This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2021-12660C

Topology Identification with Smart Meter Data Using Security Aware Machine Learning

Cody Francis, Vittal S. Rao
Department of Electrical and Computer Engineering
Texas Tech University
Lubbock, TX, USA
cody.francis@ttu.edu, vittal.rao@ttu.edu

Rodrigo D. Trevizan
Energy Storage Technology & Systems
Sandia National Laboratories
Albuquerque, NM, USA
rdtrevi@sandia.gov

***Abstract*—Distribution system topology identification has historically been accomplished by unencrypting the information that is received from the smart meters and then running a topology identification algorithm. Unencrypted smart meter data introduces privacy and security issues for utility companies and their customers. This paper introduces security aware machine learning algorithms to alleviate the privacy and security issues raised with unencrypted smart meter data. The security aware machine learning algorithms use the information received from the Advanced Metering Infrastructure (AMI) and identifies the distribution systems topology without unencrypting the AMI data by using fully homomorphic NTRU and CKKS encryption. The encrypted smart meter data is then used by Linear Discriminant Analysis, Convolution Neural Network, and Support Vector Machine algorithms to predict the distribution systems real time topology. This method can leverage noisy voltage magnitude readings from smart meters to accurately identify distribution system reconfiguration between radial topologies during operation under changing loads.**

Keywords—Advanced metering infrastructure, distribution system topology identification, linear discriminant analysis, security aware.

I. Introduction

The accurate description of a distribution system topology is crucial in areas of the power distribution system (PDS) operations such as distributed energy resources management, state estimation, power flow analysis, conservation voltage reduction (CVR), load management, demand response, volt/VAR optimization (VVO), to name a few [1]. Therefore, there exists an interest in performing Distribution System Topology Identification (DSTI) which is the estimation of switch statuses which allows the inference of the real time topology of the PDS.

All PDS switch statuses may not accurately be captured by the Energy Management System (EMS), which oversees the operation of a given distribution system. It is often cost-prohibitive to include real-time telemetering in all PDS switches. Alternative methods to gather switch status involve sending crews to the field, which can introduce human error.

Several DSTI methods have been proposed to overcome these limitations. State estimation-based schemes where switch positions are represented by state variables have been proposed [2][3]. A normalized residual test can be used to identify errors in switch status assumptions. However, real-time measurements are commonly scarce, therefore state estimation and DSTI cannot be realized due to lack of observability [4][5].

The increasing deployment of advanced metering infrastructure (AMI) and Phasor Measurement Units (PMUs) produce an abundance of data, thus presenting opportunities for addressing the DSTI problem using data-driven approaches. In [6][7], change in high-resolution voltage magnitudes and angles captured by PMUs were used to detect and identify the reconfiguration of PDS. Detection is achieved by monitoring changes in the norm of trend vectors, while signature matching of system reconfiguration is used for topology identification [6][7]. Deep Neural Networks trained with positive and negative sequence voltages, as well as active power injections from distributed energy resources (DER), have also been proposed for DSTI [8]. In [9], a method based on support vector machine (SVM) leveraging data from relay measurements was proposed to predict the status of distribution grid switches.

One of the downfalls of those approaches to DSTI is that they typically rely on PMU and/or SCADA measurements, which are typically in short supply in PDS. In spite of several proposed applications [10], PMU deployments in PDS are still insufficient for DSTI. Also, lack of integration between DER and EMS limits the applicability of some DSTI methods [8].

On the other hand, AMI measurements are much more abundant and they have been used for billing, low-resolution load forecasting, load management, and connection verification [11]. In [12], a method using smart meter energy usage data is employed in the context of graph theory to obtain low voltage system topology under the assumption of small to no load variation. This approach aims to provide a graph-theoretic interpretation of Principal Component Analysis and energy conservation.

Sharing AMI data with third-party application software has cybersecurity implications. Smart meter data is currently encrypted using Advanced Encryption Standard (AES), which

This work was supported by the U.S. Department of Energy, Office Electricity, Energy Storage program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy National Nuclear Security Administration under contract DE-NA-0003525.This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.
SAND 2021-XXXX X.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

is standardized by the American National Standards Institute (ANSI) under the code for electric metering ANSI C12.22-2012. The current AES is called non-homomorphic. This means that this encryption scheme does not allow for mathematical operations to be performed on the encrypted information. So the smart meter data must be unencrypted for it to be used in a machine learning algorithm [13]. Unencrypting the smart meter data can introduce unintended issues with security and privacy for both the utility company and their customers [14]. Unprotected data can allow unintended or bad actors to determine personal behavior patterns, which would allow them to target certain demographics such as latch key children or the elderly for home invasion [13].

A candidate solution to the AMI privacy problem is homomorphic encryption. This form of encryption allows performing computation on encrypted data, so data privacy is preserved. This technique has been intensely researched in the medical and financial fields to extract vital information and trends from patient medical or financial information using various machine learning (ML) algorithms while still preserving the confidentiality of the patients medical or financial information. In [15] the Cheon-Kim-Kim-Song (CKKS), also known as Homomorphic encryption for Arithmetic of Approximate Numbers (HEAAN), is used in conjunction with SVM to show that it is plausible to preserve privacy of data while using the data in a ML algorithm by applying their technique to various datasets containing sensitive data, such as medical and financial information. In [16] the authors also perform CKKS encryption but train and predict to a high degree of accuracy with the convolutional neural network (CNN) algorithm on two publicly available datasets: the Modified National Institute of Standards and Technology (MNIST) dataset, which are greyscale handwritten single digits between 0 and 9, and the Canadian Institute for Advanced Research (CIFAR-10) dataset which consists of 60,000 color images from 10 classes

In this paper, we propose a method for DSTI that uses ML algorithms associated with homomorphic encryption. This scheme introduces an additional security to data collected and used by the utility company for situational awareness. Since multiple mathematical operations will be needed to perform the LDA, CNN, and SVM algorithms for topology identification, fully homomorphic encryption schemes were chosen for this paper since somewhat homomorphic encryption (SHE) only allows a limited number of operations. Fully homomorphic encryption was also used because it supports both addition and multiplication operations unlike partially homomorphic encryption (PHE) schemes which exclusively support either addition of multiplication, but not both.

The remainder of the paper is organized as follows in Section II the DSTI problem is described in detail including the machine learning algorithms and encryptions methods employed. Section III presents the results of LDA, CNN, and SVM predictive success both with and without encryption when applied to a simulated distribution system then a comparison between the results of the topology identification algorithms. The conclusion is presented in Section IV giving the authors' final thoughts about the paper's intent.

II. Topology Identification Techniques

There are T radial and connected topologies in the PDS so predicting the transition of any two topologies $\mathbb{T}_j$ and $\mathbb{T}_k$ becomes a classification problem, where $j, k \in \{1,2, \ldots, T\}$, including $j = k$, i.e., no reconfiguration. For this classification problem class C will represent these pairs of topologies and for a system with T topologies, the problem will have $T^2 = \mathrm{C}$ classes.

The DSTI methods used in this paper utilizes time-series voltage measurements with 15-minute interval from 91 different smart meters on the distribution system to obtain the feature vectors used in the classification problems. A transitional change occurs throughout the distribution system in the voltage magnitudes when there is a topology change due to a switching event. The topology is identified by three separate algorithms LDA, CNN, and SVM with the results of the identification process compared in the results section.

A. Machine Learning Algorithms

a) Linear Discriminant Analysis (LDA): The LDA predicts the current topology of the distribution grid by projecting the trend vector into a preexisting library of vectors created using one-versus-all LDA classifiers. Each class of the problem corresponds to a distinct transition between any two radial topologies obtained by system reconfiguration by switch operations.

The training of the LDA classifier uses simulated time-series data to create a library of projection vectors used to identify the switching transition. The input to the model is the trend vectors assembled from time-series voltage magnitude measurements and then sorted into the coinciding class $\mathrm{c} \in \mathrm{S} = \{1,2,\ldots,\mathrm{C}\}$.

A one-versus-all approach was chosen for this algorithm to lower the computation expense of classification during the training and validation portions of the simulation. The one-versus-one method would require a binary classification for each possible class combination.

The linear discriminant analysis (LDA) described in [17], requires several simplifying statements the first of which is that the voltage time-series measurements are synchronized with small synchronization errors relative to the 15-minute sampling rate and that this would not greatly harm the performance of the classifier if all voltage measurements were taken either before or after the topology transition occurred. Another assumption is that the topology of the system is always radial, and all customers are being served. Therefore, the scenario of interest in this paper is when a distribution grid reconfiguration has been successfully completed. Finally, it is assumed that there is a training dataset containing voltage magnitude data that is correctly labeled for distribution topology.

b) Support Vector Machine (SVM): A multiclass nonlinear one versus all SVM was trained and used to predict the topology of a simulated distribution grid. SVM is a supervised machine learning model which requires a training portion in which the topology is already known for each set of smart meter trend vectors collected.

The features used are the trend vectors $\vec{\delta}(t)$ which the length

of corresponds to the number of smart meter readings available and are used for training to learn the values of α_i and b for each class which in this case is the different combination of switches making up the topologies of the distribution grid. To create a nonlinear SVM model a linear SVM is transformed into a dual Lagrangian problem with Karush-Kuhn-Tucker (KKT) condition with its dual quadratic optimization expressed as [15]:

$$L(\alpha) = \sum_{j=1}^{S} \alpha_j - \frac{1}{2}\sum_{j,i=0}^{S} \alpha_i\alpha_j y_i y_j K\left(\vec{\delta}_i,\vec{\delta}_j\right) \quad (1)$$

$$\text{Subject to } \alpha_i \geq 0,\ i = 1, \ldots, S \qquad \sum_{j=1}^{S} \alpha_j y_j = 0$$

Solving the above optimization problem in the training portion for each set of AMI trend vectors with corresponding known topology the coefficients α_j are obtained allowing the prediction phase which is accomplished with the decision function:

$$y_{pred} = \text{sign}\left(\sum_{j=1}^{S} \alpha_j y_j K\left(\vec{\delta}_i,\vec{\delta}_j\right) + b\right). \quad (2)$$

The predicted label of the nonlinear system is found with the equation above with S being the set of support vectors with α and b being the Lagrange multiplier found for each class. Nonlinearity of the SVM is accomplished using a polynomial kernel K shown below with d representing the degree of the polynomial and b being a constant, with a larger d making the nonlinear decision boundary more complex.

$$K\left(\vec{\delta}_i,\vec{\delta}_j\right) = \left(\vec{\delta}_i \cdot \vec{\delta}_j + b\right)^d \quad (3)$$

In the above equation the SVM kernel shows that the dot product is used between $\vec{\delta}_i$ and $\vec{\delta}_j$ to create a nonlinear classification boundary based on a polynomial. The parameter b trades off correct classification of training examples against maximization of the decision functions margin which means that for larger values of b, a smaller margin will be accepted if the decision function is better at classifying all training points correctly and a lower b will create a larger margin therefore a simpler decision function at the cost of training accuracy.

c) Convolution Neural Network (CNN): Another supervised learning algorithm used in this paper to predict distribution grid topology is the CNN. Each time series vector column of trend vectors $\vec{\delta}$ is used for the input layer for the CNN. Using the training data, the CNN performs backwards propagation which calculates error and updates the parameters of the network to maximize the output of correct predictions of topology.

The parameters that are updated through the backpropagation of training are the weights and biases, which are initialized using random values. To achieve a nonlinear boundary for classification in a CNN, a rectified linear activation (ReLU) function is used. The ReLU activation function is represented mathematically by max (0.0, x) meaning that any negative value is returned as a zero [16].

B. Encryption Methods

a) NTRU Encryption: Nth-Degree Truncated Polynomial Ring Unit (NTRU) encryption is a fully homomorphic (FHE) which supports many different operations an unlimited number of times which will be necessary to complete the LDA, CNN, and SVM algorithms while keeping the trend vector data in an encrypted state.

In NTRU encryption [18], transmitting ciphertext from the smart meter to the utility company requires the creation of a public and a private keys. The public key is known by both the smart meter and the utility company, and the private key is only known by the utility company. Two polynomials f and g with coefficients of -1, 0, and -1 and containing the highest degree N – 1 are needed to generate the key pair as in [18]. The polynomial f must be chosen so that inverses modulo q and p exist which are computed so $f \cdot f_p = 1$ (mod p) and $f \cdot f_q = 1$ (mod q) are true. When a f that is not invertible is picked by the utility company the utility company must choose another f. The utility company's private keys are f, f_p, and g, while the public key h is generated computing (4).

$$h = pf_q \cdot g \ (mod\ q) \quad (4)$$

To send encrypted AMI data to the utility, the information is put into the form of a polynomial m representing the smart meter data with coefficients in $\left[-\frac{p}{2}, \frac{p}{2}\right]$. After creating the polynomial, which is a representation of the information being sent, the smart meter randomly chooses a polynomial r with small coefficients not restricted to -1,0,1 as before. The polynomial r is meant to obscure the information being sent by the smart meter. With the utility companies public key h the encrypted message e is computed by the smart meter as shown:

$$e = r \cdot h + m \ (mod\ q). \quad (5)$$

This system encrypts the AMI data so that it can be securely sent to the utility company. As shown in [18] r must not be revealed by the AMI or utility company because anybody knowing r could compute the message m by evaluating $e - r \cdot h$. In addition to the publicly available information, the utility company knows its own private key and obtains m by first multiplying the encrypted message e and part of their private key f

$$a = f \cdot e \quad (mod\ q) \quad (6)$$

By rewriting (6) above, (7) is constructed.

$$a = pr \cdot e + f \cdot m \quad (mod\ q) \quad (7)$$

As shown in [18] coefficients of a between 0 and q-1 are chosen in the interval $\left[-\frac{q}{2}, \frac{q}{2}\right]$ to ensure that the original smart meter information is properly recovered since the smart meter chooses the number of the information m in the interval $\left[-\frac{p}{2}, \frac{p}{2}\right]$. This implies that all coefficients of $pr \cdot g + f \cdot m$ already lie within the interval $\left[-\frac{q}{2}, \frac{q}{2}\right]$ because the polynomials

r, g, f and m and prime p have coefficients that are small compared to q. Coefficients are left unaltered during reducing modulo q and the utility company recovers the original message correctly. The next step will be to calculate the modulo p of a as shown below.

$$b = a\ (mod\ p) = f \cdot m\ \ (mod\ p) \tag{8}$$

With the above equation correct because $pr \cdot g\ (\text{mod}\ p) = 0$ the utility company can use the other part of their private key f_p to recover the smart meters encrypted information by multiplication of b and f_p

$$c = f_p \cdot b = f_p \cdot f \cdot m\ (mod\ p), \tag{9}$$

so that

$$c = m\ \ (mod\ p). \tag{10}$$

With the above equation being correct because the property $f \cdot f_p = 1\ (\text{mod}\ p)$ was required for f_p.

It is important to note that only encryption carried out by the smart meter is performed in this simulation so that the trend vectors are encrypted for every time interval and then this encrypted information is sent to the utility company where the LDA, CNN, and SVM training and validation is performed on the encrypted data.

b) CKKS Encryption: CKKS encryption is also FHE, meaning it supports different mathematical operations an unlimited number of times, which will be necessary to complete the LDA, CNN, and SVM algorithm while keeping the trend vector data encrypted. CKKS also known as HEAAN (Homomorphic Encryption for Arithmetic of Approximate Numbers) was used to encrypt the smart meter trend vector data and while encrypted the LDA, CNN, and SVM algorithms was run to identify the topology of the power grid. CKKS is based on Learning-with-Errors (LWE) method which intentionally injects a small amount of random error represented by *e* into the ciphertext which renders solving/learning the secret key s_k extremely difficult.

In [15] it shows the trend vectors measured by AMI to be encrypted belong to the vector $[p]$, which are in the plaintext space $\mathcal{P}$ with the set of integers bounded by $\frac{p}{2}$ denoted by

$$[p] : = \left\{ i \in \mathbb{Z} : -\frac{p}{2} \leq i < \frac{p}{2} \right\}. \tag{11}$$

The smart meter trend vector will be the message m in plaintext and $\boldsymbol{m}$ in ciphertext. The message *m* is an element of plaintext space $m \in \mathcal{P}$. *C* is ciphertext space which **m** is an element of C so that $\boldsymbol{m} \in C$.

$\mathbb{Z}_q$is a set of integers modulo q where q = L p with L typically being a power of 2. First a p value is chosen so that $|m| < \frac{p}{2}$ and then a secret key s_k is chosen with integer vector of size N such that s_k is an element of $\mathbb{Z}_q$ with modulo q with N elements such that $s_k \in \mathbb{Z}_q^N$.

To encrypt the smart meter readings, a column vector is constructed, which is a collection of smart meter trend vectors collected from N different smart meters so that $m \in \mathcal{P} = [p]^n$ with n elements in $[p]$. Now that the trend vectors from all smart meters during the same 15-minute time interval are collected in m they are encrypted into a new random column vector, A $\in \mathbb{Z}_q^{n\ x\ N}$ a new vector e = [r] ^{n} is randomly sampled where r < L so that each component e_i satisfies $|e_i| < \frac{L}{2}$ for i = 1,…, n. Then, the trend vector is encrypted using this vector e, as shown in (12).

$$b \leftarrow (-A \cdot s_k + Lm + e)\ (mod\ q) \tag{12}$$

The formula above shows the process of retrieving a ciphertext from plaintext message column vector so that the ciphertext $\boldsymbol{m}$ is an element of the ciphertext space *C* as shown below.

$$\boldsymbol{m} = [\text{b, A}] \in C = \mathbb{Z}_q^{n\ x\ (N+1)} \tag{13}$$

To decrypt the trend vectors of the smart meters which have been arranged in a column vector that corresponds to a certain 15-minute interval the formula below is applied to $\boldsymbol{b}$

$$\left[\frac{(m \cdot s)\ mod\ q}{L}\right] = \left[\frac{Lm\ +\ e}{L}\right] \rightarrow m \tag{14}$$

III. Case Study

To validate the performance of the DSTI algorithms and FHE presented in Section II, several simulations were performed using an IEEE 123-bus system., The results of those simulations are presented below. To gauge the effectiveness of the FHE topology identification algorithm presented in this paper it the results are compared with the DSTI algorithm using unencrypted data with different noise levels added to the voltage time-series data.

A) Description of Distributed Power System

The IEEE 123-bus system used in the simulation has 8 switches with their locations is shown in Fig. 1. This test feeder was used for training and validation of the proposed algorithms for DSTI without encryption and DSTI with encryption. This system has 5 radial topologies, therefore there exist 20 possible topology transitions. Additionally, there are 5 classes that represent unchanged topology between measurement scans. In total there are 25 possible switch transitions that represent the classes for the DSTI problem.

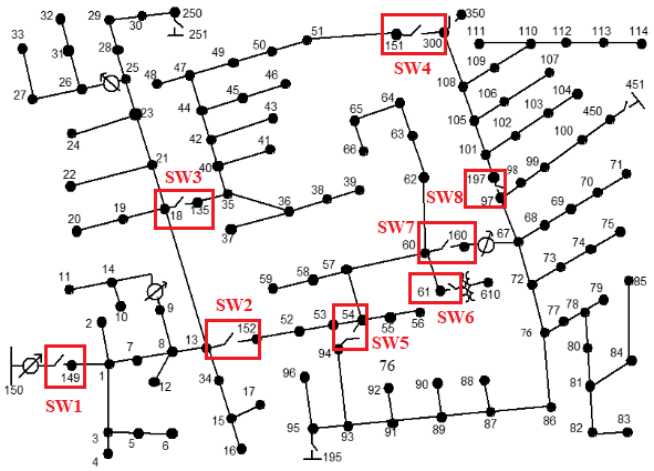

Fig. 1. IEEE 123-bus test feeder showing its 8 switches [20].

A combination of the programs Open DSS Version 7.6.5.91, MATLAB 2020a, and a MATLAB Toolbox GridPV Version 2.2, were used to generate the synthetic data and simulate the grid operations, construct the library, and validate its successful predictions for the DSTI algorithm with and without FHE. The yearly sequential time simulation of the distribution system is generated using OpenDSS and the classification algorithms are implemented in MATLAB using GridPV as an interface.

The sequential time simulation is yearlong to show the DSTI is robust not only to load variations throughout a day but also seasonal load variations. For each time series, a random switching event can occur programmed randomly, and the voltage measurements collected are used to either train the algorithm or predict the current topology of the distribution system. The training of the one-versus-all LDA, SVM, and CNN classifiers were constructed using 4 months, January through April (11,519 trend vectors), of voltage magnitude measurements collected at 15-minute intervals from the 91 smart meters of the test system that contained a load. During the 4-month training, a total of 4,000 switching events occurs randomly spanning the 20 classes that are actual switching events. The remaining 8 months of data were used for validation (23,520 trend vectors) during which 150 switching events occur randomly also spanning the 20 classes that are actual switching events.

Varying amounts of noise was added to the voltage magnitude measurements received from the smart meters to determine the robustness of the algorithms to noise. The noise is added to the voltage magnitude before any encryption of the smart meter data is completed with four different variations of Gaussian noise added 0%, 1%, 2%, 3%.

TABLE I. LDA UNDER DIFFERENT Encryption and Noise SCENARIOS

Noise Level	LDA Without Encryption			
	Correct	*Wrong*	*% Correct*	*False Positive*
None	150	0	100	0
0.01	150	0	100	0
0.02	149	1	99.3	5
0.03	145	5	96.7	8
	LDA With NTRU Encryption			
None	47	103	31.3	35
0.01	55	95	36.7	52
0.02	35	115	23.3	73
0.03	30	120	20.0	113
	LDA With CKKS Encryption			
None	128	22	85.3	5
0.01	102	48	68.0	15
0.02	83	67	55.3	53
0.03	88	62	58.7	89

B) Summary of Results

This simulation includes 150 switch changes randomly dispersed over a nine-month period and the 20 classes representing an actual topology transition.

In this simulation a false positive is considered a prediction in the validation portion that there has been a topology change when there in fact was no topology change. The results of the simulation are shown below in Tables I, II, and III.

In the simulation the standard deviation of Gaussian noise added to the voltage magnitude measurements is equivalent to 0, 1, 2, and 3 percent of the voltage magnitudes measured. As can be seen from TABLE I. both the LDA without encryption and the LDA with encryption did both exhibit a decrease in accuracy prediction as well as an increase in false positive as the noise in the voltage measurements was increased from zero to 3 percent standard deviation of the noise.

TABLE II. CNN UNDER DIFFERENT ENCRYPTION AND NOISE SCENARIOS

Noise Level	CNN Without Encryption			
	Correct	*Wrong*	*% Correct*	*False Positive*
None	150	0	100	0
0.01	148	2	98.7	2
0.02	148	2	98.7	2
0.03	146	4	97.3	5
	CNN With NTRU Encryption			
None	147	3	98.0	3
0.01	147	3	98.0	5
0.02	145	5	96.7	12
0.03	143	7	95.3	15
	CNN With CKKS Encryption			
None	149	1	99.3	0
0.01	149	1	99.3	3
0.02	146	4	97.3	3
0.03	145	5	96.7	8

TABLE III. SVM UNDER DIFFERENT ENCRYPTION AND NOISE SCENARIOS

Noise Level	Nonlinear SVM Without Encryption			
	Correct	*Wrong*	*% Correct*	*False Positive*

None	147	3	98.0	0
0.01	146	4	97.3	0
0.02	144	6	96.0	1
0.03	142	8	94.7	5
	Nonlinear SVM With NTRU Encryption			
None	142	8	94.7	3
0.01	142	8	94.7	5
0.02	135	15	90.0	16
0.03	138	12	92.0	12
	Nonlinear SVM With CKKS Encryption			
None	145	5	96.7	0
0.01	143	7	95.3	1
0.02	143	7	95.3	8
0.03	136	14	90.7	12

These simulation results show a lower accuracy for the correct transition classification when NTRU encryption is employed, and this low accuracy seems to be further degraded by the addition of Gaussian noise to the smart meter measurements.

Another reason for the loss of accuracy for the LDA, SVM, and CNN when the data is encrypted is due to the noise that is introduced by the different arithmetic operations that are performed on the data [19]. Every time a mathematical operation is performed additional noise is added to the ciphertexts this noise is cumulative and grows with each mathematical operation performed on the encrypted data. Typically noise growth from addition of homomorphically encrypted data is modeled by the formula below.

$$\text{Noise}\,(a' + b') = \text{Noise}\,(a') + \text{Noise}\,(b') \tag{15}$$

Noise growth from performing multiplication on homomorphic encrypted data is modeled by the formula below.

$$\text{Noise}\,(a' * b') = \text{Noise}\,(a') * \text{Noise}\,(b') \tag{16}$$

When (15) and (16) are compared the multiplicative operation shows a much larger buildup of noise compared to the addition operation.

IV. Conclusion

A DSTI method by using a one-versus-all LDA, SVM, and CNN classifiers using only 15-minute voltage magnitude time series data as input has been presented in this paper and was compared to the same classifiers with a FHE added to the data before being used by the classifiers. The classifiers without encryption appeared to function with few errors predicted of the transitions between radial topologies in the IEEE 123-Bus test system over an 8-month testing period with a prior 4-month training period even with the addition of varying noise added to the voltage magnitude measurements. These methods have shown robustness to load variations and noise in measurements. When plaintext data is used as the input to the classifiers, LDA-based method has shown results that are marginally superior to those of CNN and SVM. The LDA classifier with encryption does not seem to possess an accuracy that would promising for use of this algorithm in real world scenarios, with this accuracy being even further degraded by increasing measurement noise.

The two simulations with noise added to the smart meter measurements and no noise added allow us to gauge the deterioration in the predictive ability of the topology identification algorithms with and without encryption. Even though the performance of the algorithm without encryption is degraded by noise, these methods would still have viable real-world application for the identification of the distribution grid topology since the performance remains high and more accurate meters could be used. These proposed methods do not require precise knowledge of distribution system parameters but do require a dataset of voltage magnitudes correctly labeled for each topology.

In future research the use of different fully homomorphic encryptions to gauge their effect on the algorithms predictive accuracy may give rise to an encryption scheme with an accuracy high enough for real world deployment.

Acknowledgment

This material is based upon work supported by the fellowship Graduate Assistance in Areas of National Need (GAANN). The authors would like to thank Dr. Imre Gyuk, Director of the Energy Storage Program, for his continued support. The authorsalso like to thank Drs. James Obert and Alvaro Furlani Bastos from Sandia for their technical suggestions.

References

[1] L. Blakely, M.J. Reno, and J. Peppanen, “Identifying Common Errors in Distribution System Models,” *IEEE Photovoltaic Specialists Conf. (PVSC)*, 2019, pp. 3132-3139.

[2] G. N. Korres and N. M. Manousakis, "A state estimation algorithm for monitoring topology changes in distribution systems," *2012 IEEE Power and Energy Society General Meeting*, San Diego, CA, 2012, pp. 1-8.

[3] M.E. Baran, J. Jung and T.E. McDermott, "Topology error identification using branch current state estimation for distribution systems," in *Proc. 2009 T&D Conf. & Expo.: Asia and Pacific*, Seoul, 2009, pp. 1-4.

[4] D. Deka, S. Backhaus and M. Chertkov, "Learning topology of the power distribution grid with and without missing data," 2016 European Control Conf. (ECC), Aalborg, Denmark, 2016, pp. 313-320.

[5] D.Deka, M.Chertkov, “Learning topology of distribution grids using only terminal node measurements,” in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Nov. 2016, pp. 205-211.

[6] G. Cavraro, R. Arghandeh, G. Barchi and A. von Meier, "Distribution network topology detection with time-series measurements," *IEEE PES Innovative Smart Grid Tech. Conf. (ISGT)*, Washington, DC, 2015, pp. 1-5.

[7] G. Cavraro, R. Arghandeh, K. Poolla and A. von Meier, "Data-driven approach for distribution network topology detection," in *Proc. 2015 IEEE Power & Energy Society General Meeting*, Denver, CO, 2015, pp. 1-5.

[8] M. Jafarian, A. Soroudi and A. Keane, "Distribution System Topology Identification for DER Management Systems Using Deep Neural Networks," *in Proc. IEEE PES General Meeting*, Montreal, QC, Canada, 2020, pp. 1-5.

[9] B. Poudel, D. Ruiz Garcia, A. Bidram, M. J. Reno, and A. Summers, “Circuit Topology Estimation in an Adaptive Protection System,” *IEEE North American Power Symp. (NAPS)*, 2021.

[10] A. von Meier, E. Stewart, A. McEachern, M. Andersen and L. Mehrmanesh, "Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications," in *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2926-2936, Nov. 2017.

[11] Y. Wang, Q. Chen, T. Hong and C. Kang, "Review of Smart Meter Data

Analytics: Applications, Methodologies, and Challenges," in *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125-3148, May 2019.

[12] S. J. Pappu, N. Bhatt, R. Pasumarthy and A. Rajeswaran, "Identifying Topology of Low Voltage Distribution Networks Based on Smart Meter Data," in IEEE Tran. Smart Grid, vol. 9, no. 5, pp. 5113-5122, Sept. 2018.

[13] T. Graepel, K. Lauter and M. Naehrig,"ML confidential: Machine learning on encrypted data," in *Proc. Int. Conf. Inform. Security and Cryptology*, 2012. pp. 1-21.

[14] S. Finster and I. Baumgart, "Privacy-Aware Smart Metering: A Survey," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 2, pp. 1088-1101, 2015.

[15] S. Park, J. Byun, J. Lee, J. H. Cheon and J. Lee, "HE-Friendly Algorithm for Privacy-Preserving SVM Training," in *IEEE Access*, vol. 8, pp. 57414-57425, 2020.

[16] T. Ishiyama, T. Suzuki and H. Yamana, "Highly Accurate CNN Inference Using Approximate Activation Functions over Homomorphic Encryption," *2020 IEEE Int. Conf. on Big Data (Big Data)*, 2020, pp. 3989-3995.

[17] C. Francis, V. Rao, R. D. Trevizan and M. J. Reno, "Topology Identification of Power Distribution Systems Using Time Series of Voltage Measurements," *IEEE Power & Energy Conf. at Illinois (PECI)*, 2021, pp. 1-7.

[18] J. Pipher. Lectures on the NTRU encryption algorithm and digital signature scheme: Grenoble, June 2002.

[19] P. Rama,"Exploring the Effectiveness of Privacy Preserving Classification in Convolutional Neural Networks," M.S. Thesis, Dept. Comp. Eng., Rochester Inst. of Tech, Rochester, NY, 2019.

[20] W. H. Kersting, "Radial distribution test feeders," IEEE Trans. Power Syst., vol. 6, no. 3, pp. 975-985, Aug. 1991.