An Adaptive abnormal flow detection method for new energy stations based on HHT algorithm

*YIN Liang¹, CHEN Xiaoqian¹

Electric Power Research Institute of State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750001, Ningxia, China

¹Corresponding author’s e-mail: yinliang@dky.nx.sgcc.com.cn

Abstract. With the development of new energy technology, new energy stations are becoming more intelligent and data-based, and cyber-attacks on new energy stations are increasing year by year. In response to the continuous threats brought by malicious traffic to the network of new energy stations, this paper researches on the traffic anomaly detection technology based on network communication characteristics. An adaptive abnormal traffic detection method for new energy stations based on HHT algorithm is proposed, which improves the efficiency of identifying abnormal network traffic and more accurately identifies network attacks against new energy stations. It is verified through experiments that compared with mainstream classifiers, the method studied in this paper can achieve adaptive detection while adaptively determining the threshold, and the detection accuracy can reach 95%, the false alarm rate is lower than other methods, it can provide more accurate identification results for new energy field station network cyber-attacks detection.

1. Introduction

New energy refers to various energy other than traditional energy, mainly including wind energy, solar energy, biomass energy, etc., which is the general trend of future grid development. New energy power generation has good development prospects and practical value. As more and more new energy plants and stations are constructed and connected to the grid, the network security risks of new energy plants and stations have become increasingly prominent.

The security problems of new energy plants and stations may pose a threat to the security of the entire power grid. On the one hand, attacks from the network of new energy plants and stations may penetrate upwards, seriously threatening the security and stability of the entire grid-side dispatch control system. For example, if a new energy plant is attacked, it may cause a large-scale off-grid accident of the unit, which will eventually affect the stability of the grid system voltage and frequency; on the other hand, the network of the new energy plant has firewalls, encryption machines, isolation devices, etc. The necessary security protection capabilities, but its monitoring system, control system and networked information system are also very likely to be penetrated and attacked from the vertical boundary of the network, resulting in unpredictable two-way security consequences. Therefore, real-time security threat detection and security protection at the vertical network boundary of new energy plants and stations is an inevitable and urgent need; in response to this need, research on cyberspace security monitoring frameworks, abnormal detection algorithms, and security based on large-scale traffic real-time analysis Monitoring and attack early warning systems are bound to become an important trend in the development of power production network security technology.
2. Network abnormal traffic detection technology

The detection of abnormal network traffic is a branch of the anomaly detection technology. The main function is to detect possible abnormal traffic, abnormal behavior, and abnormal status from the seemingly normal traffic, and discover abnormal security events. In this topic, the detection of abnormal traffic is not only used to directly discover abnormal security events, but also used for joint detection and correlation analysis of network layer and application layer attacks.

The development trend of anomaly detection theory is to use the frequency domain decomposition method to perform fine-grained decomposition and observation of the original signal, and then display the energy intensity of the decomposed signal in the form of an energy spectrogram, and find the anomalies hidden in the original signal according to the instantaneous energy intensity. In terms of network anomaly detection, there are many domestic and foreign researches, but there are obvious defects. At present, the mainstream method of traffic anomaly detection that is generally recognized is: a method that combines wavelet transform and PCA, and performs wavelet transform on network traffic that has been selected by features. Then analyse each wavelet coefficient matrix to select the principal components, and perform inverse wavelet transformation on the principal components. Finally, all the eigenvectors after inverse wavelet transformation are reconstructed into a new matrix, and PCA is performed to detect network abnormalities. Although wavelet transform can be used alone to detect network anomalies, it can also be combined with PCA to overcome the main defect of PCA, that is, subspace pollution, but wavelet transform requires basis functions, and higher parameters, which will affect the final result performance. Other existing abnormal detection methods such as mathematical statistics, digital signal processing methods based on wavelet analysis, data mining methods, neural network methods, etc., have the need to provide traffic training sets, the detection results depend on parameter settings, and the real-time performance is not strong. It does not have shortcomings such as adaptability.

2.1. Background and Existing abnormal traffic detection technology

In the past few decades, we have seen a series of cyber-attacks on CNI and SCADA. The 2010 computer worm Stuxnet attack caused damage to one-fifth of the centrifuges near Iran’s nuclear facilities [1]. In 2011, five global energy and oil companies were attacked, including social engineering, Trojan horses, and Windows-based attacks [1]. In 2012, a malware named Flame was found to have been running on many websites in the Middle East and North Africa for at least two years [1], [3].

Miller and Rowe’s analysis showed [1] that the number of cyber-attacks against CNI has been increasing over time. At the same time, the number of SCADA-related security incidents is also steadily increasing. In 2010, the Industrial Security Incident Library (RISI) recorded 161 incidents, and approximately 10 new incidents were added every quarter [4]. In 2013, the RISI company database had recorded 240 events from 2001 to the end of 2012 [5]. In addition, on the basis of extensive research on the network security status of a large number of SCADA systems, interviews with a large number of experts confirmed that the cyber threats in SCADA systems are escalating and they are "real and expanded" [6].

In view of the significant impact of the electric energy system on national security, in order to improve the network security of critical infrastructures, and for the energy field network security, in 2014, the US Department of Energy proposed a general network security capability maturity model in the US Department of Homeland Security. On the basis of (Cybersecurity Capability Matural Model, C2M2) [7], an ES-C2M2 (Electricity Subsector Cybersecurity Capability Matural Model) model specifically for the electric power field [8] is proposed.

Abnormal traffic detection technology. In the context of big data, machine learning algorithms are widely used in network security fields such as traffic anomaly detection, intrusion detection, and Trojan horse detection. At the same time, deep learning, integrated learning and other technologies are gradually being used in the security field. In terms of abnormal traffic detection, Tian Wang[9] et al. analysed abnormal traffic event detection through multi-frame optical flow information. Guanglu
Wei[10] et al. proposed the application of deep learning algorithms in abnormal traffic recognition, Stanimir Kabaiwanov[11] Et al. proposed a hybrid deep learning algorithm for abnormal network traffic detection. Chen Jilei et al. [12] proposed a deep learning model based on a sparse autoencoder, which improved the detection rate of intrusion detection algorithms and reduced the false alarm rate. However, when the amount of data is small, the detection effect is not ideal; Zhao Ying et al. [13] applied the Markov model to network traffic classification, and improved the semi-supervised learning network traffic classification method through the method of similarity calculation. Wu Jinlong et al. [14] used the random forest method for Trojan detection, which greatly improved the accuracy of Trojan detection, but did not demonstrate the detection of other abnormal traffic; Chen Lulu et al. [15] used gradient projection The negative matrix factorization method reconstructs the multi-dimensional entropy matrix, which has a good effect on the recognition of continuous abnormal traffic such as DDOS, but the effect of general anomaly detection is relatively unsatisfactory; Paweł[16] et al. proposed a communication network anomaly detection based on multifractal analysis Method to reduce the false alarm rate; Fatemeh Safara [17] et al. proposed an improved communication network intrusion detection method based on association rule mining and artificial neural network to detect and mine abnormal traffic in network communication; Ying Zhong [18] et al. proposed a new network anomaly detection model based on heterogeneous integrated learning, which can support more network protocol anomaly detection.

2.2. HHT algorithm introduction and steps

The core of the HHT algorithm in the field of anomaly detection is two algorithmic processes:

(1) Empirical Mode Decomposition (EMD) is performed on the time series of the original flow characteristic index value, and the original signal is decomposed into Eigen mode functions of different frequencies (Signal component) and a residual trend item;

(2) Perform Hilbert Transform on the signal component, and use the concept of frequency spectrum to show the characteristics of the signal, and find anomalies in the way of energy spectrogram. The following describes the algorithm process:

1> Empirical mode decomposition (EMD)

Assuming that the original signal is x(t), m(t) is the mean function of the upper and lower envelopes, let s(t)=x(t), h(t) is the intermediate variable of signal decomposition, and c(t) is IMF function. The signal decomposition process is as follows:

1-1> Find all the maximum and minimum points of the function x(t), construct the upper and lower envelopes through the cubic spline interpolation function, and calculate the mean function m(t);

1-2> Subtract the mean function m(t) from the function x(t) to get h(t), that is, h(t)=x(t)-m(t);

1-3> Judge whether h(t) meets the IMF conditions, if not, set x(t)=h(t), repeat steps 1, 2, and 3 for the x(t) function, otherwise go to step 4.

1-4> Set \( \text{imf}_i(t) = h(t) \), \( s(t) = s(t) - h(t) \), Judge whether s(t) satisfies the condition of the residual trend item, if it is satisfied, let r(n)=s(t), and the algorithm ends; otherwise, let x(t)=s(t), repeat the steps 1-4, find IMF component of order n \( \text{imf}_n(t) \) and r(n).

The original signal can be expressed as:

\[ s(t) = \sum_{i=1}^{n} \text{imf}_i(t) + r_n(t) \]

Among them, s(t) represents the original signal, \( \text{imf}_i(t) \) Represents the i-th eigenmode function, \( r_n(t) \) Represents residual trend items.

Taking a certain indicator "TCP flow size" as an example, the TCP flow size signal can be decomposed into multiple components, as shown in the figure below (the first sub-picture EEMD_TCP_ORIGINAL_PART3 is the original signal):
Hilbert transform

The Hilbert transform of a real signal $x(t)$ is defined as:

$$H[x(t)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau$$

Where $P$ stands for Cauchy principal value,

The analytical signal $z(t)$ is defined as:

$$z(t) = x(t) + iH[x(t)] = a(t)e^{j\theta(t)}$$

Where the instantaneous amplitude is:

$$a(t) = \sqrt{x^2(t) + H^2[x(t)]}$$

The instantaneous phase is:

$$\theta(t) = \arg\left(\frac{H[x(t)]}{x(t)}\right)$$

The instantaneous frequency is:

$$\omega = \frac{d\theta(t)}{dt}$$

Then the original signal $x(t)$ can be expressed as
The Hilbert amplitude spectrum is defined as:

\[ x(t) = \text{Re} \left( \sum_{j=1}^{n} a_j(t)e^{i\omega_j t} \right) + r_n(t) = \text{Re} \left( \sum_{j=1}^{n} a_j(t)e^{i\omega_j t} \right) + r_n(t) \]

The Hilbert amplitude spectrum is defined as:

\[ H(\omega,t) = \text{Re} \left( \sum_{j=1}^{n} a_j(t)e^{i\omega_j t} \right) \]

Indicates the instantaneous amplitude distribution on the frequency-time plane, when \( a_j(t) \) in \( H(\omega,t) \) is a square term, called Hilbert Energy Spectrogram, as shown below:

Figure 2. Hilbert Energy Spectrogram

In the figure, the brighter the colour, the higher the energy, and the greater the probability that the energy-concentrated part indicates the flow is abnormal, so the abnormal time of the original signal can be determined from the time when the energy is abnormally concentrated.

Compared with traditional wavelet analysis and other methods, the advantages of HHT method are as follows:

1. HHT can analyze nonlinear and non-stationary signals

In traditional data processing methods, there are strict requirements on signal processing. The more popular processing methods include Fourier transform and wavelet transform, among which wavelet transform can theoretically process nonlinear and non-stationary signals, but in practice like the Fourier transform processing in the algorithm, it can only be a linear non-stationary signal. HHT can handle any non-linear and non-stationary signal, which has great advantages in signal analysis.

2. HHT is adaptive

In the process of signal processing by HHT, components can be generated adaptively, while most other algorithms need to set preconditions, such as Fourier trigonometric function, wavelet basis, etc., which are uncertain. However, HHT does not need to set function conditions in advance, and decomposes the signal adaptively according to the characteristics of the signal to obtain components with actual physical meaning, which is of great significance for adaptive anomaly detection.

3. HHT is not restricted by Heisenberg's uncertainty principle

According to Heisenberg's uncertainty principle, it can be known that certain physical quantities of a microscopic particle cannot have a definite value at the same time, that is, the more certain one quantity is, the greater the uncertainty of the other quantity. HHT is not restricted by Heisenberg's uncertainty principle, so that it can achieve high accuracy at the same time as time and frequency, while traditional signal processing methods are restricted by this. These methods cannot achieve the best accuracy at the same time in terms of time and frequency in principle.

Due to the above advantages, the HHT algorithm has been highly valued and applied in the field of anomaly detection. Several research results have also shown its advantages in the field of anomaly detection over signal analysis, data mining, neural networks and other methods. Based on this theory,
this subject proposes and studies the adaptive detection technology of abnormal network traffic based on HHT algorithm.

### 3. Adaptive anomaly detection technology for new energy stations based on HHT algorithm

This paper proposes an adaptive anomaly detection technology based on network traffic feature extraction and HHT algorithm. The anomaly detection technology can be divided into three levels: the first level is the expression of characteristic indicators, namely: real-time collection of large-scale network traffic, and through the frequency domain decomposition of various characteristic indicators, cycle calculation, variance analysis, energy analysis and other processes, Multi-dimensionally accurately portray the situational characteristics of network traffic. The second level is the expression of abnormal situation events, that is, through abnormal detection of characteristic indicators, microscopic abnormalities of the situation are found and abnormal events are triggered. The third level is event correlation and combined output, that is, correlate micro-level abnormal situation events to trigger higher-level macro-level abnormal situation alarms and attack alarms. And how to use signal processing methods to dig out hidden and possible abnormal signals from the time series of characteristic indicators, and to detect and warn them to realize abnormal discovery of unknown hidden situations.

The overall anomaly detection technology roadmap for the above 3 levels is shown in the figure below:

![Network space situation quantification and anomaly detection technology based on online traffic deep mining](image)

As can be seen from the figure, the quantitative expression of network traffic security situation is divided into three steps, namely: (1) traffic feature collection and situation feature characterization; (2) situation feature-oriented adaptive learning and anomaly analysis; (3) abnormality Situation detection and warning. The following are described separately:

1. **Flow collection and situation feature index extraction**

First, use traffic probes to capture and analyse traffic 7x24 hours to obtain situational characteristic indicators in multiple dimensions, and realize online real-time collection, logging, and storage of characteristic indicators. Every 1-minute (or a specified period) for the entire network, each subnet and a single IP host generate an analysis record, which contains the current values of all characteristic indicators. The project team has designed more than 40 characteristic indicators for network traffic in the early stage, and all of them are collected, stored and visualized in real time.
Based on the above-mentioned basic processes such as flow collection, flow classification, and flow tracking, this project proposes a multi-dimensional situation feature extraction technology. The purpose is to design and collect multi-dimensional feature indicators in the flow, and describe the operating situation of the network space and network entities (host, host, Users, etc.).

According to the traffic of the whole network and the traffic of a single network entity, this paper proposes 6 categories and more than 40 dimensions of situation characteristic indicators, and respectively designs corresponding online mining algorithms to obtain the above-mentioned dimensional indicators from the traffic. The six categories are described as follows:

- Traffic statistics feature indicators: This category of indicators provides a means to describe the summary statistical features of network macro or micro entity traffic. Can be used to detect cyber-physical attacks and large-scale active attacks.
- Traffic morphology characteristic indicators: designed to characterize the traffic morphology of customer networks, including scale, user composition, application behavior patterns, etc. It can be used to detect large-scale active attacks, internal attacks, and cyber-physical attacks.
- Characteristic indicators of encrypted communication behaviors: designed to provide a description and verification basis for encrypted sessions, encryption machines, and protocol compliance in encrypted networks. Can be used to find insider attack threats against encryption machines.
- General behavior index: It is designed to describe the general behavior of the entire network, IP subnet, and host/user, and can be used to describe the situation of basic network services. Can be used to detect active/inside attack threats.
- Network spatial structure type characteristic index: It aims to characterize the network form, mainly for the whole network routing characteristics, server group and key backbone link routing characteristics. It can be used to detect cyber-physical attacks and specific types of active attacks.
- Application access behaviour characteristic indicators: designed to characterize the application behaviour of users and application servers. Can be used to detect specific types of internal user attacks.

The characteristic indicators of the above multiple dimensions can be automatically uploaded to a variety of big data platforms to provide storage and subsequent analysis, and online real-time monitoring can also be provided. The project will mine various indicators related to anomalies from the network traffic and form a log record every N minutes. The preliminary design of relevant indicators is as follows (the indicator system will be further expanded during the investigation of new energy plants):

| Characteristic index | Meaning of characteristic index |
|----------------------|---------------------------------|
| The following indicators are global (or subnet, or basic host information) | |
| MPID | Measuring point number |
| BEGINTIME | Statistics start time |
| SAVETIME | Statistics deadline |
| IP | IP address |
| MAC | MAC address |
| BLOGUSER | user account |
| The following indicators are related to traffic statistics | |
| ONLINE USERS | Number of online users |
| IP_INBPS | Average incoming IP traffic (bps) |
| IP_OUTBPS | Average outgoing IP traffic (bps) |
| Characteristic index | Meaning of characteristic index |
|----------------------|---------------------------------|
| TCP_INBPS            | TCP average incoming traffic (bps) |
| TCP_OUTBPS           | TCP average outgoing traffic (bps) |
| UDP_INBPS            | UDP average incoming traffic (bps) |
| UDP_OUTBPS           | UDP average outgoing traffic (bps) |

The following indicators are related to traffic patterns:

- **TCP_FLOWS**: TCP sessions
- **TCP_PEERS**: Number of TCP peers
- **PKTS_PER_TCPFLOW**: Average number of packets per TCP session
- **AVGLLEN_IN_TCPFLOW**: TCP session average incoming data packet length (bytes)
- **AVGLLEN_OUT_TCPFLOW**: TCP session average outgoing data packet length (bytes)
- **UDP_FLOWS**: UDP sessions number
- **UDP_PEERS**: Number of UDP peers
- **PKTS_PER_UDPFLOW**: Average number of packets per UDP session
- **AVGLLEN_IN_UDPFLOW**: Average packet length of UDP sessions (bytes)
- **AVGLLEN_OUT_UDPFLOW**: Average packet length of UDP sessions (bytes)

The following indicators are related to abnormal encrypted communication behaviour:

- **IPSEC_FLOWS**: Number of encrypted sessions
- **IPSEC_HOSTS**: Number of encrypted session hosts
- **PKTS_PER_IPSECFLOW**: Average number of data packets per encrypted session
- **AVGLLEN_IN_IPSECFLOW**: Average data packet length of encrypted session (bytes)
- **AVGLLEN_OUT_IPSECFLOW**: Average packet length of encrypted session (bytes)

The following indicators are related to general abnormal behavior:

- **DNS_QUERYs**: Number of DNS requests issued
- **SYN_OUTCOUNTER**: Number of TCP SYN sent
- **SYN_INACKS**: Number of SYN+ACK responses received
- **SYN_INCOUNT**: Number of TCP SYN received
- **SYN_OUTACKS**: Number of TCP SYN+ACK sent
- **ICMP_INPPS**: ICMP incoming average traffic (pps)
- **ICMP_OUTPPS**: Average outgoing ICMP traffic (pps)
- **OTHERIP_INPPS**: Average incoming traffic of other IP packets (pps)
- **OTHERIP_OUTPPS**: Average outgoing traffic of other IP packets (pps)

The following indicators are related to changes in the network structure and frequently visited targets by users:

- **TTLSERVER_MAX**: Maximum routing hops on the server
- **TTLSERVER_MIN**: Minimum routing hops on the server side
- **TTLCLIENT_MIN**: Client minimum routing hops
- **TTLCLIENT_AVG**: Average routing hops of the client
| Characteristic index                                      | Meaning of characteristic index                                                   |
|-----------------------------------------------------------|----------------------------------------------------------------------------------|
| The following indicators are related to common application behaviours of web users | HTTP GET Number of requests                                                      |
| HTTP_GETS                                                | Number of emails sent                                                             |
| MAIL_SENT                                                | Number of received mail                                                           |
| MAIL_RECV                                                | Number of media streams                                                           |
| MEDIA_FLOWS                                              | FTP Number of downloads                                                           |
| FTP_DOWNLOADS                                            | FTP Number of uploads                                                             |
| FTP_UPLOADS                                               | Number of packets with successful multi-pattern matching                          |
| PATTERN_MATCH_MAX                                        | The number of suspected retransmissions of the same type of payload               |
| PAYLOAD_REQUESTS_MAX                                     | Number of protocol compliance detection alarm messages                            |

(2) Self-adaptive learning and abnormal analysis for situation characteristics

On the basis of flow collection and situation feature extraction, an online learning algorithm is designed to learn the index data obtained by flow feature collection. The key technology of this step is adaptive learning and anomaly analysis technology. That is: how to adaptively learn, identify the normal state and normal pattern in the network, form a continuously updateable learning result, and apply the learning result for subsequent abnormality analysis and detection. The output of this step is the learning knowledge base of a single situation indicator and the abnormal analysis result of the current indicator value.

Compared with the traditional wavelet signal decomposition method, the advantages of the HHT method in this paper include: (1) It can analyse nonlinear and non-stationary signals. In traditional data processing methods, there are strict requirements on signal processing. The more popular processing methods include Fourier transform and wavelet transform, among which wavelet transform can theoretically process nonlinear and non-stationary signals, but in practice like the Fourier transform processing in the algorithm, it can only be a linear non-stationary signal. (2) It is adaptive. In the process of HHT processing signals, components can be generated adaptively, and Fourier and wavelet transforms need to set preconditions. For example, the function base set by Fourier is a trigonometric function, and wavelet transforms select wavelets that meet certain conditions. Basis, and the result of wavelet transform will vary with the selected wavelet basis transform, which has uncertainty. However, HHT does not need to set the function conditions in advance, and decomposes the signal adaptively according to the characteristics of the signal to obtain components with actual physical meaning, which is of great significance for analysing the signal. (3) HHT is not restricted by Heisenberg's uncertainty principle. (4) The superiority of frequency solving method.

(3) Network traffic abnormal situation warning

On the basis of the above learning and anomaly analysis, online alarms of abnormal situations are carried out. The online abnormal alarm output is divided into three categories:

(a) The absolute value of a single characteristic index is abnormal. Once the current sampling value of any characteristic index breaks through the threshold (the upper limit of the learning range), an alarm record will be generated. This project intends to output alarms for all more than 40 situation characteristic indicators.

(b) A single characteristic indicator suddenly jumps to the alarm. The jump scale of the previous sample value and the current sample value of any characteristic index, once the jump threshold (upper limit of the learning range) is broken, an alarm record will be generated. This type of alarm also includes 40 characteristic indicators.
Multiple characteristic indicators are associated with alarms. According to the associated knowledge base of multiple features, the common anomaly of multiple features often represents an intelligible anomaly with a clear physical meaning (for example, DDOS attacks, etc., defined as "understandable anomalies"). Therefore, if the current sampling values or sudden jump scales of multiple characteristic indicators meet the requirements of an item in the associated knowledge base, and the association of multiple indicators can trigger a higher-level "understandable abnormality" alarm.

4. Experiment analysis

4.1. Algorithm steps

Based on this algorithm, this paper proposes a flow modelling method based on multi-dimensional quantitative features of digital signals. The method includes the following steps:

<1> First, collect the actual flow of the target environment of the new energy plant online;
<2> Secondly, carry out multi-dimensional feature extraction of traffic (extract more than 40 dimensions of traffic feature indicators such as statistical, morphological, encrypted communication, general behaviour, spatial structure, and behaviour), and analyse the network space Real-time quantification of the traffic situation;
<3> Then, the Empirical Mode Decomposition (EEMD) method is used to decompose each characteristic index in the frequency domain to obtain signal components of different frequencies, and calculate the signal strength, period, variance, energy and other characteristic values of different components;
<4> Finally, the multiple characteristic values of the above-mentioned multiple dimensional characteristic indicators together constitute the flow characteristic characterization indicator system, which is the flow model of the new energy plant.

The advantages of the modelling method in this article are:

<1> Comprehensiveness, using multiple dimensional feature indicators to describe the characteristics of traffic more comprehensively, while traditional modelling methods only describe simple indicators such as traffic size, and can only complete low-dimensional descriptions;
<2> Adaptability, the modelling process does not need to select a model or adjust parameters, and almost all existing models need to manually select models and adjust parameters, and do not have full adaptive modelling capabilities;
<3> Real-time, the flow modelling process and flow collection process are completed in real time simultaneously, without offline analysis and training process of various data sets;
<4> Practicality. The goal of this flow model is to accurately describe the flow characteristics and provide an accurate reference for anomaly detection, while traditional flow models are mostly for the purpose of flow mechanism research, flow fitting, and flow prediction, and are practical in anomaly detection Not strong.

In this paper, the real-time quantification process of network space traffic situation is realized by the method of network boundary traffic collection and characteristic characterization index system. This method continuously mines multiple dimensional feature indicators in real time from network traffic, draws network space and its traffic characteristics in real time. On the one hand, it can ensure the real-time or quasi-real-time monitoring, early warning, and emergency response requirements of the network situation, and on the other hand a small amount of log scale enables fine-grained characterization of network traffic characteristics. It can provides a high-quality basic information source for subsequent traffic anomaly analysis, detection and security warnings. Compared with traditional methods, it has obvious advantages in real-time, depiction accuracy, data scale, and data quality.
4.2. Analysis of results
The experiment in this paper collects the flow of the vertical network boundary of the new energy plant for a week, and uses part of the DARPA2000 intrusion detection data set provided by Lincoln Laboratory, which contains samples of different types of attacks and abnormal traffic. At the same time, it is compared with the wavelet analysis method to verify the effectiveness of this method.

Taking a certain indicator "TCP flow size" as an example, the TCP flow size signal can be decomposed into multiple components, as shown in the following figure (Figure 1 EEMD_TCP_ORIGINAL_PART3 is the original signal), and Figure 2 and Figure 3 are the slaves of the original traffic after EEMD decomposition. 11 flow components from high frequency to low frequency, r(12) is the residual quantity.

![Figure 4. Raw traffic](image1)

![Figure 5. Network traffic after EEMD decomposition](image2)
At the same time, the traditional wavelet analysis method is used to decompose the flow, and the decomposed components are analysed for the period. The wavelet analysis and decomposition period comparison of the EEMD decomposition period is shown in the following table:

| Component | Wavelet Cycle (min) | EEMD Cycle (min) |
|-----------|---------------------|------------------|
| IMF1      | 3.20                | D1 2.96          |
| IMF2      | 6.58                | D2 5.52          |
| IMF3      | 13.5                | D3 11.18         |
| IMF4      | 26.95               | D4 22.40         |
| IMF5      | 55.42               | D5 44.54         |
| IMF6      | 124.93              | D6 90.44         |
| IMF7      | 508.34              | D7 191.45        |
| IMF8      | 1053                | D8 387.95        |
| IMF9      | 1474.2              | D9 737.1         |
| IMF10     | 3685.5              | D10 1474         |
| IMF11     | 7371                | D11 2948.4       |
| r12       | 65535               | d12 65535        |

After analysis, it can be concluded that the component period after wavelet decomposition increases gradually at a rate of twice, and the components decomposed by EEMD have no obvious multiple growth relationship. The period of the IMF4 component is about 30 minutes, and the wavelet component D4 is about 30 minutes. Select component 4 of wavelet analysis and EEMD decomposition for further analysis. Select the detection window as the component period, that is, the detection window size of IMF4 is 27min, and the detection window size of D4 is 22min. Calculate the variance of the components obtained by the EEMD method and the wavelet transform method, and perform probability statistics to fit the probability distribution. The fitting method adopts the commonly used curve close to the variance distribution, and the fitting function is relatively simple Gaussian fitting. According to the obtained fitting curve, the point with the largest curvature of the curve is used as the dividing point, and the variance corresponding to the dividing point is calculated. The variance is used as the threshold. If it is greater than the threshold, it is considered that there is abnormal traffic, otherwise it is considered that the traffic in the detection window belongs to normal traffic. After many tests and comparisons, the test results are shown in the following table:

| Test Result | Method of this article | Wavelet method |
|-------------|-----------------------|---------------|
| Number of alarms | 20                    | 20            |
| Correct times    | 19                    | 14            |
| Number of errors  | 1                     | 6             |
| Correct rate     | 95%                   | 70%           |
| False alarm rate | 5%                    | 30%           |

It can be seen from the chart that the method studied in this paper can achieve 95% detection accuracy when the threshold is adaptively determined, and the false alarm rate is lower than wavelet.

5. In conclusion

This paper deeply identifies the security threats of new energy plants, analyzes the threat characteristics of the vertical network boundary of new energy plants, establishes the active monitoring and early warning capabilities of the network-related vertical network boundary attack threats of new energy plants, and real-time monitors and identifies the network from the plant side. It can further threaten the security of the power grid dispatching control system, prevent network information security attacks against the new energy power plant monitoring system, prevent power security accidents caused by power monitoring system security incidents, and ensure the security of the new energy power plant and the power grid as the source of the power grid stable operation.
The adaptive abnormal traffic detection technology for new energy farms and stations based on the HHT algorithm proposed in this paper improves the efficiency of identifying abnormal network traffic, and more accurately identifies cyber-attacks against new energy farms and new energy farms. Site’s cyber-attacks. The experimental results show that the method studied in this paper can achieve adaptive detection while adaptively determining the threshold, and the detection accuracy can reach 95%, and the false alarm rate is lower than that of wavelet, which can be used for new energy fields. Network detection provides more accurate recognition results.

Acknowledgment
Fund Project: Supported by the Science and Technology Project of State Grid Ningxia Electric Power Co., Ltd. (5229DK18002V)

References
[1] D. R. B. Miller, “A survey SCADA of and critical infrastructure incidents,” in Proceedings of the 1st annual conference on research in information technology, ACM, 2012.
[2] J. H. Guan J., J. Graham, “A digraph model for risk identification and management in SCADA systems,” in 2011 IEEE international conference on intelligence and security informatics (ISI), IEEE CP., 2011, pp. 150–155.
[3] H. J. A. Nicholson, S. Webber, S. Dyer, T. Patel, “SCADA security in the light of cyber-warfare,” Comput Secur., vol. 31, no. 4, pp. 418–436, 2012.
[4] M. F. Z. Tudor, “What went wrong? A study of actual industrial cyber security incidents,” in Industrial Control Systems Joint Working Group (ICSJWG) spring conference, 2010.
[5] RISI, “Industry attacks growing.” 2013. 
[6] H. Morgan, “Cyber security risk management in the SCADA critical infrastructure environment,” Eng Manag J, vol. 25, no. 2, pp. 38–45, 2013.
[7] “Cybersecurity Capability Maturity Model (C2M2) Program,” 2014 .
[8] “Electricity Subsector Cybersecurity Capability Maturity Model (ES-C2M2),” 2014.
[9] Tian Wang, Meina Qiao, Aichun Zhu, Guangcun Shan, Hichem Snoussi. Abnormal event detection via the analysis of multi-frame optical flow information[J]. Frontiers of Computer Science, 2020, 14(2).
[10] Guanglu Wei, Zhonghua Wang. Adoption and realization of deep learning in network traffic anomaly detection device design[J]. Soft Computing, 2020 (prepublish).
[11] Stanimir Kabaivanov, Kabaivanov Stanimir, Markovska Veneta. Hybrid deep-learning analysis for cyber anomaly detection[J]. IOP Conference Series: Materials Science and Engineering, 2020, 878(1).
[12] CHEN Jilei, QI Yunsong. Intrusion detection method based on deep learning[J]. Journal of Jiangsu University of Science and Technology (Natural Science Edition), 2017, (6): 795-800.
[13] Zhao Ying, Han Chunhao. Application and research of Markov model in network traffic classification[J]. Computer Engineering, 2018, 44(5):291-295.
[14] Wu Jinlong, Shi Xiaofei, Xu Jia et al. Research on Hierarchical Trojan Detection Technology Based on Random Forest[J]. Communication Technology, 2016 (04): 475-480.
[15] Chen Lulu, Guo Wenpu, He Hao. An abnormal flow detection method based on ME-PGNMF[J]. Computer Engineering, 2018(1):165-170.
[16] Pawel Dymora, Miroslaw Mazurek. An Innovative Approach to Anomaly Detection in Communication Networks Using Multifractal Analysis[J]. Applied Sciences, 2020, 10(9).
[17] Fatemeh Safara, Alireza Souri, Masoud Serrizadeh. Improved intrusion detection method for communication networks using association rule mining and artificial neural networks[J]. IET Communications, 2020, 14(7).
[18] Ying Zhong, Wenqi Chen, Zhihaliang Wang, Yifan Chen, Kai Wang, Yahui Li, Xia Yin, Xingang Shi, Jiahai Yang, Keqin Li. HELAD: A novel network anomaly detection model based on heterogeneous ensemble learning[J]. Computer Networks, 2020, 169.