if the meter was bugged or faulty by the time it rolled out from factories itself. For example, Puerto Rico had experienced electricity theft amounting to annual losses for the utility estimated at $400 million [50]. Its electricity Company, Puerto Rican Electrical Power Authority (PREPA) had also filed bankruptcy once [51]. Due to that, according to a report by Brian Krebs [52] and [33], the company had asked the FBI to investigate large-scale thefts of electricity-related to its smart meters. The Cyber Intelligence Section of the FBI found that the former employees of the smart meter manufacturer company had hacked and bugged the smart meters that were then sold to the customer for $300 − $700 each [53]. From this case, we could derive that the usage of hashing, digital signatures, and cryptographic algorithms are very important in making sure that access and control during the manufacturing process are given to only trusted parties [54].

Meanwhile, on the network level, the main concern is on how to maintain availability and authenticity of end-to-end transfer of genuine smart meter data. While Firewall and common network protection such as Virtual LAN (VLAN) and demilitarized zones (DMZ) are a must, this is particularly challenging in AMI as it uses a lot of network technology in its infrastructure [40]. For instances, HAN technology such as ZigBee Protocol allows an attacker to join the HAN network without proper authentication and further impersonating or send a command to other devices connected [40]. This issue also happens to GSM technology, where the GSM tower requires only one-way authentication of the device to tower, providing loopholes of the attacker to exploit the system using the fake tower. For this issue, shared key authentication (SKA) should be used. For example, [55] introduces a security solution that is designed to ensure privacy protection, data authenticity, confidentiality and data integrity of smart meter data.

Another threat level that both impacts but also a subset of all three levels mentioned above are legal/political threat. As described by [56], due to various limitation experienced by developing countries, quality of standards and audits regarding information security is not quite on par and rigorously done as most developed countries.

2.4 Anomalies Detection using Data Analytics

All of the security features mentioned earlier are of intrusion prevention systems (IPSs). It prevents threat events from occurring [49]. Another feature that could further secure the AMI is by using intrusion detection systems (IDSs) and network intrusion detection systems (NIDS). It comes in action during and after the occurrence of threat events [57]. They are divided into two types; ‘Signature-based IDSs’ and ‘Anomaly-based IDS’. Signature-based IDS monitor system behavior or network traffic(NIDS) for predefined attack patterns, known as signatures [49] [57]. Meanwhile, anomaly-based IDS detect abnormal events that deviate from the normal behavior of the system [49] [57].

Anomalies and outliers technically refer to the same thing and usually used interchangeably by researchers. Anomalies can be categorized as point anomaly [58][59], contextual anomaly [60][59] and collective anomaly[61][60].

As smart meter and smart grid deployments continue to gain momentum, demand for data analytics capability to better manage grid operations and plan network investment is at an all-time high. It is important to manage high volume real-time data from smart meters and identifying innovative analytics to grab any potential opportunities presented by the data generated from smart meters.

The availability of large smart meter consumption data in AMI has the potential to enable new insights and better decisions [62]. Currently, extensive research is being done to carry out data analytics [63], [64]. The research is mainly focused in forecasting the accurate and efficient power consumption/supply [65], predicting peak demand [62], identifying the impact of temporal data granularity on the accuracy of electricity consumption [62], analyzing security risk [66], and detecting future attacks [67]–[69].

A forensic investigation procedure on electric meters has been done by [64]. In their finding, an increasing number of committed crimes with advanced technologies and sophisticated methods. Another study only focuses on power system disturbance criminal case [70]. However, very limited studies were conducted on digital data in AMI. Which is a loss as [71] in his literature review has compiled how data mining is used in predicting crime and legal judgements.

In this study, predictive analytics will be utilized through applied mathematical techniques to uncover explanatory and predictive models.

3. PROPOSED TECHNIQUE

The proposed technique consists of five major steps that are step 1: training data clustering, step 2: the interquartile range (IQR) range comparing, step 3: cluster updating and Mean recalculation 4: IQR pattern modulation test and 5: flagging and reporting with the output from each step serving as the input to the subsequent step in the process.

Firstly, for distance or pattern-based anomaly detection, reference pattern must first be set as a guide to classify the following data. This data set is called "training data". The training data will go through the clustering process using K-Means technique.

After the training step has been done, a consumption unit of a smart meter is read, and the counter is started. For counter less than 24, IQR value is calculated for each cluster, and the new point value will be compared with each IQR threshold ranges.

If the point fits in 'normal' cluster's IQR range, then the value will be updated into the cluster. Meanwhile, if the point fits in 'less-than-normal' cluster's IQR range, then the value will be updated into the cluster. Else, it will be updated in 'higher-than-normal'. In addition, for every point that does not fit in 'normal' cluster IQR, a flag that will be reflected in the output report will be given accordingly.

At the point where the counter hit 24 (one day reading), the mean of the clusters will be recalculated, and the cluster will change in mean and points membership accordingly. This step is done for every 24 readings instead of for every updated point is to save on performance cost.

After recalculation of mean happened, the new IQR is calculated and compared with the past four IQR records collected. This step is to ensure there is no attempt of injecting false data little-by-little, enough to fit in normal IQR to manipulate the cluster's pattern. For constant drop or rise in readings, the flag is given.

This process will be repeated until all data points are finished, or it finds Null value. Once, a null value is found, a report is generated. And the report can be given to a physical inspection team or forensic team for further investigation.

3.1 K-Means

K-Means will be used in this study to categorize the training data set into three clusters and during the cluster's Mean centroids recalculation that occurs every 24 readings (a day consumption). Firstly, the number of the centroid is chosen according to how we want to partition the data and the value is initially randomized. In this study, we set the 'K' number of cluster to K=3.

The first cluster is 'normal' cluster, which is expected to be populated with the un-modified consumption reading. The next cluster will be 'less-than-normal' cluster. This cluster is expected to be of data with deductive-ly manipulated data. In contrast, the other one, that is expected to consist of additive-ly modified data is a 'higher-than-normal' cluster.

Secondly, the distance between each of the data points with each of the centroids is calculated. Euclidean distance is used as the distance metric to find which cluster the points belong to. The formula is:

$$dist((x, y), (a, b)) = \sqrt{(x - a)^2} + (y - b)^2 \quad (1)$$

After figuring out the distances, the data point is then grouped to the nearest centroid using below mathematical representation (description Table 1):

$$S(t)i=\{xp: \| \| xp-m(t)i \| \| 2\leq \| \| xp-m(t)j \| \| 2\forall j,1\leq j\leq k\} \quad (2)$$

Table 1: Formula Description.

Symbol	Description
S1 ,..., Si	Cluster
m	Centroid value (mean)
x	Data Point
xp	Distance of Data Point(x) to the mean(m)
Symbol	Description

Which means Set i contains all data points (x) in which the distance from the data point (xp) to the mean of i (mi) is smaller than that of the distance from the data point to the mean of all other centroids (mj).

This process is repeated for every data points until data points finish and similar data points get grouped into clusters. While this process happens, it repeatedly calculates the Mean value for each cluster and reassigns the centroid. This step can be represented using below mathematical representation:

$$m(t+1)i=1/|S(t)i|/\sum xj\in S(t)ixj \quad (3)$$

New centroid value of cluster Set I (mi) is equal to Mean (sum of value divide by total number) of cluster i members. Now that we are done with clustering, we need to iteratively Reassign data points to new clusters, refining the centroid values until they do not change much.

$$S(t+1)i=\{xp:\|\|\|xp-m(t+1)i\|\|\|2\leq\|\|\|xp-m(t+1)j\|\|\|2\,\forall j,\ 1\leq j\leq k\} \quad (4)$$

Cluster i contains data point where its distance to newly calculated Mean m(t+1) i is nearest compared to other newly calculated Mean m(t+1) j.

3.2 IQR

The interquartile range (IQR), is somewhat similar to Z-score in terms of finding the distribution of data and then keeping some threshold to identify the outlier. But it uses Median-the middle value, as a point of reference compared to Mean in Z-Scores. In such settings, the median is typically considered to be more robust to outliers [72].

The main reason for introducing IQR technique in this study is to leverage performance issues caused by repetitive iteration of K-Means. IQR works faster as it does not need recalculation of Means for every new data point arrival, but will just sorts the value and compares the data point value fitness to the value within the quartile range. Anomaly will be indicated per data point that lies outside the overall distribution of the dataset.

IQR implementation consist of five general step. The first one is to arrange the data in increasing orders. Secondly the data is divided into four equal parts with each part comprises of quarter data. The four quarters will result in five boundary lines namely Q0, Q1, Q2, Q3 and Q4 as shown in Figure 1.

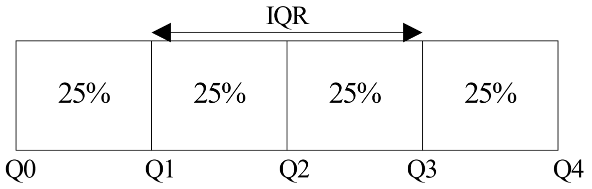

Figure 1: IQR Representation

Thirdly, the values of Q1, Q2 and Q3 are calculated. Q2 is the median of the data set, Q1 is the middle number between the smallest number and the median of the data set and Q3 is the middle value between the median and the highest value of the data set.

Next, IQR is calculated using a formula "IQR = Q3 − Q1" to get the "range between the between upper and lower quartiles". Datapoint that falls outside of 1.5 times of an interquartile range above the 3rd quartile and below the 1st quartile is considered anomalies.

$X = Q1-(1.5xIQR) \text{ or } Q3+(1.5xIQR)$ (5)

4.CONCLUSION AND FUTURE WORKS

From the review of AMI infrastructure, type of anomalies and techniques of anomaly detection done above, this study proposes to implement K-Means and IQR to detect falsified data. K-Means and IQR are chosen due to the suitability of the smart meter consumption data set and its ability to produce the expected end result. Furthermore, IQR is chosen to aide K-Means in terms of performance.

The proposed technique is expected to (i) identify compromised smart meter (ii) enables quick identification for larger sized smart meter consumption data in AMI environment.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education Malaysia (MoHE) and Universiti Tenaga Nasional (UNITEN) for funding this study under the Fundamental Research Grant Scheme (FRGS) of Grant No. FRGS/1/2017/ICT03/UNITEN/03/1.

REFERENCES

1. CyberSecurity Malaysia, **CNII Portal**, *About CNII*, 2017. [Online]. Available: http://cnii.cybersecurity.my/main/about.html.
2. Edison Electrical Institute, **Smart Meters and Smart Meter Systems : A Metering Industry Perspective**, 2011.
3. L. Wenpeng, **Advanced metering infrastructure**, *South. Power Syst. Technol.*, vol. 3, no. 2, pp. 6–10, 2009.
4. W. Tong, L. Lu, Z. Li, J. Lin, and X. Jin, **A Survey on Intrusion Detection System for Advanced Metering Infrastructure**, *2016 Sixth Int. Conf. Instrum. Meas. Comput. Commun. Control*, pp. 33–37, 2016.
5. M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, **Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study**, *IEEE Syst. J.*, vol. 9, no. 1, pp. 31–44, 2015.
6. L. Montanari and L. Querzoni, **Critical Infrastructure Protection : Threats , Attacks and Countermeasures**, 2014.
7. C. Lopez, A. Sargolzaei, H. Santana, and C. Huerta, **Smart Grid Cyber Security: An Overview of Threats and Countermeasures**, *J. Energy Power Eng.*, vol. 9, no. 7, pp. 632–647, 2015.
8. G. M. Shiffman and R. Gupta, **Crowdsourcing cyber security: A property rights view of exclusion and theft on the information commons**, *Int. J. Commons*, vol. 7, no. 1, pp. 92–112, 2013.
9. P. Beaumont, **Iran nuclear experts race to stop spread of Stuxnet computer worm**, *The Guardian*, 2017. [Online]. Available: https://www.theguardian.com/world/2010/sep/26/iran-stuxnet-worm-nuclear.
10. B. M. Clayton, **Stuxnet malware is "weapon" out to destroy ... Iran' s Bushehr nuclear plant ?**, *Christian Science Monitor*, 2010. [Online].

Available: http://www.csmonitor.com/USA/2010/0921/Stuxnet-malware-is-weapon-out-to-destroy-Iran-s-Bushehr-nuclear-plant.

11. A. Davitt, **Smart Grid Security: Is the United States Safe from Cyber Attacks?**, 2013. [Online]. Available: https://www.energyacuity.com/blog/bid/293990/Smart-Grid-Security-Is-the-United-States-Safe-from-Cyber-Attacks.
12. R. Mahmud, R. Vallakati, A. Mukherjee, P. Ranganathan, and A. Nejadpak, **A survey on smart grid metering infrastructures: Threats and solutions**, in *2015 IEEE International Conference on Electro/Information Technology (EIT)*, 2015.
13. S. Goel and Y. Hong, **Security Challenges in Smart Grid Implementation**, in *Smart Grid Security*, 2015, pp. 247–281.
14. F. Skopik, Z. Ma, T. Bleier, and H. Grüneis, **A Survey on Threats and Vulnerabilities in Smart Metering Infrastructures**, *Int. J. Smart Grid Clean Energy*, pp. 22–28, 2012.
15. J. Hu, A. V Vasilakos, and S. Member, **Energy Big Data Analytics and Security : Challenges and Opportunities**, vol. 7, no. 5, pp. 2423–2436, 2016.
16. C. Lopez, A. Sargolzaei, H. Santana, and C. Huerta, **Smart Grid Cyber Security: An Overview of Threats and Countermeasures**, *J. Energy Power Eng.*, vol. 9, no. 7, pp. 632–647, 2015.
17. H. Almakrami, **Intrusion detection system for smart meters**, in *Smart Grid (SASG), 2016 Saudi Arabia*, 2016.
18. J. C. Foreman and D. Gurugubelli, **Cyber Attack Surface Analysis of Advanced Metering Infrastructure**, *Electr. J.*, vol. 28, no. 1, pp. 94–103, 2016.
19. F. Skopik, Z. Ma, T. Bleier, and H. Grüneis, **A Survey on Threats and Vulnerabilities in Smart Metering Infrastructures**, *Int. J. Smart Grid Clean Energy*, pp. 22–28, 2012.
20. F. Aloul, A. R. Al-ali, R. Al-dalky, and M. Al-mardini, **Smart Grid Security : Threats , Vulnerabilities and Solutions**, *Smart Grid Clean Energy Smart*, no. 971, pp. 1–6, 2012.
21. E. D. Knapp and R. Samani, *Applied Cyber Security and the Smart Grid*. Elsevier Inc., 2013.
22. S. McLaughlin, D. Podkuiko, and P. McDaniel, **Energy Theft in the Advanced Metering Infrastructure**, in *Lecture Notes in Computer Science*, 2010, pp. 176–187.
23. A. Almajali, **Modeling Worm Propagation in the Advanced Metering Infrastructure**, in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2016, pp. 0–3.
24. A. AlMajali, E. Rice, A. Viswanathan, K. Tan, and C. Neuman, **A systems approach to analysing cyber-physical threats in the smart grid**, *2013 IEEE Int. Conf. Smart Grid Commun. SmartGridComm 2013*, pp. 456–461, 2013.
25. U. Greveler, B. Justus, and D. Loehr, **Forensic content detection through power consumption**, *IEEE Int. Conf. Commun.*, pp. 6759–6763, 2012.
26. K. Lyko, M. Nitzschke, and A.-C. N. Ngomo, *Big Data Acquisition*. 2016.
27. ABI, **Smart Electricity Meters to Total 780 Million in 2020 , Driven by China's Roll-out**, 2015.
28. IBM Corporation, **Managing big data for smart grids and smart meters**, 2012.
29. G. D'Acquisto, J. Domingo-Ferrer, P. Kikiras, V. Torra, Y.-A. de Montjoye, and A. Bourka, **Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics**, 2015.
30. S. Zawoad and R. Hasan, **Digital Forensics in the Age of Big Data : Challenges , Approaches , and Opportunities**, *High Perform. Comput. Commun. (HPCC), 2015 IEEE 7th Int. Symp. Cybersp. Saf. Secur. (CSS), 2015 IEEE 12th Int. Conf. Embed. Softw. Syst. (ICESS), 2015 IEEE 17th Int. Conf.*, vol. IEEE, pp. 1320–1325, 2015.
31. R. A. Manaf and N. A. Mohamad, **CyberCSI 2nd Half Year 2012, Summary Report**, 2013.
32. J. R. Batmetan, Q. C. Kainde, M. Nur, T. Komansilan, and S. Kumajas, **Information Security Governance in Small Cities in Developing Countries**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5, pp. 182–191, 2019.
33. S. Bhattacharjee, A. Thakur, S. Silvestri, and S. K. Das, **Statistical Security Incident Forensics against Data Falsification in Smart Grid Advanced Metering Infrastructure**, *Proc. Seventh ACM Conf. Data Appl. Secur. Priv. - CODASPY '17*, pp. 35–45, 2017.
34. V. Tudor, M. Almgren, and M. Papatriantafilou, **The influence of dataset characteristics on privacy preserving methods in the advanced metering infrastructure**, *Comput. Secur.*, vol. 76, pp. 178–196, 2018.
35. F. Abdul Rahim, A. Abu Bakar, S. Yussof, R. Ismail, and R. Ramli, **Privacy preservation framework for advanced metering infrastructure**, in *Proceedings of the 6th International Conference on Computing and Informatics, ICOCI 2017*, 2017, no. 128, pp. 744–749.
36. D. Bugden and R. Stedman, **Energy Research & Social Science A synthetic view of acceptance and engagement with smart meters in the United States**, *Energy Res. Soc. Sci.*, vol. 47, no. January 2018, pp. 137–145, 2019.
37. R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, **A practical feature-engineering framework for electricity theft detection in smart grids**, *Appl. Energy*, vol. 238, no. December 2018, pp. 481–494, 2019.

38. S. Wan Yen, S. Morris, M. A. G. Ezra, and T. Jun, **Electrical Power and Energy Systems E ff ect of smart meter data collection frequency in an early detection of shorter-duration voltage anomalies in smart grids**, *Electr. Power Energy Syst.*, vol. 109, no. December 2018, pp. 1–8, 2019.
39. D. B. Avancini, J. J. P. C. Rodrigues, S. G. B. Martins, and R. A. L. Rab, **Energy meters evolution in smart grids : A review**, *J. Clea*, vol. 217, pp. 702–715, 2019.
40. Y. Li, R. Qiu, and S. Jing, **Intrusion detection system using Online Sequence Extreme Learning Machine (OSELM) in advanced metering infrastructure of smart grid**, *PLoS One*, vol. 13, no. 2, pp. 1–16, 2018.
41. R. Blom, M. Korman, R. Lagerstrom, and M. Ekstedt, **Analyzing attack resilience of an advanced meter infrastructure reference model**, *IEEE Proc. 2016 Jt. Work. Cyber-Physical Secur. Resil. Smart Grids, CPSR-SG 2016 - This Work. is Part CPS Week 2016*, 2016.
42. A. A. C??rdenas, R. Berthier, R. B. Bobba, J. H. Huh, J. G. Jetcheva, D. Grochocki, and W. H. Sanders, **A framework for evaluating intrusion detection architectures in advanced metering infrastructures**, *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 906–915, 2014.
43. J. Horalek and V. Sobeslav, **ANALYSIS OF COMMUNICATION PROTOCOLS FOR SMART METERING**, vol. 10, no. 3, pp. 1438–1446, 2015.
44. Z. Shah, Z. Tari, A. Anwar, A. Y. Zomaya, and A. N. Mahmood, **A Spatio-temporal Data Summarization Paradigm for Real-time Operation of Smart Grid**, *IEEE Trans. Big Data*, vol. 14, no. 8, pp. 1–1, 2017.
45. N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, **Comprehensive feature selection for appliance classification in NILM**, *Energy Build.*, vol. 151, pp. 98–106, 2017.
46. J. Mendel, **Smart Grid Cyber Security Challenges : Overview and Classification**, *e-mentor*, vol. 1, no. 1, 2017.
47. K.Rajendra, **CONTROLLING OF POWER THEFT AND REVENUE LOSSES BY USING WIRELESS TECHNIQUES**, vol. 4, no. 1, pp. 395–399, 2016.
48. M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim, and Z. A. Khan, **Minimizing electricity theft using smart meters in AMI**, *Proc. - 2012 7th Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput. 3PGCIC 2012*, pp. 176–182, 2012.
49. P. Jokar, **Detection of Malicious Activities Against Advanced Metering Infrastructure in Smart Grid**, University of British Columbia, 2015.
50. V. B. Krishna, G. A. Weaver, and W. H. Sanders, **PCA-based method for detecting integrity attacks on advanced metering infrastructure**, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9259, no. Qest, pp. 70–85, 2015.
51. Reuters, **Puerto Rican Power Utility PREPA Files For Bankruptcy | Fortune**, *http://fortune.com/2017/07/02/puerto-rico-prepa-energy-bankruptcy/July 3, 2017*, 2017. .
52. Brian Krebs, **FBI: Smart Meter Hacks Likely to Spread — Krebs on Security**, 2012. .
53. S. Bhattacharjee, A. Thakur, and S. K. Das, **Towards Fast and Semi-supervised Identification of Smart Meters Launching Data Falsification Attacks**, pp. 173–185, 2018.
54. D. Andeen, **Ensuring the Complete Life-Cycle Security of Smart Meters Urgent Need for Secure Smart Meters**, pp. 1–8, 2014.
55. T. Ahmad, D. Qadeer, U. Hasan, and S. Zada, **Non-Technical Loss Detection, Prevention and Suppression Issues for AMI in Smart Grid**, *Int. J. Sci. Eng. Res.*, vol. 6, no. 3, 2015.
56. V. R. Sayoc, T. K. Dolores, M. C. Lim, L. Sophia, and S. Miguel, **International Journal of Advanced Trends in Computer Science and Engineering Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse68832019.pdf Computer Systems in Analytical Applications**, vol. 8, no. 3, pp. 195–200, 2019.
57. G. M. L. Ii, **Security Analytics : Using Deep Learning to Detect Cyber Attacks**, 2017.
58. B. Turnbull and S. Randhawa, **Automated event and social network extraction from digital evidence sources with ontological mapping**, *Digit. Investig.*, vol. 13, pp. 94–106, 2015.
59. P. Gogoi, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, **A survey of outlier detection methods in network anomaly identification**, *Comput. J.*, vol. 54, no. 4, pp. 570–588, 2011.
60. B. Dougherty, J. White, and D. C. Schmidt, **Future Generation Computer Systems**, *Inf. Syst.*, vol. 28, pp. 371–378, 2012.
61. B. Amen, **Collective Anomaly Detection Using Big Data Distributed Stream Analytics**, no. September 2018, 2019.
62. K. Grolinger, A. L'Heureux, M. A. M. Capretz, and L. Seewald, **Energy forecasting for event venues: Big data and prediction accuracy**, *Energy Build.*, vol. 112, pp. 222–233, 2016.
63. H. Mohammed, N. Clarke, and F. Li, **An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data**, *J. Digit.*, 2016.
64. R. A. De Faria, K. V. Ono Fonseca, B. Schneider, and Sing Kiong Nguang, **Collusion and Fraud Detection on Electronic Energy Meters - A Use Case of Forensics Investigation Procedures**, *2014 IEEE Secur. Priv. Work.*, pp. 65–68, 2014.

65. P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, **Big Data Analytics for Dynamic Energy Management in Smart Grids**, *Big Data Res.*, vol. 2, no. 3, pp. 94–101, 2015.
66. S. Abraham and S. Nair, **A Predictive Framework for Cyber Security Analytics using Attack Graphs**, *Int. J. Comput. Networks Commun.*, vol. 7, no. 1, pp. 1–17, 2015.
67. G. Jangla and D. A. Amne, **Big Data Analytics and Hadoop for Detecting Targeted Attacks**, *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 2257–2263, 2016.
68. F. Stouten, **Big data analytics attack detection for Critical Information Infrastructure Protection**, Luleå University of Technology, 2016.
69. J. Chris Foreman and D. Gurugubelli, **Identifying the Cyber Attack Surface of the Advanced Metering Infrastructure**, *Electr. J.*, vol. 28, no. 1, pp. 94–103, 2015.
70. A. Ukil and R. Zivanovic, **Automated analysis of power systems disturbance records: Smart Grid big data perspective**, *2014 IEEE Innov. Smart Grid Technol. - Asia, ISGT ASIA 2014*, pp. 126–131, 2014.
71. R. Mothukuri and B. B. Rao, **Data mining on prediction of crime and legal judgements: A state of an art**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 6, pp. 3670–3679, 2019.
72. J. Kawale, S. Chatterjee, A. Kumar, S. Liess, M. Steinbach, and V. Kumar, **Anomaly construction in climate data: Issues and challenges**, *NASA Conf. Intell. Data Underst.*, pp. 189–203, 2011.