Electric Power Systems Research 196 (2021) 107207

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Data source authentication of synchrophasor measurement devices based on 1D-CNN and GRU

Shengyuan Liu [a,b], Shutang You [b], Chujie Zeng [b], He Yin [b], Zhenzhi Lin [a,*], Yuqing Dong [b], Wei Qiu [b], Wenxuan Yao [b], Yilu Liu [b,c]

[a] Department of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
[b] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA
[c] Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

ARTICLE INFO

Keywords:
Synchrophasor measurement device (SMD)
Data authentication
One-dimensional convolutional neural network (1D-CNN)
Gated recurrent unit (GRU)
FNET/GirdEye

ABSTRACT

Synchrophasor measurement devices (SMDs) have been widely deployed to support real-time monitoring and control of power systems. In the meantime, data spoofing has emerged in recent years. Therefore, it is of great importance to study data authentication algorithms for detecting and defending the data spoofing effectively. In this work, a one-dimensional convolutional neural network (1D-CNN) is utilized to extract temporal signatures hidden in frequency, voltage angle and amplitude data; then the gated recurrent unit (GRU) employs these temporal signatures for data source authentication. In case studies, the performances of different algorithms are tested in large-scale power systems with numerous SMDs for the first time, and comparisons among different algorithms show that the proposed algorithm can achieve a higher accuracy of data source authentication with a shorter time window.

1. Introduction

In the context of the smart grid industry [1], several technologies such as data communication [2] and power management [3] have been greatly developed in the Internet of Things (IoT)-environment. For example, the real-time data such as frequency, voltage amplitude and angle can be measured with the help of IoT devices, which allow operators to monitor the real-time situation of power systems and take control measures in a timely manner. However, the measured data are also valuable for intentional spoofing and attack, therefore, it is important to perform data authentication for part of IoT devices in power systems, i.e., synchrophasor measurement devices (SMDs) [4].

In fact, data spoofing can be divided into several types and this work focuses on the model-free detection for data source mixing, which was preliminarily studied by our previous research [5]-[8]. In [5], an *L*-level Daubechies wavelet-based feature extraction method is employed for electrical network frequency (ENF) and a feed-forward artificial neural network (F-ANN) is utilized to identify the data source locations. In [6], the mathematical morphology (MM) method is used to decompose the ENF into several intrinsic components for time-frequency sparsity analysis and a random forest-based algorithm is used for correlating the correct data sources of SMDs. In [7], the multi-grained cascaded forest (gcForest) algorithm is further presented and combined with time-frequency sparsity analysis to achieve better performance and more detailed analyses and discussions are given as well. However, the aforementioned algorithms require a relatively long-time window segment for feature extraction and subsequent data authentication, which would then delay data spoofing detection. Hence, a two-layer ANN, whose input features are fed by ensemble empirical mode decomposition (EEMD) and fast Fourier transform (FFT), is constructed in [8]. The algorithm can achieve high accuracy with a 20-second window and can be applied for SMDs located quite close to each other. Nevertheless, the algorithm presented in [8] requires high-reporting rate data (i.e. 1.44kHz) collected by advanced universal grid analyzers (UGAs) which have not been widely deployed in the current stage, and it cannot be applied directly for existing frequency disturbance recorders (FDRs). Besides, all algorithms in [4]-[8] are lacking testing for the situations with numerous SMDs and their performance would decrease quickly if more SMDs are to be authenticated.

In light of this, this work proposes a one-dimensional convolutional neural network (1D-CNN) and gated recurrent unit (GRU)-based algorithm for data source authentication of FNET/GridEye. The contributions of this work can be summarized as follows.

* Corresponding author.

https://doi.org/10.1016/j.epsr.2021.107207
Received 25 October 2020; Received in revised form 27 March 2021; Accepted 29 March 2021
Available online 6 April 2021

i) In the past, only frequency data are utilized for data authentication while 1D-CNN is employed in this work to extract the measurement's features from multiple data series constituted by frequency, voltage angle and amplitude for the first time.
ii) The GRU network is utilized to authenticate the source of measured data. Compared with traditional machine learning methods such as ANN and gcForest, the proposed GRU network can consider the temporal information embedded in the time series data. Compared with other time series models, such as the long short-term memory (LSTM) network, the GRU network is more computationally efficient.
iii) The proposed 1D-CNN-GRU-based algorithm not only can deal with the high-reporting rate SMDs such as UGAs located closely to each other which is quite difficult for other algorithms to distinguish, but also can deal with the large number of commonly deployed FDRs in bulk actual power systems. Besides, the window required for the proposed algorithm is shorter than other algorithms.

2. 1D-CNN-GRU-based data source authentication algorithm for FNET/GridEye

FNET/GridEye is a power system monitoring system being deployed worldwide [9] and can be regarded as a part of IoT in the smart industry. The basic IoT devices (i.e., SMDs) in FNET/GridEye are FDRs with a 10Hz reporting rate and UGAs that can achieve a 1.44kHz reporting rate. This work aims to authenticate the data source, in other words, to identify which SMDs the data collected from. Therefore, this work aims to extract the features of each SMD based on historical measurement first, and then classify the online measured data into the different sources of SMDs based on the extracted features. Therefore, the data source authentication problem in this work can be abstracted as a supervised classification problem and it is critical to find a suitable classification algorithm embedded with a feature extraction method, which can minimize the final identification error (i.e., maximize the accuracy of data source authentication). Due to its advantages in analyzing time series data, 1D-CNN has been widely employed for sequence models such as natural language processing and human activity recognition [10]. Since the data of SMDs are also time-stamped, 1D-CNN is very suitable for extracting inherent features from them. Generally, three types of measurements, i.e., frequency, voltage angle and amplitude data can be collected by FNET/GridEye, and they need to be filtered (i.e. denoised) and detrended before feature extraction to improve extraction effectiveness. It is assumed that the denoised and detrended measurement data in a single time window can be denoted as $\mathrm{P}_{3\times N_{\mathrm{data}}}$, where N_{data} is the number of samples in a single time window. Then, the 1D-CNN can be employed for feature extraction as shown in Fig. 1.

In Fig. 1, the i^{th} convolution kernel $\mathbf{K}_i$ with size L_{CK} would slide from samples 1 to N_{data} to extract features and the l^{th} feature graph $\mathbf{S}^l$ can be output as

$$\mathbf{S}^l = \boldsymbol{f}\left(\sum_{i=1}^{N_{\mathrm{CK}}}\left(\mathbf{S}^{l-1} * \mathrm{K}_i + \boldsymbol{b}_i^l\right)\right) \tag{1}$$

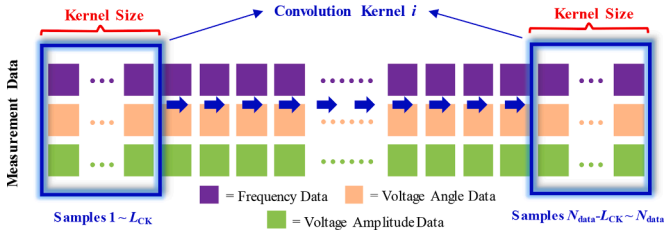

Fig. 1. Illustration of 1D-CNN for feature extraction of measurement data.

where $\boldsymbol{f}(\cdot)$ is activation function, $\boldsymbol{b}_i^l$ is the i^{th} bias for the l^{th} feature graph and N_{CK} is the number of convolutional kernels used in this convolutional layer. In this work, $\mathrm{P}_{3\times N_{\mathrm{data}}}$ can be regarded as $\mathbf{S}^0$, and the rectified linear unit (ReLU) is usually selected as the activation function. Since the features extracted by 1D-CNN follow the time series, the temporal information embedded in measurement data is retained and would be fed as inputs for GRU layers. GRU is developed from LSTM network and both of them can mitigate the issues of gradient explosion and gradient vanishing during the training process when compared with the traditional recurrent neural network (RNN) [11]. In fact, the motivations of using GRU in this work are: i) GRU can remember the previous status during training processes and fits well for time series analysis; ii) compared with LSTM, GRU only has two gates (i.e. the update gate and the reset gate), so using GRU is more computationally efficient and can reach convergence much faster [11]. It is noted that the stochastic gradient descent (SGD) method can help to improve the convergence of neural network-based algorithms and minimize the loss function as small as possible, so the SGD method is utilized for the proposed 1D-CNN-GRU-based algorithm. It is also worth mentioning that it is hard to directly give a qualitative conclusion about the convergence and optimality of a neural network-based algorithm since the activation functions of the neural network-based algorithm are highly nonlinear and nonconvex. However, the convergence and optimality of a neural network-based algorithm can be reflected by the training time and accuracy, and comparisons among different algorithms can also help to describe the capacity of convergence and optimality indirectly. For example, the results in Table 1 and Table 2 show that the proposed 1D-CNN-GRU-based algorithm can achieve general convergence and best optimality when compared with other algorithms. The internal structure of GRU is shown in Fig. 2; where $\boldsymbol{x}_t$ denotes the input sequence of GRU, which is obtained by reshaping the feature graph $\mathbf{S}^l$; and $\boldsymbol{h}_t$ denotes the output sequence, which is the predicted value of GRU. Besides, $\boldsymbol{r}_t$, $\boldsymbol{z}_t$, and $\boldsymbol{d}_t$ are the intermediate sequences and they are respectively determined as

$$\boldsymbol{r}_t = \boldsymbol{\sigma}(\mathbf{W}_r\boldsymbol{x}_t + \mathbf{U}_r\boldsymbol{h}_{t-1} + \boldsymbol{b}_r) \tag{2}$$

$$\boldsymbol{z}_t = \boldsymbol{\sigma}(\mathbf{W}_z\boldsymbol{x}_t + \mathbf{U}_z\boldsymbol{h}_{t-1} + \boldsymbol{b}_z) \tag{3}$$

$$\boldsymbol{d}_t = \mathbf{tanh}[\mathbf{W}_d\boldsymbol{x}_t + \mathbf{U}_d(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}) + \boldsymbol{b}_d] \tag{4}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \odot \boldsymbol{d}_t \tag{5}$$

where $\boldsymbol{\sigma}(\cdot)$ and $\mathbf{tanh}(\cdot)$ denote the vector format of sigmoid and hyperbolic tangent functions, respectively. $\odot$ denotes the pair-wise operation. $\mathbf{W}_r$, $\mathbf{U}_r$, $\mathbf{W}_z$, $\mathbf{U}_z$, $\mathbf{W}_d$ and $\mathbf{U}_z$ are weight matrices to be trained, and $\boldsymbol{b}_r$, $\boldsymbol{b}_z$ and $\boldsymbol{b}_d$ are the bias vectors to be trained, respectively. In this work, the cross-entropy is selected as the loss function for training. In other words, the optimization model of the proposed 1D-CNN-GRU-based algorithm can be represented as

$$\min f_{\mathrm{loss}}\left(\mathbf{y}, \widehat{\mathbf{y}}\right) = \frac{1}{N_{\mathrm{batch}}}\sum_{j=1}^{N_{\mathrm{data}}}\sum_{i=1}^{N_{\mathrm{SMD}}}\left[-y_{ji}\ln\left(\widehat{y}_{ji}\right) - \left(1 - y_{ji}\right)\ln\left(1 - \widehat{y}_{ji}\right)\right] \tag{6}$$

$$\text{s.t.}(1) - (5)$$

where $\mathbf{y}$ and $\widehat{\mathbf{y}}$ are the actual and predicted class matrices of SMDs,

Table 1
Authentication accuracy by different algorithms on testing set.

Algorithm	DWT-BP [4]	MM-gcForest [6]	MM-RFC [7]	EEMD-FFT-BP [8]	1D-CNN-LSTM	1D-CNN-GRU
Accuracy	68.5%	59.6%	60.1%	78.3%	84.6%	88.2%
Required Window Length	10min			20s	10s	10s

Table 2
Authentication accuracy of different algorithms on the testing set and their training time.

Algorithm	DWT-BP [4]	MM-gcForest [6]	MM-RFC [7]	EEMD-FFT-BP [8]	1D-CNN-LSTM	1D-CNN-GRU
Accuracy	42.5%	35.6%	46.5%	74.8%	78.9%	80.2%
Training Time	1,685s	408s	862s	1,436s	4,653s	2,751s

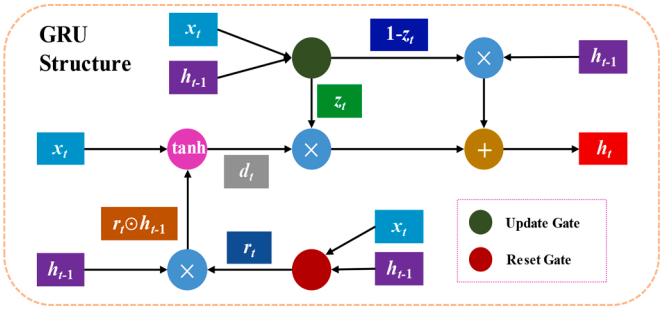

Fig. 2. Internal structure of GRU.

respectively; y_{ji} and $\hat{y}_{ji}$ are elements of the matrices, respectively. N_{batch} is the number of batches in the training process and N_{SMD} is the number of SMD's data sources to be identified. Finally, the sources of measurement data can be identified by minimizing the loss function $f_{\text{loss}}(\mathbf{y}, \hat{\mathbf{y}})$. In summary, the system model of the proposed 1D-CNN-GRU networks is shown in Fig. 3, where four 1-D convolutional layers, two pooling layers, two GRU layers, two dropout layers, one reshape layer, one fully connected layer and one softmax layer are utilized together. The objective function of this work is $f_{\text{loss}}(\mathbf{y}, \hat{\mathbf{y}})$ and this work aims to minimize it by optimizing the variables $\mathbf{K}_i$ and $\boldsymbol{b}_i^l$ in (1) and $\mathbf{W}_r$, $\mathbf{U}_r$, $\mathbf{W}_z$, $\mathbf{U}_z$, $\mathbf{W}_d$, $\mathbf{U}_z$, $\boldsymbol{b}_r$, $\boldsymbol{b}_z$ and $\boldsymbol{b}_d$ in (2)-(5).

3. Case studies

To demonstrate and compare the performance of the proposed 1D-CNN-GRU-based algorithm for data source authentication of SMDs, two cases with actual measurement data are illustrated in this section. It should be mentioned that the sizes of all convolutional kernels are tuned as L_{CK}=10 samples with experiences, thus the shape of each layer would also be determined in order according to the given input data. The first case is aiming to demonstrate the performance of the proposed algorithm for SMDs with high-reporting rate located very closely, and the second case is for demonstrating the situation of numerous SMDs in bulk actual power systems. All tests are performed on the Windows 10 platform with an Intel Core i7-9700 processor and 16GB RAM using Tensorflow backend combined with Keras in Python environment.

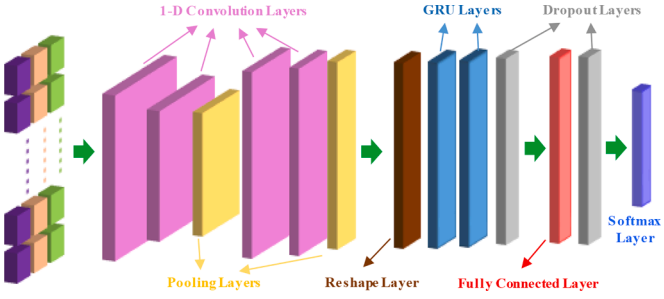

Fig. 3. Structure of the proposed 1D-CNN-GRU networks.

3.1. Case studies for high-reporting rate synchrophasor measurement devices located closely to each other

UGAs can collect measurement data up to 1.44kHz although they are not widely deployed currently. As mentioned in [8], there are three trial UGAs deployed in Knoxville, TN, USA now, and they are located closely to each other (with distances of 3.54km, 8.85km and 11.26km), which means that they are quite difficult to be distinguished. To demonstrate the algorithm proposed in this work, the measured data from 2019/07/17 to 2019/07/18 (i.e., $2 \times 24 \times 3600 \times 1440 \times 3 \approx 7.5 \times 10^8$ points) are used for verifying and comparing the performances of different algorithms, and the testing results are given in Table 1. It is noted that the results are based on 5-fold cross-validation.

It can be seen that the proposed 1D-CNN-GRU-based and 1D-CNN-LTSM-based algorithms achieve the highest and the second-highest accuracies (i.e., 88.2% and 84.6%), which shows the effectiveness of the 1D-CNN feature extraction method and the importance of considering inherent correlations among time-series data. The other four algorithms can only obtain accuracies lower than 80%. Besides, the algorithms based on DWT-BP, MM-gcForest or MM-RFC require a 10-minute window, and the EEMD-FFT-BP-based algorithm requires a 20-second window. For the proposed algorithm, a 10-second window is selected as the trade-off between accuracy and time delay of data authentication. Therefore, it can be concluded that the proposed algorithm outperforms other algorithms (i.e., higher accuracy and shorter window length) for high-reporting rate SMDs even located very closely.

3.2. Case studies for numerous synchrophasor measurement devices in large-scale power systems

Although the proposed algorithm shows good performance in Section 3.1, the number of UGAs involved in testing is too small due to their limited deployment. Therefore, case studies on FDRs, which are with a relatively lower reporting rate (i.e., 10Hz) but deployed worldwide, are employed here for further demonstrations. In this case, the measurement data of 178 FDRs from 28 countries are available and this case aims to authenticate the data sources concerning countries, which may have potential applications for forensic analysis [12] across countries. The performances of the proposed algorithm and the other algorithms are given in Table 2 together for comparisons.

It can be seen that: i) The accuracies obtained by all algorithms decrease with the increase of the number of FDRs required to be authenticated, and the accuracy of the algorithms based on DWT-BP, MM-gcForest, MM-RFC deteriorate sharply while the accuracies of the other three ones are still acceptable; ii) The proposed 1D-CNN-GRU-based algorithm outperforms the others with respect to accuracy (i.e., 80.2%) although it spends the second-longest training time (i.e., 2,751s). It should be clarified that the model training can be done in the off-line stage, so the increase in training time has little impact on practical applications as long as online authentication time is short enough. In fact, all these algorithms require less than 0.1s for online authentication. Therefore, it can be concluded that the proposed 1D-CNN-GRU-based algorithm can be applied for large-scale power systems with numerous SMDs and performs better than other algorithms.

4. Conclusions

This work proposes a data source authentication algorithm for synchrophasor measurement devices based on 1D-CNN and GRU, which can be implemented in the IoT-environment for the smart industry and achieve higher accuracy with a shorter time window compared with existing algorithms. 1D-CNN is utilized for extracting the temporal features contained in measured data and these features are fed as sequence input for GRU network. Compared with previous data authentication algorithms, the proposed one can achieve higher accuracy with the shorter window length.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Funding for this research was provided by the NSF Cyber-Physical Systems (CPS) Program under award number 1931975. This work also made use of Engineering Research Center shared facilities supported by the Engineering Research Center 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] Ö. Tuttokmağı, A. Kaygusuz, Smart grids and industry 4.0, in: Proceedings of International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1–6.
[2] N.A. Qarabsh, S.S. Sabry, H.A. Qarabash, Smart grid in the context of industry 4.0: an overveiw of communication technologies and challenges, Indonesian J. Electr. Eng. Comput. Sci. 18 (2) (2020) 656–665. May.
[3] D. Çelik, M.E. Meral, Current control based power management strategy for distributed power generation system, Control Eng. Pract. 82 (2019) 72–85. Jan.
[4] H. Liu, J. Li, J. Li, J. Tian, T. Bi, K.E. Martin, Q. Yang, Synchronised measurement devices for power systems with high penetration of inverter-based renewable power generators, IET Renew. Power Gener. 13 (1) (2019) 40–48. Jan.
[5] W. Yao, J. Zhao, M.J. Till, S. You, Y. Liu, Y. Cui, Y. Liu, Source location identification of distribution-level electric network frequency signals at multiple geographic scales, IEEE Access 5 (2017) 11166–11175. May.
[6] Y. Cui, F. Bai, Y. Liu, Y. Liu, A measurement source authentication methodology for power system cyber security enhancement, IEEE Trans. Smart Grid 9 (4) (2018) 3914–3916. Jul.
[7] Y. Cui, F. Bai, Y. Liu, P.L. Fuhr, M.E. Morales-Rodriguez, Spatio-temporal characterization of synchrophasor data against spoofing attacks in smart grids, IEEE Trans. Smart Grid 10 (5) (2019) 5807–5818. Sep.
[8] S. Liu, S. You, H. Yin, Z. Lin, Y. Liu, W. Yao, L. Sundaresh, Model-free data authentication for cyber security in power systems, IEEE Trans. Smart Grid 11 (5) (2020) 4565–4568. Sep.
[9] Y. Liu, W. Yao, D. Zhou, L. Wu, S. You, H. Liu, L. Zhan, J. Zhao, H. Lu, W. Gao, Y. Liu, Recent developments of FNET/GridEye - a situational awareness tool for smart grid, CSEE J. Power Energy Syst. 2 (3) (2016) 19–27. Sep.
[10] L. Cao, Y. Wang, B. Zhang, Q. Jin, A.V. Vasilakos, GCHAR: An efficient group-based context—aware human activity recognition on smartphone, J. Parallel Distrib. Comput. 118 (2018) 67–80. Aug.
[11] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, 2014 arXiv: 1412.3555.
[12] C. Grigoras, Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis, Forensic Sci. Int. 167 (2) (2007) 136–145. Apr.