Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4252766, 8 pages
https://doi.org/10.1155/2022/4252766

Research Article

Network Flow Anomaly Detection Based on Improved Echo State Network

Mingzhong Chen,[1] **Bin Qiu**,[2] **and Jie Ji**[3]

[1]*Department of Mechanical and Electrical Engineering, Shantou Polytechnic, Shantou 515078, China*
[2]*Departments of Computer, Shantou Polytechnic, Shantou 515078, China*
[3]*Network and Information Center, Shantou University, Shantou 515063, China*

Correspondence should be addressed to Mingzhong Chen; cmzgjjx@163.com

Received 20 March 2022; Accepted 23 June 2022; Published 8 July 2022

Academic Editor: Pierre-Martin Tardif

Aiming at the problems existing in the current network flow anomaly detection, a network flow prediction model based on echo state network of double loop reserve pool is designed, which solves the problem of reserve pool generated randomly in traditional echo state network and improves the accuracy and instantaneity of network flow prediction. Then, an anomaly detection method based on dynamic threshold is proposed, which takes the difference between the predicted value and the real value as the basis for judging the occurrence of anomalies. Simulation results show that the improved prediction model and anomaly detection method can effectively detect the abnormal behavior of network flow, and the detection effect is better than other models.

1. Introduction

With the continuous development of information technology, there are more and more applications appearing on the Internet. At present, network has become the main correspondence carrier. But the network attacking behavior is increasing day by day, and the network security problem is becoming more and more prominent. Users' network behavior is reflected in network flow, which is characterized by nonlinear, timeliness, suddenness, and diversity [1]. Traditional linear prediction methods have been unable to adapt to the requirements of modern network development. In recent years, nonlinear prediction models such as neural network, grey model, support vector machine, and hybrid model are widely used in scene prediction with good nonlinear mapping and flexible learning method. However, there are still problems such as low prediction accuracy and slow prediction speed [2]. Under this situation, echo state network (ESN), a new recursive neural network, emerges at the historic moment. With a reserve pool structure, ESN shows excellent nonlinear processing ability and fast prediction speed. However, traditional ESNs use a single-ring reserve pool, and the structure and weights of the reserve pool are randomly generated. Even reserve pools with the same parameters may exhibit significant performance differences [3]. Moreover, the training time complexity is high. These factors will affect the nonlinear characterization and real-time prediction of network flow by ESN.

Therefore, in order to ensure the nonlinearity and real-time performance of network flow prediction, this paper designs a network flow anomaly detection method based on ESN of double loop reserve pool (ESN-DLRP). First, network flow is normalized and denoised by wavelet. Second, the multiring reserve pool ESN is used to construct network flow prediction model, which solves the reserve pool problem generated randomly by traditional ESN and improves the accuracy and instantaneity of network flow prediction. Then, we can get the diagram of true value and predicted value of the network flow. Finally, an anomaly detection method based on dynamic threshold is proposed to extract packets with large deviation between the true value and the predicted value. Simulation results show that this model

can detect abnormal network flow behavior effectively, and the detection effect is better than other models.

2. Correlation Theory and Technology Study

2.1. Wavelet Transformation and Wavelet Denoising. The wavelet transform is to transform the infinite length of trigonometric function basis into a finite length of decaying wavelet basis. It is shown in Figure 1.

Therefore, wavelet function is expressed as follows:

$$W_{a,\tau} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \psi\left(\frac{t-\tau}{a}\right) dt. \quad (1)$$

The formula has two variables: one is a (scale) to control the stretch of wavelet basis; the other one is τ (translation) controls of wavelet base translation. The wavelet transform of a given signal is to expand the signal according to a small cluster of wavelet functions, that is, the signal is expressed as a series of linear combinations of wavelet functions with different scales and different translations. The coefficients of each term are called wavelet coefficients.

Due to the influence of many factors, there may be some noise in network flow collection. Noise is a kind of useless signal, which not only wastes storage space and transmission time but also causes interference to network flow prediction. In general, useful network flow corresponds to the wavelet coefficient with large amplitude, while noise corresponds to the wavelet coefficient with small amplitude [4]. So we choose a threshold λ to compare with wavelet coefficients. If wavelet coefficients is lower than λ, it can be considered as noise and must be removed to obtain the network flow after denoising [5]. Equation (2) below is the threshold function.

$$\widehat{W}_{a,\tau} = \begin{cases} \operatorname{sgn}(W_{a,\tau}) \times \left(|W_{a,\tau}| - \dfrac{2\lambda}{1 + e^{|W_{a,\tau}|-\lambda}}\right), & |W_{a,\tau}| \geq \lambda, \\ 0, & |W_{a,\tau}| < \lambda. \end{cases} \quad (2)$$

$W_{\alpha,\tau}$ is original wavelet coefficients. $\widehat{W}_{\alpha,\tau}$ is denoised wavelet coefficients. Threshold function is a continuous function, so when $|W_{a,\tau}| \geq \lambda$, the denoised network flow can be expressed as follows:

$$f(x) = \operatorname{sgn}(x)\left(|x| - \frac{2\lambda}{1 + e^{|x|-\lambda}}\right). \quad (3)$$

2.2. Traditional Echo State Network

2.2.1. Structure of Traditional Echo State Network. Echo state network (ESN for short) is a new recursive neural network composed of input layer, reserve pool, and output layer. As the core of ESN, the structure of the reserve pool is randomly generated and contains large-scale sparsely connected neurons. These neurons contain the operating state of the system and have strong nonlinear learning ability, which can predict the future value according to the known value

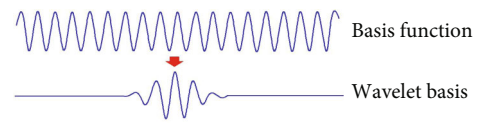

Figure 1: Transformation from basis function to wavelet basis.

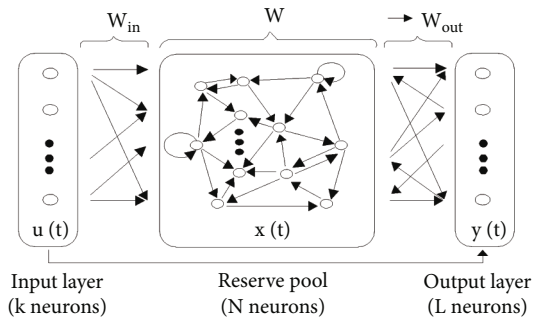

Figure 2: Traditional ESN's basic structure.

of network flow time series [6]. Figure 2 shows traditional ESN's basic structure.

If the input layer has K neurons, reserve pools has N neurons, and output layer has L neurons, when time is t, the signals constructed by the input layer, reserve pool, and output layer are the following equations, respectively:

$$u(t) = [u_1(t), u_2(t), \cdots, u_K(t)]T, \quad (4)$$

$$x(t) = [x_1(t), x_2(t), \cdots, x_N(t)]T, \quad (5)$$

$$y(t) = [y_1(t), y_2(t), \cdots, y_L(t)]T. \quad (6)$$

If the connection matrix between the input layer and reserve pool is $W_{\text{in}}\,(N \times K)$, the connection matrix inside the reserve pool is $W\,(N \times N)$, and connection matrix between the reserve pool and output layer is $W_{\text{out}}\,(L \times N)$. If the input at the current time is $u\,(t+1)$, the state equation of the reserve pool is

$$x(t+1) = f_{\text{in}}(W_{\text{in}} \times u(t+1) + Wx(t)). \quad (7)$$

The state equation of the output layer is

$$y(t+1) = f_{\text{out}}(W_{\text{out}} \times x(t+1)). \quad (8)$$

In Eq. (7), f_{in} is the excitation function of the reserve pool. $x\,(t+1)$ is the status of the reserve pool at the current time. $x\,(t)$ is the state of the reserve pool at the last moment. In Eq. (8), f_{out} is the excitation function of the output layer, and $y\,(t+1)$ is the output status at the current time.

2.2.2. ESN's Training. The data of ESN's network flow is divided into training samples and verification samples. Through the input connection matrix W_{in}, training samples $\{u\,(t+1), y\,(t+1)\}$ get into the reserve pools. In Eqs. (7) and (8), W_{in} and W are generated randomly before the training. The only factor needs to be trained is W_{out}, which

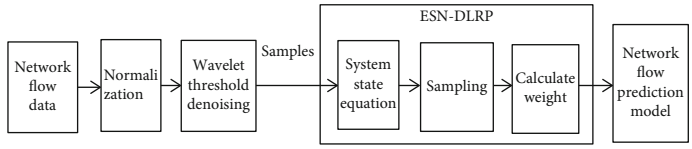

Figure 3: The construction process of network flow model prediction.

means that the training process of ESN is the process of determining W_{out} [7]. In Eq. (7), as W_{in} and W are randomly generated, we can get $x(t+1)$ immediately. In Eq. (8), W_{out} can be calculated by $x(t+1)$ and $y(t+1)$.

Compared with the traditional neural network, ESN has a stronger ability to predict network flow. However, the single-ring reserve pool is adopted, and the reserve pool structure and weight are randomly generated, which affects the nonlinear characterization and real-time prediction of network flow by ESN to a certain extent [8].

3. Design of Network Flow Prediction Model Based on Improved ESN

In order to ensure the nonlinearity and real time of network flow prediction, this paper designs a network flow prediction method with ESN of double-loop reserve pool structure. After collecting the original network flow for a Web system, it is normalized and denoised to get the denoised network flow data. Then, the network flow prediction model is constructed by using the ESN of double-loop reserve pool. Such process is shown by Figure 3.

3.1. The Prepossessing of Network Flow Data

(1) The collected original network flow is normalized to make the network flow data between [0.1,1]. Let x_{min} and x_{max} represent the minimum and maximum value of network flow, respectively, and x' represents the normalized data

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \times 0.8 + 0.1. \quad (9)$$

(2) Wavelet threshold method is used to denoise the normalized network flow and eliminate the noise to avoid interference to normal signals

(3) The network flow after denoising is divided into training sample and verification sample, and the training sample is input into ESN-DLRP for learning

3.2. Design and Prediction Method of ESN-DLRP

3.2.1. The Design of ESN-DLRP. Echo state network based on double loop reserve pool (ESN-DLRP for short) is shown in Figure 4. Each neuron in the reserve pool is connected with its neighbors in a circular manner to form the first ring. Then, the first neuron A is taken as the starting point to connect with neuron B at interval d ($d = 3$ in this paper) in a circular manner, and then B is connected with neuron C of subsequent interval d, and so on to form the second ring, thus constructing ESN-DLRP-d (d is the neuron interval of the second ring). This double-loop reserve pool structure avoids the randomness of the reserve pool generated by traditional ESN, enhances the connectivity between neurons in the pool, and improves the ability of nonlinear characterization of network flow [9].

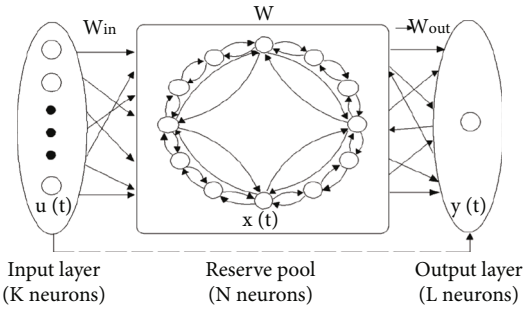

Figure 4: The structure of ESN-DLRP.

ESN-DLRP's prediction principle is using input layer's network flow sequence at time t to predict the output layer's network flow data $tr(t+h)$ at time $t+h$. At time t, input layer's input vector is $u(t) = [tr(t-K), tr(t-K+1), \cdots, tr(t-1), tr(t)]^T$, and output layer's output vector is $y(t) = [tr(t+h)]^T$. h is predictive length. So at time t, the reserve pool's inside state vector is

$$x(t) = f_{in}(W_{in}u(t) + Wx(t-1)). \quad (10)$$

In Eq. (10), f_{in} is the excitation function in the inside reserve pool. $x(t-1)$ is the reserve pool's state at the previous time. At time t, the output layer's state vector is

$$tr(t+h) = y(t) = W_{out}x(t). \quad (11)$$

In Eq. (11), f_{out} is the output layer's excitation function.

3.2.2. ESN-DLRP's Training and the Prediction Model Construction

(1) *Set Up Connection Matrix W in the Reserve Pool.* Given neuron $i, i = 1, 1+d, 1+2d, \cdots, 1+(N/d-2) \cdot d$. Set the elements of the reserve pool connection matrix to $W_{i,i+d} = r$, $W_{i+d,i} = r$. When $i = 1+(N/d-2) \cdot d$, set elements $W_{i,1} = r$, $W_{1,i} = r$. The weight value $r \in (0, 1)$, and N is the neuron's number

(2) According to the predictive characteristics of ESN-DLRP, training samples' input-output pairs $\{u_{train}(t),$

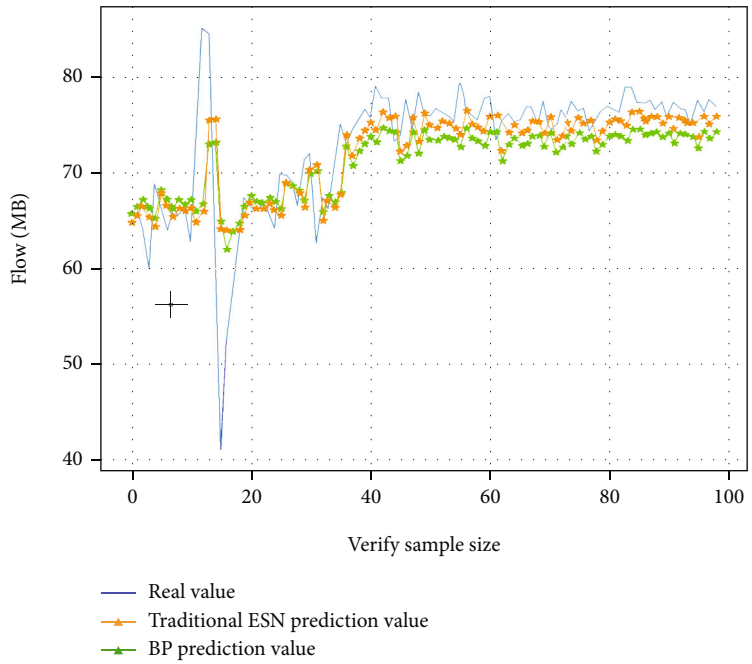

Figure 5: Comparison between BP and traditional ESN prediction effect.

$y_{\text{train}}(t+h), t = 1, 2, \cdots, T\}$ are set up and enter the reserve pool by inputting connection matrix W_{in} [10]

(3) *The Sampling Stage*. Start collecting the state vectors $x(t)$ in the reserve pools and relevant sample data $y_{\text{train}}(t+h)$ from time t_0 until the end of time T, so as to set up the state matrix $X(N, T-t_0+1) = [x(t_0), x(t_0+1), \cdots, x(T)]$ and expected output matrix $Y(L, T-t_0+1) = [y_{\text{train}}(t_0+h), y_{\text{train}}(t_0+1+h), \cdots, y_{\text{train}}(T+h)]$

(4) *Weight Calculating*. According to state matrix X and expected output matrix Y to get the output weight W_{out}, that is, $Y \approx W_{\text{out}}X$, because the actual output matrix $\widehat{Y} = [y(t_0+h), y(t_0+1+h), \cdots, y(T+h)]$ is in linear relation with state matrix X, which means $\widehat{Y} = W_{\text{out}}X$. The training purpose is to approach the expected output matrix Y by the actual input matrix $\widehat{Y}$ and get $Y = W_{\text{out}}X$. Then, $W_{\text{out}} = Y \bullet X^{+}$. Where X^{+} is X's pseudoinverse matrix

(5) Put computed W_{out} in Eqs. (10) and (11) to get network flow's prediction mode

3.2.3. Predict the Output of the Validation Samples. According to the network flow's prediction model, predict the output of the validation samples and set up input-output pairs [11].

4. Forecasting Trial and Performance Evaluation of the Prediction Model

In this paper, MATLAB in business mathematics software is used for a Web application system, every 5 minutes for network flow data sampling, 100 samples are, respectively, input four models: BP (back propagation) neural network, traditional ESN, ESN-DLRP-3, and ESN-DLRP-8, which record the predicted value for a future period of time and compare the predicted value with the real value. The predictive results of the four models are shown by Figures 5, 6, and 7.

The below is to analyze four models' predictive performance by root mean square error (RMSE) and mean absolute percent error (MAPE) [12].

(1) *RMSE*. The sum of the square of the deviation between time series data's predictive value and the real value. Divide this sum by the number of samples. Get the square root of the quotient. It can reflect the predictive accuracy of the predictive models. If the value is more small, the prediction is more accurate

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}}. \tag{12}$$

(2) *MAPE*. Subtract time series data's predictive value and the real value to get the difference. Divide the difference by the real value. Get the average value of the quotient's absolute value. If the average value is more small, the prediction is more accurate

$$\text{MAPE} = \frac{1}{\text{n}}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{\widehat{y}_i}\right| \times 100\%. \tag{13}$$

y_i is the predictive value, $\widehat{y}_i$ is the real value, and n is the sample number.

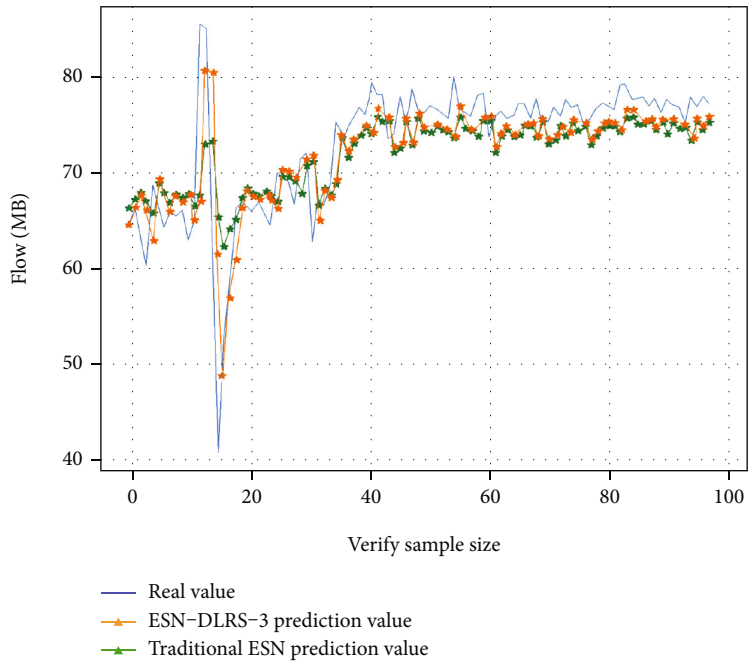

Figure 6: Comparison between traditional ESN and ESN-DLRP prediction effect.

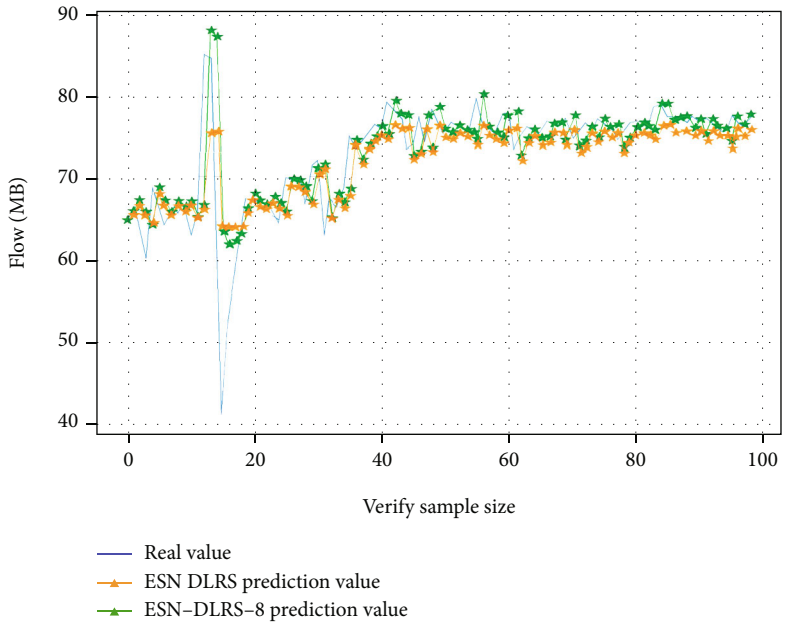

Figure 7: Comparison between ESN-DLRP-3 and ESN-DLRP-8 prediction effect.

According to the data and test time from Figures 4 to Figure 6, four models' predictive performance is shown by Table 1. In terms of the prediction error, the traditional ESN is smaller than the BP neural network, and ESN-DLRP is obviously smaller than the BP neural network and the traditional ESN. In the case of double loop, the larger the neuron spacing in the second ring, the smaller the prediction error, but the difference was not significant. That is, the prediction error of ESN-DLRP-8 is slightly smaller than that of ESN-DLRP-3. From the test time of 100 samples, ESN-DLRP is obviously smaller than that of BP neural network and traditional ESN. In the case of double loop, the test time of ESN-DLRP-8 is slightly shorter than that of ESN-DLRP-3. That is to say, using ESN-DLRP-3 model can achieve better prediction accuracy and test rate.

5. Detection of Abnormal Network Flow

When the Web system runs normally, network flow can be accurately predicted by using the ESN-DLRP, that is, the predicted value of network flow is close to the real value, but when abnormal events occur in the system, there is a large deviation

Table 1: Comparison of network flow prediction performance between different models.

Predictive model	RMSE (%)	MAPE (%)	Test time (ms)
BP neural network	5.19	5.60	45.23
Traditional ESN	5.01	5.30	30.29
ESN-DLRP-3	4.65	4.11	22.95
ESN-DLRP-8	4.57	3.98	20.34

Table 2: Table of detection effects under different thresholds.

Serial number	Threshold value	Detection rate	False report rate
1	10%	97%	39%
2	20%	95%	34%
3	30%	92%	24%
4	40%	90%	22%
5	50%	88%	8.6%
6	60%	87%	7.0%
7	70%	86%	2.8%
8	80%	85%	0.7%
9	90%	80%	0.5%
10	100%	76%	0.2%

between the predicted value and the real value [13]. In this paper, the sliding window mechanism is adopted, and the number of data frames sent or received each time is called window length. The data point is judged whether abnormal or not according to the difference between the real value and the predicted value of network flow in the window.

If current time is t, real value is $\widehat{y}(t)$, predictive value is $y(t)$, the length of the sliding window is l, and we can estimate the difference between the real value and predictive value according to the absolute deviation value $D(t)$, average deviation value u, and relative deviation value $P(t)$.

(1) *Absolute Deviation Value* $D(t)$. The absolute value of the difference between the data point's real value and predictive value

$$D(t) = |\widehat{y}(t) - y(t)|. \tag{14}$$

(2) *Average Deviation Value u*. The average value of the absolute deviation value in the sliding stage

$$u = \frac{1}{l}\sum_{t=0}^{l} D(t). \tag{15}$$

(3) *Relative Deviation Value* $P(t)$. The proportion of the data point's absolute deviation value to average deviation value. Then proportion value times 100%.

$$P(t) = \frac{|\widehat{y}(t) - y(t)|}{u} \times 100\%. \tag{16}$$

Anomaly is essentially the behavior of a data point deviating from normal value. In this paper, each sliding window's average value u is the standard of the normal value. If data point's absolute deviation value (t) is not higher than to the average value u, it is considered normal [14]. If it is higher than u, we need to further estimate if it is higher than threshold value K. If it is not higher than K, it is considered normal, otherwise, it is abnormal. The anomaly detection process for each sliding window is as follows:

Step 1. Get the real value and predicted value pairs of each data point in the sliding window.

Step 2. Calculate the absolute deviation value $D(t)$ of each data point.

Step 3. Count average deviation value u of data series within the sliding interval.

Step 4. Compare data point's absolute deviation value $D(t)$ with u. If $D(t) \le u$, the data is normal; otherwise, step 5 needs to be followed.

Step 5. Determine the dynamic threshold.

Select the data in a sliding range as the detection sample set, and record the number of normal data, abnormal data, and total data in the sample set. The effect of anomaly detection can be measured by two indexes: detection accuracy (detection rate) and false report rate. Detection rate refers

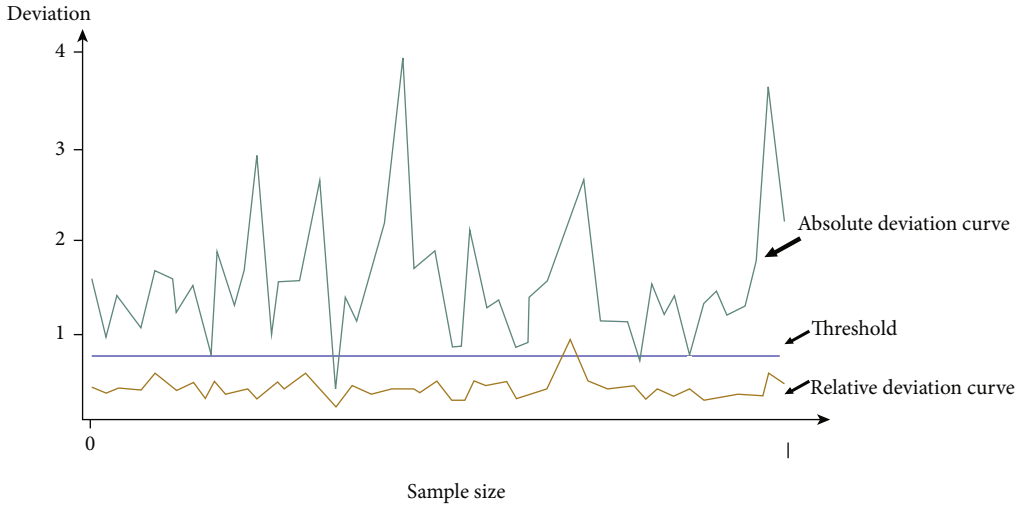

Figure 8: Absolute deviation and relative deviation value curves.

to the ratio between the correct amount and the total amount in the abnormal amount detected. False report rate refers to the ratio of incorrect detection times to total detection times [15]. When different thresholds are selected, the detection rate and false report rate of data in this sample set are shown in Table 2.

Table 2 shows that when the threshold value is 80%, this detection method has a high detection rate and a low false report rate, so the threshold value of anomaly detection is determined as 80% in this paper. Figure 8 shows the absolute deviation and relative deviation curves of each data point in the sample set. It can be seen from the figure that the relative deviation is used to solve the problem that absolute deviation is easy to produce false detection at peak value and valley value. By comparing the relative deviation value with the threshold value, abnormal data in the Web system can be accurately detected.

Step 6. The relative deviation $P(t)$ of each data points was calculated and compared with the threshold K. If $P(t) \leq K$, the data is normal; otherwise it is abnormal.

6. Conclusion

Network flow anomaly detection is a hotspot in the field of network security. This paper mainly focuses on three aspects of research: first, it analyzes the problems faced by traditional ESN in network flow prediction. The second is to design a network flow prediction model based on the ESN-DLRP. Through experiments, the performance of the neural network, traditional ESN, ESN-DLRP-3, and ESN-DLRP-8 model is compared. The results show that the ESN-DLRP-3 model can achieve good prediction accuracy and test rate. Third, proposing an anomaly detection method based on dynamic threshold value. The threshold value is determined dynamically. If the relative deviation between the real value of data point and the predicted value exceeds the threshold value, the data will be judged as abnormal data.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This work was supported by a grant from the Universities Characteristic Innovation Project of Guangdong Province (no. 2020KTSCX305).

References

[1] K. Mengxuan, J. Song, P. Fan, B. Gao, X. Zhou, and Z. Li, "The study of network flow prediction based on deep learning," *Computer Project and Application*, vol. 57, no. 10, pp. 1–9, 2021.

[2] W. Jian, *The Study of the Key Technology of Abnormal Network Detection*, Nanjing University Of Posts And Telecommunications, 2020.

[3] D. Zhen, L. Ma, and G. Shun, "Abnormal network flow detection based on wavelet analysis," *Computer Science and Technology*, vol. 46, no. 8, pp. 178–182, 2019.

[4] L. Zhida and H. Lv, "Network flow prediction system based on wavelet coefficient perception," *TERA-HERTZ Science and Technology and Electronic Information Journal*, vol. 17, no. 1, pp. 131–135, 2019.

[5] N. Mingzhu and J. Gang, "State identification of photoelectric detection system of wavelet denoising and echo state network," *Laser Magazine*, vol. 42, no. 5, pp. 143–146, 2021.

[6] Y. Wei and J. Zhang, "Abnormal network flow detection based on time series analysis," *Jilin University Journal (Science Version)*, vol. 55, no. 5, pp. 1249–1254, 2017.

[7] L. Yan, "Abnormal network flow detection based on time series analysis," *Modern Electronic Technology*, vol. 40, no. 7, pp. 85–87, 2017.

[8] L. Cheng, *The Study and Realization of Prediction System Based on Echo Neural Network's Time Series*, Jiangsu University, 2020.

[9] W. Zongjiang, S. Qi, M. Shang, and D. Shen, "Microgrid equivalent modeling based on advanced echo state network," *Metrology Journal*, vol. 42, no. 7, pp. 923–929, 2021.

[10] Y. Xinyan, *The Study of the Network Flow Prediction Based on Advanced ESN*, Nanjing University Of Posts And Telecommunications, 2019.

[11] S. Tiening, *The Study of the Optimization and Application of Echo State Network*, Guangxi Normal University, 2019.

[12] Z. Rrenjun and W. Wang, "Network flow prediction model simulation based on deep neural network," *Computer Simulation*, vol. 38, no. 6, pp. 475–479, 2021.

[13] L. Yingqi, Y. Huang, and X. Sun, "Network flow prediction model based on deep echo state network," *Nanjing University Of Posts And Telecommunications Journal (Natural Science Version)*, vol. 38, no. 5, pp. 85–90, 2018.

[14] Y. Zeyong, *The Study of the Key Technology of Network Flow's Identification and Prediction System*, Xidian University, 2018.

[15] J. Hua, H. Zhang, Y. Luo, and X. Wang, "The detection of abnormity of adaptive thresholding network flow based on KL length," *Computer Engineering*, vol. 45, no. 4, pp. 108–113, 2019.