RESEARCH ON ATTACK DATA MINING ALGORITHM OF INTERNET OF THINGS IN BIG DATA ENVIRONMENT

Fuguang Yao
Information Center, Chongqing University of Education, Chongqing, 400065, China

ABSTRACT: In big data environment, in order to ensure the security of Internet of things, it is necessary to mine attack data accurately. An attack data mining algorithm is proposed based on envelope spectrum feature extraction in big data environment. The time-domain model of attack data distribution in big data environment is constructed by using the non-stationary time-varying sequence analysis method, and the signal modeling of the attack data in the Internet of things is carried out by combining the signal fitting method. Envelope feature detection method is used to analyze the envelope features of the attack data in the Internet of things, and the filtering detection of the attack data in the Internet of things is realized by using the matched filtering method. The effective mining of the offensive data of the Internet of things under the environment of big data is completed. The simulation results show that the algorithm is accurate and the probability of attack spectrum detection is improved. It has a good application value in the security design of Internet of things.

KEYWORDS: Big data environment; Internet of things; attack; network security

1 INTRODUCTION

In the big data environment, a large number of data are transmitted through the Internet and the Internet of things. People pay more attention to the security of the network. When the network is attacked, the network security is designed to the users' privacy information security and the network's own network security. Information leakage and network paralysis will occur, seriously affecting the normal network operation. It is very important to study the data mining method of network attack in big data environment. Under the environment of big data, the Internet of things is vulnerable to network attack, and the attack forms on the Internet of things are diversified, the intensity and hidden characteristics of the attack are enhanced, and the effective and accurate detection of the attack data of the Internet of things is achieved. It can ensure the safety of the Internet of things and improve the survival ability of the Internet of things[1].

Traditionally, the main algorithms for feature extraction and detection of Internet of things attack data are time-frequency analysis based network attack detection algorithm, empirical mode decomposition attack detection method, wavelet analysis based network attack detection algorithm, and so on[2,3]. With the rapid development of signal processing technology and information processing technology, modern signal and information processing theory is introduced into network security detection to realize signal analysis and attack data feature extraction, and to improve the security of Internet of things. In reference [4], a data mining and feature extraction algorithm for Internet of things attacks based on blind separation of time-frequency features and adaptive cascading filtering in multi-sensor networks was proposed, which was based on the time-frequency features of attack data and virus data. The association rule mining method is used to realize the data mining of Internet of things attacks. However, the computation cost of the method in attacking data mining is high, and the real-time performance of security detection is not good. In reference [5], a data mining algorithm based on fractional Fourier transform and post-focusing of higher-order cumulants is proposed. The dual-threshold energy detection is used to extract the beam-feature of IOT attack data. The directivity of attack detection is improved, but the probability of false alarm is higher when SNR is low.

Aiming to solve the above problems, this paper proposes a data mining algorithm for Internet of things attacks in big data environment based on envelope spectrum feature extraction. The time-domain model of Internet of things attack data distribution in big data environment is constructed by using non-stationary time-varying sequence analysis method, and the signal model of Internet of things attack data is built by combining signal fitting method. The envelope feature of the attack data is analyzed by using the envelope feature.
detection method, and the filtering detection of the attack data of the Internet of things is realized by using the matched filtering method. The effective mining of the attack data of the Internet of things is completed under the environment of big data. Finally, the simulation results show that the proposed method can improve the capability of Internet of things attack data mining in big data environment.

2 INTERNET OF THINGS ATTACK AND PRE-PROCESSING OF ATTACK DATA

2.1 Network structure analysis and Internet of things attack data model construction

In order to detect the attack data of the Internet of things (IOT) with full directionality, it is necessary to fit and analyze the network structure attack node and the interference node under the multi-source distributed attack. Build the model of Internet of things attack data. The network environment studied in this paper is a cross-platform network environment. Cross-platform network system supports users to share the high-speed operation and data management of hundreds of thousands of servers in any terminal and any location[6]. At the same time, it also brings loopholes to the network intrusion. In the cross-platform network environment, the communication protocol is very different, resulting in the attack source has multi-distributed attributes. In the data mining of the Internet of things attack data, the distributed model of the attack node in the cross-platform network environment is analyzed, as shown in figure 1.

![Fig.1. Distributed model of the attack nodes](image)

Combined with figure 1, the undirected graph model of the application support layer of the Internet of things is given, where node \( G = (V, E) \) is any node \( v \in V \) in the application service layer, that is any connected edge in the network, that is \( e \in E \). The attack source of the Internet of things is located in the attack of the Internet of things. The channel bandwidth during the attack is \( T = K f_T \), and the high order cumulant structure of attack object middleware is defined as \( G_z, G_z = (V, E) \), and each cluster head node frame is divided into \( N \times N \) unit matrix. With multi-source distributed virus plants, the access service to the Internet of things is interrupted, and the purpose of virus invasion is achieved. In the Internet of things environment, the middleware of multi-source attack distributed feature is \( G_z, G_z = (V_z, E_z) \), \( G_z \subseteq G_i \). Based on the analysis of the distributed model of the multi-source attack nodes mentioned above, the attack data model of the Internet of things is constructed under the background of color noise[7]. The attack data of the Internet of things is a set of multi-source broadband signals with time-frequency scale coupling. Assuming that the network multi-source attack system is a three-dimensional continuous MIMO system, the attack signal is time-varying and non-stationary. Under the condition that the phase \( \phi(t) \) of the multi-source attack node is uniformly sampled, the estimation is based on the time-frequency joint distribution characteristic estimation. The single component signal model of Internet of things attack is as follows:

\[
g(t) = \sqrt{s} f(s[t - \tau])
\] (1)

Where, \( f(t) \) is the single component main frequency feature of the Internet of things attack, \( s = (c - \nu)/(c + \nu) \), called time scale factor. Wavelet scale transform is used to decompose the time scale, and a normalized scale parameter is obtained, which represents the Internet of things. Attack data scaling changes. In addition, the attack delay \( \tau = 2R/c \), \( R \) is the Euclidean distance between the dormant node and the attack node in the Internet of things[8].

Continuous wavelet transform is used to decompose the data of Internet of things attack by empirical mode characteristic, a set of two-dimensional functions of time scale \( a \) and time shift \( b \) which represent the interior details of the attack data are obtained, and the wavelet function is used[9]. The number is the mother wavelet, and the transformation process expression of wavelet decomposition for the Internet of things is obtained as follows:
When the collected attack data \( x(k) = s(k) + w(k) \) is a quasi-stationary random signal, under the action of limited SNR, wavelet scale decomposition combined with affine change is used to construct the mathematical model of the attack data of the Internet of things. The distribution characteristic makes the attack data of the Internet of things keep the best impulse function in the time frequency space of wavelet basis function. If the attack data of the Internet of things is taken as the mother wavelet function, the multi-source wave of the attack data of the Internet of things can be guaranteed. The focusing directivity of beam lays a foundation for feature extraction and accurate detection of Internet of things attacks[10].

### 2.2 Matched filter detection

Based on the above data model of Internet of things attack, it can be found that the Internet of things attack is hidden in the legitimate signal, and the noise interference is large, so it is difficult to achieve accurate feature extraction in the case of low SNR. The feature extraction and signal detection of Internet of things attacks need to be filtered. In this paper, the structure of second-order lattice notch filter is designed by using the principle of self-matched filtering. Because adaptive IIR notch can effectively realize spectrum enhancement and noise cancellation of the attack data of IOT, it has a better performance in the filtering of IOT attack[11], the block diagram of matched filter structure for attack data mining is shown in figure 2.

![Block diagram of matching filter for attack data mining](image)

Fig.2. Block diagram of matching filter for attack data mining

Assume that the input attack data \( x(k) \) is a set of multi-source broadband FM sequences, which is composed of the noise represented by the Internet of things attack data and the legitimate network data, and \( y(k) \) is the Internet of things attack data through the second-order lattice trapping. When the notch frequency point of the filter falls within the range of prediction errors in \( s(k) \), noise cancellation is carried out by convolution processing of the prediction error of the preceding term or the prediction error of the latter term[12]. Adaptive weight is \( \hat{w}(0) = 0 \). The iterative formula of the filter is:

\[
\theta_1(k + 1) = \theta_1(k) - \mu \text{Re}[y(k)\hat{\phi}^*(k)]
\]  

(3)

Based on the above analysis, the transfer function of the notch filter is obtained as follows:

\[
H_z(z) = \frac{(1 + \sin \theta_1)}{\cos \theta_2} + \frac{1 - z^{-2}}{1 + \sin \theta_2}(1 + \sin \theta_2)z^{-1} + \sin \theta_2z^{-2} \]

Where

\[
G(z) = 1 - \sin \theta_1
\]

(5)

The input signal \( u(k) \) of the Internet of things attack data is processed by adaptive IIR to minimize the error between the output signal and the expected signal, as:

\[
\epsilon(k) = d(k) - y(k) = d(k) - \sum_{i=1}^{\infty} W_i x(k - i)
\]

(6)

The mathematical expectation is taken from both sides of the upper formula to realize the interference filtering of attack data, which is helpful to improve the detection accuracy of Internet of things attacks.

### 3 OPTIMIZATION OF ATTACK DATA MINING ALGORITHM

#### 3.1 Envelope feature extraction of Internet of things attack data

The time-domain model of attack data distribution in big data environment is constructed by using the non-stationary time-varying sequence analysis method, and the signal modeling of the attack data in the Internet of things is carried out by combining the signal fitting method. Envelope feature detection method is used to analyze the envelope characteristics of the attack data[13]. If the sampling data of N IOT attack data is given \( x(n) \), \( n=1,2,...,N \), the rectangular envelope of the attack data is calculated, and the bilinear transformation of the attack data is obtained as follows:

\[
s(t) = \sum_{n=1}^{N} p_i \sin(\omega_i n + \Phi_i) + \zeta(n)
\]

(7)

Where, \( \Phi_i \) is the instantaneous frequency, \( \zeta(n) \) is the Euclidean distance between two attack points,
$p_k$ is the scale parameter, and the spectrum distortion part is estimated as:

$$\frac{1}{2\pi s} \sum_{k=-q/2}^{q/2} b_k \phi(n+c_k) = \hat{f}_k(n)$$

(8)

The multi-source beamforming processing of the high dimensional feature vector is carried out, and the multi-source beamforming envelope of the outputted IOT attack data is obtained by using adaptive beamformer:

$$u(t) = \frac{1}{\sqrt{T}} \text{rect}(1/T) \exp(-j[2\pi K \ln(1/2) - t])$$

(9)

where $\text{rect}(t) = 1, |t| \leq 1/2$. The envelope characteristics and limiting amplitude-frequency characteristics of multi-source beam are further analyzed. Under wavelet transform, the frequency modulation law of the envelope feature of the beam is obtained as a hyperbolic function, that is:

$$f(t) = \frac{K}{t_0 - t}, |t| \leq T/2$$

(10)

where $K = T f_{s0} f_{\text{min}}/B, t_0 = f_0 T/B$, envelope feature detection method is used to analyze the envelope features of the attack data in the Internet of things, and the filtering detection of the attack data in the Internet of things is realized by using the matched filtering method[14].

3.2 Attack data mining

The directional beam characteristics of multidimensional phase space of Internet of things attack data is used. The scale affine transformation of IOT attack data is obtained as follows:

$$\psi_{a,b}(t) = [U(a,b)\psi(t)] = \frac{1}{\sqrt{|a|}} \psi(t-b/a)$$

(11)

By affine transformation, the attack data is matched with the kernel function in a multi-source attack phase. The time domain $f(t)$ of the attack data is taken as the generating wavelet function $\psi(t)$ of energy accumulation, $a = 1/s, b = \tau$, and the variables are replaced in the whole time-frequency plane:

$$f_{s0}(t) = [U(1/s, \tau)f(t)] = \sqrt{|a|} f(s(t-\tau))$$

(12)

The non-uniform sampling output of the amplitude-frequency response characteristics of the Internet of things attack data is obtained by detecting the envelope features of the independent variables:

$$\hat{f}_k(t, \tau) = \frac{1}{2\pi \tau} \sum_{k=-q/2}^{q/2} b_k \phi(t + c_k \tau)$$

(13)

According to the random linear separation invariance of attack signal and the time-frequency characteristic of eigenwave output after the formation of multi-source wave, the spectrum of attack data of Internet of things is detected[15].

When the autocorrelation variables $X$ and $S$, are random independent, $i = 1, 2, L, N$. In combination, the variance and mean value of these randomly separated variables are obeyed to the Gaussian distribution, and the filtering detection of the attack data of the Internet of things is realized by combining the matched filtering method, and the effective mining of the attack data of the Internet of things under the big data environment is completed.

4 SIMULATION EXPERIMENT AND RESULT ANALYSIS

In order to test the performance of this method in implementing Internet of things attack data mining in big data environment, the simulation experiment is carried out. The algorithm is designed and implemented by using Matlab 7 mathematical simulation language. The test set of Internet of things attack data is collected in MIT Lincoln Lab KDD Cup2015 virus database. The discrete sampling rate of attack sample data is $f_s = 10^s f_0 Hz = 10kHz$. The length of the system is 1024, the number of visits to the virus database is 2000, the number of attack signal samples is 1200, the signal model is constructed through the above data acquisition results, the bandwidth of the attack data of the Internet of things is obtained, according to the above simulation environment and Parameters, the data mining of Internet of things attack is carried out, and the mining result is shown in figure 3.

![Figure 3](image-url)
The analysis of Figure 3 shows that this method is used to exploit the data mining of the Internet of things. The data mining output has better envelope directivity. It has good anti-interference performance, effectively restraining the interference of the sidelobe beam, improving the directivity of the beam, improving the ability to detect the attack of the Internet of things, and testing the number of attacks on the Internet of things in different ways. According to the accuracy of mining, the comparison of detection performance curves is shown as shown in Figure 4. Analysis figure 4 shows that this method has higher accuracy and better performance in the Internet of things attack data mining.

5 CONCLUSION

In this paper, an attack data mining algorithm is proposed based on envelope spectrum feature extraction in big data environment. The time-domain model of attack data distribution in big data environment is constructed by using the non-stationary time-varying sequence analysis method, and the signal modeling of the attack data in the Internet of things is carried out by combining the signal fitting method. Envelope feature detection method is used to analyze the envelope features of the attack data in the Internet of things, and the filtering detection of the attack data in the Internet of things is realized by using the matched filtering method. The effective mining of the offensive data of the Internet of things under the environment of big data is completed. The simulation results show that the algorithm is accurate and the probability of attack spectrum detection is improved. This method has good application value in the security design of Internet of things.

6 ACKNOWLEDGEMENTS

This paper supported by the special issue “Future School (preschool education)” of NATIONAL CENTER FOR SCHOOLING DEVELOPMENT PROGRAMME (NO.CSDP18FC3204); Project supported by the major issue of Chongqing Education Science "13th Five-Year” Plan in 2017 (No.2017-GX-139); Scientific Research Platform of Chongqing University of Education (NO.2017XJPT07); Scientific Research Program of Chongqing University of Education (No. KY2016TZ02).

7 REFERENCES


