Study on the Network Traffic Abnormal Detection Based on EEMD

Zhigang Zhao

Department of Information & Automation ,Tianjin Tianshi College, Tianjin, 301700, China

296186511@qq.com

Keywords: Network Traffic; Abnormal Detection; EEMD; RBF neural network

Abstract. Network traffic abnormal directly reflects the health status of the network and real time detection of network traffic abnormal is very important. Hereby a method of network traffic abnormal detection method based on ensemble empirical mode decomposition (EEMD) was proposed. The historical network traffic data was collected and analysis by EEMD and the radial basis function (RBF) neural network prediction model is established for network traffic abnormal detection. The historical network traffic data removes the abnormal data by analysis of intrinsic mode function (IMF), which is used as the input of the RBF neural network for prediction. If the error between prediction value and the actual value is larger than a threshold, then this point of network traffic can be judged as an abnormal data. The proposed method takes advantages of network traffic prediction and EEMD analysis, which can implement real time detection of network traffic abnormal.

Introduction

Effective extraction and identification of network traffic abnormal characteristics can help to find the network attack and network abnormal in time [1], which can provide the network security information for administrator to take security decisions. The essence of the location and identification of network traffic abnormal characteristics is to extract the regular abnormal signals from the massive network traffic data, and the abnormal characteristics is processed by classifier to implement the extraction and identification of network traffic abnormal. Therefore, the network traffic abnormal detection can be classified as signal processing and feature recognition. When the network is under attack, the load of the network in the target, even near the target will be increased dramatically, which reflect in the network traffic will be abnormal changes. Although some extraction and identification of network traffic abnormal methods have been proposed recently, such as wavelet analysis [2][3], series analysis [4] and phase transition phenomena analysis [5], the real time detection of network traffic abnormal is still no ideal solution.

The network traffic exhibits self similarity on the large time scale of second level, and has multiple fractal properties on the small time scale of the millisecond scale. Therefore, the network traffic can be described as a complex nonlinear system. However, the reflection of network attacks or network abnormal in the network traffic is not so obvious, which results in the detection of network traffic abnormal becomes a data mining problem. Network traffic prediction can establish a stable health network model, which needs the historical network traffic data has no abnormal data. Although using the network traffic prediction method to analyze network traffic abnormal cannot implement the classification of abnormal reasons, the detection results can provide warning to the administrator to take appropriate inspection and preventive measures. Hereby, a method of network traffic abnormal detection method based on EEMD was proposed. The historical network traffic data was analyzed by IMF, and the prediction model is established by RBF neural network. The abnormal data in the historical network traffic data can be removed by analysis of IMF, which is used as the input of the RBF neural network for prediction, and then the prediction model can reflect the health status of network traffic. The detection of network traffic abnormal can be judged according to the output of the prediction model by the error between the actual network traffic value and the prediction output value, in which a threshold is set by the administrator. If the actual network traffic value serious deviation from the predicted value, then the point at current must be taken as abnormal condition, and then the actual network traffic value is instead of predicted value for next point prediction. The proposed method can issue a warning to the administrator in time and implement real time detection of the network traffic abnormal. Although it cannot provide the reason of the abnormal, it can give the security early warning information in advance to avoid further damage to the network.

The Basic Principle of EEMD and RBF Neural Network

EEMD is an improved method of empirical mode decomposition (EMD) which proposed by N.E. Huang, which can obtain robust decomposition results by using Gaussian white noise [6]. EMD is an adaptive signal decomposition method which decomposes a signal to a series of IMFs, and it is very suitable for the analysis of nonlinear and non-stationary signals. EEMD adds Gaussian white noise several times, and carries on EMD for the signal after adding noise, and then ensemble average processing is done for all the decomposition results, which can avoid mode mixing at a certain extent. Let the signal is $x(t)$ and the Gaussian white noise is $\omega_j(t)$, then the mixed signal can be given by

$$x_j(t) = x(t) + \omega_j(t) \tag{1}$$

Then EMD is carried on $x_j(t)$ and the mixed signal can be expressed by the IMFs as follow

$$x_j(t) = \sum_{i=1}^{n} C_{ij}(t) + r_{jn}(t) \tag{2}$$

Where $C_{ij}(t)$ is the IMF set and $r_{jn}(t)$ is the residual component of EMD. Carrying on EMD for all the mixed signals, then the IMFs set can be obtained. The range of signal noise ratio (SNR) of each time adding white noise is different, and then the IMFs for each EMD of the mixed signal is different. The final IMFs set can be got by ensemble average processing as follow

$$C_j(t) = \frac{1}{M} \sum_{i=1}^{M} C_{ij}(t) \tag{3}$$

Where $C_j(t)$ is the IMF set of EEMD and M is the times of EMD for mixed signals, and then the signal $x(t)$ can be expressed as follow

$$x(t) = \sum_{j} C_j(t) + r(t) \tag{4}$$

Where $r(t)$ is the residual component of EEMD.

If M is large enough, the adding Gaussian white noise tends to 0 in each IMF of EEMD, and the mode mixing can be suppressed at some extent. Therefore, EEMD is robust for analysis of nonlinear and non-stationary signal.

The network traffic can be equivalent to a nonlinear system, and then it is suitable for EEMD processing. The network traffic abnormal data can be analyzed in the IMFs, and the IMFs can be taken as the input of RBF neural network for prediction model establish after the network traffic abnormal data removed. RBF neural network was proposed by Moody J and Darken C, which can approximate any continuous function with arbitrary precision [6]. As other neural network models, RBF neural network is constructed by input layer, hidden layer and output layer, which model is shown Fig.1. The transfer function of RBF neural network is usual Gaussian function and the input of hidden layer is the product of the input signal and the weights, sometimes the threshold is used for adjust the fitting accuracy. The input of the hidden layer is given by

$$r_j(t) = \exp(-(\left\| w_{ij} - x(t) \right\| \times b_j)^2) \tag{5}$$

Where w_{ij} is the weight between the input layer and the output layer and b_j is the threshold. The output of RBF neural network is the sum of the output of the hidden layer after weighted mean, the transfer function of the output layer is usual linear function, and then the output of RBF neural network can be given by

$$y(t)=\sum_{j=1}^{m} w_j r_j(t) \quad (6)$$

Where w_j is the weight between the hidden layer and the output layer. The basic idea of RBF neural network is that RBF is taken as the basis to construct the space of hidden layer. Therefore, the input can be mapped to hidden layer space directly without weights. Once the center of RBF is determined, the mapping relation can be determined.

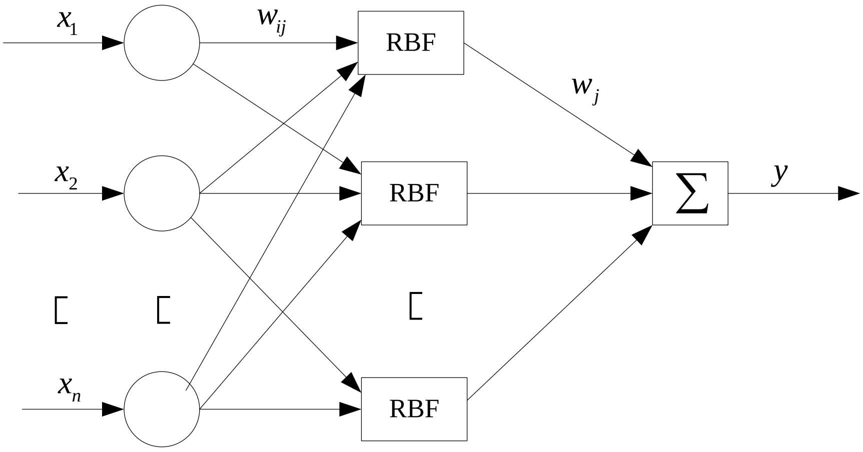

Fig.1 RBF neural network structure

Network Traffic Abnormal Detection Based on EEMD

Network traffic prediction relies on historical data, and then the abnormal data would influence the prediction accuracy. If the abnormal data is removed from the historical data by EEMD, the RBF neural network can establish the ideal network traffic prediction model, and then the prediction model is used to detection the neural network abnormal data in time. The network traffic historical data is decomposed by EEMD and cubic spline interpolation [7] carries on the IMFs to remove the abnormal data, and then the IMFs is taken as the input of the RBF neural network for prediction model establishment. The current collection data of network traffic is compared with the real time prediction value, if the prediction value is severe inconsistence with the actual value, then we can judge the network traffic data abnormal. The implementation process is shown in Fig.2.

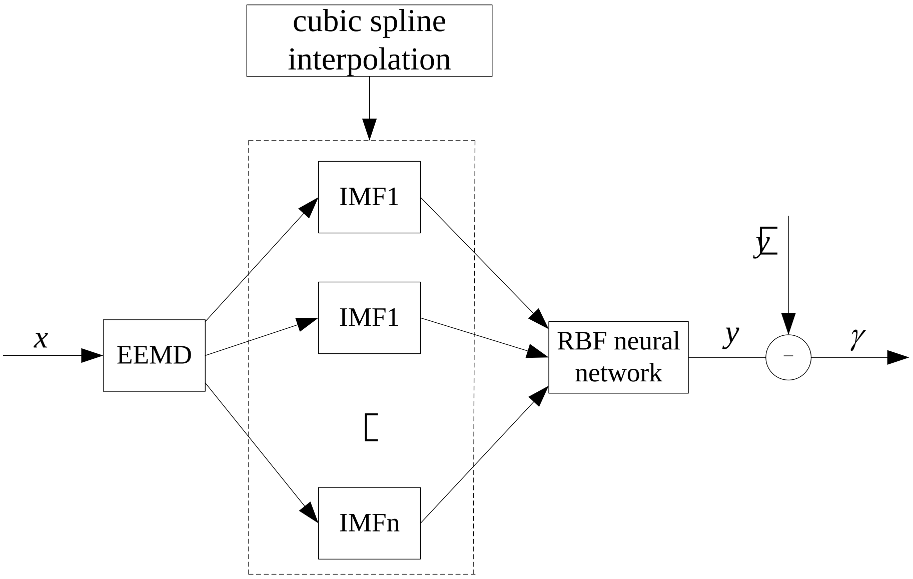

Fig.2 implementation process of network traffic abnormal detection

The network traffic abnormal data in time detection can be implemented by the comparison of γ and the threshold δ. If $|\gamma| \geq \delta$, the current network traffic data is abnormal, and the current data is instead of the prediction data for next point prediction. If the abnormal data last for quite a while, then the administrator must take necessary measures for the network security and analysis the reason of the abnormal.

Computer Simulation and Analysis

In the simulation, the network traffic data is collected from a campus network monitoring center and the sampling time interval is 1 minutes. The RBF neural network prediction model use a day of continuous collection of data as a training sample. The collected data is decomposed by EEMD and carried cubic spline interpolation on the IMFs. The result of decomposition of the network traffic data during 5 minutes by EEMD is shown in Fig.3.

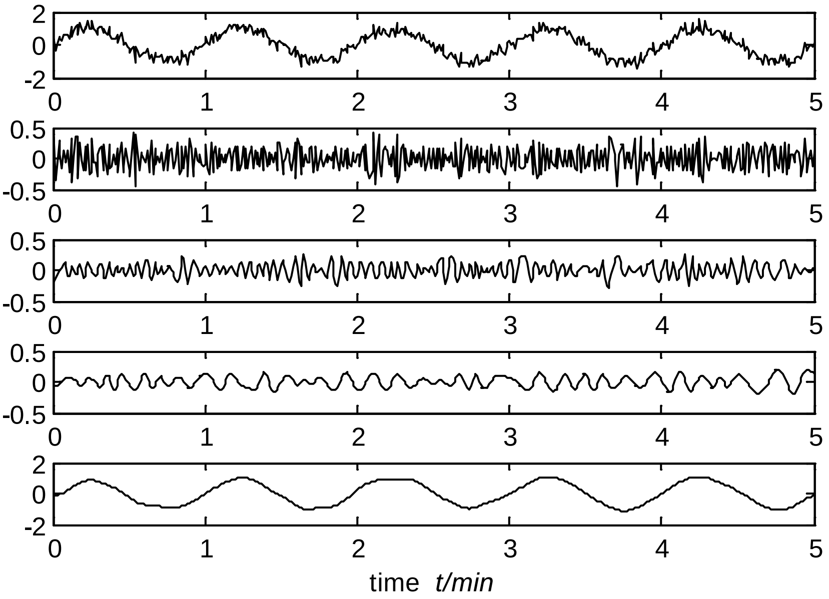

Fig.3 The IMFs of EEMD (5 minutes)

The network traffic abnormal data is detected by the RFB neural network traffic prediction model by using the network traffic data of the next day. The detection result is shown in Fig.4. From Fig.4 can see that, the network prediction error is larger than the threshold $\delta = 0.5$, which shows it is abnormal in the network.

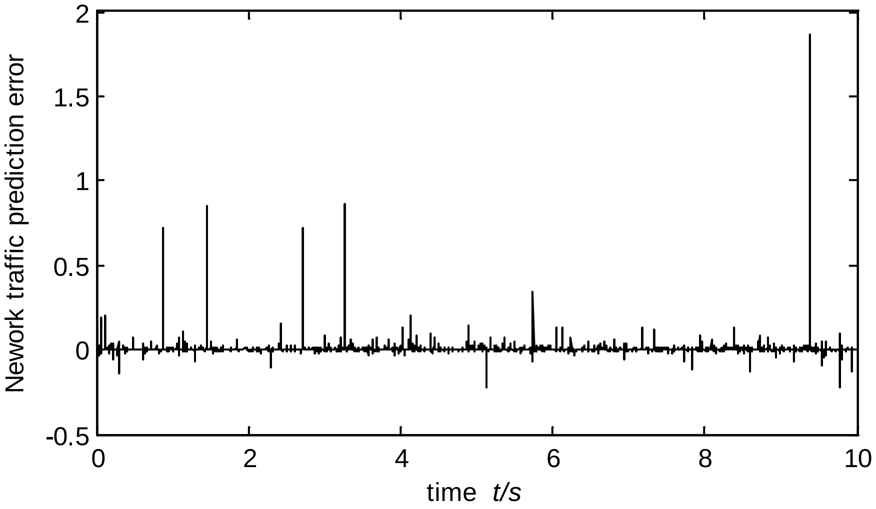

Fig.4 The prediction error

Conclusion

This work proposed a data detection method of network traffic abnormal based on EEMD, which takes the advantages over the network prediction and the EEMD. The network traffic data is analyzed by EEMD combined with cubic spline interpolation in advance, and the RBF neural network is acted as the prediction model with the IMFs after preprocessing. The current network traffic data can be detected by comparing with the prediction value. This method can implement the network abnormal data detection in real time. However, further analysis and research are needed for the abnormal reasons.

References

[1] Mahajan R., Bellovin S. M., Floyd S., et al. Controlling high bandwidth aggregates in the network [J]. Computer Communication Review, 2002, 32(3): 62-73.

[2] SHI H. S., YANG Z. F. Optimization of decision tree principal component feature tracking and transient abnormal feature extraction [J]. Information Technology, 2014, (7): 158-162.

[3] YANG J. P., LIU X. C. Study on the network abnormal detection based on the wavelet transforms [J]. Journal of Shandong Agriculture University (Nature Science), 2012, 43(1): 95-99.

[4] CHENG H., SHAO Z. Q., FANG Y. Q. Log-infinitely divisible cascades analysis of abnormal network traffic [J]. Computer Engineering, 2006, 32(10): 9-11+14.

[5] WANG X., DING L. Study on phase transition phenomena in network traffic anomalies [J]. Journal of Communications, 2006, 27(2): 184-188.

[6] XIE Z. P., LIU J.H., WANG S.T. RBF network learning algorithm using robust least-squares [J]. Control and Decision, 2010, 25(4): 502-506.

[7] LIU Q., LI S.S., GAO W., et al. Application study with the method of curve formula based on the kinds of difference of three times’ functions [J]. Journal of Safety and Environment, 2009, 9(4): 71-74.