Highlights

An Intrusion Detection Method for Industrial Control Systems Based on Bidirectional Simple Recurrent Unit

Jie Ling, Zhishen Zhu, Yu Luo, Hao Wang

- An intrusion detection method for ICS based on BiSRU is proposed.
- Skip connection is employed to alleviate the vanishing gradient problem.
- Bidirectional structure optimization is used to improve the training effectiveness.
- The proposed method has higher accuracy and shorter training time than other methods.

An Intrusion Detection Method for Industrial Control Systems Based on Bidirectional Simple Recurrent Unit

Jie Ling[a], Zhishen Zhu[a], Yu Luo[a,*], Hao Wang[b]

[a]*Guangdong University of Technology, China*
[b]*Norwegian University of Science and Technology, Norway*

Abstract

With the development of computer and network technologies, the original security of industrial control systems (ICSs) has been compromised, and security issues have become increasingly prominent. Effective intrusion detection methods for ICSs have been proposed. Recently, intrusion detection methods based on deep learning, such as long short-term memory and gated recurrent units, have immensely improved the detection accuracy compared with traditional methods. However, there are still problems that remain to be solved, such as vanishing gradient and low training efficiency. Therefore, this study proposed an intrusion detection method based on a bidirectional simple recurrent unit (BiSRU). With skip connections employed, the optimized bidirectional structure in the SRU neural network is able to alleviate the vanishing gradient problem and improve the training effectiveness. Two standard industrial datasets from Mississippi State University are used in the simulation. The results show that the proposed method is

*Corresponding author.
Email address: yuluo@gdut.edu.cn (Yu Luo)

more accurate and requires less training time than other methods.
Keywords: industrial control system, intrusion detection, deep learning, neural network, bidirectional simple recurrent unit

1. Introduction

Early industrial control systems (ICSs) use special networks and operating systems that have no connection to the Ethernet or Internet, and there are basically no network security issues [1]. With the development of computer and network technologies, many ICSs have used the Ethernet, wireless network equipment, and general operating systems to connect with management systems and remote terminals [2]. The connections between ICSs and the Internet have become increasingly concentrated, and security issues have become increasingly prominent. Meanwhile, new vulnerabilities are increasingly discovered in ICSs, including supervisory control and data acquisition (SCADA) systems, distributed control systems, and programmable logic controllers. Attacks on ICSs via the Internet continue to occur, causing potentially serious safety hazards to the industrial Internet [3]. Therefore, the security capability of ICSs should be improved.

The intrusion detection of ICSs has been widely investigated in recent years. Some scholars have proposed intrusion detection methods for ICSs based on recurrent neural network (RNN). However, time-series algorithms have two shortcomings. First, shallow architectures cannot correctly identify minority class examples with complex features. With the increase in network layers, vanishing gradients seriously degenerate the model and make it difficult to converge, resulting in low accuracy of intrusion detection. Second, scaling recurrent networks, such as long short-term memory (LSTM) and gated recurrent unit (GRU), suffer from the time dependence of state computations, i.e., the computation of each step is suspended until the complete execution of the previous step. This sequential dependency causes recurrent networks to be slower than other models and limits their parallelizability. This paper improves the simple recurrent unit (SRU) [4] to the bidirectional SRU (BiSRU) model to solve the two above problems. Skip connection [5] are applied to alleviate vanishing gradients, and bidirectional structure optimization improves the accuracy of ICS intrusion detection. BiSRU is compared with LSTM, GRU, CNN and three traditional machine learning methods through simulations with the gas pipeline and water storage tank standard industrial datasets of the Mississippi State University Center for critical infrastructure protection [6].

The rest of this paper is organized as follows: Section 2 introduces the related works. Section 3 presents the existing problems of RNNs. Section 4 proposes an intrusion detection method for ICS based on BiSRU. Section 5 shows the simulation details and results. Finally, we conclude our work in Section 6.

2. Related Work

In this section, we briefly survey the relevant works, including traditional intrusion detection methods and recently proposed deep learning-based methods.

Traditional machine learning-based methods used to be popular for ICS security protection. Researchers usually focus on one of the two steps in traditional methods, namely, feature extraction and classification. Shin et al. [7] first studied intrusion detection methods for wireless industrial sensor networks and designed a hierarchical framework for detection and data processing. Dai et al. [8] used different discretization and feature selection algorithms to extract the differences among multiple optimal feature subsets. Liang et al. [9] proposed an industrial network intrusion detection algorithm based on a multifeature data clustering optimization model, which selects a node with a high security coefficient as the cluster center and matches the multifeature data around the center into a cluster. The above methods pay more attention to feature selection, while some methods focus on the classification algorithm. Nader et al. [10] proposed a one-class classification for intrusion detection in SCADA systems by using the support vector data description. Ren et al. [11] proposed a detection model based on weighted naive Bayes that was optimized with the particle swarm optimization algorithm. Ponomarev et al. [12] proposed an approach to detect intrusions in network-attached ICSs by using a reduced error pruning tree (REPtree). Although machine learning-based methods have achieved good performance in recent years, they still have their own inherent defects; for example, SVMs experience a bottleneck as the number of samples grows, naive Bayes methods are not suitable for data with related attributes, and decision trees have poor generalization ability. Thus, there is an urgent need to study the intrusion detection problem and propose a method with a higher detection rate.

Fortunately, in recent years, the emerging deep learning method has achieved great success in various fields, especially in computer vision [13] and speech recognition [14]. Such success has encouraged many scholars in the security field to pursue security solutions for ICSs based on deep learning. Wei et al. [15] proposed a data traffic prediction model based on an autoregressive moving average using time series data. Wang [16] proposed a network intrusion detection system using the naive Bayes classifier and deep neural network (DNN). Yang et al. [17] proposed a deep-learrning-based network intrusion detection system and used the convolutional neural network (CNN) to extract the features. Abassi et al. [18] proposed an attack detection model that leverages DNN and decision tree classifiers to detect cyber-attacks from the new representations.

Due to the time-series attributes in network traffic data, RNNs seem to be a good choice. Fang et al. [19] proposed an intrusion detection model based on a hybrid CNN and RNN model, which can accurately identify the type of network traffic, to solve the advanced persistent threat in power information networks. Yu et al. [20] presented an ICS intrusion detection method based on LSTM to improve the insufficient timing memory capability of RNN. Xu et al. [21] introduced a novel intrusion detection system consisting of an RNN with GRU to simplify the memory unit structure of LSTM and reduce the calculation time of the algorithm while maintaining the classification accuracy. Most of the above time-series RNN methods have problems of vanishing gradients and low parallelizability, which limit their performance in ICS intrusion detection.

3. Existing Problems of RNNs

3.1. Time-consuming problems in RNN training

An RNN is a special kind of neural network with self-connections in the field of deep learning. The network state of the previous moment can be transferred to the current moment, and the state of the current moment can be transferred to the next moment through a self-recurrent connection in the hidden layer, which makes

RNNs suitable for time-series problems. Compared with CNNs, this sequential dependency makes LSTM, GRU, or other RNNs unable to be parallelized, thereby limiting the training speed of the model.

Taking LSTM as an example, the calculation process of LSTM is as follows.

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right), \qquad h_t = o_t \odot \tanh\left(c_t\right). \tag{1}$$

where x_t, c_t and o_t are the input, memory unit and output gate of the network at time step t, respectively; W_o and b_o are the weight and bias parameters; h_{t-1} and h_t are the output of the previous layer and the current layer, respectively; σ is the activation function; and $\odot$ is the dot product. From Equation (1), it can be seen that output h_t of the current moment indirectly depends on output h_{t-1} at the previous time, which limits the parallelizability and increases the training time. Similarly, the same sequential dependence situation occurs in GRU.

3.2. Vanishing Gradient Problem

During the gradient descent calculation, the chain rule is used to conduct error backpropagation to obtain the minimum partial derivative of the loss function of the hidden state. The specific recurrence formula is expressed as follows.

$$\frac{\partial l}{\partial h_0} = \left(\frac{\partial h_t}{\partial h_0}\right)^T * \frac{\partial l}{\partial h_t} = \left(\sum_{i=1}^{r} m_i^t u_i v_i^T\right)^T * \frac{\partial l}{\partial h_t} = \sum_{i=1}^{r} m_i^t u_i v_i * \frac{\partial l}{\partial h_t}. \tag{2}$$

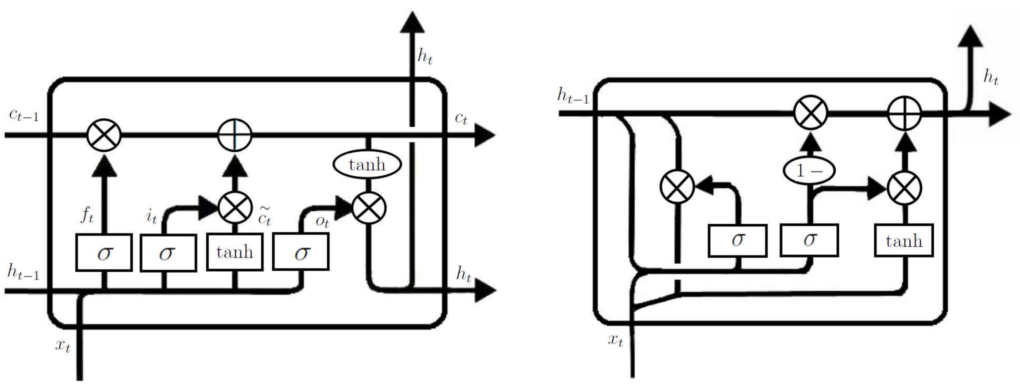
where l is the loss function, h_0 is the hidden state, h_t is the output gate of the network at time step t, and m, u, and v are variables during singular value decompositions. When t is large, the partial derivative value of the loss function to the cell state only depends on the maximum singular value m_i. The t power of m_i tends to be infinitesimal and causes the gradient to vanish when $m_i < 1$. The vanishing gradient problem invalidates the gradient descent algorithm and reduces the long-distance dependence of the NN.

Existing Solution for Vanishing Gradients: To alleviate gradient vanishing in the back propagation of RNNs, a complex structure composed of gates is introduced to control the information flow to hidden neurons and to ensure that the feedback path can enable timely and effective gradient calculation feedback. L-STM and GRU are two typical examples.

(1) In LSTM, input, output, and forget gates are added to the neurons of an RNN. A forget gate can alleviate the vanishing gradient when it propagates backward with the time series. A forget gate controls the self-connection unit and determines which parts of the historical information should be discarded, and the calculation is as follows.

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \,. \tag{3}$$

where c_{t-1} and c_t are the memory units at the current unit and previous unit, respectively, and i_t, f_t and $\widetilde{c_t}$ are the input gate, the forget gate and the new status information, respectively. The LSTM cell is illustrated in Figure 1(a).



(a) LSTM (b) GRU

Figure 1. Structure of cells

(2) The GRU model is a simplified structure of LSTM. The input and forget gates of LSTM are combined to form a new control gate, that is, the update gate. Update gate z_t can determine whether the current data are important to the entire model and whether to ignore the current input data. The calculation is as follows.

$$z_t = \sigma\left(W_z[h_{t-1}, x_t]\right). \tag{4}$$

where W_z is the weight parameter of the update gate, h_{t-1} is the output of the previous step, x_t is the input of the current step, and σ is the activation function. The GRU cell is illustrated in Figure 1(b).

Relying on the complex structure composed of gates, the gradient information can provide feedback, and to some extent, the LSTM and GRU models can alleviate the vanishing gradient problem. However, in the industrial control intrusion detection tasks, when increasing the depth of LSTM and GRU, the vanishing gradient problem still exists, which makes the convergence of the final model difficult and decreases the accuracy of intrusion detection.

4. Proposed Approach

As mentioned above, most of the existing RNN methods have the problems of vanishing gradients and low effectiveness in model training. To address the above two problems, in this paper, we propose a bidirectional simple recurrent unit (BiSRU)-based intrusion detection model in which a simple recurrent unit (SRU) is used to replace the LSTM and GRU to reduce the training time. The skip connection strategy is employed to alleviate the vanishing gradient problem, and moreover, the bidirectional structure can be used to better extract the sequence feature information.

4.1. SRU

An SRU is designed to facilitate the training of deep models with highly parallelized implementation [4]. Due to the efficiency of SRU, it is utilized to replace LSTM and GRU to improve the efficiency of model training in RNN. The main improvements of SRUs are twofold: the dependence of the current time step on the previous time step is completely eliminated, and the use of parallel computations accelerate the training of the model.

The SRU mainly includes a forget gate and a memory unit. The forget gate, which indicates the importance of the previous step to the current state, is used to adjust the memory unit. The memory unit is used to calculate the final output state. Typically, a single layer of an SRU involves the following computations:

$$\widetilde{x}_t = W_x x_t, \tag{5}$$

$$f_t = \sigma\left(W_f x_t + b_f\right), \tag{6}$$

$$c_t = f_t \odot c_{t-1} + (1 - f_t) \odot \widetilde{x}_t, \tag{7}$$

$$h_t = g\left(c_t\right), \tag{8}$$

where the subscript t is the time step, x_t is the input, W_x and W_f are the weight parameters, b_f is the bias and σ and g are the activation functions. $\widetilde{x}_t$ in Equation (5) is the temporary state. f_t in Equation (6) is the forget gate, which indicates the importance of the previous step to the current state. c_t in Equation (7) is the memory unit. h_t in Equation (8) is the output of the network. As seen from the above equations, the conversion between the gate control unit and the input only depends on the input of the current time step. Thus, the matrix operation with a large amount of calculations can be processed in parallel. Although the calculation of memory unit c_t still depends on the previous time step, the calculation of c_t and h_t in the SRU only involves point multiplication with low computational cost.

4.2. Skip Connection

Skip connections [22] from hidden layers to output layers have long been used in NNs and can alleviate the vanishing gradient problem. Thus, we apply it to the final state output calculation of the SRU to alleviate the vanishing gradient in deep NNs. First, reset gate r_t is set as follows:

$$r_t = \sigma\left(W_r x_t + b_r\right). \tag{9}$$

Then, output state h_t is calculated by skip connections:

$$h_t = r_t \odot g\left(c_t\right) + \left(1 - r_t\right) \odot x_t. \tag{10}$$

$(1 - r_t) \odot x_t$ allows the gradient to directly propagate to the previous layer with skip connections, which is equivalent to adding one to the partial derivative of the cell-state loss function: $\frac{\partial l}{\partial h} = \frac{\partial(f+h)}{\partial h} = 1 + \frac{\partial f}{\partial h}$. The method can effectively back propagate the error and alleviate the vanishing gradient problem, although the value of the derivative is extremely small. The SRU with skip connections is shown in Figure 2.

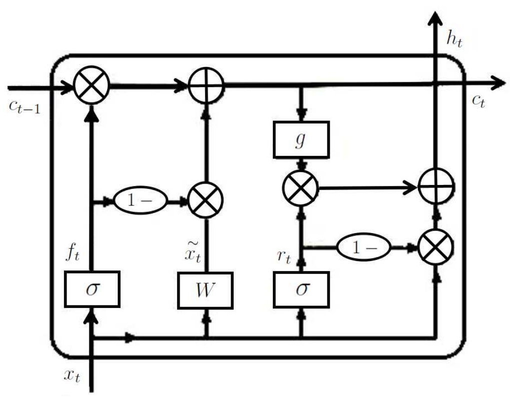


Figure 2. SRU cell

4.3. *BiSRU*

The traditional time-series model, which usually reads the sample sequence from front to back, can obtain the forward information of the sample sequence. However, this method is unsuitable for sample information with complex sequences and uncertain correlations and affects the subsequent sample analysis.

Therefore, this paper uses a bidirectional structure to effectively obtain the sequence feature information in the intrusion detection samples. The structure of the BiSRU model is shown in Figure 3.

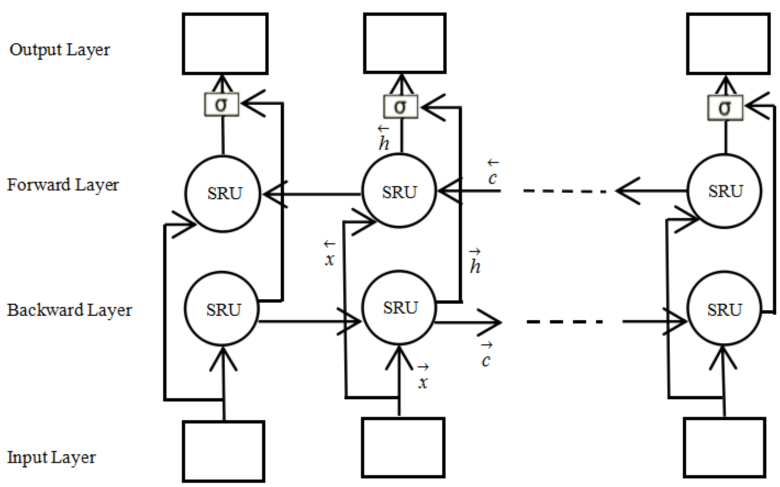


Figure 3. Structure of BiSRU

In Figure 3, $\overrightarrow{x}$ and $\overleftarrow{x}$ are the forward and reverse readings of the sample sequence, respectively. $\overrightarrow{c}$ and $\overleftarrow{c}$ are the memory units of the forward and reverse SRUs, respectively. $\overrightarrow{h}$ and $\overleftarrow{h}$ are the output states of the forward and reverse SRUs, respectively.

The overall flow of ICS intrusion detection based on BiSRU is shown in Figure 4.

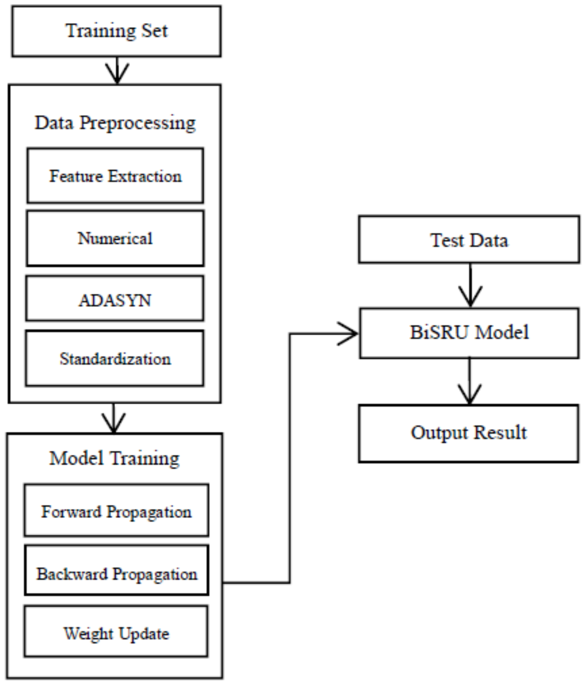


Figure 4. Flowchart of the intrusion detection model

5. Simulation Experiments and Results Analysis

5.1. Implementation Details

The proposed method is compared with three traditional machine learning-based methods (Naive Bayes [11], SVM [10] and REPtree [12]) and three deep-learning-based methods, including two methods with RNN structures (LSTM [20] and GRU [21]) and one CNN model [17]. Experiments were conducted on a workstation with an AMD Ryzen 5 2600 six-core processor@ 3.85 GHz, 16 GB

RAM, GTX 1660Ti@6G GPU and a Windows 10 64-bit operating system. We used the latest version of Keras packages for the implementation of the BiSRU model. The specific parameters of the simulation platform are presented in Table 1.

Table 1

Experimental parameters

Parameter name	**Description**	**Value(Gas)**	**Value(Water)**
depth	Hidden layer size	4	5
optimizer	Gradient descent algorithm	Adam	Adam
activation	Activation function	softmax	softmax
epochs	Iteration size	20	8
batch_size	Samples per epoch	128	128
unit	Hidden unit size	128	128
dropout	Random deactivation rate	0.1	0.1

5.2. Dataset

Table 2

Description of datasets

attack type	Describe	Number(Gas)	Number(Water)
Normal	Normal data	61156	172415
NMRI	Naive malicious response injection attack	2763	9187
CMRI	Complex malicious response injection attack	15466	24920
MSCI[1]	Malicious state command injection attack	782	1833
MPCI	Malicious parameter command injection attack	7637	3725
MFCI	Malicious function command injection attack	573	1320
DoS	Denial-of-service attack	1837	1237
RECO	Reconnaissance attack	6805	34002

The gas pipeline and water storage tank standard industrial datasets are used, which were proposed by Mississippi State University in 2014. As a relatively complete datasets, they have been used for simulation experiments of ICS intrusion detection in recent years [24, 25]. The first dataset is collected from a set of gas pipeline systems based on Modbus-TCP, which has a similar composition and structure as the SCADA system in the actual production environment. The gas pipeline dataset contains large-scale samples of normal data and seven types of attack data (61156 benign samples and 35863 malicious samples). The water storage tank dataset contains normal data and seven types of attack data and has

[1]The number of MSCI samples is small, and the characteristics of MSCI are complex and easy to be mistakenly detected as detected as CMRI; we used ADASYN [23] to synthesize those data.

enough samples (172415 benign samples and 76224 malicious samples); the details can be seen in Table 2. Before feeding into the model, the data need to be preprocessed by min-max standardization and one-hot encoding. The dimensions of the input vector for the two datasets are 26 and 23, respectively.

5.3. Benchmarking Metrics

The overall accuracy (ACC), true positive rate (TPR), false positive rate (FPR), and false negative rate (FNR) are used as key performance indicators to evaluate the proposed method. Because the datasets of gas pipeline systems and water storage tank systems are imbalanced, we introduced the Matthews correlation coefficient (MCC) to evaluate the performance. The calculations of the five metrics are as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \tag{11}$$

$$TPR = \frac{TP}{TP + FN}, \tag{12}$$

$$FPR = \frac{FP}{FP + TN}, \tag{13}$$

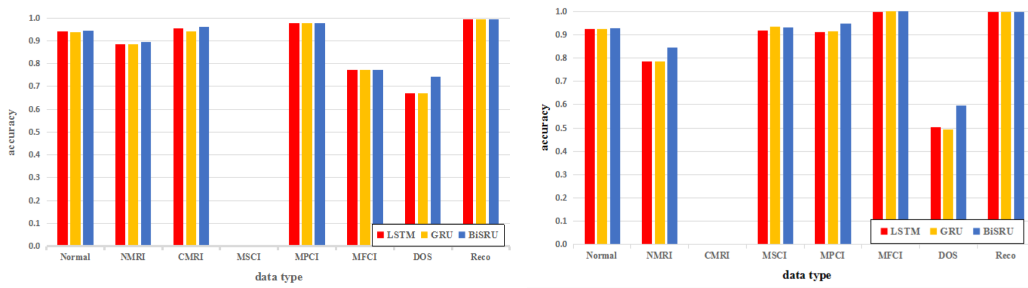
$$FNR = \frac{FN}{TP + FN}, \tag{14}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \tag{15}$$

where TP represents the number of detected benign samples. TN denotes the number of detected malicious samples. FP is the number of malicious samples detected as benign, and FN indicates the number of benign samples detected as malicious.

5.4. Selection of network layers

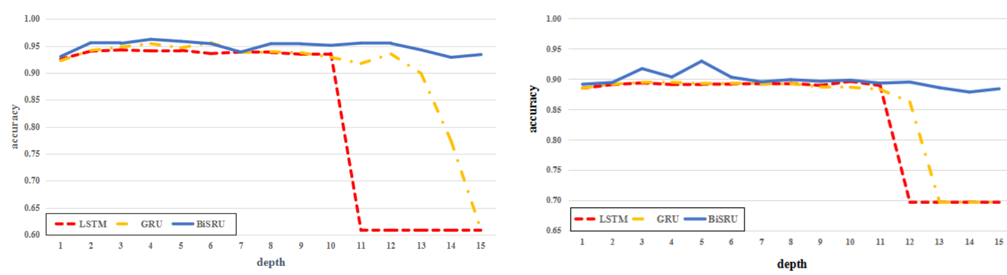
To verify the effectiveness of BiSRU in ICSs, we conduct an ablation study by comparing the proposed BiSRU with two other RNN structures, LSTM [20] and GRU [21]. When selecting the number of network layers, accuracy and computational time are the main concerns. Since the time consumption will increase with an increase in the number of network layers, NNs with a low number of layers are selected when accuracy is ensured. Therefore, a set of experiments to determine the influence of the number of hidden layers in an RNN is designed, and the NN structure as the number of hidden layers is varied from 1 to 15 is tested.



(a) gas pipeline, depth=1

(b) water storage tank, depth=1

Figure 5. Detection results of normal data and various types of attack data



(a) gas pipeline

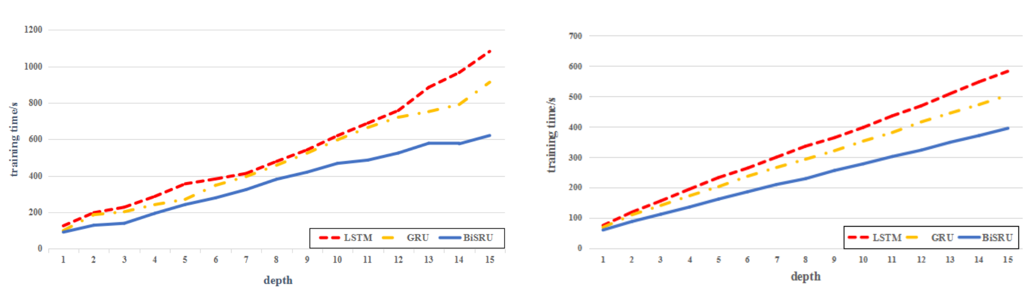
(b) water storage tank

Figure 6. Accuracy comparison of different depths

As shown in Figure 5(a) and Figure 5(b), the recognition rates of the three algorithms for MSCI data in the gas pipeline dataset and CMRI data in the water storage tank dataset in the one hidden layer NN are low and are nearly 0%. The complexity of the model should be improved, and a NN with a deeper hidden layer should be used to detect such data.

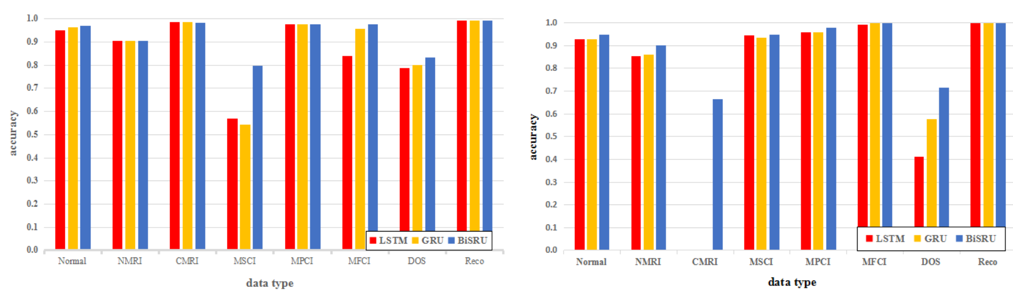
As shown in Figure 6, the three network models with different depths are compared. Overall, the three models achieve the highest accuracy in the network with 4 - 5 hidden layers on the gas pipeline dataset and water storage tank dataset. Meanwhile, the vanishing gradient in LSTM and GRU appears when using 11 and 13 hidden layers, respectively, on the gas pipeline dataset and 11 and 12 hidden layers, respectively, on the water storage tank dataset, causing the accuracy to drop to approximately 60%. Meanwhile, BiSRU does not exhibit a vanishing gradient until the 15 hidden-layer NN, and its accuracy remains higher than 92% on both datasets. This verifies the robustness of BiSRU, as its performance is not affected by different depths of the hidden layers, and it can effectively alleviate the vanishing gradient problem.

As shown in Figure 7, BiSRU has the shortest model training time compared with LSTM and GRU. The results of the proposed method exceed other techniques with a flat curve in all metrics on the gas pipeline and water storage tank dataset.



(a) gas pipeline (b) water storage tank

Figure 7. Training time comparison of different depths



(a) gas pipeline, depth=4 (b) water storage tank, depth=5

Figure 8. Detection accuracy for normal data and various types of attack data

5.5. Results

Experiments were conducted with the same hardware, software environment, and algorithm parameters. The ratio of the training set to the test data was 8:2. The results were compared in terms of the metrics ACC, TPR, FPR, FNR and

MCC for each classification algorithm.

As shown in Figure 8, the accuracy of BiSRU for MSCI data in the gas pipeline dataset and CMRI data in the water storage tank dataset are higher than those of the other algorithms, and BiSRUs recognition rate for other samples is basically the same as those of LSTM and GRU.

Table 3

Benchmarking metrics for the different algorithms (gas pipeline)

	ACC/%	TPR/%	FPR/%	FNR/%	MCC
Naive Bayes	93.52	96.55	11.52	3.44	86.11
SVM	95.35	96.79	7.45	3.00	89.99
REPtree	92.85	94.98	9.21	5.02	85.80
CNN	95.97	94.66	1.74	5.34	91.58
LSTM	94.09	95.23	7.67	4.77	87.61
GRU	95.43	94.04	2.13	5.96	90.46
BiSRU	96.23	97.28	5.91	2.31	92.15

Table 4

Benchmarking metrics for the different algorithms (water storage tank)

	ACC/%	TPR/%	FPR/%	FNR/%	MCC
Naive Bayes	58.91	43.85	0.33	56.14	41.37
SVM	92.70	96.33	16.49	3.67	81.72
REPtree	92.10	99.60	26.92	0.39	80.43
CNN	89.95	87.41	0.10	12.59	76.42
LSTM	89.15	86.54	0.10	13.45	74.53
GRU	89.27	86.67	0.10	13.32	74.82
BiSRU	92.94	96.00	13.56	4.00	83.60

As shown in Table 3, the benchmarking metrics for the 8 algorithms on the gas pipeline dataset and water storage tank dataset were compared. Compared with other methods, BiSRU has the highest ACC, TPR and MCC and the lowest FNR on the gas pipeline dataset, and its FPR is slightly higher than that of the CNN and GRU. Through the comprehensive evaluation, BiSRU obtains the best intrusion detection effectiveness on this dataset. As shown in Table 4, because the water storage tank dataset is imbalanced, we should pay more attention to the MCC data. Although other methods have one or more benchmarking metrics (TPR, FPR or FNR) that are better than those of our proposed method, BiSRU maintains consistent results in all 5 metrics, especially for MCC, which shows that BiSRU has the best performance on the imbalanced dataset. Therefore, BiSRU is suitable for large-scale high-dimensional network traffic data generated by the SCADA system.

6. Conclusion

Vanishing gradients and model training inefficiency emerge when the recurrent neural networks deal with large-scale network traffic data in industrial control systems that are high-dimensional and time-series. This study proposed an intrusion detection method for industrial control systems based on Bidirectional simple recurrent unit, which introduces skip connections and bidirectional structure optimization, to solve these problems. Two datasets proposed by the key infrastructure protection center of Mississippi State University are used in the simulation experiments. The results show that the proposed model has superior performance to the other six companion methods. Additionally, compared to the other two recurrent neural networks, long short-term memory and gated recurrent unit, the proposed model has higher accuracy and shorter training time. In future work, the optimization of neural network performance considering false positive rate and the recognition of unknown attack types will be studied.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 61702112), the Key Areas Research and Development Program of Guangdong Province (grant No. 2019B010139002) and the project of Guangzhou Science and Technology (grant No. 201902020006 & No. 201902020007).

References

[1] Wang, X., Li, J., Tan, Z., Ma, L., Li, F., Huang, M.. The state of the art and future tendency ofinternet+oriented network technology. Journal of Computer Research and Development 2016;53(4):729–741. doi:10.7544/issn1000-1239.2016.20151146.

[2] Suo, Y., Wang, S., Qin, Y., Li, Q., Feng, D., Li, J.. The state of summary of security technology and application in industrial control system. Computer Science 2018;45(4):25–33. doi:10.11896/j.issn.1002-137X.2018.04.004.

[3] Zhang, Q., Zhou, C., Tian, Y., Xiong, N., Qin, Y., Hu, B.. A fuzzy probability bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems. IEEE Transactions on Industrial Informatics 2017;14(6):2497–2506. doi:10.1109/TII.2017.2768998.

[4] Lei, T., Zhang, Y., Wang, S.I., Dai, H., Artzi, Y.. Simple recurrent units for highly parallelizable recurrence. arXiv preprint arXiv:170902755 2017;doi:10.18653/v1/D18-1477.

[5] Orhan, A.E., Pitkow, X.. Skip connections eliminate singularities 2018;.

[6] Morris, T., Gao, W.. Industrial control system traffic data sets for intrusion detection research. vol. 441. 2014, p. 65–78. doi:10.1007/978-3-662-45355-1_5.

[7] Shin, S., Kwon, T., Jo, G.Y., Park, Y., Rhy, H.. An experimental study of hierarchical intrusion detection for wireless industrial sensor networks. IEEE Transactions on Industrial Informatics 2010;6(4):744–757. doi:10.1109/TII.2010.2051556.

[8] Dai, Y., Chen, X., Chen, H., Ye, J., Lin, J., Guo, W.. Feature selection based approach to network intrusion detection. Application Research of Computers 2017;34(8):2429–2433. doi:10.3969/j.issn.1001-3695.2017.08.043.

[9] Liang, W., Li, K., Long, J., Kui, X., Zomaya, A.Y.. An industrial network intrusion detection algorithm based on multifeature data clustering optimization model. IEEE Transactions on Industrial Informatics 2020;16(3):2063–2071. doi:10.1109/TII.2019.2946791.

[10] Nader, P., Honeine, P., Beauseroy, P.. l p -norms in one-class classification for intrusion detection in scada systems. IEEE Transactions on Industrial Informatics 2014;10(4):2308–2317. doi:10.1109/TII.2014.2330796.

[11] Ren, X., Jiao, W., Zhou, D.. Intrusion detection model of weighted navie bayes based on particle swarm optimization algorithm. Computer Engineering and Applications 2016;52(7):122–126. doi:10.3778/j.issn.1002-8331.1405-0141.

[12] Ponomarev, S., Atkison, T.. Industrial control system network intrusion detection by telemetry analysis. IEEE Transactions on Dependable and Secure Computing 2016;13(2):252–260. doi:10.1109/TDSC.2015.2443793.

[13] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al. A survey on deep learning in medical image analysis. Medical image analysis 2017;42:60–88. doi:10.1016/j.media.2017.07.005.

[14] Chandna, P., Miron, M., Janer, J., Gomez, E.. Monoaural audio source separation using deep convolutional neural networks 2017;(2017):258–266. doi:10.1007/978-3-319-53547-0_25.

[15] Wei, M., Kim, K.. Intrusion detection scheme using traffic prediction for wireless industrial networks. Journal of Communications and Networks 2012;14(3):310–318. doi:10.1109/JCN.2012.6253092.

[16] Wang, K.. Network data management model based on nave bayes classi-

fier and deep neural networks in heterogeneous wireless networks. Computers and Electrical Engineering 2019;75:135–145. doi:10.1016/j.compeleceng.2019.02.015.

[17] Yang, H., Cheng, L., Chuah, M.C.. Deep-learning-based network intrusion detection for scada systems. In: 2019 IEEE Conference on Communications and Network Security (CNS). 2019, p. 1–7. doi:10.1109/CNS.2019.8802785.

[18] Al-Abassi, A., Karimipour, H., Dehghantanha, A., Parizi, R.M.. An ensemble deep learning-based cyber-attack detection in industrial control system. IEEE Access 2020;8:83965–83973. doi:10.1109/ACCESS.2020.2992249.

[19] Fang, Y., Li, M., Wang, P., Jiang, X., Zhang, X.. Intrusion detection model based on hybrid convolutional neural network and recurrent neural network. Journal of Computer Applications 2018;38(10):2903–2907,2917. doi:10.11772/j.issn.1001-9081.2018030710.

[20] Yu, B., Wang, H., Yan, B.. Intrusion detection of industrial control system based on long short term memory. Information and Control 2018;47(1):54–59. doi:10.13976/j.cnki.xk.2018.0054.

[21] Xu, C., Shen, J., Du, X., Zhang, F.. An intrusion detection system using a deep neural network with gated recurrent units. IEEE Access 2018;6:48697–48707. doi:10.1109/access.2018.2867564.

[22] Srivastava, R.K., Greff, K., Schmidhuber, J.. Training very deep networks. arXiv: Learning 2015;.

[23] He, H., Bai, Y., Garcia, E.A., Li, S.. Adasyn: Adaptive synthetic sampling approach for imbalanced learning 2008;:1322–1328doi:10.1109/IJCNN.2008.4633969.

[24] Khan, I.A., Pi, D., Khan, Z.U., Hussain, Y., Nawaz, A.. Hml-ids: A hybrid-multilevel anomaly prediction approach for intrusion detection in scada systems. IEEE Access 2019;7:89507–89521. doi:10.1109/access.2019.2925838.

[25] Haghnegandar, L., Wang, Y.. A whale optimization algorithm-trained artificial neural network for smart grid cyber intrusion detection. Neural Computing and Applications 2020;32(13):9427–9441. doi:{10.1007/s00521-019-04453-w}.

Biography

Jie Ling is a professor at Guangdong University of Technology, Guangzhou, China. He received the Ph.D. degree from The Sun Yat-Sen University, China, in 1998. His research interests include computer applications and intelligent video processing technology.

Zhishen Zhu is a graduate student at Guangdong University of Technology, Guangzhou, China. His current research interests include machine learning and intrusion detection in industrial systems.

Yu Luo is a lecturer at Guangdong University of Technology, Guangzhou, China. She received the Ph.D. degree from The South China University of Technology, Guangzhou, China, in 2016. Her research interests include image recovery, dictionary learning, and pattern recognition.

Hao Wang is currently an Associate Professor with the Norwegian University of Science and Technology, Norway. He received the Ph.D. degrees from The South China University of Technology, Guangzhou, China, in 2006. His current research interests include big data analytics, industrial internet of things, high performance computing, safety-critical systems, and communication security.