Analysis on NSAW Reminder Based on Big Data Technology

Ziheng Jin*

[1.] The 5th Business Division Nanjing Research Institute of Electronic Engineering, Nanjing, Jiangsu, China
[1] jinziheng100@ricetgc.edu.cn
* corresponding author

(Received: April 23, 2022; Revised: May 15, 2022; Accepted: July 12, 2022; Available online: September 30, 2022)

Abstract

NSAWS is an intelligent and real-time large database management system. By analyzing the user identity data and access rights contained in the collected information, the NSAWS finds out the potential risks and issues an alarm notice in a timely manner. This paper mainly studies how to strengthen the prevention of network attacks in BD environment from the following aspects. This paper first introduces the common technology on the Internet in our country and its application status; secondly, it expounds the architecture, deployment mode and operation mode of the large database system based on the basic security facilities such as cloud computing platform and firewall. Then the traditional NSAWS is analyzed and the simulation platform is tested. The results show that the platform has high accuracy and stability in alerting network security hazards and can effectively protect network security.

Keywords: Big Date Technology, Network Security, Security Early Warning, Early Warning Reminder

1. Introduction

With the wide application of computer network, a variety of malicious network attacks continue to emerge, and develop to the direction of large-scale, automation and co-assimilation, the harm of attacks is increasing. In order to ensure network security to the greatest extent, people widely use firewall, antivirus software, intrusion detection system and other security facilities, but they can only meet some network security needs. Users also hope that according to the current network security situation, predict the future network attacks, identify the final attack intention of the attacker, and early warning of the upcoming attacks [1, 2].

In recent years, with the rapid development of cloud computing, Internet and Internet T, the number of data collection terminals has increased rapidly. Many scholars have made many achievements in the network security of BD processing, and the rapid growth of network BD has brought severe challenges to the storage and computing performance of traditional hardware. According to a report released by International (International Data Corporation,IDC) estimated 1.8ZB, by 2020 and summarizes network BD as 5V features: Mass (Volume), Variety (Variety), Identification (Veracity), Fast (Velocity), and Low Value Large Density (Value). Network BD has the characteristics of large scale, variety (structured, semi-structured and unstructured), sudden, emergent, making it difficult for researchers to evaluate and predict their changing states [3, 4].

This paper breaks the limitations of the traditional NSSA model, takes a variety of distributed algorithms as the core of the model, combined with the high-dimensional, large number and rough characteristics of network big data, puts forward a NSAW reminder model to deal with network danger scenarios [5, 6].

2. Overview of NSAW Reminder Based on BD Technology

2.1. The Concept of Network Big Data

In recent years, with the rapid development of cloud computing, Internet of things, Internet of things and other information technology, the number of data acquisition terminals has increased rapidly. The era of BD has arrived, and the rapid growth of data has become a challenge and opportunity for many industries. Network BD refers to the large-scale data produced by the blending of real life, computer network, mobile Internet, Internet of things and other pluralistic worlds. At present, the rapid growth of network BD has brought severe challenges to the storage and computing performance of traditional hardware. Network BD has large scale and many kinds (structured, semi-structured and unstructured). It is difficult for researchers to evaluate and predict the state of change because of its sudden and surging characteristics. The development of science and technology is twofold, and the network BD is no exception, which brings value and leads to a lot of security problems at the same time [7, 8].

This paper proposes a variety of NSSA models based on big data, including the cleaning and extraction of network big data, the storage and management of network big data, and the mining and calculation of network big data. The purpose of this paper is to construct a new network BD analysis and processing model by analyzing the fundamental characteristics of network big data, such as huge scale and complex relationship, which can comprehensively and effectively deal with the network security problems in BD environment and solve all kinds of known and unknown problems in network security from the point of view of network security situation. At present, the NSSA technology based on BD is not mature, and even the precise definition still lacks a unified standard. Therefore, a set of new theories and methods are needed to guide the research of network big data. This paper breaks the traditional data analysis method and network security situation awareness processing mode, puts forward a new idea of network security situation awareness under BD environment, and finds an effective network security situation awareness model from the point of view of network data traffic characteristics [9, 10].

2.2. Network Security Threats

As the Internet evolves, cyber-security threats have also become increasingly diverse. From early simple system vulnerabilities, web counterfeiting, to increasingly complex and difficult to guard against intrusion penetration, DoS attacks, etc. Reviewing a more famous event: on July 18, 2009, the online auction of private car quota in Shanghai was attacked by a clearly targeted DoS in the last five minutes of the first bid, causing many bidders unable to bid normally and were forced to cancel. Back in February 2000, similar methods of attack rose from theory to practice and began to become popular. Almost all kinds of well-known websites at home and abroad have been attacked by the same attacks. Yahoo, eBay, Amazon CNN and other websites in the United States forced the president to ask personally. Many examples show that the DoS attack has become an important threat to the current Internet network security. Especially with the development of attack technology and the gradual integration of reflection attacks, botnets, and even peer-to-peer networks (P2P) technologies, traditional DoS attacks have developed into more diverse and harmful distributed denial of service (DDoS) attacks, creating unprecedented difficulty to protect important websites and information systems [11, 12].

2.3. NSAW

NSAW mainly finds signs of intrusion according to network abnormal traffic, network abnormal behavior, virus threat and so on. When the ultimate goal of the attacker is not achieved, the intrusion process is matched by the pre-built attack model to judge the possible attack behavior of the attacker in the next step, and the impact of the attack on the network and the threat that will be posed by the attack are evaluated. The purpose of NSAW is to take proactive defense measures such as network isolation, attack blocking and so on before the attacker further endangers the system.

2.4. Related Algorithms of BD Network Security Technology

2.4.1. Association analysis algorithm

The FP-Growth algorithm employs the strategy of growing frequent patterns to mine frequent sets without generating candidate patterns. The algorithm presses the data into memory while building the book, which only needs to scan the dataset twice, greatly less I/0 overhead, and thus has a great advantage in processing big data. Define the following symbol, and Rmin represents support (the proportion of records containing specific attributes).

$$R = \frac{Count\,(w \cup e \cup t)}{n} \tag{1}$$

2.4.2. Dynamic Time Warping Algorithm

The standardization of each attribute can effectively avoid the influence of different units on the distance calculation caused by the disunity of data and the different distribution of attributes. The standardized formulas for reference templates and test templates are as follows:

$$E = \frac{e_i - r_m}{r_a} 1 \in [1, 9] \qquad (2)$$

E represents their standard difference, ei represents the mean of 4 properties, and rm represents their standard difference.

3. System Design of NSAW and Reminder Based on BD Technology

3.1. Architecture of NSAWS

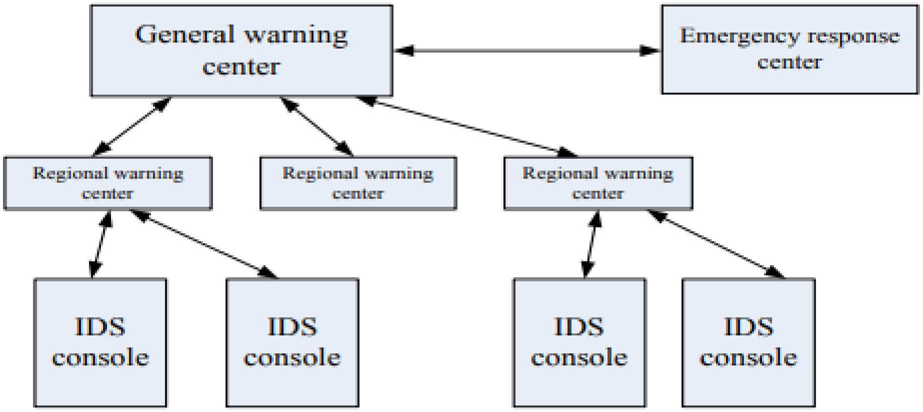

Figure 1. Distributed NSAW architecture

NSAW system is a decision support system for open information sources, which provides threat assessment and early warning for network attacks. In order to achieve the purpose of NSAW, researchers put forward different NSAWS architectures. In this paper, a distributed early warning architecture based on Web is proposed by using multi-level structure.

In Figure 1, the system will issue early warning based on risk assessment and aggregate the data to the Advanced early warning Center to conduct a comprehensive scattered data threat assessment to determine whether early warning is required. In the event of an attack, an early warning should be issued to the emergency call center, which does not fall within the scope of the NSAWS. Early warning centers at all levels can effectively exchange information to ensure reliable, accurate and timely early warning.

3.2. Architecture of the System

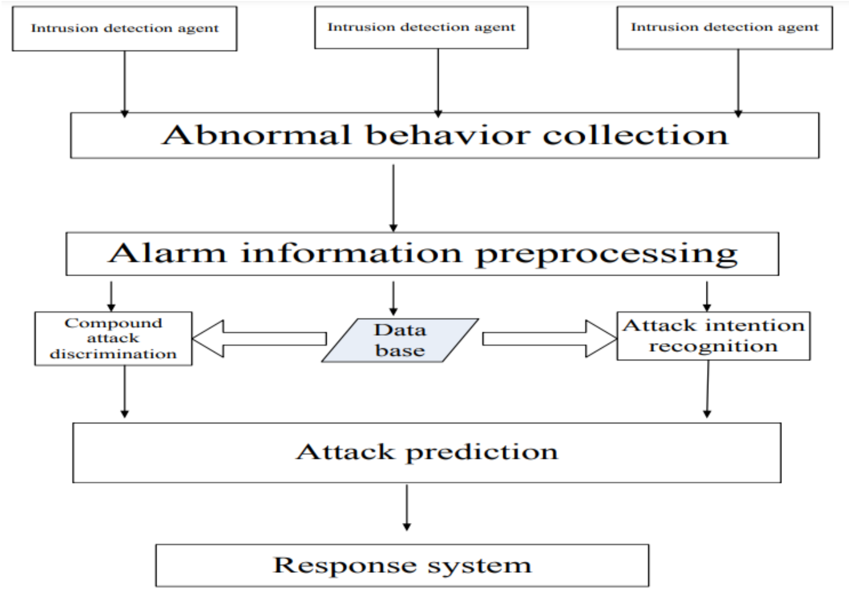
Intrusion detection agent
Intrusion detection agent
Intrusion detection agent
Abnormal behavior collection
Alarm information preprocessing
Compound attack discrimination
Data base
Attack intention recognition
Attack prediction
Response system

Figure 2. System architecture

The intrusion process shows complexity, diversity and distribution. In order to achieve the final purpose of the attack, the attacker cannot complete the task by simple attack alone, and needs a composite attack composed of multiple simple attacks. General intrusion detection system only alarms simple attacks, in order to effectively warn against composite attacks, according to the requirements of NSAW, the prototype system adopts distributed NSAW architecture. Figure 2 is the specific structure of the system.

3.3. Functional Composition of the System

3.3.1. Abnormal behavior collection

NSAW requires real-time tracking of abnormal behavior on the network, timely and accurate detection of attacker intention and the next attack. Abnormal behavior collection is an important part of NSAWS, which is mainly responsible for receiving a large number of low-level alarm information generated by intrusion detection agents. These alarm information is collected and sorted out by intrusion detection agents through the information collection and analysis of sensor agents distributed in different network segments, from which we can find out whether there are any violations of security policies and signs of attack in the network. The module provides the original data input to the NSAW system.

3.3.2. Alarm information pre-processing mechanism

In a complete compound attack, each attack step may trigger the intrusion detection system to generate multiple alarm information, some of these alarm information may be false alarm information, some may be repeated alarm information. When such alarm information is collected, these alarm information should be filtered and merged in order to identify the attacker's attack intention and predict the next possible attack.

3.3.3. Compound attack discrimination

When the attacker carries on the compound attack, the alarm information generated by different intrusion detection systems on the network, after preprocessing the alarm information, only retains the necessary alarm information. According to the alarm information sequence, the hidden Markov model of each model to generate the alarm information sequence is calculated based on the pre-defined hidden Markov model of the compound attack scene. And the hidden Markov model of the compound attack scene with the highest probability is selected, which represents the most likely compound attack scene of the attacker.

3.3.4. Attack prediction module

This module is an important part of the prototype early warning system. It is responsible for predicting the action route of the attacker according to the alarm information and the attack intention that the attacker has completed. The action route of the attacker is the core content of the early warning system.

In general, the course of action of an attacker consists of two parts:

1) Judging the possible attack intention of the attacker: judging the possible attack intention of the next step is the next step of the attacker to grasp the attacker's course of action, which is the first step of attack prediction.
2) Judging the possible attack mode of the attacker in the next step: it will predict the specific attack mode to be adopted by the attacker in order to obtain the attack target, that is, the ultimate purpose of the attack prediction.

4. Test of NSAW Reminder Based on BD Technology

4.1. Overview of the Simulation Platform

Table 1. Main steps and procedures of the experiment

Step	Experimental process
The first step	Generate training datasets and model the activities of the data during the offline learning phase to generate templates for each abnormal activity
Step two	The incoming network data flow is preprocessed, and the feature attributes of each data record are extracted by a novel data dimension reduction method, which makes the data conducive to the subsequent analysis.
Step three	Once the data preprocessing is completed, the template is used to match the data to complete the recognition of active features, and each data record is classified according to the attribute features.

Step 4	A novel unsupervised method is used to cluster analysis of data records whose characteristics are Not obvious or do not belong to known types. Store the resulting new exception in the exception library.
Step 5	Once the new abnormal library students achieve online learning. The template library is updated with the new exception, and the weight is assigned to each template, and each weight is updated by error feedback learning. This can effectively reduce the false detection rate

In order to verify the effectiveness of the proposed scheme, a NSAW and reminder simulation platform based on BD technology is designed and built for experimental needs. The Spark provides a simpler programming model than the Hadoop Map Reduce. These features of Spark provide a good, stable, and convenient experimental environment for the experiments in this chapter. There are 4 types of 25 different abnormal activities in the dataset, and the different types of abnormal behavior contain different attribute characteristics, and we establish the corresponding exception behavior template by analyzing these attributes. With the number and variety of abnormal behaviors, there must be the missing types of templates, adding new discovered exceptions to the exception library through unsupervised online learning. The work to be done is to make the model constantly detect the datasets, learn to discover various abnormal activities, accurately identify the specific abnormal types through the matching of the template, improve the detection efficiency and reduce the error detection rate. The experiment is divided into five steps, as shown in Table 1.

4.2. Analysis of Simulation Results

In solving the BD problem, in the face of massive data, the hardware configuration of a single computer can no longer be processed in an effective time. Moreover, the scale of the network and the transmission rate of data are also very fast. In this case, the network situational awareness put forward new requirements, not only to be able to deal with big data, but also to be real-time relatively high. The following experiments are carried out on the open source BD platform. The data is divided into logical data stream in the unit of time slice, and then the data blocks scattered on each node in the cluster are processed, and the final manifestation is to process the data of each time slice in the form of batch processing. Figure 3 shows the false detection rate in the detection.

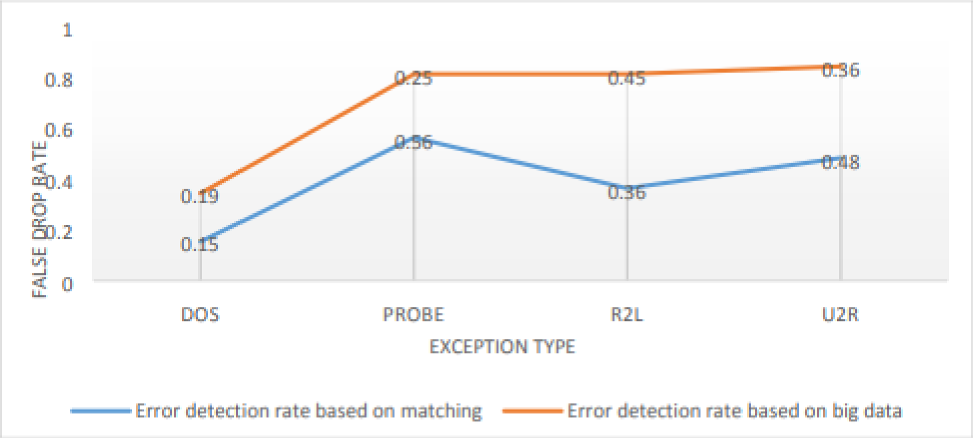

Figure 3. Misdetection rate

Under the condition of different amounts of data, the stability and detection rate of the model are also slightly different. With the continuous increase of data processing capacity, the stability of the traditional methods is very poor, the detection rate of each anomaly type is very unstable, and the volatility is great. Through the optimization of dynamic time warping algorithms and the adaptive adjustment strategy based on weights, the model proposed in this paper has good stability in dealing with large amounts of data. And with the increase of data, the accuracy of training adjustment of each weight is higher, and the detection rate as a whole shows an upward trend.

5. Conclusion

To sum up, the most fundamental challenge for us in the BD era is how to condensed understandable knowledge from the data. There are many challenges and difficulties in the research of NSSA model under the background of big data. Classical algorithms and models are faced with many challenges, such as large amount of data, complexity, high dimension, redundancy, noise and so on. How to parallelize or improve the existing algorithms and models, and then put forward a new model has become the focus of research. How to apply these BD processing technologies to the NSSA model to form a general framework to deal with network big data, and to improve the accuracy, performance, efficiency and accuracy of each stage of NSSA is also the key part of the research.

References

[1] Cui Ping. Design of laboratory security early-warning system based on wireless network monitoring technology. Modern electronics technology, 2019,042 (012): 37-39, 44.

[2] Lu Guosheng, Tian Lin, Chen Junhao. Early warning and quantitative analysis of power network security risk. Modern scientific instruments, 2019,000 (001): 101-103145.

[3] Wang Z. Research on information security early warning and decision support system based on risk control. Boletin Tecnico/Technical Bulletin, 2017, 55 (20): 581-590.

[4] Li C Y, Zheng L. Analysis of Tai Chi Ideological and Political Course in University Based on BDand Graph Neural Networks. Scientific Programming, 2021, 2021 (1): 1-9.

[5] Yi M, Xu X, Xu L. An Intelligent Communication Warning Vulnerability Detection Algorithm Based on IoT Technology. IEEE Access, 2019, 7 (99): 164803-164814.

[6] Yuan T, Zhang Y X, Ma S Y, et al. Combining the BDanalysis and the threat intelligence technologies for the classified protection model. Cluster Computing, 2017, 20 (2): 1-12.

[7] ShengliZhou, XinWang, ZeruiYang. Monitoring and Early Warning of New Cyber-Telecom Crime Platform Based on BERT Migration Learning. China Communications: the English version, 2020 (3): 140-148.

[8] YAN, Yan, YANG, et al. Disaster reduction stick equipment: A method for monitoring and early warning of pipeline-landslide hazards. Journal of Mountain Science, 2019, v.16 (12): 4-17.

[9] Einy S, Oz C, Navaei Y D. The Anomaly- and Signature-Based IDS for Network Security Using Hybrid Inference Systems. Mathematical Problems in Engineering, 2021, 2021 (9): 1- 10.

[10] Xue Y, Zhu L, Wang W, et al. Research on fault analysis and positioning technology of distribution network based on BD. Journal of Physics: Conference Series, 2019, 1176 / 062029.

[11] Huang J C, Ko P C, Fong C M, et al. Statistical Modeling and Simulation of Online Shopping Customer Loyalty Based on Machine Learning and BDAnalysis. Security and Communication Networks, 2021, 2021 (3): 1-12.

[12] Yi L, Niu D, Wang H, et al. Assessment Analysis and Forecasting for Security Early Warning of Energy Consumption Carbon Emissions in Hebei Province, China. Energies, 2017, 10 (3): 391.