Research on Governance of Bad Information Based on Big Data Technology

Wei Wu

China academy of information and communications, 100037, Beijing, China

Wsidy@126.com

Keywords: Network security, governance, big data.

Abstract: Aiming at the problem of data turbulence existing in the traditional information security filtering system in the big data Web environment, a design of a big data dynamic bad information security filtering system in the Web environment is proposed. Adopting C / S system architecture, the control end of the front-end host is well optimized, which provides a good hardware reserve for subsequent filtering calculations. Through the gate-gate data information filtering system, the traditional filter system avoids the phenomenon of mismatch calculation, and effectively solves the problem of data information shock. The weighted random adaptive algorithm is optimized to ensure that bad data information can be completely filtered in the big data dynamic Web environment. In order to verify the effectiveness of the big data dynamic bad information security filtering system designed in this paper, a comparative simulation test was designed. The experimental data shows that the big data dynamic bad information security filtering system designed in this paper can effectively deal with bad data. Filter information.

1. Introduction

With the development of IT technology and communication technology, computing power and network bandwidth have been rapidly improved. Emerging services such as cloud computing, Internet of Things, and social networks have emerged. Data is growing and accumulating at an unprecedented rate. The era of big data has arrived. Big data technology has the ability to quickly obtain valuable information from huge amounts of data, complex structures, and many types of data. It can reveal content and changes that cannot be seen by traditional means. It is the current concern of academia, industry, and even governments of various countries. hot spot. Big data technology brings new opportunities for the development of the information security industry. Traditional feature-based information security analysis technology has been widely used in malicious code detection, intrusion detection, etc. However, with the increase in data scale and the emergence of some emerging threats, higher requirements have been placed on security analysis and detection technology. The application of big data analysis technology to analyse information security issues has become a hot research topic in the industry. Gartner clearly stated in the 2012 report that "information security is becoming a big data analysis problem." With the help of big data security analysis methods, it can not only solve the collection and storage of massive data, but also respond more actively and flexibly to new and complex violations and unknown and variable risks based on machine learning and data mining methods. BDSA (Security Big Data Analysis) came into being [1].

2. Necessity and importance of using big data to enhance network security defines capabilities

On the one hand, from the perspective of enhancing the effectiveness of cyber ideological security risk response, it is necessary to pay close attention to the risk perception ability of big data. Xi Jinping pointed out at the National Cyber Security and Informatization Work Conference, "To maintain cyber security, we must first know where the risks are, what kind of risks, and when they occur." First, big data can perceive "where" the risk source of network ideology. With the development of the "fifth space" of the Internet, ideological security risks gradually extend from the real-life field to the network virtual field. By analysing multivariate heterogeneous data, big data can locate missing security information and correlate single-point abnormal behaviour from it, thereby finding clues that endanger ideological security and "tracing the roots" of network security threats. Second, big data can identify "what kind of risks" exist in cyberspace. Affected by the rapid iterative information technology, the cyberspace public opinion game is becoming more intense, the hostile thoughts continue to ferment, and the ideological attack methods have rapidly evolved. There are various forms of ideological security risks. By implanting a semantic search engine, big data can identify words related to ideology and scientifically assess the cyberspace risk situation. Based on this, big data can monitor complex variables in the field of ideology in real time, scientifically predict high-risk areas, and prepare emergency plans in advance to prevent the spread of security risks. On the other hand, from the perspective of improving the scientific governance of network ideology security, it is necessary to give full play to the precise identification capabilities of big data. With the increasing popularity of Internet technology, network ideological security governance has become "an important part of socialist ideological work in the new era."

As a new stage of informatization development, big data is an important basis for network ideological security governance. First, big data can grasp the overall situation of network ideology, and provide a reference basis for firmly grasping the right to speak of the security governance of network ideology. The second is that big data can accurately grasp the public opinion and help to formulate targeted governance programs. "The netizens come from ordinary people. When the people go online, the public opinion will go online." Using big data to widely collect and comprehensively analyse large-scale data can more accurately grasp the key points, difficulties and public concerns and expectations of security governance. In order to improve the predictability and scientific of the governance plan, let the advantages of big data information technology benefit the people's livelihood to the maximum, and finally achieve the function of "collecting people's will", "collecting people's wisdom" and "gathering people's hearts"[2].

3. Key technologies for network security processing

3.1 Big data processing technology

The calculation mode of big data can be divided into batch computing and streaming computing. The AMP laboratory at Berkeley University has proposed a software stack for data analysis. From the perspective of big data computing models, big data processing technologies are divided into three types: batch data processing technology, streaming data processing technology, and interactive data query technology. Batch calculation first stores the data, and then performs centralized calculation on the stored static data, as shown in Figure 1. Batch data processing technology achieves high-throughput processing of large-scale static data, and is characterized by its large throughput. Complex batch data processing usually has a time span of tens of minutes to hours. Because batch data processing technology performs extremely well when dealing with large amounts of persistent data, it is often used in historical data analysis such as network full traffic analysis, log analysis, and fraud detection and APT detection in the field of network security.

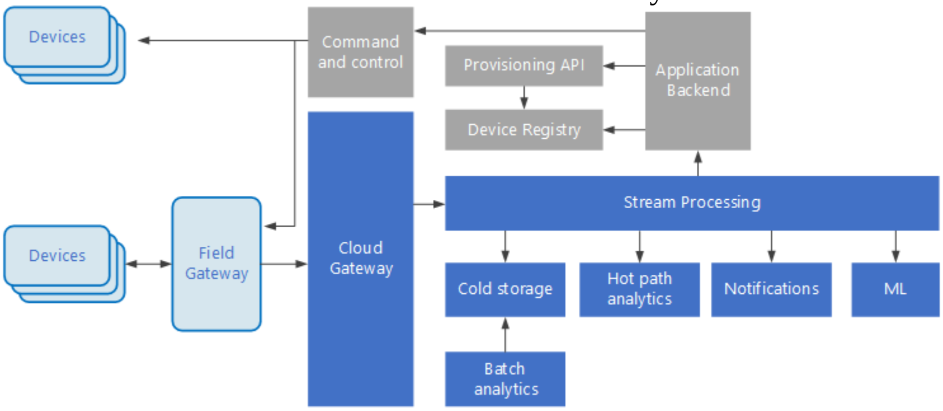

Figure 1. Schematic diagram of batch data processing

3.2 Big data security analysis technology

The general steps of network security visualization: First, determine the concerns of network security analysts, that is, what data is there, and what information needs to be obtained from the data; then, design the visualization structure to represent the data, and establish the data to the visualization structure Mapping; finally, design human-computer interaction functions such as zooming, focusing, playback, association, and updating to realize the communication between people and visual tools. Unlike traditional network security visualization, big data security visualization faces two problems: on the one hand, the scale of network security data, that is, how to propose new visualization methods to help security analysts analyse large-scale, high-latitude, multi-source, Dynamically evolve cybersecurity data and make decisions in real time. On the other hand, creating a visual representation of big data that conforms to the psychological image of network security analysts can enable security analysts to discover the security issues implicit in big data at a glance. In recent years, some scholars have begun to gradually and deeply discuss how to quickly process large-scale flow time series data and how to visualize the flow changes of large-scale network monitoring objects [3].

4. Design of network bad information processing platform under big data technology

4.1 Overall structure

In this paper, the design of the big data dynamic bad information security filtering system under the Web environment adopts the C / S system model, which abandons the traditional N / S system mode, which is convenient for system maintenance and system upgrade. The filter system is faced with a large amount of information data, so the C / S mode is more suitable. The C / S mode hardware system is mainly composed of three levels, as shown in Figure 2. The first layer is the front-end control layer, which is the command control centre of the big data dynamic bad information security filtering system in the Web environment; the second layer is the operating system, which includes calculators, data regulators, databases, and data sorters. The execution system is mainly to analyse and filter the data in the network; the third layer is the user end, including CVDO, file driver, perception runner, etc. Mainly identify users and order and transfer commands [4].

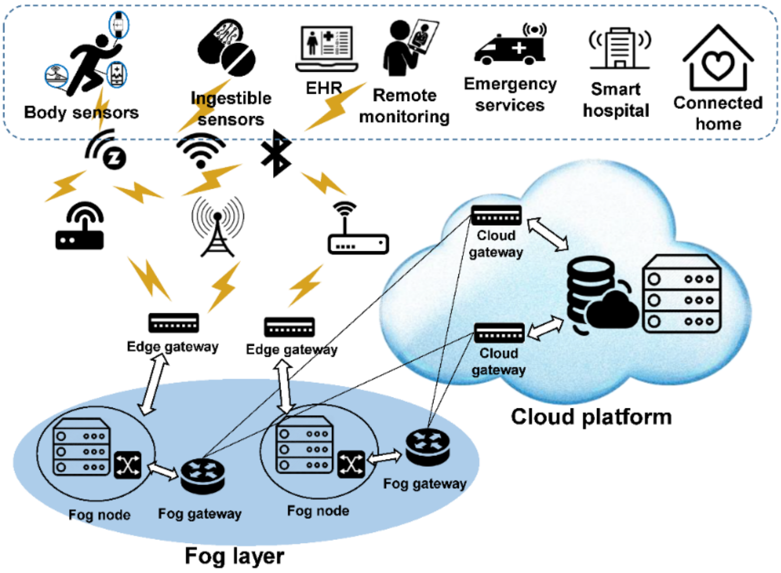

Figure 2. Network bad information processing platform

In order to filter data information more accurately, the big data dynamic bad information security filtering system designed in the WEB environment is optimized for the front-end host control terminal. The optimization of the front-end host control terminal greatly improves the filtering capacity of the system. Improve the system's logical computing power, thereby improving the system's screening ability. To filter massive data, it is necessary to ensure that the conventional data can be run freely, and the information must be controlled while filtering, which is a great test for the logical computing ability of the system. Therefore, the optimal selection of the front-end host control end improves the system's logical computing power. The data filtering process of the big data dynamic bad information security filtering system under the WEB environment designed in this paper is shown in Figure 3.

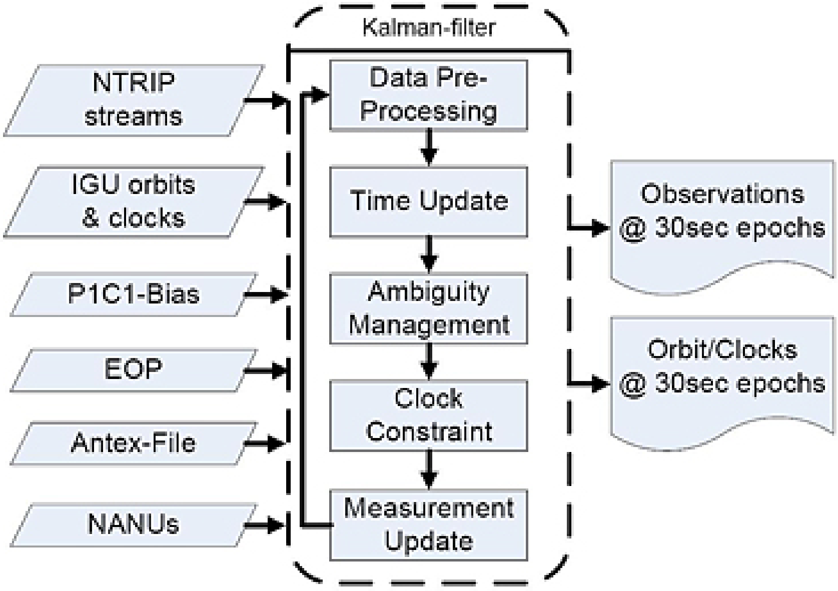

Figure 3. Data filtering process

4.2 Analysis of data mining algorithms

The big data dynamic bad information security filtering system designed under the Web environment in this article uses the gatekeeper filtering system to effectively filter bad data. The problem of unbalanced calculation of mismatch in traditional methods can be avoided, and the computing power of the system can be improved. The gatekeeper filtering system will change the symbolic attributes for different data information. The selection of different symbolic attributes of the gatekeeper filtering system is shown in Table 1.

Table 1. Symbolic attributes corresponding to different data information

Data information	Numeric attribute	Symbolic attribute
1	2	X
2	5	Z
3	9	Y
4	4	Z
5	7	X
6	6	Y

The gatekeeper filtering system effectively classifies different symbolic attributes, so that more accurate filtering can be performed on different data. Assume that F (u, v) is the filtering threshold of the filtering system, f (x, y) is the discrimination attribute covered by the data in the corresponding WEB network, and uπ is the reference ratio of the information extraction coefficient, so that the symbol attribute The establishment of the equation is:

$$F\left(u,v\right)=\frac{c\left(u\right)c\left(v\right)}{4}\sum_{x=0}^{n}\sum_{y=0}^{n}f\left(x,y\right)+\boldsymbol{cos}\frac{\left(2x+1\right)u\pi}{16}\boldsymbol{cos}\frac{\left(2y+1\right)v\pi}{16} \tag{1}$$

Through the above calculation, the information data in the Web network can be effectively distinguished, which can reduce the workload of a part of the filtering system and improve the accuracy of the filtering system. After primary filtering, it enters the condition matrix for matrix filtering.

5. Big data analysis based on mining and association

The goal of early statistical analysis is to convert mixed big data into small data that can be used for subsequent safety analysis. In this analysis process, it is an important task to form and establish a series of IP-based black and white lists as soon as possible. Through the whitelist mechanism, you can actively adjust the early network traffic, selectively discard a lot of data that is meaningless for security analysis, and reduce the pressure on later storage and analysis processing. Through the blacklist mechanism, you can focus on areas of interest, track their changes and development trends, and conduct more elaborate and in-depth security monitoring. For sudden major security vulnerabilities, big data analysis can quickly give security warnings and large-scale security assessments. In recent years, waves of ApacheStruts2 vulnerabilities have had a profound impact, and there are dozens of application information systems based on this Java framework in the campus network. How to quickly find out the Struts2 development framework from thousands of websites, query some feature fields (. action and .do, etc.) of the URL in the metadata through HTTP query through HIVE and analyse it with Shell script, which can be analysed within half an hour Get an accurate answer. Using Python scripts combined with the corresponding series of vulnerability POCs, you can quickly scan these websites and determine how many are affected by the returned results [5].

6. Conclusion

Big data technology has gradually penetrated into the processing and solutions of many cyberspace security problems, changed the research pattern of cyberspace security, and improved the level of cybersecurity defines technologies such as advanced cyberattack detection and information security risk perception analysis and processing. However, under the new situation of cyberspace security, in the fierce confrontation of offense and defines, we still need to continue to use big data technology, comprehensively use multi-source data, explore the containment of complex cyber-attacks, perceive network information security risk perception, early warning and disposal, Research and judgment of new technologies to improve the ability of big data to support network security.

References

[1] Liu, H., Morstatter, F., Tang, J., & Zafarani, R. The good, the bad, and the ugly: uncovering novel research opportunities in social media mining. International Journal of Data ence and Analytics, 1(4) (2016) 137-143.

[2] Hernando, A. Bobadilla, J. Ortega, F. & A. Gutiérrez. Method to interactively visualize and navigate related information. Expert Systems with Application, 111(11) (2018) 61-75.

[3] Chin, W. L. Li, W. & Chen, H. H. Energy big data security threats in iot-based smart grid communications. IEEE Communications Magazine, 55(10) (2017) 70-75.

[4] Murthy, S. C. & Blackstone, E. H. Research based on big data: the good, the bad, and the ugly. The Journal of thoracic and cardiovascular surgery, 151(3) (2015) 629-630.

[5] Udegbe, E. Morgan, E., & Srinivasan, S. Big data analytics for seismic fracture identification using amplitude-based statistics. Computational Geoences, 23(6) (2019) 1277-1291.