Design and Implementation of Network Malicious Traffic Screening and Analysis System Based on Big Data Platform

Xu Gang

College of Information Engineering, Yunnan Technology and Business University, Kunming, Yunnan, China

Keywords: large data platform; network malicious traffic; design of screening and analysis system; distributed storage

Abstract: With the continuous progress of science and technology, the development of Internet technology has also risen. The development of network technology promotes the continuous expansion of network scale and the realm that can be touched by the Internet has also been continuously expanded. However, due to the lack of corresponding management technology, the problems of network security management have become increasingly prominent, and the solution of network security problems is also now. Nowadays, one of the urgent problems in the field of network security needs to be solved. Of course, the emergence of network security problems is not only an external cause, similar to hacker attacks, virus implantation and other issues, there are also certain problems within the network, such as the emergence of security vulnerabilities system, the existence of their own hidden dangers and related problems. Therefore, in the face of network security problems, public vulnerability detection system is limited to external network attacks, the application of this system has been unable to meet the application requirements of modern networks.

1. Introduction

In view of the current situation of the network system, in addition to coping with attacks from outside, the relevant network security management departments should actively find out the problems and shortcomings of the network itself, and relevant detectors should timely feedback and analyze the abnormal situation. At present, the common detection technologies are mainly related to statistical analysis, machine learning methods, neural networks and agents. Although some special detection technologies have obvious characteristics, there are also obvious shortcomings.

2. System objectives

Network Malicious Traffic Screening and Analysis System is a powerful support of large data platform. It collects and classifies all kinds of information on the network platform, uses this intelligent way to understand the normal use of network behavior, and then uses the baseline network behavior and real-time. By analyzing and comparing the network access behavior, the hidden danger of abnormal network attack can be identified. The relevant network security supervision departments can use visual means to warn users who have abnormal network behavior and visually see the emergence of abnormal network behavior and abnormal network path, so as to timely. Processing network security information [1].

3. System Architecture and Processing Flow

3.1. Technical Architecture

The structure of the system is mainly composed of network data collector for information acquisition, then transmission to the distributed real-time data transmission channel, further data analysis in the distributed flow processing platform, and finally network data storage and network information exchange in the large data platform. The network malicious traffic screening and analysis system has three main functions: first, it can use distributed real-time data transmission

  DOI: 10.25236/aisct.2019.024

channels to transmit network data. The specific application of the network data collector is to integrate the data on the network platform, through the distributed real-time data transmission channel, and finally transmit the integrated data packet to the distributed stream processing platform. The second is the real-time processing of network data transmitted by real-time data transmission channel by distributed stream processing platform. The platform parses the transmitted network data and matches the project through the feature library of the large data platform. If the abnormal data packet appears in the matching process, it will be stored in the large data platform. Thirdly, the large data platform carries out clustering analysis and classification training projects for anomalous data packets transmitted by distributed stream processing platform, so that the network data protocol feature library can be updated in time. This is also a complementary process [2].

3.2. System Processing Flow

The network security management platform transits the distributed real-time data transmission channel to the distributed flow processing platform according to the network results of the above process analysis, while the distributed flow processing platform real-time analyses the network data packet, for the abnormal data appearing in the network data packet, matches the characteristics of the feature library, and finally in the big data. The platform carries out data clustering analysis and classification training, and finally updates the network data feature database in real time. The processing flow of the system is mainly divided into the following four aspects: First, network data acquisition. Under the premise of the normal operation of network system, the use of switches transfers the network data information of one or more ports to a fixed port and monitors its network. The use of network data collector can dynamically expand the inherent characteristics of the data itself - the number of data, and increase the throughput of data for the network data collector itself. Secondly, real-time transmission of collected data. All kinds of data collected on the network platform can be transmitted to the distributed stream processing platform in real time, and the data can be transmitted through the unique characteristics of the real-time data channel itself (queuing characteristics and caching characteristics), so as to realize the sequencing of network data acquisition and the scalability of management.

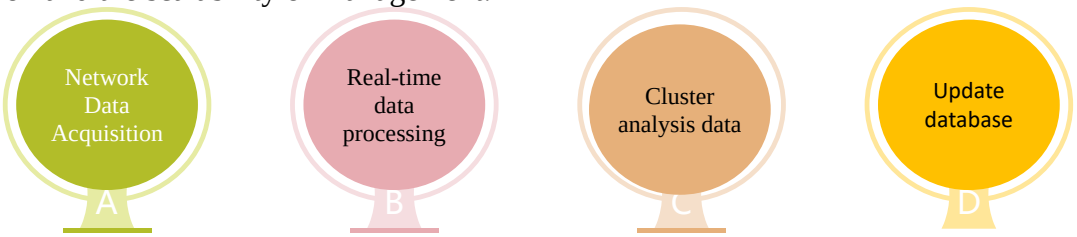

Fig 1 Processing Flow

The third aspect is real-time processing of collected information through distributed process platform. Through data analysis, data feature matching and data access statistics of collected network data. We can analyze it according to the network protocol, so that we can collect the relevant protocols, sending addresses and sending ports, destination addresses and destination ports of network data packets, and then we can carry out feature matching projects and access statistics projects according to the aspects included in the parsed network data packets. The participation of the signature matching project is that the normal network data source and the abnormal network data source are diagnosed by the classification of the network database. If the abnormal network data package is diagnosed, the abnormal data can be stored on the large data platform. The last one is to update the network data of the network data protocol feature library. We should know that the network data protocol feature library includes two main aspects: the normal characteristics and abnormal characteristics of network data traffic. According to the type characteristics of the network data protocol feature library itself, the network detection department can judge whether the network data conforms to the basic characteristics of the network data by examining the sending address, sending port, receiving address and receiving port. The appearance of abnormal features is mainly the abnormal and priority of matching expressions. Different, matching expressions are checked by

logical operation. Of course, another aspect is to classify various network data[3].

4. Systematic Experiments and Results

In order to prove the validity and feasibility of malicious traffic screening and analysis system, this paper designs a distributed security network project. The specific topology will be described in detail after the next meeting. First of all, access is detected, through the intervention of top-level routers, the connection of top-level switches is used, and finally the network data is distributed to different computer rooms. The system only needs to collect network data from the top switch ports, and then decode, merge and sort them. Finally, according to the established process, the tested network data packets are classified and processed to ensure that the sorted network data packets can be stored on the large data platform. The application of the acquisition card is to convert the initial information and data. Network data can be transformed into descriptive information through the acquisition card, which can be directly analyzed and interpreted by the platform software. Although the amount of data is reduced, the amount of information has been completely stored, which is to some extent reduced. The work intensity of network platform data analysis is analyzed. Here, we can analyze the six modules and their functions of the acquisition card.

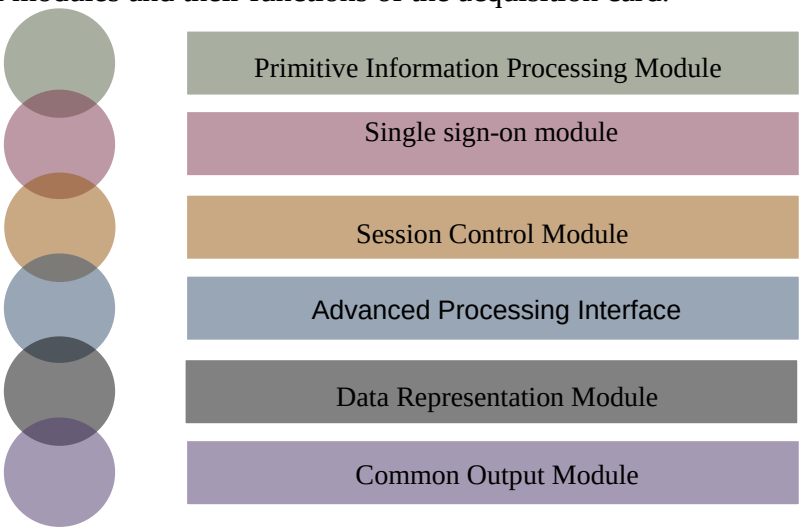

Fig. 2 Six modules of the acquisition card

First, the module of original information processing, which can receive the transmitted message, promotes the emergence of multi-interface concurrency. The second function is the single sign-on module, which promotes the sequence of messages and other related content. The third one is session control module. The fourth one is advanced processing interface module, which can analyze and combine network data information. The fifth one is data representation module, and the last one is public output module[4].

5. Key Technologies Used

5.1. Large Data Storage Technology

This technology is an effective and feasible retrieval method for massive network data information, so the analysis and warning network data of common pages on the network will be put on the big data platform through this technology.

5.2. Real-time streaming event processing technology

This is a large data throughput and fault-tolerant processing system for real-time network data. The system itself can receive a variety of network data sources. The network data source of this system uses kafka, which dissolves a large amount of collected network data information into various small places by shunting and dissolving it. In project management, it can detect information anomaly effectively for every small module.

5.3. Large Data Mining and Analysis Technology

This is based on the use of Spark MLlib technology, through summarizing classification, statistics, frequency analysis and other means to learn the open data ports, summed up the rules of the ports, through the detection and comparison of large data platforms, in connection frequency, connection range, number of traffic and other aspects of timely and effective analysis of abnormal network data. With the emergence and wide spread of information, relevant network detection departments can solve the emergence of malicious traffic from the root.

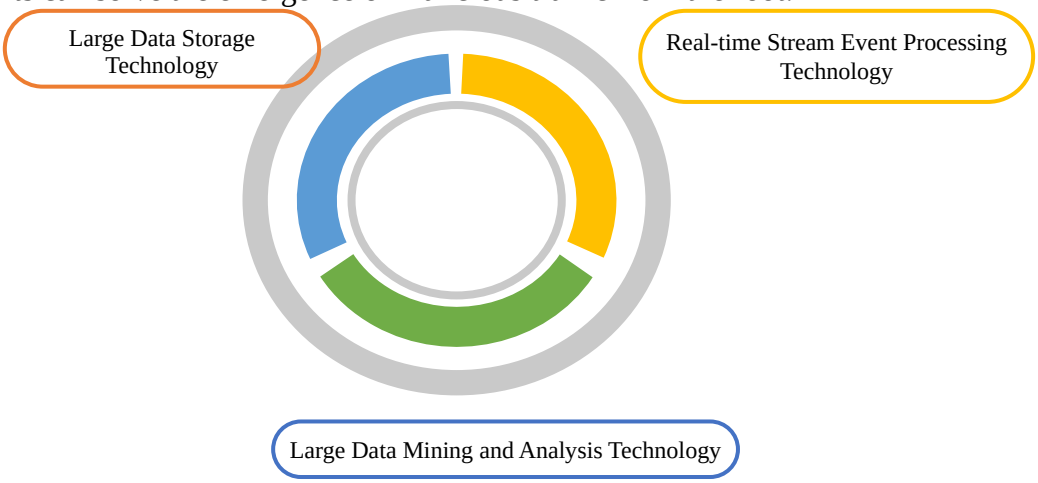

Figure 3 key technologies

6. System Verification

A large enterprise computer room has made a pilot study on the application after system development. The collection service can run steadily and detect the malicious traffic of the server, illegal website access, malicious attack of the external network and so on. It can also retrieve the source of the external network attack in the further monitoring, which is spam.

7. Conclusion

As we all know, the traditional network malicious traffic detection and management methods are inherent and old-fashioned, and can not adapt to the current analysis and detection of abnormal network information data. Therefore, in view of the existing problem of wanton expansion of malicious network traffic under large data, the relevant network security detection departments It should be analyzed and summarized. Only by constantly exploring and updating the analysis technology, can we avoid the wanton invasion of malicious network traffic on various network platforms, ensure the information security of network users, guarantee the security of network environment, and promote the healthy development of the network field.

References

[1] Cheng Weihua, Zhao Jun, Wu Peng. Network traffic detection and analysis based on large data flow [J]. Journal of Nanjing University of Technology, 2017, 41 (3): 294-300.

[2] Yang Qing. Network abnormal traffic detection based on large data analysis [J]. Mechanical design and manufacturing engineering, 2018, 47 (11): 83-86.

[3] Xiang Chaojun, Luo Wangdong, Zhang Hao, et al. Large Data Analysis System for Internet Traffic Flow Based on DNS and Flow Data [J]. Telecom Technology, 2018, 534 (09): 35-40.

[4] Xu Pengfei, Yang Qian. Design of Channel Traffic Flow Acquisition System Based on Large Data Analysis [J]. Ship Science and Technology, 2017 (20): 88-90.