

Spatio-Temporal Characterization of Synchrophasor Data Against Spoofing Attacks in Smart Grids

Yi Cui, *Member, IEEE*, Feifei Bai, *Member, IEEE*, Yilu Liu, *Fellow IEEE*, Peter Fuhr, *Senior Member, IEEE*, Marissa Morales-Rodriguez, *Member, IEEE*

Abstract—“Source ID Mix” has emerged as a new type of highly deceiving attack which can alter the source information of synchrophasor data measured by multiple phasor measurement units (PMUs), thereby paralyzing many wide-area measurement systems (WAMS) applications. To address such sophisticated cyber attacks, we have exploited the spatio-temporal characteristics of synchrophasor data for authenticating measurements’ source information. Specifically, the source authentication is performed when the measurements are subjected to three types of spoofing attacks. Some practical difficulties in applying the proposed method on real-time authentication caused by insufficient measurement data have also been solved. Experimental results with real synchrophasor measurements have validated the effectiveness of the proposed method in detecting such complicated data spoofing and improving power systems’ cyber security.

Index Terms—Cyber security, machine learning, phasor measurement unit (PMU), spoofing attack, wide-area measurement systems (WAMS).

I. INTRODUCTION

Wide-area monitoring systems (WAMS) have been supporting the smart grids by providing a vastly critical information and communication functionalities which enable power administrators to sense, monitor, and manage the power flows throughout the network. While the accelerated cyber physical integration may enhance the control performance and communication efficiency of the power network, it also brings about system vulnerabilities to potential cyber attacks, such as denial of service (DoS) [1] and synchronization spoofing [2], etc. Therefore, extensive concerns have arisen among the service providers to address various malicious deceptions on WAMS data and enhance the cyber security of power infrastructures [3].

While there is a seemingly onslaught of cyber attacks occurring on digital systems, of particular note to cyber and power engineering research activities is malware termed “Source ID Mix”. “Source ID Mix” attacks represent a new type of complex, highly-deceiving false data injection attack (FDIA). In the power sector, such attacks can exchange the source information of measurement data among different PMUs without altering the measurement values. This results in placing the measured data into wrong positions in associated data servers. This, in turn, causes the WAMS-based applications to, at best, act confused due to this changing set of measurements that are interpreted as originating from incorrect grid locations. A worse case arises when this incorrect geo/grid location causes the WAMS to become simply paralyzed due to the incongruity of the measurements [4].

Cyber attacks on PMUs’ location (source) information can be achieved through different ways. One such way is via Global Position System (GPS) spoofing where artificial GPS signals are produced by the adversary, resulting in erroneous location information being associated with the PMUs due to the PMUs’

commercial-grade GPS receivers failing to distinguish the spoofed from legitimate GPS signals. While a simple correction may entail incorporation of a PMU-based GPS encryption and/or authorization procedure [5]. Another form of source information spoofing may occur in the communications between PMUs and the associated data server. Such a spoofing attack is possible because while most PMUs utilize the IEEE C.37.118 standard to define the data frame. However, this standard does not include security mechanisms in its data specifications, thereby making it vulnerable to cyber attacks [6-7]. In the presented paper, we consider the communication medium of PMU data is TCP/IP protocol [8]. However, if the communication medium is direct point-to-point fibres, such data spoofing attacks could be mitigated [9].

A. Impact of “Source ID Mix” Spoofing on Frequency Data

Two case studies are presented to demonstrate the impacts of “Source ID Mix” spoofing on the WAMS-based applications especially when such attack happens in the frequency data.

Case study 1: Effect on electromechanical disturbance localization. In this case study, the disturbance is initiated by a generation trip in the Eastern Interconnection (EI) in North America. Fig. 1(a) shows 20-second frequency signals measured by four relevant PMUs (PMU1 to PMU4 in Fig. 1(b)) which are close to the distribution location. It is assumed that the “Source ID Mix” data spoofing happens on PMU1 and PMU4 by swapping the frequency signals from these two PMUs. The disturbance location is estimated by the wave-front arrival time (WAT)-based method in [10]. Fig. 1(b) shows the confirmed disturbance location (denoted by black star) provided by North American Electric Reliability Corporation (NERC). The estimated disturbance locations using frequency signals before and after the data spoofing are also presented. As can be seen from Fig. 1(b), the WAT-based method can correctly estimate the disturbance location (denoted as red circle) using non-spoofed frequency signals. However, the estimated disturbance location (denoted as blue circle) is 100 miles far away from the confirmed location once the frequency signals are subjected to the “Source ID Mix” data spoofing.

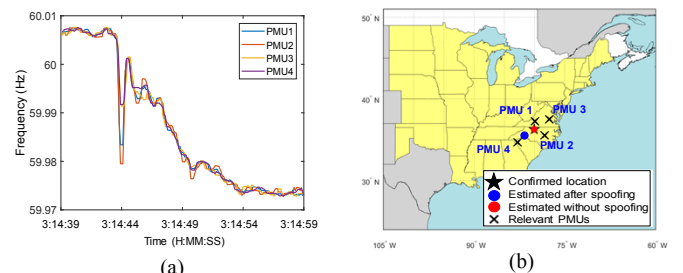


Fig. 1 (a) Frequency measurements from four PMUs, (b) confirmed and estimated disturbance location before and after spoofing on frequency data.

Case study 2: Effect on wide-area damping control. In order to demonstrate the impact of the data spoofing of frequency measurements on wide-area damping control, the classical four-machine two-area system is utilized as shown in Fig. 2. The control actuation location is at G1. The parameters of the system and the controller are described in [11].

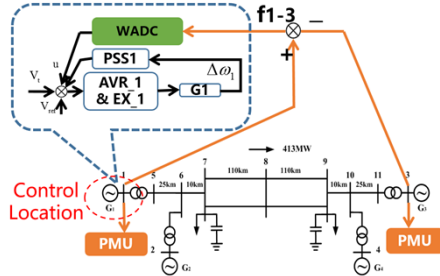


Fig. 2 Four-machine two-area system with damping control.

In this case study, the disturbance is the 65% generation trip (happens at 1s) at G4 and the frequency responses of four generators after this disturbance are shown in Fig. 3(a). The difference of the frequency measured at Bus-1 and Bus-3 is used as the input of the controller. It assumes that ‘‘Source ID Mix’’ data spoofing happens at Bus-1 and Bus-3 by swapping the frequency signals of these two buses. Fig. 3(b) compares the control performance (frequency deviation in per unit value) before and after data spoofing of measured frequency signals. From Fig. 3(b) it is observed that the controller makes frequency deviation between G1 and G3 fast damped (blue curve) using non-spoofed input frequency signals while the frequency deviation increases (red curve) once the input frequency signal of the controller is spoofed. This has demonstrated that the ‘‘Source ID Mix’’ data spoofing on the frequency signals has significant effect on the performance of wide-area damping controller.

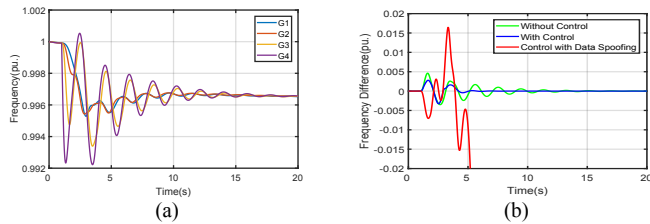


Fig. 3 (a) Frequency of each generator after disturbance, (b) damping control performance comparison.

B. Related Work

Extensive studies have been performed to detect the spoofing attacks on PMUs’ data as shown in TABLE I. It summarizes the

most related works on PMU data spoofing detection and performs a qualitative comparison among different methodologies with respect to four properties, i.e., (1) Source localization (SL): This feature indicates the ability of the methodology to identify the locations of the spoofed PMUs; (2) Cope with large networks (CLN): This index indicates the capability in dealing with large networks without computational difficulties or complexity; (3) Cope with topology change (CTC): This property evaluates the ability to detect attacks after system topologies change. (4) Requirement of external source (RES): This index indicates whether external data sources are required for a proper functionality. From TABLE I it is observed that the presented methods can be roughly classified into two types, i.e., state estimation based methods (SEM) and correlation coefficient based methods (CCM). One common flaw of SEM is that they are exclusively applicable to a particular system configuration, which means the influence of network topologies changes on the spoofing detection are not considered. On the other hand, such SEM may not be suitable to detect the specific ‘‘Source ID Mix’’ attacks on the frequency data, as most published state estimation-based algorithms use the voltage magnitude, phase angle, active and reactive power of buses for detection while the frequency signals measured by different PMUs can be replaced although the bus voltage and power flows still maintain basic power principles, which may not produce outliers in the estimation results. In addition, the correlation coefficient based detection techniques may also fail in detecting the ‘‘Source ID Mix’’ data spoofing since the spoofed data are still within an acceptable (measurement) range and highly correlated with each other.

C. Summary of Contributions

To overcome the above limitations, this paper proposed a machine learning-based measurement source authentication framework by exploiting the spatio-temporal characteristics of frequency signals recorded by WAMS. Specifically, a Time-Frequency (TF) sparsity mapping is proposed to extract the distinctive spatio-temporal characteristics from the frequency measurements at different locations. Then the extracted spatio-temporal characteristics are integrated with advanced multi-Grained Cascade Forest (gcForest) algorithm [12] for source information authentication. The proposed approach has been firmly verified by using real PMU data subjected to a variety of ‘‘Source ID Mix’’ spoofing scenarios. Moreover, potential solutions to practical difficulties in real-time source authentication caused by insufficient measurement data are also presented.

TABLE I
COMPARISON OF THE STATE OF THE ART METHODOLOGIES FOR PMU DATA SPOOFING DETECTION

Type	Ref No.	Spoofing detection mechanism	SL	CLN	CTC	RES
State Estimation	[13]	Kullback–Leibler distance (KLD) of the spoofed data will show a significant increase compared with normal conditions. Certain threshold is required based on empirical knowledge from the experts.	Yes	Yes	No	No
	[14-15]	Supervised machine learning algorithms are trained for data spoofing detection, including k-Nearest Neighbor (kNN) classifier, multi-layer neural networks (NN) and support vector machines (SVMs).	Yes	Yes	No	No
	[16-17]	Discrepancies between the Markov graph of the phase angle are used to achieve spoofing detection.	Yes	Yes	No	No
	[18]	Kalman filter is used for deception detection. The identification accuracy is highly dependent on the predefined threshold of Euclidean detector.	Yes	Yes	No	No
	[19]	Online information (such as load prediction, power grid operation schedules, and real-time measurements) apart from traditional SCADA systems is required to identify anomalies.	Yes	Sort of	Yes	Yes
Correlation Coefficient	[20]	Correlation coefficient between the spoofed signals and normal measurement data becomes lower within a specific time interval, but it cannot handle ‘‘Source ID Mix’’ data spoofing.	Yes	Yes	Yes	No

The findings of the investigation improve the understanding of whether the power frequency data measured

by WAMS embrace distinctive spatio-temporal signatures. From a cybersecurity-grid resiliency perspective, this determination leads to the possibility of potentially further malicious exploitation regarding manipulation of accurate source information authentication. In addition, the investigation also plays a fundamental role in developing solutions to more difficult problems associated with mitigating “Source ID Mix” spoofing on PMU data and improving the cyber security of power systems.

II. LOCATION-DEPENDENT FREQUENCY SIGNALS DATASET

In this study, we particularly focus on the frequency measurements from the EI power grid. All the frequency signals for source authentication experiments are retrieved from FNET/GridEye, which is a GPS-synchronized wide-area frequency monitoring system mainly deployed at the distribution level of the power grid [8]. For now, 233 PMUs have been installed within the EI to continuously record power network information (i.e., frequency, voltage and phase angle) and achieve a real-time condition monitoring of the power grid. In this paper, frequency signals (reporting rate is 10Hz) from 12 PMUs at different intra-grid locations within the EI (as shown in Fig. 4(a)) are collected. The collected signals cover frequency variations within the summer season (June, July and August in 2017) over broad geographical areas. Specifically, the three-month period is first divide into a series of non-overlapping time windows with 10-minute length. Then frequency signals of the above 12 locations in the EI within randomly selected 1000 time windows are collected to construct a location-dependent dataset.

An example of recorded frequency segments at three cities within the EI is shown in Fig. 4(b), where major difference is observed from the variations of the frequency signals. It appears smaller variations are observed in the frequency signal measured at generation/load centers (high inertia). The frequency fluctuations at E1 and E2 seem to be the most controlled, with frequency signals exhibiting high similarity in the manner of their variations. Conversely, signals from E10 appear to drift more before returning to the nominal value. It is speculated that this variation is due to different local power grids which may have different control mechanisms and power supply capabilities interacting. The net result being impacting the effectiveness and manners in which they are controlling the frequency variations. Considering the randomness of local grid condition variations and the uniqueness of local grid characteristics, such variations in the frequency signals may possess unique spatio-temporal characteristics that can be used for measurement source authentication.

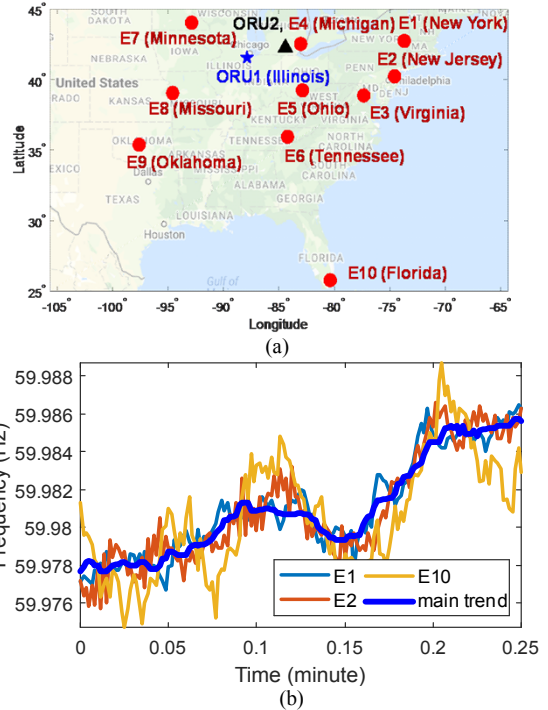


Fig. 4 (a) Locations of frequency measurements collected within the EI and (b) example of frequency signals recorded by three PMUs.

III. PROPOSED MEASUREMENT AUTHENTICATION APPROACH

Based on the above discussions, “Source ID Mix” attacks on PMU frequency data can be formulated as follows: Let column vector $S_i = [s_{i,1}, s_{i,2}, \dots, s_{i,m}]^T$ be the time series frequency segment measured by i -th the PMU from time 1 to time n , where $1 \leq i \leq m$. Let $S = [S_1, S_2, \dots, S_m]$ be the $n \times m$ matrix representing frequency segments measured by all the PMUs in the system, from time 1 to time n . Given an arbitrary frequency segment S_k , where $1 \leq k \leq m$, it is considered as being subjected to the “Source ID Mix” data spoofing when all or only a portion of data of S_k within a specific time range have been replaced by the data of S_i ($i \neq k$) during the same time interval.

In order to detect “Source ID Mix” spoofing attacks on PMU data, a measurement source authentication approach is proposed in this section by exploring the spatio-temporal characteristics of the frequency signals. It mainly contains four steps as follows:

- 1) Extract the variations of each measured frequency signal through a weighted high pass filter - Eq. (1);
- 2) Implement Mathematical Morphology (MM) [21] to decompose the extracted frequency variations in 1) into multiple levels and obtain the intrinsic mode function (IMF) at each level (total 30 levels are used) - Eq. (2) to Eq. (5);
- 3) For each decomposed IMF, the sparsity value is calculated (Eq. 6) and then a sparsity trend is constructed by combining the sparsity values from decomposed IMFs over 30 levels. Further, the roughness index of each frequency variations is calculated (Eq. 7). Both sparsity trend and roughness index are used as unique spatio-temporal signatures of the measured frequency signal;
- 4) The extracted spatio-temporal signatures extracted from non-spoofed frequency segments are used to train the gcForest [12]

algorithm to learn the underlying relationship between the informative features and each measurement location. Then the above trained gcForest algorithm is subsequently used to make identification on the source information of the frequency signal of interest. Finally, the source authentication results (authenticated/spoofed) are provided as the output of the proposed approach.

The flowchart of the described methodology and the corresponding schematic diagram are shown in Fig. 5. Moreover, each step in the above source authentication will be elaborated in the remaining contexts of Section III.

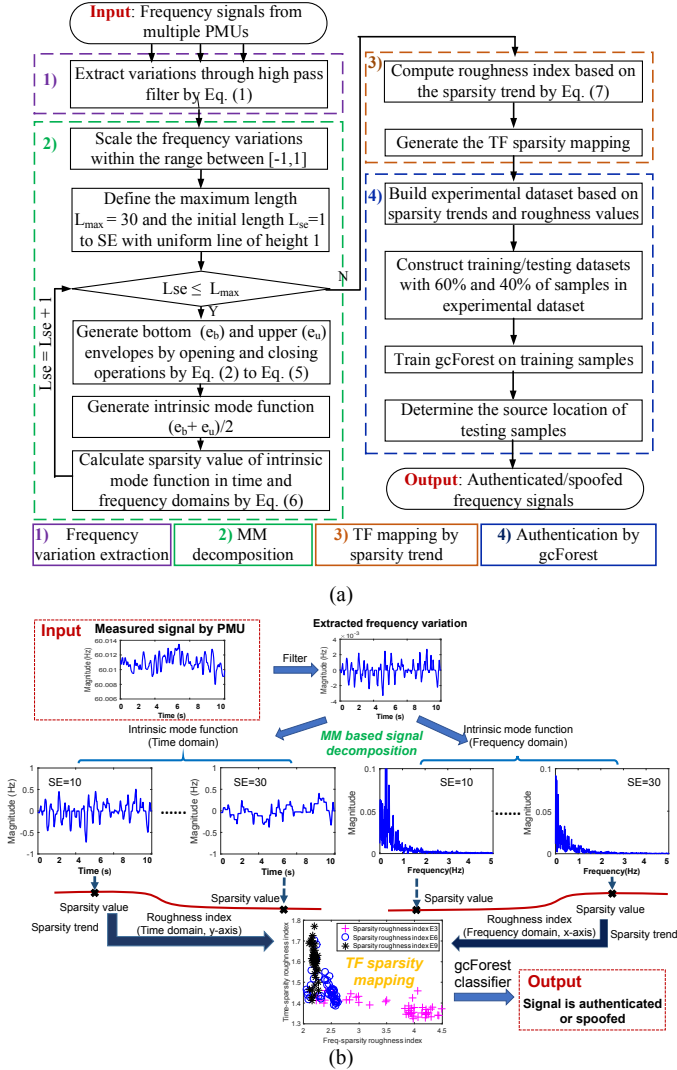


Fig. 5 (a) Flowchart and (b) schematic diagram of the proposed source authentication methodology.

A. Extracting Frequency Variation

To identify the source locations of frequency segments, necessary data pre-treatments, such as checking data continuity and eliminating measurement outliers, are initially conducted on the raw data in the location-dependent frequency dataset. Then each frequency segment is smoothed by a specifically designed high pass filter (HPF). Finally, such smoothed segment is removed from the raw frequency segment to capture the frequency variations as (1).

$$S_{hp}^n(t) = S^n(t) - \sum_{i=-(K-1)/2}^{(K-1)/2} \omega_i S^n(t-i) \quad (1)$$

where $S^n(t)$ and $S_{hp}^n(t)$ are the t -th sample in the raw frequency segment and extracted frequency variations measured by n -th PMU, ω_i is the weight of HPF. It is defined as the convolution of the vector $\begin{bmatrix} 0.5 & 0.444 & 0.4 & 0.333 & 0.25 \\ 0.444 & 0.4 & 0.333 & 0.25 & 0.167 \end{bmatrix}$. K denotes the order (odd number) of the HPF.

B. Mathematical Morphology(MM)-Based Decomposition

Followed by the frequency variation extraction, MM is subsequently employed to decompose the extracted frequency variations into a series of IMFs at multiple time and frequency scales, where each IMF reveals the underlying nonlinearity and non-stationarity characteristics of the original signal. The MM performs the signal decomposition by conducting various integral geometry operations between a structural element (SE) $P(n)$ and the signal $S(n)$ [21]. The SE is a pre-defined geometric shape (such as uniform, sinusoidal, triangular, etc.) with finite length, by which the signal decomposition is performed with the MM. As for the MM-based method, researchers proved that compared with the length of structure elements (SE), the shapes of SE do not affect signal analysis much [22]. For the application of MM on frequency measurements, uniform line shape SE is used due to its simplicity in processing without considering its amplitude. Dilation (\oplus) and erosion (\ominus) are two preliminary operations of MM, which are mathematically defined as (2) and (3).

$$S(n) \oplus P(n) = \max[S(n-m) + P(m)], n-m \geq 0, m \geq 0 \quad (2)$$

$$S(n) \ominus P(n) = \min[S(n+m) - P(m)], n+m \geq 0, m \geq 0 \quad (3)$$

Based on (2) and (3), opening (\circ) and closing (\bullet) provide other two combined operations, which are shown as (4) - (5).

$$S(n) \circ P(n) = [S(n) \ominus P(n)] \oplus P(n) \quad (4)$$

$$S(n) \bullet P(n) = [S(n) \oplus P(n)] \ominus P(n) \quad (5)$$

TABLE II summarizes the properties of MM operator to the frequency variation signals. Fig. 6 shows an example of the signal processing by four operations of MM with uniform line SE. The length of SE is set to two with uniform height of one. It is clear that among the four operations, closing and opening can preserve positive and negative peaks of original signals respectively (shown in Fig. 6(c) and Fig. 6(d)). The magnitudes of positive and negative peaks of the original signals remain unchanged after they are executed by closing and opening operations. In this paper, closing and opening operations are employed for analyzing the peak distribution of measured frequency signals.

TABLE II
PROPERTIES OF THE MM OPERATOR TO FREQUENCY VARIATION

MM operator	Positive peak	Negative peak
Dilation	Smoothing	Reducing
Erosion	Reducing	Smoothing
Opening	Reducing	Preserving
Closing	Preserving	Reducing

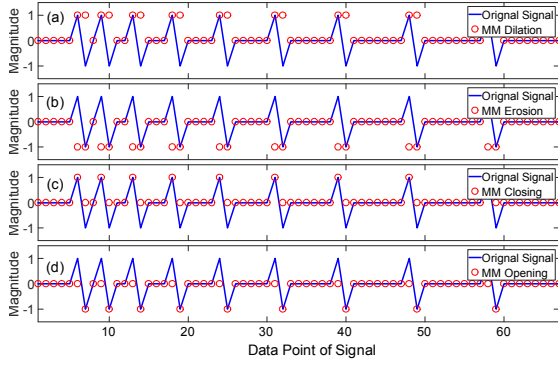


Fig. 6 Original (blue line) and processed (red dots) signals by four operations of mathematical morphology, (a) dilation, (b) erosion, (c) closing, (d) opening.

The length of the SE also affects the MM signal processing. Fig. 7 shows the processed signals by closing and opening of MM using uniform line SE with variable length (i.e., 1, 4, 6, and 8 data points). The height of SE is one. It can be seen from Fig. 7 that closing and opening of MM can cover as many as positive and negative peaks only when the length of SE is larger than the interval (number of data points marked in Fig. 7a) of two adjacent peaks in the original signal. When the length of SE increases, more peaks from the original signal can be covered by the upper envelop (marked as red cross) and lower envelop (marked as black circle), which can decompose the signals into multiple time and frequency resolutions.

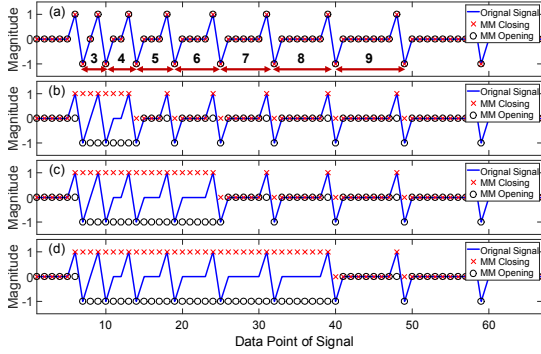


Fig. 7 Original signals (blue line) and processed signal by closing (red cross) and opening (black circle) of MM using uniform line SE with variable length of (a) 1, (b) 4, (c) 6, (d) 8 data points.

As shown within the green rectangle in Fig. 5(a), MM based signal decomposition is performed through four steps:

- 1) Scale the extracted frequency variations within the range between $[-1,1]$ and define the SE as uniform line shape with initial length and height of one.
- 2) Based on the current length of SE, implement the closing and opening operations of MM on the frequency variations to build the upper and bottom envelopes;
- 3) Determine the IMF by calculating the mean values of upper and bottom envelopes.
- 4) Increase the length of the SE by one and repeat step 2) to 4) until the maximum SE length (30 in this paper) is reached.

Fig. 8 depicts an example of 10-second IMFs in time domain extracted by MM decomposition using uniform line shape SE with variable length, and the corresponding frequency spectrum computed by Fast Fourier Transformation (FFT). As can be

seen from Fig. 8, by recursively performing opening and closing operations, the IMFs are extracted from the original frequency data by MM at multiple time and frequency scales. In addition, the frequency component with maximum magnitude also shifts from high frequency region to low frequency band when the length of SE grows. This enables the MM-based signal decomposition similar to a low pass filter (LFP).

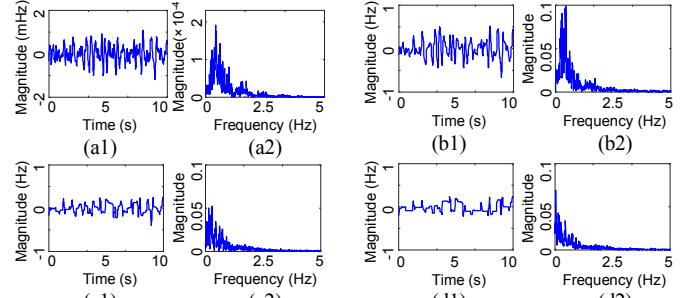


Fig. 8 Intrinsic mode functions (IMFs) derived by MM decomposition in time domain (left subplot) and corresponding frequency spectrum (right subplot), (a) original frequency segment, (b)-(d) IMFs using uniform line SE with length of 10, 20 and 30.

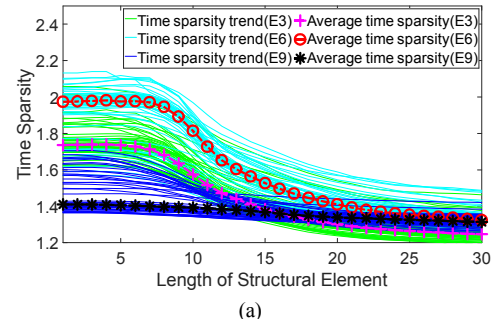
C. Time-Frequency (TF) Sparsity Mapping

Based on the extracted IMFs from the previous step, sparsity values (SP) of each IMF are subsequently computed in both time and frequency domains. The sparsity value can quantitatively evaluate the peak distribution of a signal where flat space is placed between two adjacent peaks as (6).

$$SP_l = \sqrt{\frac{\sum_{j=1}^N S_{IC}^l(j)^2}{\sum_{j=1}^N |S_{IC}^l(j)|}} \quad (6)$$

where S_{IC}^l is the extracted IMF in time (or frequency) domain with SE length l . N is the data length of S_{IC}^l . Since each frequency variation signal is decomposed into 30 scales ($l=1,2,L,30$), a series of sparsity values of the IMFs are computed to generate a sparsity trend, which describes the variations of the peak distribution.

Fig. 9 shows the derived sparsity trends with SE length increasing from 1 to 30. In Fig. 9, frequency segments are recorded from three PMUs (E3, E6 and E9) and for each unit, 50 frequency segments are presented. It is observed that the sparsity trends of frequency signals from the same location show a similar distribution and they tend to merge as a cluster in both time and frequency domains. A higher sparsity value indicates more peaks (valleys) are observed in the extracted IMF. The mean values of sparsity trends from each unit are calculated and they are also shown as traces with markers in Fig. 9.



(a)

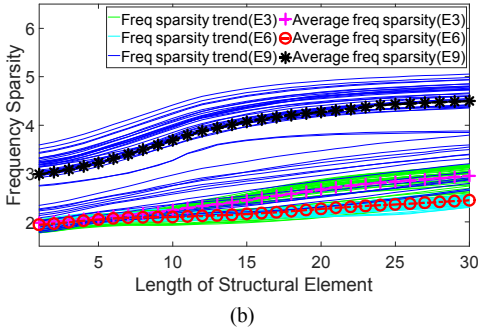


Fig. 9 Sparsity trends of IMFs from three PMUs in (a) time and (b) frequency domain. Each solid curve denotes the sparsity trend of the extracted IMF. Trace with markers shows the mean values of the sparsity trends from each unit.

By using the derived sparsity trends in Fig. 9, the roughness index (defined as an average of absolute values) of each sparsity trend in both time and frequency domains is calculated as (7) to quantify the variation of sparsity trends and they are projected onto a TF plane to construct a TF sparsity mapping (shown in Fig. 10). A higher roughness index indicates more peaks (valleys) with large magnitude are observed in the frequency variation signal over all decomposed levels (30 is used in the paper).

$$R = \frac{1}{P} \sum_{i=1}^P |SP_i| \quad (7)$$

where R is the roughness index and $P = 30$ is the data size of each sparsity trend. By observing Fig. 10 it is clear that the frequency segments from different source locations can be separated with clear boundaries by mapping the roughness index onto a TF plane. Therefore, both sparsity trends and roughness index are selected as distinctive spatio-temporal signatures and they are further input into the gcForest to recognize the source location of measured frequency segments. TABLE III summarizes the extracted location-sensitive signatures for frequency measurements authentication, where 62 features are derived from each 10-min frequency segment.

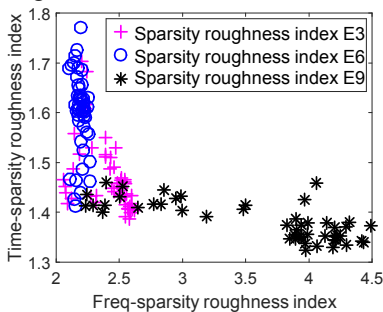


Fig. 10 TF sparsity mapping based on sparsity trends in Fig. 9.

TABLE III

DERIVED SPATIO-TEMPORAL SIGNATURES BY TF SPARSITY MAPPING

Spatio-temporal signatures	Dimension
Sparsity trend of decomposed IMFs in time domain with variable SE length (1 to 30)	30
Sparsity trend of decomposed IMFs in frequency domain with variable SE length (1 to 30)	30
Roughness index of sparsity trends in time domain	1
Roughness index of sparsity trends in frequency domain	1
Overall	62

D. Source Authentication Using gcForest

Based on the informative spatio-temporal features of each

frequency signal, gcForest [12] is trained to construct a mathematical model that approximates the relationship between the spatial characteristics and the source locations of the frequency segments. Upon the arrival of new frequency segment, the trained algorithm is further invoked to authenticate the source information of the measured frequency signal of interest. It should be notable that there may exist other methods for signal decomposition, informative feature extraction and source identification but MM, TF sparsity mapping and gcForest are initially selected in this paper for good reasons. For example, MM-based signal decomposition does not require the pre-selection of SE shape and it can be applied to the real situations in which the types and properties of frequency variation signals are unknown. In terms of the superiority of the gcForest, it possesses good representation learning abilities by generating a deep forest ensemble with a cascade structure and multi-grained scanning. In addition, it has much fewer easily tuned hyper-parameters and exhibits high generalization capability, which can avoid the potential over-fitting difficulties during the training.

IV. CASE STUDIES AND RESULTS ANALYSIS

In this section, the developed TF sparsity mapping is implemented to authenticate the source information of frequency measurements when the measured signals are subjected to malicious “Source ID Mix” attacks. The purpose of the numeric experiments is to examine whether the spoofed frequency segments can be recognized by using the underlying spatio-temporal characteristics.

A. Numeric Experiment Configuration

The performance of the proposed authentication method is evaluated through nine case studies when signals are subjected to three types of “Source ID Mix” spoofing attacks, including complete replacement of frequency signals (case study I), non-repetitively partial replacement of frequency signals (case study II) and repetitively partial replacement of frequency signals (case study III). Each type of spoofing attack is simulated through three scenarios (i.e., S1: spoofing without “out of region” unit (ORU); S2: spoofing with a distant ORU and S3: spoofing with a close ORU). It should be noted that for all case studies, the gcForest is only trained by using the non-spoofed frequency signals from the aforementioned ten units (E1-E10 in Fig. 4). Moreover, all PMUs in Fig. 4 have been calibrated to ensure the extracted location-sensitive signatures are due to the randomness of the local power system condition variations instead of the discrepancies of configurations among different units.

For each case study, the source authentication of frequency segments is performed by the following procedures: The database containing spatio-temporal signatures extracted from tens of thousands of frequency fragments is divided into a training dataset and a testing dataset with sample ratio of 60%/40%. Then multiple folds cross validation is performed on the training samples to determine the optimal configuration of the gcForest, i.e., the size and number of multi-grained scanning windows. Once the values of the above optimal hyper-parameters are determined, gcForest is trained on the whole training samples. Then the well-trained classifier is invoked to

determine the source locations of testing frequency examples. Finally, the overall match accuracy of all testing samples (i.e., ratio between the number of correctly identified samples in the testing dataset and the number of all samples in the testing dataset), match accuracy of spoofed frequency segments (i.e., ratio between the number of correctly identified samples belonging to the spoof class in the testing dataset and the number of all samples belonging to the spoof class in the testing dataset) and false-negative and false-positive rates [23] are recorded to evaluate the identification performance.

TABLE IV summarizes the sample distribution (size of training and testing samples in each class and graphical compositions of spoofed segments) in the above nine case studies. All numeric experiments were executed on a 3.2 GHz, 8GB RAM computer. The source identification of a single frequency segment only requires 0.03 second which is considerably short time for online application.

TABLE IV
SAMPLE CONFIGURATION OF NINE CASE STUDIES

Case No.	Testing		Training
	Graphic notation of sample composition in spoof class	Class number	
I-S1	-	10 classes (E1-E10)	10 classes (E1-E10) 600 segments for each class
I-S2	1	11 classes (E1-E10 + spoof class)	
I-S3	1		
II-S1	(1-P) P		
II-S2	(1-P) P	400 segments for each class	
II-S3	(1-P) P		
III-S1	(1-P)/2 P/2 (1-P)/2 P/2		
III-S2	(1-P)/2 P/2 (1-P)/2 P/2		
III-S3	(1-P)/2 P/2 (1-P)/2 P/2		

*number on the graphical notations indicates data portion ($P \in [0.1, 0.9]$) of each PMU contributed to the spoofed frequency segments.

B. Case Study I: Complete Replacement of Frequency Signals

In case study I-S1, it is assumed the adversary alters the frequency data by replacing the whole segment without knowing the measurement information from other “out of region” units, which means the frequency fragments in the testing dataset are all from the same group of PMUs in the training dataset. The identification results by using the statistical characteristics method in [24] are also calculated for comparison purpose. The extracted statistical characteristics include mean, variance and variation range of the measured frequency segment, variance of nine level detail IMF signals through wavelet decomposition and residual signal after wavelet decomposition, two coefficients (β_1, β_2) in autoregressive (AR) model and variance of innovative signal after AR modeling. The simulation results show that the proposed method can correctly recognize all the frequency segments in the testing dataset with 100% accuracy, while the overall identification rate by using the statistical characteristics based method is relatively low, which is 94% as shown in TABLE V. Further examination on the match accuracy of each class shows that the identification accuracy of frequency segments from E1 and E2 is significant lower among the ten classes. A possible reason for the low identification rate is that when the frequency segments are recorded at multiple intra-grid locations, the measured frequency signals exhibit similar

variations among different units (see E1 and E2 in Fig. 4(b)), which makes the mean values and variance of measured signals from certain locations almost identical. Moreover, the differences in the variance of IMFs through wavelet transformation among multiple PMUs also become minimal, thus the classifier may not be able to determine the source locations of corresponding signals with high accuracy by using above statistical characteristics.

Similar to author’s previous experiment in [25-26], the gcForest model is first trained using frequency segments measured in June, and then it is tested using frequency segments collected in July. It is found the overall match accuracy by using the proposed method is almost 100%. Similar identification results can be observed by collecting the training and testing frequency segments in other seasons (for example, trained on January, tested on February, etc.). This indicates the spatio-temporal characteristics of the frequency measurements at the above ten locations in the EI are quite stable within two adjacent months. Therefore, the gcForest model does not have to be trained frequently and it can be trained at regular time, for example, once in every month.

TABLE V
MATCH ACCURACY BY STATISTICAL BASED METHOD FOR CASE STUDY I-S1

Class	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Accuracy(%)	81	88	96	96	96	100	100	96	92	96

In case study I-S2, it is assumed the adversary introduces frequency segments from one distant “out of region” unit (ORU1 in Fig. 4(a)) and adds them into the original testing dataset in case study I-S1 to form a spoof class. The ORU1 also resides within the same interconnection of the aforementioned ten units. Since the gcForest is only trained on the frequency segments from E1-E10 to learn the underlying spatio-temporal characteristics, the classifier cannot directly recognize the samples from the ORU1. For each testing sample, the classifier provides a probability giving its confidence in its decision on the source of recording. By making use of this feature of the classifier, a confidence threshold (T_h) is defined to make a final decision on the source location of the testing sample, i.e., if the classification confidence of a sample assigned by gcForest is lower than T_h , we advance this sample as the spoof class.

Fig. 11 shows the dependency between the match accuracy of frequency examples in the spoof class and T_h by using the proposed method and statistical characteristics based method. It is observed that there is a “trade off” on the match accuracy between the spoof class (blue curve) and the rest ten classes (purple curve). Near-perfect identification rate of the spoof class is attained by using high confidence threshold (above 0.8), while the overall match accuracy of the rest ten classes declines slightly. By tracking the overall match accuracy (black curve) of all classes, an optimal confidence threshold is empirically selected when the peak value of the overall match accuracy is attained. TABLE VI compares the identification performance of two approaches with optimal threshold. From Fig. 11 and TABLE VI it is observed that by using the proposed method, the gcForest can identify the frequency segments from the spoof class with 100% accuracy, which outperforms the statistical characteristics based method (81%).

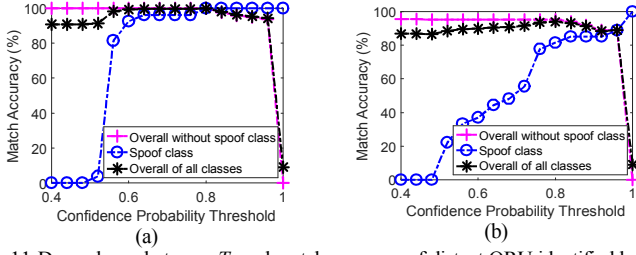


Fig. 11 Dependency between T_h and match accuracy of distant ORU identified by (a) the proposed and (b) statistical characteristics based method.

TABLE VI

COMPARISON OF IDENTIFICATION PERFORMANCE FOR CASE STUDY I-S2

Identified by the proposed method					Identified by statistical characteristics				
OAcc*	SAcc	FN	FP	T_{opt}	OAcc	SAcc	FN	FP	T_{opt}
100	100	0	0	0.8	94	81	7	0.7	0.8

*OAcc: overall match accuracy of all classes, SAcc: match accuracy of spoof class, T_{opt} : optimal threshold value, FN: false-negative rate, FP: false-positive rate.

In case study I-S3, frequency segments from one close “out of region” unit (ORU2 in Fig. 4(a)) have been added into the testing dataset in case study I-S1 to form a spoof class. By comparing Fig. 12(a) with Fig. 11(a) it appears that when the threshold is less than 0.8, the identification accuracy of frequency signals from ORU2 becomes lower. This is because ORU2 and E4 reside within the same local grid and the distance between these two units is less than 70 miles, which makes the frequency segments from ORU2 quite similar to those from E4. Once the gcForest tries to identify the source location of a frequency fragment from ORU2, it may assign this segment into E4 with high confidence probability. Therefore, a higher threshold is required to achieve high identification accuracy of the samples in the spoof class. TABLE VII compares the identification performance of two approaches with optimal thresholds. It indicates the overall match accuracy of the proposed method attains 98% by using 0.84 threshold and the identification accuracy of the spoof class also falls in high range of 96%. However, the identification rate of the spoofed class is less than 50% by using the statistical characteristics based approach. This implies the differences in the statistical features of the measured frequency signals from multiple close intra-grid locations become minimal which may not fully represent the spatial signatures of each location. Such minimal differences in the statistical feature values have confused the classifier when determining the class boundaries and could result in more mistakes in the source identification.

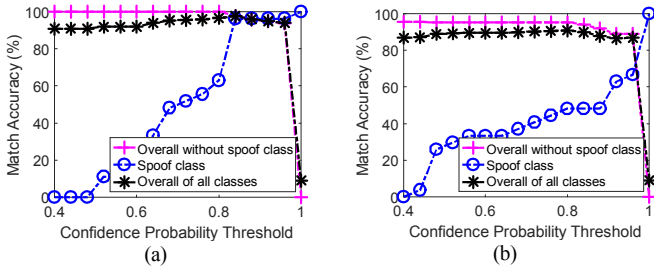


Fig. 12 Dependency between T_h and match accuracy of close ORU identified by (a) the proposed and (b) statistical characteristics based method.

TABLE VII

COMPARISON OF IDENTIFICATION PERFORMANCE FOR CASE STUDY I-S3

Identified by the proposed method					Identified by statistical characteristics				
OAcc	SAcc	FN	FP	T_{opt}	OAcc	SAcc	FN	FP	T_{opt}
98	96	2	0.2	0.84	90	48	10	1	0.8

C. Case Study II: Non-repetitive Partial Replacement of Frequency Signals

In case study II-S1, the frequency segments from the original ten units are not completely replaced by other units, but only a portion of the data have been replaced, which produces the spoofed frequency signals in the testing dataset. In the first scenario (case study II-S1), the spoofed frequency signals are generated by exchanging part of the consecutive data from two out of ten PMUs within E1-E10. Given many permutations in selecting arbitrary two PMUs (total $C_{10}^2 = 45$ combinations) to generate the spoofed segments, to avoid the unbalanced sample distribution in the testing dataset and achieve an easy comparison on the identification performance with other case studies, each combination is evaluated separately. Specifically, at each time, frequency data from only two PMUs are mixed up and they are put into the original testing dataset to form a spoof class for source identification. Then the evaluation metrics, such as the overall match accuracy, match accuracy of the spoof class as well as the optimal threshold candidate are determined for this specific configuration. The evaluation metrics from all 45 combinations are averaged out as final evaluation scores while the optimal threshold candidate with maximum occurrence rate among 45 combinations is selected as the final optimal threshold.

Fig. 13 compares the overall match accuracy of the frequency segments by mixing two PMUs’ data when the spoofed data portion varies from 10% to 90%. The contour maps of the match accuracy are also presented with numbers indicating the specific accuracy levels. From Fig. 13(a) it is observed that both overall match accuracy and match accuracy of the spoof class roughly present an axial symmetry distribution with 50% spoofed data portion as the central line. The overall match accuracy attains its peak at 99% by using the proposed method when the spoofed data portion is 50%. Moreover, the overall match accuracy shows a decrease trend when the spoofed data portion deviates from 50%. This indicates given that two PMUs are involved in generating the spoofed frequency segments, the differences of the spatial characteristics between the spoofed and valid frequency segments become significant when half of the data have been replaced. TABLE VIII compares the identification performance of two approaches and the optimal threshold with spoofed data portion varying from 10% to 50%. It appears that the identification rate of the spoof class is no more than 30% by using both two methods when only 10% (1 minute) data have been replaced. The low accuracy is mainly due to the small size of the replaced data. In addition, the optimal threshold also decreases slightly when the spoofed data portion grows from 10% to 50%. To solve the above problem, further discussions on improving the identification accuracy by using short frequency segments will be presented in Section IV-E.

TABLE VIII

COMPARISON OF IDENTIFICATION PERFORMANCE AND OPTIMAL THRESHOLD WITH VARIABLE SPOOFED DATA PORTION FOR CASE STUDY II-S1

Spoofed data portion	Identified by the proposed method					Identified by statistical characteristics				
	OAcc	SAcc	FN	FP	T_{opt}	OAcc	SAcc	FN	FP	T_{opt}
10%	94	30	6.4	0.6	0.8	88	22	12	1.2	0.84
30%	99	89	1	0.1	0.8	92	63	7.7	0.7	0.8
50%	99	93	0.6	0.06	0.76	94	82	6	0.6	0.8

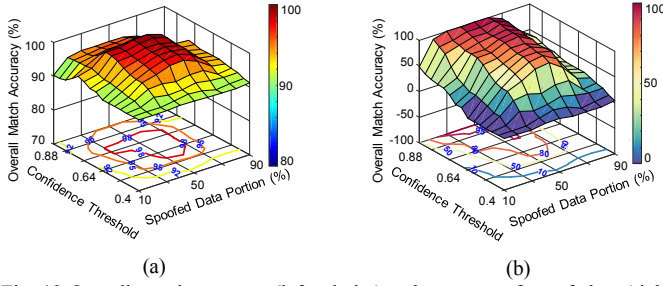


Fig. 13 Overall match accuracy (left subplot) and accuracy of spoof class (right subplot) by using the proposed method in case study II-S1.

In case study II-S2, it is assumed that the adversary generates the spoofed frequency segments by replacing partial data from E4 with ORU1. As shown in Fig. 14, both overall match accuracy and match accuracy of the spoof class exhibit an increase trend when the spoofed data portion increases to 90%, which is different from the previous results in Fig. 13. The maximum identification accuracy of the spoof class is 96% when 90% of data have been replaced by ORU1, demonstrating the spatio-temporal signatures from this “out of region” unit become dominant which are easier to be recognized by gcForest. Followed by previous case study, in case study II-S3, the spoofed frequency segments are produced by replacing partial data from E4 with ORU2. The overall match accuracy and match accuracy of the spoof class with variable spoofed data portion are shown in Fig. 15. Compared with Fig. 14, the overall match accuracy and the accuracy of the spoofed class by using the proposed method is relatively low, but it is still higher than the statistical approach. The maximum identification accuracy of the spoof class is 93% given 90% data are replaced. Similar to the results from case study I-S3, a high threshold 0.84 is generally required for the gcForest to maintain such high identification accuracy.

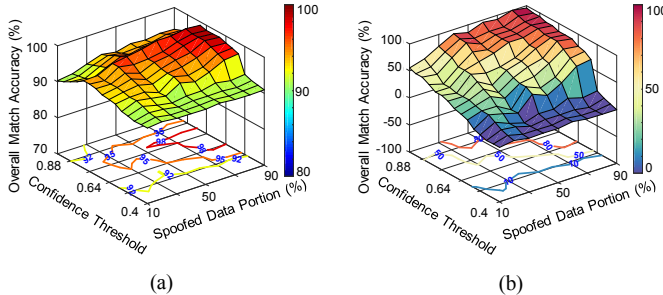


Fig. 14 Overall match accuracy (left subplot) and accuracy of spoof class (right subplot) by using the proposed method in case study II-S2.

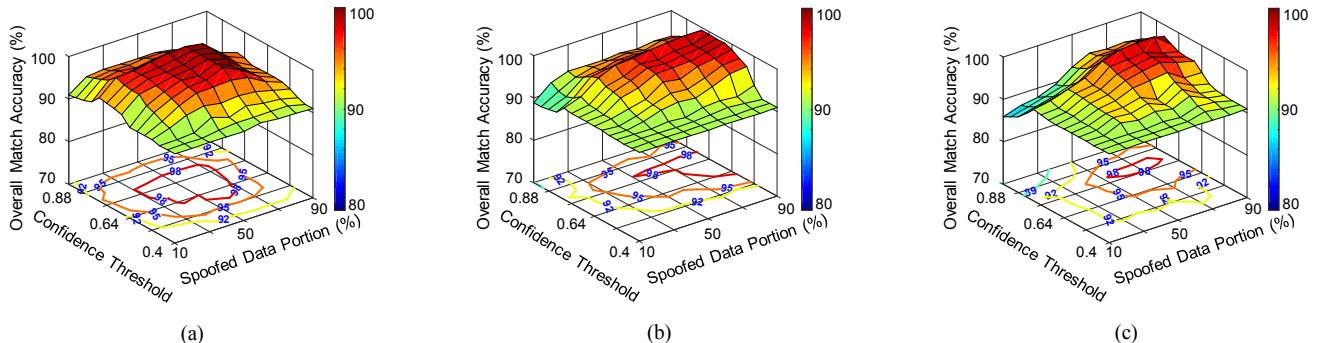


Fig. 16 Overall match accuracy by using the proposed method in case study (a) III-S1, (b) III-S2 and (c) III-S3

TABLE IX
COMPARISON OF IDENTIFICATION PERFORMANCE AND OPTIMAL THRESHOLD WITH VARIABLE SPOOFED DATA PORTION FOR CASE STUDY II-S2

Spoofed data portion	Identified by the proposed method					Identified by statistical characteristics				
	OAcc	SAcc	FN	FP	T_{opt}	OAcc	SAcc	FN	FP	T_{opt}
10%	94	38	5.7	0.5	0.8	89	26	11	1	0.8
50%	98	82	1.7	0.17	0.8	92	56	8.4	0.8	0.8
90%	99	96	0.003	3×10^{-4}	0.8	94	81	6	0.6	0.8

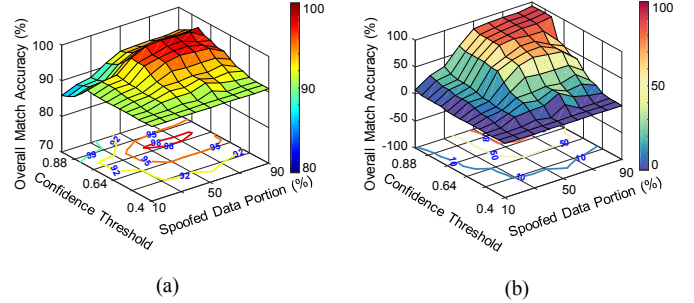


Fig. 15 Overall match accuracy (left subplot) and accuracy of spoof class (right subplot) by using the proposed method in case study II-S3.

TABLE X
COMPARISON OF IDENTIFICATION PERFORMANCE AND OPTIMAL THRESHOLD WITH VARIABLE SPOOFED DATA PORTION FOR CASE STUDY II-S3

Spoofed data portion	Identified by the proposed method					Identified by statistical characteristics				
	OAcc	SAcc	FN	FP	T_{opt}	OAcc	SAcc	FN	FP	T_{opt}
10%	91	4	8.7	0.8	0.8	81	4	19	0.2	0.96
50%	98	81	1.7	0.2	0.8	89	26	11	1.1	0.8
90%	98	93	2.7	0.3	0.84	91	44	9.4	0.9	0.8

D. Case Study III: Repetitively Partial Replacement of Signals

To investigate whether the integrity of the replaced data will affect the spatio-temporal characteristics and subsequently influence the source authentication accuracy, this case study assumes the data spoofing attack becomes more complicated where the spoofed segments are generated by periodically replacing part of the data from the validated units with other units. Similar to the previous case study II, three data spoofing scenarios are simulated. Fig. 16 shows the overall match accuracy under three data spoofing scenarios. By comparing the results from Fig. 16 with each counterpart in case study II, it shows that the overall match accuracy follows the similar distribution as results in each data spoofing scenario of case study II. This demonstrates the extracted spatio-temporal signatures are likely to keep consistent no matter the spoofed data are from the whole piece of consecutive data block or they are assembled by a number of short data fragments.

E. Source Authentication Using Short Frequency Segments

Due to the difficulties in identifying the source locations of frequency segments using short data fragments, two interpolation techniques, i.e., spline and fractal interpolation [27] are implemented to upsample the short frequency segments and further increase the identification accuracy of the proposed source authentication method. Fractal is kind of phenomena which exists in a wide range of signals. Such signal usually possesses a scale invariant structure as the structure repeats itself on subintervals of the signal. Generally, fractal structures are characterized by calculating Hurst exponent (H) using multifractal detrended fluctuation analysis [28]. A signal is considered to possess fractal structures when H is above 0.5. By calculating the H value of daily frequency signals from E1- E10, it appears the frequency variation contains fractal structures which can be resampled through fractal interpolation.

In this section, the source location identification is performed by using the same experimental configuration of case study I-S1 but the reporting rate of all frequency signals are boosted to 120Hz through spline and fractal interpolation. The identification performance by using frequency segments with different length (varying from 1 second to 15-min) has also been evaluated. The correlation between the overall match accuracy and data length is shown in Fig. 13. It indicates that the match accuracy of the frequency segments processed by fractal interpolation is consistently higher than the spline interpolation, especially when the segment length is less than 30 seconds. Increasing the segment length may lead to a growth in the overall match accuracy. Near 100% match accuracy is achieved given 10-minute frequency signals are used. A decrease trend is observed in the overall match accuracy when the segment length further increases. This is mainly due to the smooth effect on the sparsity trends and roughness index derived from long frequency segments. By contrast, reducing the signal length will also lower the match accuracy, which is caused by the insufficient data samples available for extracting the spatio-temporal characteristics for source identification.

To demonstrate the effectiveness of the fractal interpolation in detecting data spoofing on short frequency segments, Fig. 17(b) shows the overall match accuracy of frequency segments subjected to the data spoofing scenario I-S2 by using 10-second frequency fragments (a typical duration of a disturbance within the EI). All the 10-second segments in both training and testing datasets are interpolated by the fractal interpolation for source identification. From Fig. 17(b) it is observed that the maximum overall match accuracy has increased to 90% after fractal interpolation while the identification accuracy of the spoof class also increases to 92% by using 0.85 threshold. This demonstrates the spatial signatures extracted from the interpolated data become distinctive, which allows gcForest to identify the spoofed segments with high accuracy. It should be notable that the identification accuracy of the proposed methodology could be further improved if the original frequency signals are measured using high reporting rate PMUs.

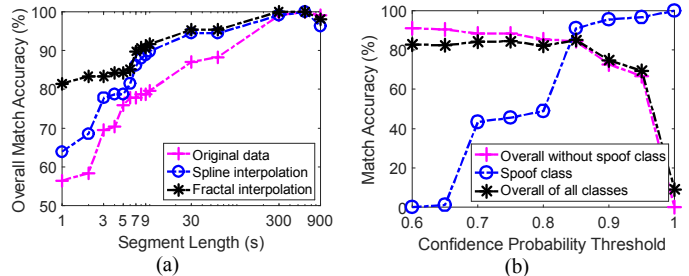


Fig. 17 Overall match accuracy of short frequency signals under data spoofing scenarios (a) I-S1 and (b) I-S2.

V. CONCLUSIONS

This paper proposes a machine learning based framework to authenticate the source information of PMUs' frequency measurements, thereby negating "Source ID Mix" cyber attacks. By leveraging the proposed TF sparsity mapping, the underlying spatio-temporal characteristics of frequency signals measured at different intra-grid locations within the EI have been revealed. The high accuracy and practicality of the proposed authentication methodology have been validated by variety case studies in detecting the manipulated PMUs' data under different spoofing scenarios. The experimental results demonstrate that the identification rate of frequency data primarily depends on the distance among measurement geolocations as well as the input frequency signals. A confidence probability threshold above 0.8 is preferred for the gcForest to successfully identify the "Source ID Mix" attacks, which happen among different cities in the same interconnection. A higher threshold has to be assigned to the classifier in order to recognize the spoofing attacks occurring at near range locations, although such data replacements normally would not cause severe damages to WAMS functionality. The overall identification rate has been demonstrated to be preferential to that of the traditional statistical characteristics based approach. Practical difficulties in source authentication due to the short data length can be potentially solved by interpolating fractal structures within the signals.

This research will prove beneficial to system operators and service providers in helping them gain valuable insights over legitimate data patterns used in detecting anomalous patterns. Moreover, findings also suggest that use of these research results can also facilitate mitigating the potential "Source ID Mix" data spoofing and further exploitation towards accurate source authentication strategies for improving power system cyber security.

ACKNOWLEDGMENT

This work was supported primarily by the Engineering Research Centre Program of the National Science Foundation and the Department of Energy under NSF Award Number EEC-1041877 and the CURENT Industry Partnership Program.

REFERENCES

- [1] C. Beasley, X. Zhong, J. Deng, R. Brooks and G. Venayagamoorthy, "A survey of electric power synchrophasor network cyber security," in *Proceedings of 5th IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, 2014, Istanbul, TURKEY, pp. 1-5.

- [2] Z. Zhang, S. Gong, A. D. Dimitrovski and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Trans. Smart Grid*, vol.4, Issue 1, pp. 87-98, 2013.
- [3] C. Glenn, D. Sterbentz and A. Wright, "Cyber threat and vulnerability analysis of the U.S. Electric sector," Idaho National Lab, Idaho Falls, No. INL/EXT--16-40692, 2016.
- [4] H. Lin, Y. Deng, S. Shukla, J. Thorp and L. Mili, "Cyber security impacts on all-PMU state estimator - a case study on co-simulation platform GECO," in *Proceedings of International Conference on Smart Grid Communications*, 2012, Tainan, Taiwan, pp. 587-592.
- [5] Y. Fan, Z. Zhang, M. Trinkle, A. D. Dimitrovski, J. B. Song and H. Li, "A Cross-Layer defense mechanism against GPS spoofing attacks on PMUs in smart grids," *IEEE Trans. Smart Grid*, vol.6, Issue 6, pp. 2659-2668, 2015.
- [6] R. Khan, K. McLaughlin, D. Laverty and S. Sezer, "IEEE C37.118-2 synchrophasor communication framework: Overview, cyber vulnerabilities analysis and performance evaluation," in *Proceedings of 2nd International Conference on Information Systems Security and Privacy*, 2016, Rome, Italy, pp. 1-10.
- [7] S. Kumar, M. K. Soni and D. K. Jain, "Cyber security threats in synchrophasor system in WAMS," *Int. J. Comput. Appl. Technol.*, vol.115, Issue 8, pp. 17-22, 2015.
- [8] Y. Liu, W. Yao, D. Zhou, L. Wu, S. You, H. Liu, L. Zhan, J. Zhao, H. Lu, W. Gao and Y. Liu, "Recent developments of FNET/GridEye - a situational awareness tool for smart grid," *CSEE Journal of Power and Energy Systems*, vol.2, Issue 3, pp. 19-27, 2016.
- [9] B. Kasztenny, N. Fischer, K. Fodero and A. Zvarych, "Communications and data synchronization for line current differential schemes," in *Proceedings of 38th Annual Western Protective Relay Conference*, 2011, Spokane, WA, pp. 1-19.
- [10] T. Xia, H. Zhang, R. Gardner, J. Bank, J. Dong, R. Zuo, Y. Liu, L. Beard, P. Hirsch, G. Zhang and R. Dong, "Wide-area frequency based event location estimation," in *Proceedings of IEEE Power Engineering Society General Meeting, June 24-28, 2007*, Tampa, FL, USA, pp. 1-7.
- [11] F. Bai, L. Zhu, Y. Liu, X. Wang, K. Sun, Y. Ma, M. Patel, E. Farantatos and N. Bhatt, "Design and implementation of a measurement-based adaptive wide-area damping controller considering time delays," *Electr. Power Syst. Res.*, vol.130, pp. 1-9, 2016.
- [12] Z. H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proceedings of Twenty-Sixth International Joint Conference on Artificial Intelligence, 19-25 August, 2017*, Melbourne, Australia, pp. 1-7.
- [13] G. Chaojun, P. Jirutitijaroen and M. Motani, "Detecting false data injection attacks in AC state estimation," *IEEE Trans. Smart Grid*, vol.6, Issue 5, pp. 2476-2483, 2015.
- [14] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Syst. J.*, vol.11, Issue 3, pp. 1644-1652, 2017.
- [15] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Networks Learn. Syst.*, vol.27, Issue 8, pp. 1773-1786, 2016.
- [16] R. Moslemi, A. Mesbahi and J. M. Velni, "A fast, decentralized covariance Selection-Based approach to detect cyber attacks in smart grids," *IEEE Trans. Smart Grid*, vol.9, Issue 5, pp. 4930-4941, 2018.
- [17] H. Sedghi and E. Jonckheere, "Statistical structure learning to ensure data integrity in smart grid," *IEEE Trans. Smart Grid*, vol.6, Issue 4, pp. 1924-1933, 2015.
- [18] K. Manandhar, X. Cao, F. Hu and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using kalman filter," *IEEE Trans. Control Network Syst.*, vol.1, Issue 4, pp. 370-379, 2014.
- [19] A. Ashok, M. Govindarasu and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Trans. Smart Grid*, vol.9, Issue 3, pp. 1636-1646, 2018.
- [20] J. Landford, R. Meier, R. Barella, X. Zhao and E. Cotillasanchez, "Fast sequence component analysis for attack detection in synchrophasor networks," in *Proceedings of 5th International Conference on Smart Cities and Green ICT Systems*, 2016, Rome, ITALY, APR 23-25, pp. 1-8.
- [21] J. Serra, *Image analysis and mathematical morphology*, Orlando, FL, USA: Academic Press, Inc, 1983.
- [22] J. C. Chan, H. Ma and T. K. Saha, "Time-frequency sparsity map on automatic partial discharge sources separation for power transformer condition assessment," *IEEE Trans. Dielectr. Electr. Insul.*, vol.22, Issue 4, pp. 2271-2283, 2015.
- [23] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inform Process Manag.*, vol.45, Issue 4, pp. 427-437, 2009.
- [24] A. Hajji-Ahmad, R. Garg and M. Wu, "ENF-based region-of-recording identification for media signals," *IEEE Trans. Inf. Forensics Secur.*, vol.10, Issue 6, pp. 1125-1136, 2015.
- [25] W. Yao, J. Zhao, M. J. Till, S. You, Y. Liu, Y. Cui and Y. Liu, "Source location identification of distribution-level electric network frequency signals at multiple geographic scales," *IEEE Access*, vol.5, pp. 11166-11175, 2017.
- [26] Y. Cui, F. Bai, Y. Liu and Y. Liu, "A measurement source authentication methodology for power system cyber security enhancement," *IEEE Trans. Smart Grid*, vol.9, Issue 4, pp. 3914-3916, 2018.
- [27] H. Sun, "A practical MATLAB program for multifractal interpolation surface," in *Proceedings of 8th International Conference on Natural Computation*, 2012, Chongqing, Sichuan, China, pp. 1-5.
- [28] E. A. F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in matlab," *Front. Physiol.*, vol.3, pp. 1-18, 2012.



Yi Cui (M'19) received the B.Sc. and M.Sc. degrees in Electrical Engineering from Southwest Jiaotong University, Chengdu, China, in 2009 and 2012, respectively, and received the Ph.D. degree in Electrical Engineering at the University of Queensland, Brisbane, Australia, in 2016.

Currently, he is a Postdoctoral Research Fellow in the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, USA. His research interests include wide-area monitoring and control, measurement based small signal stability, condition assessment and fault diagnosis of power transformers.



Feifei Bai (M'16) received the B.S. degree and Ph.D. degree in Power System and Automation from Southwest Jiaotong University, China, in 2010 and 2016, respectively. She was a joint Ph.D. student at the University of Tennessee, Knoxville, USA, from 2012 to 2014. She is currently an advance

Queensland research fellow in the School of Information Technology and Electrical Engineering, University of Queensland, Australia. Her main research interests are renewable energy integration into power grid, voltage management of distribution networks with high PV integration, small signal stability analysis and wide-area damping control.



Yilu Liu (F'04) received the B.S. degree from Xian Jiaotong University, China, and the M.S. and Ph.D. degrees from Ohio State University, Columbus, in 1986 and 1989, respectively. Dr. Liu is currently the Governor's Chair with the University of Tennessee, Knoxville (UTK), and Oak Ridge National

Laboratory (ORNL). She is a member of the U.S. National Academy of Engineering. She is also the Deputy Director of the DOE/NSF-cofunded Engineering Research Center CURENT. Prior to joining UTK/ORNL, she was a Professor with Virginia Tech. She led the effort to create the North American power grid frequency monitoring network at Virginia Tech, which is now operated at UTK and ORNL as GridEye. Her current research interests include power system wide-area monitoring and control, electromagnetic transient analysis, and power transformer modeling and diagnosis.



Peter L. Fuhr (SM'05) received the B.S. degrees in physics and mathematics from Beloit College and the M.S.E. and Ph.D. degrees in electrical engineering from the Johns Hopkins University. He has been involved in industrial wireless, sensors, and secure systems for more years than he cares to state as a

NASA Space Optical Physicist, a University Professor, a Serial Entrepreneur, and a Researcher with the U.S. National laboratory. He is currently a Distinguished Scientist with the Oak Ridge National Laboratory serving in the capacity as the Technology Director for the Unmanned Aerial Systems Research Laboratory and the Director of Grid Security. He has authored and delivered hundreds of technical journals and conference publications/presentations. His research activities are featured in the SPIE Milestone Series on Fiber Optics. He received the Presidential Award for Excellence in Research on networked sensor systems for structures. He serves as the Director of the International Society for Automation Test and Measurement Division (over 1500 members), is the co-founder and past

Chairman of the Wireless Industrial Networking Alliance, and co-chairs of the Secure Infrastructure Controls Society.



Marissa E. Morales-Rodríguez (M'17) was born in San Juan, Puerto Rico. She received the B.S. and M.S. degrees from the Department of Chemistry, University of Puerto Rico at Mayagüez, and the Ph.D. degree in energy science and engineering with entrepreneurship track from The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, in 2017. She is currently a Research Scientist with the Oak Ridge National Laboratory, Oak Ridge, TN, USA. She has been involving in the areas of chemical sciences concentrating on applications related to sensing, additive manufacturing, and document security. She has authored/co-authored over 10 publications and one patent. Her current research activities include research and development involving: sensors and systems for microdrones and cybersensors on mobile platforms for extended grid state monitoring and security. She serves as the Director-Elect for the International Society of Automation Test and Measurement Division (over 1500 members) and the American Chemical Society since 2016.