Network APT Attack Detection Based on Big Data Analysis

Guo-gen Fan[1(✉)] and Jian-li Zhai[2]

[1] Guangzhou Huali Science and Technology Vocational College, Guangzhou 511325, Guangdong, China
fanguogen@foxmail.com
[2] Huali College Guangdong University of Technology, Guangzhou 511325, Guangdong, China

Abstract. In order to improve the security of the distributed optical fiber sensing network, the self-adaptive detection of the fiber sensing network needs to be carried out, and an overlap detection algorithm under the APT attack of the distributed optical fiber sensing network based on the spectral characteristic component and the big data analysis is proposed. the large data sampling model of the network APT attack is constructed, the attack characteristics and the related properties of the distributed optical fiber sensing network virus are simulated by adopting the spectrum correlation characteristic detection and the large-data quantization characteristic coding, and the large-data fusion and feature extraction of the APT attack information are realized, the output abnormal characteristic detection of the distributed optical fiber sensing network is carried out through the feature extraction result, a distributed optical fiber sensing network intrusion large data statistical analysis model is constructed, and a narrow-band signal spectrum offset correction method is adopted, And calculating the connection probability density and the individual infection probability of the APT attack node, and improving the detection capability of the network APT attack. The simulation results show that the algorithm can effectively implement the network APT attack detection, improve the security detection capability of the network APT attack, and has a good network security protection capability.

Keywords: Big data analysis · Network · APT · Attack detection · Sensing network

1 Introduction

With the development and popularization of the network technology, the network security is becoming more and more concerned. The Internet has the characteristics of openness, no competence and non-security, and the network attack network virus is on the rise of the number and the degree of harm [1]. The present situation of network security is worrying due to the direction and non-regularity of the virus and the invisibility. according to the results of the American VIRUS official statistics, the average global average of 30 million computers has been attacked by various viruses, and the detection of network attack is an important means to protect the security of the

Y.-D. Zhang et al. (Eds.): ICMTEL 2020, LNICST 326, pp. 338–348, 2020.
https://doi.org/10.1007/978-3-030-51100-5_30

computer system, and how to establish a high-efficiency network APT attack detection method, It is a focus of many experts in this field [2]. At present, the new type of network virus's invasion and propagation is carried out in a normal way for continuous attack. It is difficult to detect, and it is necessary to effectively detect the APT attack. It is of great significance to study the detection algorithm of distributed optical fiber sensing network virus under continuous attack [3].

The aggressive behavior of the network virus is a distributed optical fiber sensing network APT attack signal in the data of the computer network, and the intrusion attack is carried out on the network user in the form of data information. The traditional method is difficult to realize the effective detection, and the main detection method is based on the detection method of the ant colony algorithm, the detection method based on the immune genetic algorithm and the detection method based on the neural network algorithm. Among them, the most common is the overlapping detection method of APT attack based on the neural network algorithm [4]. Distributed optical fiber sensing network APT attack detection algorithm with any large frequency, Which is suitable for realizing large-frequency signal detection, but has limited detection performance for multi-frequency network intrusion signals, high calculation complexity and difficulty in implementation [5]. Interference attack location method is proposed to optimize the state transition feature of the intrusion tolerant system, and the intrusion path of the virus is analyzed and analyzed. The structure level characteristic decomposition of the potential intrusion signal is realized, and a certain effect is obtained. However, the algorithm cannot effectively realize the overlapping detection of the continuous attack virus, and proposes an overlap detection algorithm under the APT attack of the distributed optical fiber sensing network based on the spectral characteristic component and the big data analysis [6]. By the improved design of the algorithm, the test performance of the network APT attack is improved, and the performance verification of the simulation experiment shows its effectiveness.

2 APT Attack Model of Distributed Optical Fiber Sensing Network

In order to extract the continuous attack signal model of virus, it is necessary to construct the continuous attack model of distributed optical fiber sensor network virus, design the virus state transition model of distributed optical fiber sensor network, describe the virus system of distributed optical fiber sensor network as five states of Markov chain HMM, and analyze the security attributes of intrusion tolerance system [7]. Spectrum correlation feature detection and big data quantitative feature coding are used to simulate the attack characteristics and related properties of distributed optical fiber sensor network virus, and big data fusion and feature extraction of APT attack information are realized. described by Langevin equation, the APT attack of distributed optical fiber sensor network is a bistable nonlinear driving multi-frequency resonance model, and the expression of Langevin equation is as follows:

$$\frac{dx}{dt} = ax - bx^2 + s(t) + \Gamma(t) \tag{1}$$

In the formula, a, b is the system parameter. By adjusting the system parameters, the interference noise can be described by mathematical model as follows:

$$f(x) = sgn\left\{ z \sum_{i=1}^{l_1} \alpha_i^{+} y_i K(x_i, x) + \sum_{i=1}^{l_2} \alpha_i^{-} y_i K(x_i, x) + b \right\} \tag{2}$$

In the process of APT attack in distributed optical fiber sensor network, assuming that the amplitude of APT attack signal is A, the amplitude adjustment coefficient of input signal is as follows:

$$x(t) = \sum_{i=0}^{p} a(\theta_i) s_i(t) + n(t) \tag{3}$$

Firstly, the local extreme point of the signal in the stochastic resonance system is determined. In order to improve the detection performance of the continuous attack, the APT attack behavior is transformed into a convex combinatorial optimization problem. Using the wrapper feature selection model and the method of marking the average value of the upper and lower envelope lines, the EMD difference component of the APT attack signal is obtained by the nonlinear mapping function $\{(x_1, y_1), (x_2, y_2), \ldots (x_i, y_i), \ldots (x_n, y_n)\}$ of the sample set $\varphi(x)$:

$$\bar{\tilde{u}}^{*}_{e|v,k} = \tilde{u}_{e|v,k} + h(x_k) \tag{4}$$

The optimal classification hyperplane structure is carried out in the high dimensional feature space:

$$f(x) = w \cdot \varphi(x) + b = 0 \tag{5}$$

In the formula, w represents a weight vector, and b represents a threshold value. In order to minimize structural risk, the optimal classification plane should meet the following constraints:

$$y_i \cdot (w \cdot \varphi(x_i) + b) \geq 1 \tag{6}$$

The label information of the APT attack file block and the file block is correlated to the S-Table, a normal resonance function is introduced, the non-negative relaxation variable ξ_i is adopted to improve the classification SVM generalization ability of the learning method, and each grid point can only be occupied by one virus intrusion adsorption chain node, then the attack detection and optimization problem of the virus under the continuous attack is changed to:

$$\begin{aligned} &\min \tfrac{1}{2} w \cdot w + c \sum_{i=1}^{n} \xi_i \\ &s.t. y_i (w \cdot x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, 2, \ldots, n \end{aligned} \tag{7}$$

In the formula, the artificial immune method is used to carry out the transmission impedance and immunity of the virus [8], and the Lagrange multiplier is introduced to transform the above optimization problem into dual form.

$$\min \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\varphi(x_i) \cdot \varphi\left(x_j\right) \right) + \sum_{i=1}^{n} \alpha_i \tag{8}$$

At the same time, the following conditions need to be met:

$$\sum_{i,j=1}^{n} \alpha_i y_i = 0, c \geq \alpha_i \geq 0 \tag{9}$$

The algorithm calculation process is realized by MPI interface based on chain growth. The APT attack model of distributed fiber sensing network is constructed. The block diagram of model design is shown in Fig. 1.

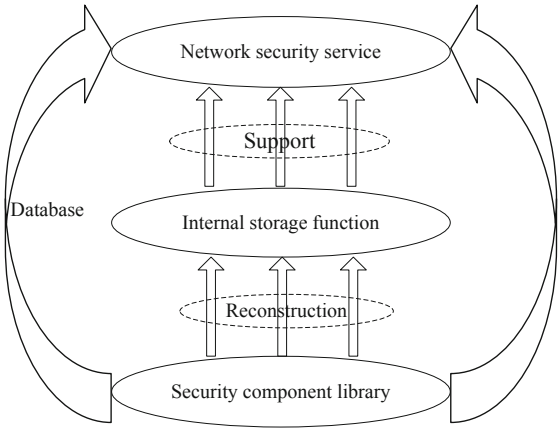

Fig. 1. APT attack detection model of distributed optical fiber sensing network

3 Propagation Probability of APT Attack in Distributed Optical Fiber Sensor Networks

The probability model of APT attack propagation in distributed optical fiber sensor network is analyzed to realize the overlapping detection of continuous attack signals. The APT attack of distributed optical fiber sensor network is connected to mobile

devices through a single connected virus. The success of transmission is related to the time of connection establishment and the propagation time of virus. The classification attribute based on information entropy is used to describe the population size of virus transmission [9]. The time-frequency analysis Viterbi algorithm is used to estimate the instantaneous frequency under APT attack, which can expand the frequency of APT attack on distributed optical fiber sensor network and reduce the attack efficiency. The single frequency signal of continuous attack signal of distributed optical fiber sensor network is described as follows:

$$s(t) = a(t)\cos\phi(t) \tag{10}$$

Since the instantaneous frequency of the distributed optical fiber sensing network is time-varying, the instantaneous spectrum corresponding to the instantaneous frequency should be present, the time-frequency analysis Viterbi algorithm is adopted, the average frequency of the continuous attack signal spectrum of the distributed optical fiber sensing network virus is equal to the time average of the instantaneous frequency, then:

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}[\arg z(t)] \tag{11}$$

Thus, the average frequency of the instantaneous spectrum is the instantaneous frequency, and the instantaneous frequency is the derivative of the phase of the analytical signal. Assuming that the time required for virus transmission is t_{virus}, the probability of successful transmission of the virus is the probability that the connection is established at t_{virus} or longer:

$$P_{t_{virus}} = 1 - F_{link}^{v1}(t_{virus}) = \frac{1}{\pi(b-a)}\int_0^{\pi}\int_0^{\frac{2R}{t_{virus}}} v\sqrt{1-\left(\frac{vt_{virus}}{2R}\right)^2}\, g(v,\phi,v1)dvd\phi \tag{12}$$

Many viruses are targeted only for specific vulnerabilities, such as for a particular bluetooth version, a particular operating system's vulnerable point, or even a particular vendor's device. Therefore, under the invasion of APT, the normal resonance of the network is suppressed, the time-frequency characteristic of the non-stationary continuous attack signal is extracted by using the time-frequency analysis method, the instantaneous frequency is estimated, and the frequency expansion is realized [10, 11]. As shown in Fig. 2, a continuous attack model from A to B in the coverage of the virus is obtained, as shown in Fig. 2, and is defined as a signal model and a propagation probability model in the distributed fiber-sensing network APT attack by the time from the node A to the node B through all the distances, And provides a signal basis for realizing the APT attack detection.

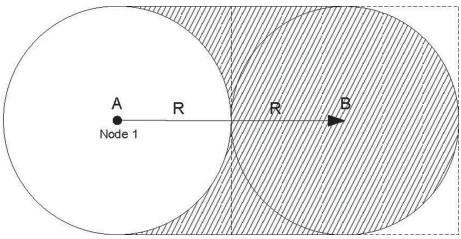

Fig. 2. Attack node distribution range of distributed optical fiber sensor networks

3.1 Improved Implementation of Detection Algorithm

The traditional detection method realizes the detection of the APT attack by using the hierarchical detection algorithm of the interference attack positioning state transfer feature extraction, and the algorithm does not adapt to the feature selection parameters under the continuous attack, and the detection performance is not good. In this paper, an overlapping detection algorithm based on a distributed optical fiber sensing network (APT) attack based on spectral characteristic component and big data analysis is presented [12, 13]. On the basis of the above-mentioned APT attack signal design, it is assumed that the amplitude of the continuous attack signal is A, and if the c_k is a fraction, the phase $\phi(t)$ is a non-uniform sample. the effect of the phase difference of the sampling interval is different, where S is the time sampling step (corresponding to Δt), and the spectrum characteristic of the narrow-band signal is as follows:

$$\begin{aligned} g(v,\phi,v1) &= \frac{u(h(v,\phi,v1)-a)-u(h(v,\phi,v1)-b)}{h(v,\phi,v1)} \\ h(v,\phi,v1) &= \sqrt{v^2+v1^2+2vv1\cos\phi} \end{aligned} \tag{13}$$

Where the F_{link}^{v1} represents the accumulated time of the connection, $v1$ represents the mobile speed of the node, the sg represents the relative movement speed of the node, and the selection of the class 1 in the model may be a "Smartphone X 'using the' Bluetooth V2". It is assumed that the virus is moving at a constant speed under a continuous attack, then:

$$v_{average} = \frac{a+b}{2} \tag{14}$$

As a kind of narrowband signal, APT attack signal produces spectrum offset through normal resonance. In this paper, the narrowband signal spectrum offset correction method is used to improve the detection performance, and the connection probability of APT attack node and the infection rate of individual can be expressed as a matrix [14–17]. Thus, the connection probability of the same class is allowed to be different, and the maximum propagation probability is defined as:

$$trans_{rate\,Max} = \frac{1}{t_{virus}} \tag{15}$$

Therefore, the connection probability can be defined as:

$$Contact_{rate} = \min\{node_{rate}, trans_{rate\,Max}\} \tag{16}$$

Finally, the dynamic system equation of the overlapping detection under the APT attack of the distributed optical fiber sensing network based on the spectral characteristic component and the big data analysis can be generalized as follows:

$$\begin{aligned}\frac{dS_k(t)}{dt} =& O_k(t)(P_{os1_k} + P_{os2_k}) + T_k(t)P_{tsk} - S_k(t)P_{spk} \\ & - S_k(t)\beta_{ak} - S_k(t)\left(\beta_{u_k(t)} + \beta_{l_k(t)}\right) - S_k(t)(P_{so1_k} + P_{so2_k})\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{dE_k(t)}{dt} =& S_k(t)\left(\beta_{u_k(t)} + \beta_{l_k(t)}\right) + X_k(t)P_{xe_k} - E_k(t)P_{ex_k} - E_k(t)P_{et_k} \\ & - E_k(t)(P_{ei1_k} + P_{ei2_k})\end{aligned} \tag{18}$$

$$\frac{dP_k(t)}{dt} = T_k(t)P_{tp_k} + S_k(t)P_{sp_k} \tag{19}$$

$$\frac{dT_k(t)}{dt} = I_k(t)P_{it_k} + E_k(t)P_{et_k} - T_k(t)P_{tp_k} - T_k(t)P_{ts_k} - T_k(t)(P_{to1_k} + P_{to2_k}) \tag{20}$$

Through the improved design of the above algorithm, the overlapping detection algorithm under the APT attack of the distributed optical fiber sensing network based on the frequency spectrum characteristic component and the big data analysis is realized, and the next step is to perform the performance verification through the simulation experiment [18–20].

4 Simulation Experiment and Result Analysis

In order to test the detection performance of this algorithm under APT attack on distributed optical fiber sensor network, the virus database is DARPA database, which is the real simulation and reproduction of distributed optical fiber sensing network data, and contains a wealth of continuous attack signal types. The decision threshold $G_T = 20\sigma^2$, adopts constant false alarm probability $p_f = 0.1$. With the collected network continuous signal, the signal is preprocessed by noise reduction filter, and the narrowband signal spectrum offset correction is used to detect and optimize the signal. The distribution of big data of the network attack is shown in Fig. 3.

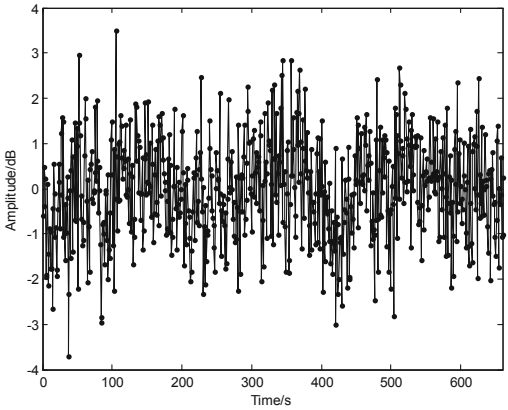

Fig. 3. Big data distribution of network attacks

The gain node tree with self-interference channel is set as $L = 2$, $N = 4$. For the convenience of calculation, the influence of APT attack probability of distributed optical fiber sensor network under different propagation time is shown in Fig. 4 through the design of the above simulation environment, the influence of APT attack infection probability of distributed optical fiber sensor network under the condition of fixed average node speed is obtained through the design of the above simulation environment.

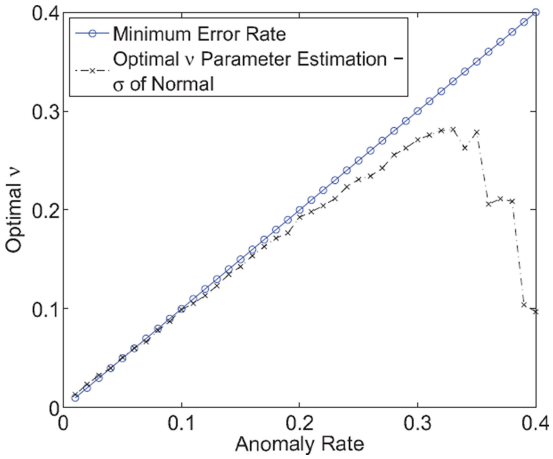

Fig. 4. Infection probability of APT attack in distributed optical fiber sensor networks

As you can see from Fig. 4, the time required for all users to infect is 50 days; when the action distance is 100 m, the time required for all people to infect is only 5 min. In order to further compare the detection performance of the algorithm under APT attack,

the overlap detection algorithm designed in this paper is compared with the traditional method, and the detection performance curve is shown in Fig. 5. It can be seen from the diagram that the proposed algorithm effectively improves the detection performance of network APT attacks and improves the overlapping detection ability of APT attacks.

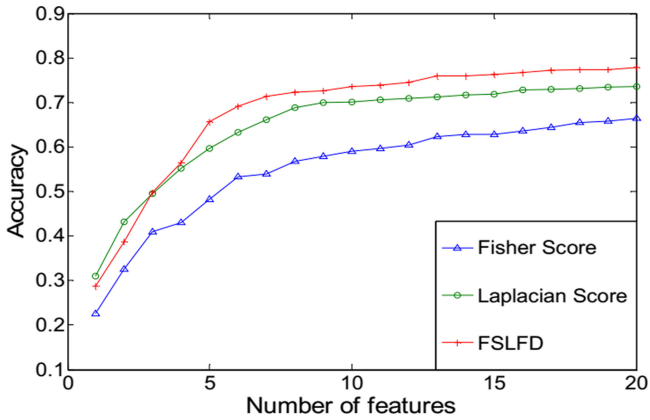

Fig. 5. Detection performance

5 Conclusions

The traditional network attack detection methods have the problems of weak security and poor detection results. An APT detection algorithm of the distributed optical fiber sensing network is proposed based on the spectral characteristic component and the big data analysis. The large data sampling model of the network APT attack is constructed, the attack characteristics and the related properties of the distributed optical fiber sensing network virus are simulated by adopting the spectrum correlation characteristic detection and the large-data quantization characteristic coding, and the large-data fusion and feature extraction of the APT attack information are realized, the output abnormal characteristic detection of the distributed optical fiber sensing network is carried out through the feature extraction result, a distributed optical fiber sensing network intrusion large data statistical analysis model is constructed, and a narrow-band signal spectrum offset correction method is adopted, And calculating the connection probability density and the individual infection probability of the APT attack node, and improving the detection capability of the network APT attack. The simulation results show that the algorithm can effectively implement the network APT attack detection, improve the security detection capability of the network APT attack, and has a good network security protection capability. This method has good application value in network security protection and virus intrusion detection.

Fund Project. 2019 Guangdong Higher Education Teaching Reform Project "Research on Network Database Learning Based on Learning Behavior Big Data Visualization"; 2019 Huali College Guangdong University of Technology Project "Research on Network Database Learning Based on Learning Behavior Big Data Visualization" (GGDHLYJZ[2019]No.32).

References

1. Huang, H., Lu, D., T., H.: Chover type law of iterated logarithm of NSD sequences. J. Jilin Univ. **56**(05), 1113–1118 (2018). Science Edition
2. Li, X., Kang, Z.: Ultra low Power and High Linear LNA based on double Cross Coupling Capacitance feedback. Autom. Instrum. **7**, 326–330 (2018)
3. Houg, X.F., Wang, H., Li, Y.: Research on efficient processing method of large amount of data based on HIVE and distributed Cluster. J. China Acad. Electron. Inform. Technol. **13** (3), 315–320 (2018)
4. Zhao, L.X.: Research and implementation of vehicle-mounted Charger based on DSP. J. Power Supply **15**(3), 158–162 (2017)
5. Guo, H.P., Dong, Y.D., Mao, H.T., et al.: Logistic discrimination based rare-class classification method. J. Chin. Comput. Syst. **37**(1), 140–145 (2016)
6. Gao, N., He, Y.Y., Gao, L.: Deep learning method for intrusion detection in massive data. Appl. Res. Comput. **35**(4), 1197–1200 (2018)
7. Zhang, Y.Z., You, R.: Wavelet variance analysis of EEG based on window function. Chin. J. Biomed. Eng. **23**(2), 54–59 (2014)
8. Yang, L., Kong, Z., Shi, H.: Multi-controller dynamic deployment strategy of software defined spatial information network. Comput. Eng. **44**(10), 58–63 (2018)
9. Liu, Y., Du, Z., Zhao, Q.: Bifurcation analysis of the ENSO recharge oscillator with time-delayed feedback. Appl. Math. Mech. **39**(10), 1128–1136 (2018)
10. Niu, W., Zhang, X., Yang, G., et al.: Modeling attack process of advanced persistent threat using network evolution. IEICE Trans. Inf. Syst. **100**(10), 2275–2286 (2017)
11. Shen, X., Qin, S.: Anomaly detection based on synthetic minority oversampling technique and deep belief network. J. Comput. Appl. **38**(7), 1941–1945 (2018)
12. Yang, Y.H., Huang, H.Z., Shen, Q.N., et al.: Research on intrusion detection based on incremental GHSOM. Chin. J. Comput. **37**(5), 1216–1224 (2014)
13. Liu, L., Liu, S.: Dynamic fuzzy clustering algorithm based on weight difference. J. Jilin Univ. **57**(03), 574–582 (2019). (Scientific version)
14. Ma, Y., Zhang, Z., Lin, C.: Research progress in similarity join query of big data. J. Comput. Appl. **38**(4), 978–986 (2018)
15. Du, Z., Zhao, Q.: Bifurcation analysis of the ENSO recharge oscillator with time-delayed feedback. Appl. Math. Mech. **39**(10), 1128–1136 (2018)
16. Xu, X., Wang, S., Li, Y.: Identification and predication of network attack patterns in software-defined networking. Peer-to-Peer Netw. Appl. **12**(1), 1–11 (2018)
17. Bang, J., Cho, Y.-J., Kang, K.: Anomaly detection of network-initiated LTE signaling traffic in wireless sensor and actuator networks based on a Hidden semi-Markov Model. Comput. Secur. **65**(6), 108–120 (2017)

18. Yin, C., Xia, L., Zhang, S., et al.: Improved clustering algorithm based on high-speed network data stream. Soft. Comput. **22**(4), 1–11 (2017)
19. Park, Y.H., Yun, I.D.: Arrhythmia detection in electrocardiogram based on recurrent neural network encoder–decoder with Lyapunov exponent. IEEJ Trans. Elect. Electron. Eng. **14**(2), 1273–1274 (2019)
20. Brito, C.J., Miarka, B., de Durana, A.L.D., et al.: Home advantage in Judo: analysis by the combat phase, penalties and the type of attack. J. Hum. Kinet. **57**(1), 213–220 (2017)