Safety Analysis of AMI Networks through Smart Fraud Detection

A H M Jakaria
Department of Computer Science
Tennessee Tech University
Cookeville, USA
ajakaria42@students.tntech.edu

Mohammad Ashiqur Rahman
Department of Electrical and Computer Engg
Florida International University
Miami, USA
marahman@fiu.edu

Md Golam Moula Mehedi Hasan
Department of Computer Science
Tennessee Tech University
Cookeville, USA
mmehediha42@students.tntech.edu

***Abstract*—Advanced metering infrastructure (AMI) is a critical part of a modern smart grid that performs the bidirectional data flow of sensitive power information such as smart metering data and control commands. The real-time monitoring and control of the grid are ensured through AMI. While smart meter data helps to improve the overall performance of the grid in terms of efficient energy management, it has also made the AMI an attractive target of cyberattackers with a goal of stealing energy. This is performed through the physical or cyber tampering of the meters, as well as by manipulating the network infrastructure to alter collected data. Proper technology is required for the identification of energy fraud. In this paper, we propose a novel technique to detect fraudulent data from smart meters based on energy consumption patterns of the consumers by utilizing deep learning techniques. We also propose a method for detecting the suspicious relay nodes in the AMI infrastructure that may manipulate the data while forwarding it to the aggregators. We present the performance of our proposed technique, which shows the correctness of the models in identifying the suspicious smart meter data.**

***Index Terms*—Advanced metering infrastructure; cybersecurity; smart grid; smart meter; energy-theft detection; machine learning.**

I. Introduction

The traditional power grids have transformed into smart grids by the rapid integration of technologies. The AMI infrastructure has revolutionized the management and development of power systems by providing newer means of data exchange. Digital smart meters, which are one of the most essential parts of the AMI infrastructure, collect electricity consumption data and report it to the utility company. They can collect fine-grained consumption data, as well as report events of malfunctions, misconfigurations, and tampering. However, they have several vulnerabilities that are exploited by cyberattackers to manipulate the collected data. Several penetration tests have revealed the vulnerabilities in smart meters [1], [2]. FBI reported an organized attempt for energy theft that may have cost over 400 million dollars annually to a utility company [3].

An energy fraud can manipulate the smart meter data in two ways: they consume more energy while report less and they report more energy for particular meters to gain secondary benefits from the utility. The data manipulation is done by physical tampering such as unauthorized tapping, bypassing the meters, firmware manipulation, as well as cyberattacks against the AMI infrastructure that is utilized to transfer data. The non-technical loss (NTL) in power grids is mainly due to electricity theft fraud, billing irregularities, and unpaid bills [4]. One of the biggest challenges for any utility company is the detection and prevention of electricity energy theft.

Machine learning techniques have been widely used to analyze big data such as consumer data collected by smart meters. They have great potentials in creating models of energy consumption behavior of different customers and detecting any anomaly in the upcoming recorded data. In this work, we utilize the benefits of deep learning to model the patterns of energy usage and identify any suspicious data based on the trained models. We also propose a novel technique to identify the malicious relay nodes in the AMI infrastructure that may be involved in the manipulation of smart meter data in terms of data integrity attacks.

The main contributions of this paper include:

1) An unsupervised machine learning model to distinguish different consumer bases according to the amount of electricity usage.
2) A deep learning approach to predict the authenticity of the incoming data recorded by smart meters.
3) An algorithmic approach to identify the malicious relay nodes that may participate in data alteration because of being compromised.
4) A thorough evaluation of the trained models using real-world data in detecting fraudulent data.

The rest of this paper is organized as follows: Section II gives an overview of AMI, and how it is used, as well as its security issues and security requirements. We discuss related works in Section III. We present our threat model in Section IV. In Section V, we present an overview of the technologies to solve the security issues, as well as techniques to prevent energy theft. We present a comprehensive taxonomy of the technologies proposed in the literature. We discuss data collection and pre-processing step in Section VI. Our experimental results are presented in Section VII. We conclude the paper in Section VIII.

II. Background

This section briefly overviews different aspects of AMI. We present an overview of the users of AMI infrastructure,

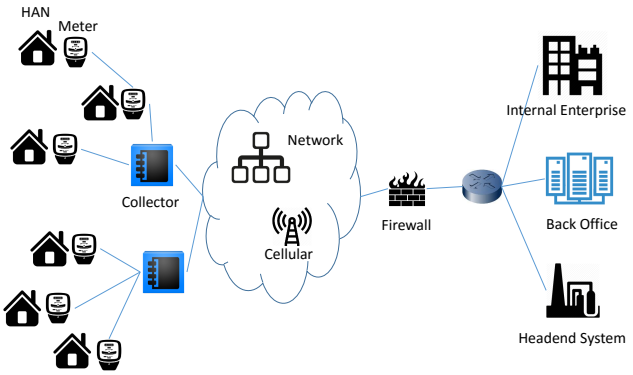

Fig. 1. A typical AMI infrastructure.

as well as the security requirements, attacker types, and threatened components of the AMI.

A. Overview of AMI

Advanced Metering Infrastructure is a network of devices that record, store, and transmit the energy usage data. It provides a suitable link between the end users and electric power utility. The AMI is an upgrade of Advanced Meter Reading (AMR) system [5], where smart meters play the most important role in collecting data [6].

B. AMI Communication Infrastructure

Fig. 1 shows a commonly used AMI communication architecture in smart grid, which is generalized from [7], [8]. AMI generally comprises the following components [9]:

- **Smart Meter:** The smart meter collects and transmits the meter consumption data periodically to the control center. This fine-grained data helps to monitor and optimize the power consumption. A smart meter generally consists of three main components: (i) a meter to record the energy generated or consumed, (ii) a computer to process and store the data temporarily, and (iii) a network device to connect to the network.
- **Gateway/Access Points:** The gateway functions as an interface between the smart meter and the control center located at the utility. It forwards the control commands sent by the control center and meter reading data collected by smart meters. It often acts as an aggregator of data when there are privacy concerns of fine-grained appliance-specific data [10].
- **Control Center:** The control center receives the real-time metering information from the network, stores it and performs data processing to generate the control commands to monitor and regulate the power generation, transmission, and distribution throughout the grid.
- **Communication Network:** The communication network can be a wide area network (WAN), neighborhood area network (NAN), or home area network (HAN), which facilitates the bidirectional communication path among the entities of AMI.

III. Related Works

The defense techniques in AMI are broadly divided into two categories: (i) passive defense, and (ii) active defense. Passive defense includes techniques such as key management between different parties in the AMI infrastructure network, secure routing protocols, secure data aggregation, secure network architecture, etc. Active defense techniques are deployed to detect attacks against the AMI infrastructure in a timely manner. Different types of intrusion detection systems (IDS) are such examples. In the literature, classification-based detection, state-based detection, and game-theory-based detection techniques have been explored.

Among all the detection techniques for energy theft, classification-based detection technique is one of the most widely used techniques. This technique is defined as the load profile classification of power consumption of a customer or a group of customers over a period of time based on the pattern of historical consumption. The basic procedure for classification-based energy-theft detection consists of different steps such as data acquisition, data pre-processing, feature extraction, model training, classification with parameter optimization, and finally, pinpointing the suspicious customer meters. The main idea of this methodology is to distinguish abnormal energy usage patterns from all energy usage patterns. Machine learning techniques are basically discussed as classification-based techniques in fraud detection in AMI.

In [11], Jumale *et al.* discussed the various machine learning algorithms and what would be more accurate in the prevention and detection of NTL with smart meter. The authors in [12] proposed a novel methodology for electricity tampering detection that uses the information obtained from the smart meter, in order to automate the detection and localization process. The smart metering infrastructure and intelligent substation are considered in the topology, which are essential for the proposed electricity theft detection.

McLaughlin *et al.* proposed both supervised and unsupervised classification techniques to create a power measurement-based anomaly detection in power usage of customers [13]. They used Naive Bayes Classification for supervised method and $k-$means algorithm for the unsupervised technique.

Joker *et al.* proposed a supervised technique based on support vector machine (SVM) classifier to detect suspicious users of electricity [14]. They utilized a combination of pattern recognition and transformer meter data to find out the suspicious meters. Depuru *et al.* [15] utilized an SVM classifier to find irregularities in customer consumption profile. Their classifier had a 98% accuracy but was based on a comparatively small training set of only 440 instances of data records. Nagi *et al.* also utilized SVM to detect fraudulent data in smart meter data [16]. They used two main features from the dataset, which were the energy consumption data and credit worthiness rating of the consumers. A distributed IDS for the smart grid was proposed in [17]. The IDS is deployed in each layer. Support vector machines (SVM) and

artificial immune algorithms are adopted for learning and classifying.

Costa *et al.* proposed an ANN-based classifier to get a 50% improvement over traditional fraud detection techniques [18]. However, none of these works perform the analysis of consumer data that has varying usage patterns during different times of the day. In our work, we not only train our classifier based on the customer type and their usage patterns during different times of the day, but also find out the suspicious network nodes that may manipulate the data while forwarding it to the data collector.

IV. Threat Model

We discuss the threat model in AMI in this section. We provide an overview of different types of attackers, their motives, and their targets in the AMI.

A. Attackers in AMI

Energy theft is one of the main motives for attacking the AMI. There are different types of attackers, such as curious and malicious eavesdroppers, intrusive data management agencies, greedy customers, and active attackers. There are three broad categories of attackers who are motivated to commit the energy theft [19]:

- Customers: Customers have been the primary adversaries with an aim to steal power. The tampering of analog meters is very common in developing countries due to the poor infrastructure, poverty, greed, and irregular metering and distributions systems. Users in developed countries steal power due to their greed, showy nature, and to hide illegal activities that utilize electricity from the grid.
- Organized crime: Professional hackers exploit the extended computing and network features in modern AMI to steal energy by creating complex software and hardware tools. They commit the crime in large scale on behalf of rogue end users who want to obtain illegal monetary benefits or interrupt the demand-response service of the utility.
- Utility company insiders: Dishonest or disgruntled employees in the utility companies may take part in the modification of data in the AMI. They may have monetary motives or simply sabotage the grid to harm the reputation of the utility company itself.

B. Targets of Threats

The main targets of attackers are the smart meters, the communication network, and the data collector. The techniques for attacking them, as well as the motivations, are discussed here in brief.

1) Smart meters: Smart meters are the most attacked components in the AMI. As smart meters are end cyber devices, users without sufficient specialized knowledge about the software and hardware properties can achieve tampering of the usage data [20]. Because of the inadequate physical tamper protections, the hackers may be able to interrupt timely collection of measurement or inject false data to the metering equipment. The hackers can capture the optical port used to communicate with smart meters and set a reader device on this port to capture the other password for other protocols after opening the meter [19].

2) Communications network: Usage data may be tampered after recording or during transmission in a smart grid by compromising the intermediate devices on the path of a smart meter data to the control center. Distributed denial of service (DDoS) attack is also common attacks against the devices in AMI, which can be launched by compromising a number of smart meters. The main purpose of DDoS attacks in AMI is to attack data collector, which prevents the normal communication between WAN and NAN [21].

3) Data collector: Data collectors may have remote disconnect functions, which can be exploited by attackers to create power outages [21]. It can be performed by installing malicious software on the data collector by using the weaknesses of network or abusing privileges by internal staff. After that, information such as IP addresses of smart meters is collected and remote disconnect command is sent to the target meters. Another way is to send the wrong command to the collectors because of wrong measurement data reported by the compromised meters.

V. Defense Technology

In this section, we discuss the methodology of the proposed fraud detection technique in AMI data. We detect the fraudulent meters that present anomalous data, as well as the network nodes that may be responsible for the data being detected as anomalous. The detection is divided into two main parts: (i) anomaly detection in the meter data using a deep learning approach, and (ii) detection of any relay node that may have participated in the alteration of the data.

A. Anomaly Detection Technique

Our anomaly detection technique consists of two stages. The first stage is an unsupervised method to cluster all meters having similar consumption patterns. Once they are clustered, a second stage of classification is applied within each cluster.

1) Unsupervised Technique: In our dataset, we have the energy consumption of users of different categories. For example, small household users have comparatively low consumption throughout the day, whereas businesses, such as corporate offices, manufacturing factories, etc. have higher consumption patterns. We first create clusters of users with similar consumption behavior. We run $k-$means algorithm on our dataset, which provides us with such clusters based on the amount of electricity consumed and the time of the data recording. We run $k-$means for different values of k, *i.e.*, different number of clusters and choose the best one based on the minimum sum of the squared distance of the data points from corresponding centroids. The sum of squared distance is calculated as follows:

$$SumOfSqDist = \sum_{m=1}^{k} \sum_{t_{m_i} \in K_m} (C_m - t_{m_i})^2$$

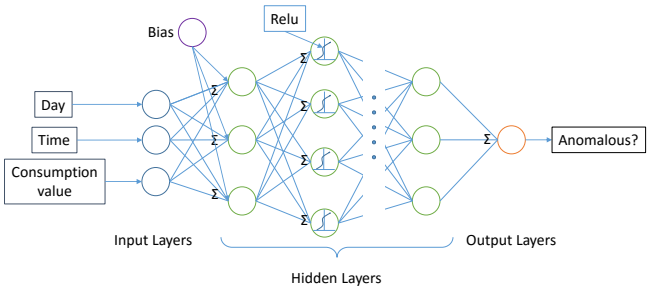

Fig. 2. Deep neural network with different layers.

Where k is the number of clusters, K_m is the set of all data points within a cluster, and C_m is the center for the corresponding cluster.

2) Supervised Technique: Within each cluster, we create a dataset for training our supervised classifier. In the dataset, we manually altered some consumption data not to match the proper consumption pattern. A 'label' attribute was introduced to the dataset, which identifies whether a data record is anomalous or legitimate.

We trained the classifier based on a multi-layer perceptron with different numbers of hidden layers with different numbers of neurons in each layer, and find the best possible model for the training purpose. We used 'relu' as the activation function, 'adam' solver for weight optimization, which is a gradient-based optimizer. The learning rate was kept 'constant', the value of alpha was set at $1e^{-4}$. We ran the model for a maximum of 200 epochs or until convergence, where in each epoch, the input samples are shuffled. In Fig. 2, we present an overview of the deep neural network, where there are multiple hidden layers. The input layer has three attributes: the day of the year, the time of the day when the data is collected, and the consumption value. The output layer defines whether a data record corresponds to normal or anomalous.

B. Suspicious Node Detection

The Next phase in our solution is the detection of suspicious nodes on the routes of the meter data. We assume that the smart meters are connected with other meters in a mesh topology, where intermediate nodes (meters) relay the data collected by its child node to the upper level, as shown in Fig. 1. The data is ultimately delivered to a local collector/aggregator, which in turn forwards the data to the control center in the utility company. Different routes may be chosen to deliver different data from a particular meter. If the intermediate nodes are compromised, they can be used to alter the legitimate meter data to launch an attack. Some malicious nodes may deliberately perform attacks on some other nodes. The compromised or malicious nodes can alter the meter data coming from other nodes. This is performed through bypassing the integrity protection schemes, if any.

Attack Model: In our attack model, we consider two strategies of a malicious node in the mesh AMI network:

1) Changing any data going through itself.
2) Changing only selective data to attack particular nodes.

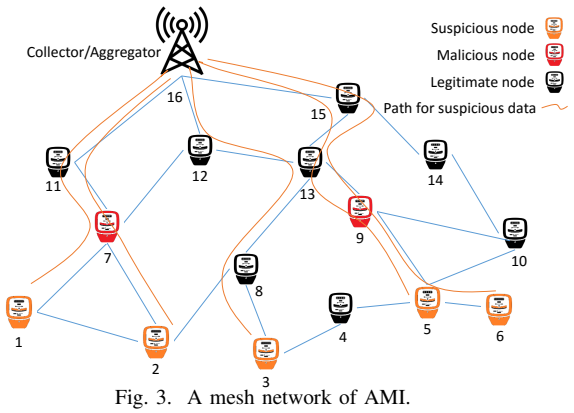

Fig. 3. A mesh network of AMI.

In the first strategy, an attacker will alter any data that is going through it to be delivered to the collector. This maximizes the goal of the attacker to create damage to the reputation of other members of the network. However, this may yield more chances of being detected as a malicious/compromised node. To prevent detection, a malicious node can alter the data incoming from certain other nodes, which is the second strategy of the attacker. We propose two different algorithms to detect the malicious nodes in both the strategies. We first collect all the data from all the meters in a particular network for a particular time period. Some of the data for a particular meter may be detected as suspicious by the smart detection system based on the fusion of supervised and unsupervised techniques discussed earlier. Fig. 3 presents a simple mesh network consisting of such meters who transmit some suspicious data towards the collector.

Algorithm 1 Suspicious Node Detection Algorithm

procedure ASSIGNSCORE()
  for each meter i **do**
    $malScore[i] \coloneqq 0$ ▷ initialize malice scores of all meters with 0.
  end for
  for each meter i **do**
    for each data $d[i]$ within time period T **do**
      $path \coloneqq$ GETPATHFROMCOLLECTOR(i, d[i])
      for each intermediate node n on $path$ **do**
        if $d[i]$ is a malicious data **then**
          $malScore[n] \coloneqq malScore[n] + 1$
        end if
      end for
    end for
  end for
end procedure

procedure ISLEGITNODE($malScore, i$)
  $probMalicious \coloneqq malScore[i]/totDataCnt[i]$
  if $probMalicious > THRESHOLD$ **then**
    return FALSE
  end if
  return TRUE
end procedure

We propose Algorithm 1 to deal with the attacks of the first type, where an attacker tries to alter any data that passes through it. The algorithm calculates the score of the nodes of being suspicious ($malScore$), based on how frequently

Algorithm 2 Suspicious Node Detection Algorithm

```
procedure CalcMalProb()
    for each meter i do
        for each other meter j do
            for all data d[j] within time period T do
                if i is on path of d[j] then
                    onPathCnt[i][j] := onPathCnt[i][j] + 1
                end if
                if d[j] is suspicious then
                    suspCount[j] := suspCount[j] + 1
                end if
            end for
            prPath[i][j] := onPathCnt[i][j]/totDatCnt[j]
            prMal[j] := suspCount[j]/totDatCnt[j]
            prMalRel[i][j] := prPath[i][j] × prMal[j]
        end for
    end for
end procedure

procedure IsLegitNode(prMalRel, i, j)
    if prMalRel[i][j] > THRESHOLD then
        return FALSE
    end if
    return TRUE
end procedure
```

they appear on the path of malicious data. We assume that the network architecture is known to the control center, as well as the path for each data up to the local collectors are known. The idea is that if a relay node appears on the paths of a number of bad data, there is a good chance for the node being compromised or malicious. The algorithm presents the methodology of pinpointing the suspicious nodes in the AMI network. If a node has more score for transmitting (relaying) malicious data than legitimate data, it shows a better chance of having been compromised and manipulating the data it is relaying. We calculate a probability, $probMalicious$, for each node based on their scores and the total number of data flowing through it ($totDataCnt$), and check it against a threshold value to mark it as legitimate or malicious.

Algorithm 2 finds out the malicious nodes in the case where an attacker targets specific nodes in the network. We calculate the count ($onPathCnt$) of a node being on the path of all other nodes. $suspCount$ is the count of a node being on paths of other nodes where the other nodes have transmitted suspicious data. We, first, calculate the probability of node i being on the path of node j ($prPath[i][j]$). Then we calculate the probability of the data of j of being malicious. Next, the probability of node i of being a malicious relay for node j's data, $prMalRel[i][j]$, is calculated. If the calculated probability is greater than a certain threshold (*e.g.*, 50%), we consider i as an attacker for node j.

Example Case Study: We implemented both the algorithms in Python and observe their performance. Table I shows an example input file for the algorithms, which corresponds to the mesh network in Fig. 3. Algorithm 1 first assigns the score of maliciousness for all the meter nodes, then tells whether a node is malicious or not based on a threshold. This algorithm works for the case where a malicious node tries to alter any data that goes through it. In our case, the algorithm returned node 7 and 9 as malicious ones. Algorithm 2, on the other hand, works in the cases where the malicious intermediate nodes target particular other nodes and alter their data. Both algorithms act fast for different network sizes. We created synthetic networks of up to 300 nodes. Both the algorithms take 1-10 ms depending on the network size.

TABLE I
SAMPLE INPUT FILE

Meter ID	Data ID	Path	Anomalous?
1	1	1 7 11 16	1
1	2	1 7 12 16	0
1	3	1 7 11 16	1
2	1	2 7 11 16	0
...	...	...	...
2	4	2 7 12 13 15 16	1
...	...	...	...

TABLE II
DATA STRUCTURE

Meter ID	Encoded Date/Time	Energy Consumption (kW-h)
1727	19548	0.140
1727	19601	0.138
...	...	...
1862	22028	1.536

VI. DATASET

In this section, we discuss the dataset collection and preparation methods in detail. First, we discuss the data collection, then we explain the data modification, and finally, we discuss the data preprocessing methods for our proposed fraud detection techniques.

A. Data Collection

We collected the electricity consumption data collected by smart meters of several households and businesses. The data is provided by the Irish Social Science Data Archive Center. We consider more than 1 million data records in building our machine learning models. Each data record has three main attributes, namely meter ID, date/time of collection, and the energy consumption data in kW-h. Each meter collects data every 30 minutes.

The first 3 digits of the date/time field denote a day starting from Jan 1, 2009. So the dataset contains data for 999 days starting from that particular day. The last 2 digits denote the time of the day when the data was collected. The values of these 2 digits vary from 01 to 48, where 01 means the data collected after the first half an hour of the day, that is, the data is collected at 00:30 hrs. 02 means the data for the second half an hour, that is, the data collected at 01:00 hrs, and so on.

Table III shows the structure of our modified dataset that we use for our unsupervised and supervised techniques. We parsed the 'Date/Time' attribute and took the first three digits and performed the modulo operation with 365 to get the day of the year. The date and time are not continuous-valued attributes, rather they are categorical values, which required one hot encoding. This essentially created several new attributes for the dataset.

TABLE III
DATA USED FOR CLASSIFICATION

Meter ID	Day of Year	Time of Day	Energy Consumption (kW-h)
1860	54	1:30 am	0.140
1860	55	1:30 am	0.138
...	...	...	...
1860	180	3:30 pm	1.536
1860	180	4:00 pm	1.742
...	...	...	...
1610	258	1:30 am	10.536
...	...	...	...
1610	265	3:30 pm	12.647
...	...	...	...

B. Data Preprocessing

As we said earlier, we scrape data for 180 days at a time for 120 customers. This provided us with a dataset of $180 \times 48 \times 120 = 1,036,800$ data records to work with. We manipulate the consumption data for our classification purpose. We assume that consumers consume more electricity from 10 am to 4 pm, which is the typical peak hour. On the other hand, electricity consumption is relatively lower between 12 am to 6 am. During the other times, consumption is average. Again, a business consumer, such as an industry, consumes much more electricity than a regular household. We adjust the data for different records manually according to our assumptions mentioned earlier.

We found any missing records corresponding to any particular time, and used the average of the preceding and succeeding record to fill in the missing value. We also consider the $z-$score of the consumption value according to the formula $x \leftarrow \frac{x-\mu}{\sigma}$ so that the variables possess approximately zero mean, which in practice, reduces computational cost while training the models.

For supervised classification, we manually modified some consumption values to create malicious entries in the dataset. We modified the existing data according to the pattern of a meter for seven continuous recorded data. In other words, to convert a particular data to malicious, it was changed to a value greater or less than the maximum or minimum values of the seven data points including the preceding and succeeding three data. We also added a 'label' attribute that signifies the legitimacy of the data. We choose 70% of the whole dataset as the training dataset and the remaining 30% as the test dataset.

C. Feature Selection

For the unsupervised clustering, we chose the time and the energy consumption value (kW-h), so that we can differentiate between different types of consumers. Each cluster gives us a group of meter IDs that share similar trends in energy consumption. Once the clustering is done, we use the data records within each cluster to train our MLP classifier separately. We use the day, time, consumption value, and the label as our training attribute.

VII. RESULTS

In this section, we present a thorough evaluation of the machine learning model that was trained with the dataset.

TABLE IV
CONFUSION MATRIX

	Predicted Negative	Predicted Positive	Total
Actual Negative	TN=231077	FP=13988	245065
Actual Positive	FN=28720	TP=26215	54935
Total	259797	40203	300000

A. Unsupervised Model Performance

With the unsupervised $k-$means technique we obtained best results in terms of the sum of squared errors of data points from their centroids for the value of $k = 7$.

B. Supervised Model Performance

$$TruePositiveRate(TPR) = \frac{TP}{TP + FN} = 0.4772$$

$$FalsePositiveRate(FPR) = \frac{FP}{FP + TN} = 0.057079$$

$$precision = \frac{TP}{TP + FP} = 0.652066$$

$$recall = TPR = 0.4772$$

$$f1 = 2 \times \frac{precision \times recall}{precision + recall} = 0.551094$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0.85764$$

We train and test our multi-layer perceptron model for each different clusters provided by the unsupervised technique. We combine the results from each of the trained models afterward. We obtain the real inspections and classified inspections by the classifier for four result types: true negative, false positive, false negative, and true positive. We show the confusion matrix for one of the clusters in Table IV. Other clusters show similar results, so we skip them here. Because the dataset is basically an unbalanced dataset, we calculate the $precision$, $recall$, and $f1$ scores for the confusion matrices. As the results show, the model has a very low false positive rate of about 5.7% and accuracy of 85.7%. The overall performance of the model is impressive.

VIII. CONCLUSION

AMI in smart grids is a very important part, which is always a target of cyber criminals for various purposes including energy theft. The Cyber defense for AMI to detect and prevent these vulnerabilities is always a challenging task. In this paper, we presented a novel technique utilizing unsupervised and supervised machine learning techniques to detect suspicious data. We also devised an algorithmic approach to find out suspicious nodes in the AMI network that may be responsible for data manipulation while forwarding energy consumption data from smart meters. Evaluation results show that the techniques perform well in detecting anomalous data, as well as malicious nodes in the network.

ACKNOWLEDGEMENT

This work was partially supported by National Science Foundation [grant number 165730].

REFERENCES

[1] "Smart meters pose security risks to consumers, utilities: Researcher," https://www.securityweek.com/smart-meters-pose-security-risks-consumers-utilities-researcher.

[2] F. Skopik, Z. Ma, T. Bleier, and H. Grüneis, "A survey on threats and vulnerabilities in smart metering infrastructures," *International Journal of Smart Grid and Clean Energy*, vol. 1, no. 1, pp. 22–28, 2012.

[3] "Fbi: Smart meter hacks likely to spread." https://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread.

[4] F. B. Lewis, "Costly throw-ups: Electricity theft and power disruptions," *The Electricity Journal*, vol. 28, no. 7, pp. 118–135, 2015.

[5] "Report to nist on smart grid interoperability standards roadmap epri," http://www.nist.gov/smartgrid/InterimSmartGridRoadmapNISTRestructure.pdf.

[6] L. Li, H. Xiaoguang, H. Jian, and H. Ketai, "Design of new architecture of amr system in smart grid," in *Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on*. IEEE, 2011, pp. 2025–2029.

[7] M. A. Rahman and E. Al-Shaer, "Formal synthesis of dependable configurations for advanced metering infrastructures," in *SmartGridComm, 2015 IEEE International Conference on*. IEEE, 2015, pp. 289–294.

[8] R. Berthier, W. H. Sanders, and H. Khurana, "Intrusion detection for advanced metering infrastructures: Requirements and architectural directions," in *SmartGridComm*. IEEE, 2010, pp. 350–355.

[9] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *IEEE Communications Surveys and tutorials*, vol. 14, no. 4, pp. 998–1010, 2012.

[10] H. Mohammed, S. Tonyali, K. Rabieh, M. Mahmoud, and K. Akkaya, "Efficient privacy-preserving data collection scheme for smart grid ami networks," in *GLOBECOM*. IEEE, 2016, pp. 1–6.

[11] P. Jumale, A. Khaire, H. Jadhawar, S. Awathare, and M. Mali, "Survey: Electricity theft detection technique."

[12] P. Kadurek, J. Blom, J. Cobben, and W. L. Kling, "Theft detection and smart metering practices and expectations in the netherlands," in *ISGT Europe)*. IEEE, 2010, pp. 1–6.

[13] 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.

[14] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in ami using customers consumption patterns," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216–226, 2016.

[15] 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*. IEEE, 2011, pp. 1–8.

[16] 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*. IEEE, 2008, pp. 1–6.

[17] Y. Zhang, L. Wang, W. Sun, R. C. Green II, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 796–808, 2011.

[18] 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.

[19] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *International Workshop on Critical Information Infrastructures Security*. Springer, 2009, pp. 176–187.

[20] I. A. Tøndel, M. G. Jaatun, and M. B. Line, "Threat modeling of ami," in *International Workshop on Critical Information Infrastructures Security*. Springer, 2012, pp. 264–275.

[21] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, W. H. Sanders, A. A. Cárdenas, and J. G. Jetcheva, "Ami threats, intrusion detection requirements and deployment recommendations," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*. IEEE, 2012, pp. 395–400.