Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080607
ISSN 2371-8838 Vol. 8 Num. 6

Design and Implementation of a Computer Network Log Analysis System Based on Big Data Analytics

Yiru Zhang[1,a,*]

[1]*Cornell Tech, 2 W Loop Rd, New York, NY 10044, USA*
[a]*yiruz315@163.com*
[*]*Corresponding author*

Keywords: Big Data, Log Collection, Mapreduce, Log Analysis, P Statistics

Abstract: This article meticulously crafts and deploys a comprehensive computer network log analysis system, leveraging advanced big data analytics technologies. Following a rigorous feasibility assessment, the system's architecture was meticulously designed, encompassing robust hardware infrastructure and software components tailored for high-performance processing. The operating environment was optimized to handle massive log data, ensuring scalability and efficiency. The core lies in the five interlocking modules: user authentication ensures secure access; log collection employs distributed techniques for seamless data aggregation; association rule mining uncovers hidden patterns and anomalies through advanced algorithms; security auditing validates log integrity and identifies potential threats; while database management ensures data storage is optimized for both speed and capacity. The system's rigorous functional testing validates its ability to maintain log data integrity, uncover intricate relationships, and bolster log analysis's authenticity and reliability. This achievement not only meets the predefined objectives but also sets a benchmark for future research endeavors in the realm of network log analysis.

1. Introduction

This paper presents the design and implementation of a sophisticated computer network log analysis system, leveraging cutting-edge big data analysis technologies to meet contemporary cybersecurity demands [1]. The feasibility of this system is thoroughly evaluated from both technical and economic perspectives, ensuring its practical applicability and efficiency in real-world environments.

From a technical point of view, the system uses SQL Server 2020 as its database configuration. Java as the programming language and Microsoft Signal Studio 2020 as the studio [2]. These tools were chosen for their reliability, versatility, and proven reliability for managing large amounts of data and complex computers. SQL Server 2020 provides powerful support for the right data management of the right data for the system and querying the right data sets. The Java network has a wide range of databases that support the development of complex algorithms to analyze online logs and detect anomalies and random security events. Microsoft Visual Studio 2020 further enhances the development lifecycle with its advanced debugging, code analysis, and testing functionalities [3], ensuring that the system not only functions as intended but also maintains high

reliability and performance under various conditions.

The system architecture is designed in a way that supports efficient big data processing, and big data and allows for in-depth analysis and interpretation of log data. Through machine algorithms and sophisticated data analysis techniques, the system can detect hidden patterns, identify potential threats, and provide practical insights for cybersecurity management [4]. We all agree that flexibility is critical in our modern, fast-paced online security environment, as is the ability to process and analyze large amounts of data in real-time to counter new threats. Using existing hardware and software, the system avoids major new investments[5][6]. Cloud-based solutions help save costs, reduce the need for expensive on-premises equipment, and facilitate efficient expansion of the economy as data demand grows. Advanced algorithms reduce the need for human intervention[7], which helps reduce operating costs and optimize productivity.

The computer network analysis system proposed in this paper provides a stable and practical solution, the system combines advanced data processing technology and strategic economics, is good at using advanced analytical methods to process and interpret Web logs, provides valuable insights for network performance and security[8], and its design improves accuracy and efficiency, but also ensures cost-effectiveness. These complex technologies combined with practical economic strategies make the system work well in handling complex network data analysis and supporting informed decision-making in network management.

2. Main Module Design of the System

The system architecture shown in Figure 1 includes several elements to enhance network security and data management. The Protocol module cargo is responsible for collecting and processing protocol types from the system. A uniform format for data, stored in a protected protocol database, ensures the integrity and availability of information. The data analysis module uses the most advanced search methods to view the collected data[9], detect possible anomalies, and extract the stored data. This information is then stored in a protected database, providing the necessary support for future security control and decision-making processes.

The certification service module plays an important role in validating professional data at a specified time according to predefined rules. The goal is to instantly detect anomalies or potential security threats and trigger instant alerts; When these problems arise. The database management module supports daily search and maintenance while providing real-time updates. This keeps the data up to date and meets changing system requirements.

Through the unobstructed integration of all its modules, the system has efficient capture, allowing analysis and management of log data. This holistic approach improves network security and integrated data management[10]. The seamless integration of all components is a harmonious collaboration between all components that enables the system to process log data efficiently and respond quickly to emerging threats. With this comprehensive integration, systems can be designed more efficiently and meet changing security requirements. Build a strong and adaptable safety net.

When building the hardware infrastructure of the system, technicians need to carefully adjust the hardware specifications to meet and guarantee the performance requirements, it works best. The application should be backed up to at least 6.0 GHz securely, ensuring more than 50 GB of storage capacity and at least 1.5 TB of storage capacity, and it can effectively support concurrent services for all four applications. The same amount of data and storage capacity requires a data server, but the storage capacity should be limited to at least 10terb for processing big data. For the customer group, it is recommended to use Intel 2022 processors, and install 10 GB of memory and at least 440 GB of hard drive to ensure smooth and efficient. This hardware is important for creating a stable and reliable foundation for system functionality.

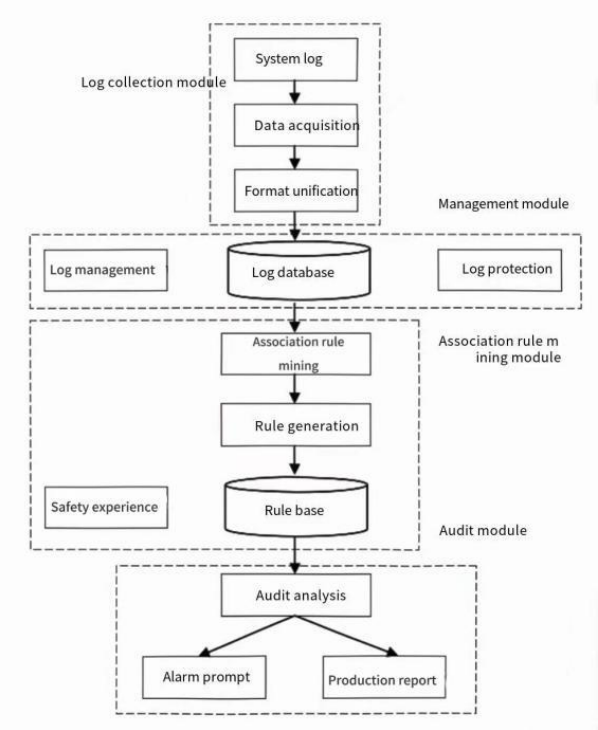

Figure 1: System Architecture Design Diagram

Optimizing the software environment is critical to successful development and continuous improvement of system functionality. This includes development tools, databases, and programming systems, and ensures compatibility with system architectures and operating systems. The optimized software environment not only supports the effective development and implementation of the system but also contributes to its further development. By seamlessly integrating these components, the system will benefit from improved performance and reliability and ultimately contribute to the creation of more efficient and adaptable operating framework conditions.

To take full advantage of Crowday technology to ensure the reliability of system functions while meeting user delivery requirements; The technical personnel must strictly adhere to the framework described in paragraph 2. It provides a clear schematic of the structure, supporting the concept and development of the system. With dual systems, teams can refine individual components for optimal performance and ensure seamless integration with the rest of the system. This careful design not only ensures the effective application of big data but also improves the overall efficiency and reliability of the system. The system, as a result, is not only efficient and effective, but at the same time, it also focuses on user-friendliness. Provide reliable functionality to meet user needs and expectations (as shown in Figure 2).

When designing user access modules, strong encryption mechanisms and correct termination procedures must be in place to avoid potential security vulnerabilities and unauthorized access to data. The module must ensure that the user data you enter through the user interface (i.e., username and password) is strictly encrypted. After entering the receipt, it is encrypted using the most advanced encryption algorithm to enhance data protection. The encrypted data is then transmitted to the server, where it is decrypted and confirmed.

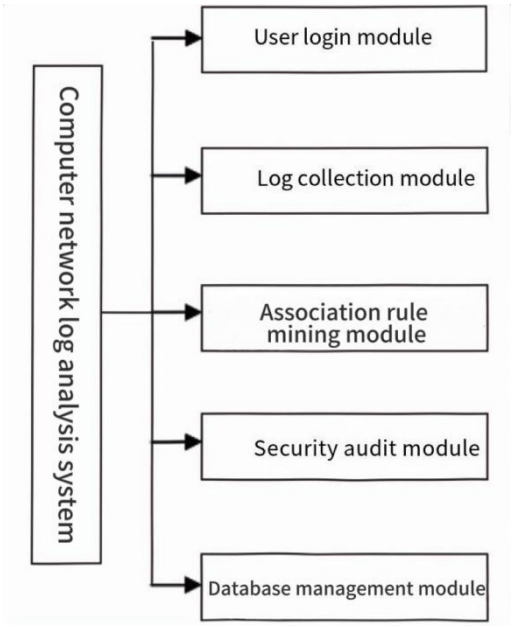

Figure 2: Schematic diagram of system functional module design

The server is used to decrypt the data so that it can be compared with the data stored in the database to verify its authenticity. When a problem is detected, the system becomes available; Without checking the certificate, the user will be prompted to check the information again and open the opportunity for correction and update. The goal of this process is to reduce the risk of unauthorized access, protect the system from potential data breaches, and maintain the integrity and confidentiality of user information.

Modules play a central role in practical applications, especially in Windows operating systems, when it comes to comprehensively capturing and organizing records in computer systems. The module is precisely designed to adhere to predefined audit policies and uses the databases needed to access and search data in the Win32 API. The logging process involves several important steps: First, you need to determine the type of logging. Ensure correct handling of records; Third, check the integrity and integrity of records; Finally, save all log data, which has been set up as an authentication system allowing the module to read, check, and ensure the protocol type in real time; Store identification data in secure storage. By maintaining a reliable ledger and application API, the module ensures the integrity and accuracy of log data, which is essential for effective monitoring and subsequent analysis of the system.

Log data collected from filtering and processing is first saved to ensure relevance and accuracy. Well-managed data is systematically stored in a database where each record precisely corresponds to a specific user log event for tracking and analysis. By using full index and database access, the system improves performance and efficiency and allows fast access to log information and requests. This process not only optimizes data processing, it also ensures that the protocol format is consistent. Standardization of format is essential for effective inspection and coordination of data, as it is the basis for consistent and reliable analysis. Figure 3 provides a detailed overview of the prediction and formatting steps, as well as a comprehensive overview of the module's operational flow and its role in maintaining system quality and data integrity.

In the association rule mining module, in the rule base, the sources of rules mainly include two methods: (1) automatic generation method. This method automatically creates rules through big data analysis technology, allowing system administrators to add and modify these rules. (2) Manual input method. This approach establishes and adds new rules by combining past security experience

and verifying the required knowledge. When auditing violations, if existing rules cannot cover all scenarios, users can develop custom rules to better support subsequent audit work.

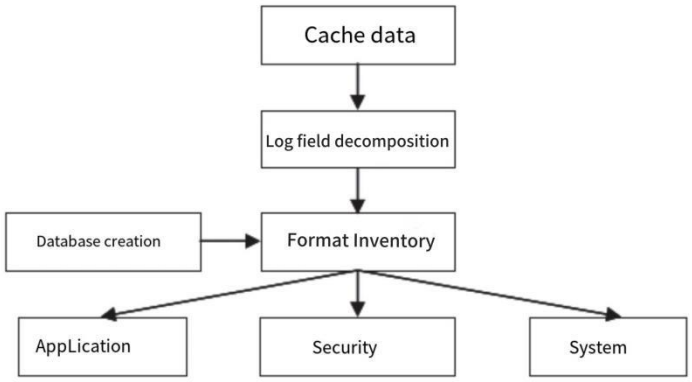

Figure 3: Unified flowchart for preprocessing and formatting

The security audit module is pivotal for the detailed examination of database rules and the matching of original data, employing sophisticated techniques to enhance network security. Initially, this module performs an exhaustive analysis of the rules within the database. It consolidates similar rules and identifies ambiguous data, which facilitates the creation of precise "event" descriptions for further investigation. A key function of this module is to detect intrusion attempts. Upon identifying any unauthorized access, the system triggers an automatic alert, ensuring prompt action.

The security audit module encompasses several critical sub-modules:

1) Real-Time Monitoring Sub-Module: This component is designed for high-performance real-time monitoring, offering intelligent oversight of both user-defined content and critical security events. It operates with minimal latency, promptly issuing alerts if any security issues are detected. This ensures that administrators are immediately notified, enabling rapid response to potential threats.

2) Security audit and warning submodule: The module uses state-of-the-art large file analysis to extract and evaluate hidden rules from log files. By addressing these rules, it can predict and mitigate potential security vulnerabilities, thereby reducing the likelihood of computer viruses and other security threats. The future viability of the modules is critical to maintaining strong security protection.

To improve the accuracy of test results, the module provides the function of an inspection box. This feature allows the collection and analysis of log data over time; Compares them to the established rules of the mechanism, flags violations in the system, and uses big data to create detailed reports. This comprehensive approach not only ensures that security testing is done correctly, it also improves the system's ability to detect and respond to new threats and increases the security and flexibility of the network.

System Management module: This module is used for core functions such as startup and access. Its main task is to ensure the validity and reliability of the database. During the operation of the system, the module supports the continuous updating and maintenance of the database, meeting the changing requirements, and improving the functionality. This module supports the integration of new rules into the host database in terms of system organization and data structure, with predefined properties. By carefully updating and storing new rules, the module ensures that the rules are applied promptly, which is essential for the accuracy and effectiveness of the database.

By reducing the possibility of data breaches and failures, the module significantly improves the overall reliability and efficiency of the system and provides a solid guarantee for its effective functioning and decision-making.

3. Experimental Results and Analysis

All functional tests show that the system does a good job of collecting and analyzing log data in depth. These two functions are essential to improve the accuracy of system log checking. During the entire operation, the system exhibits higher stability and reliability, in line with the expectations of the building.

Testing of Strong Association Rule Functionality: The system's strong association rule module has proven effective in recording and organizing extensive security events from the computer's event viewer. During testing, the module processed a substantial dataset consisting of 33,083 data entries, utilizing approximately 1,024 KB of memory. Of these entries, 15,546 were identified as valid user data, reflecting the module's ability to accurately segregate relevant information. Furthermore, the system successfully integrated 120 strong association rules into the database. These rules play a pivotal role in the detailed analysis of complex log data, aiding in the detection of intricate patterns and potential threats within the network environment.

The practicality of the one-click detection function has been verified: this function can comprehensively and quickly analyze security event logs within the time range specified by the user. During the test, a total of 1984 security logs were managed by the system, and 8 potential security risk records were successfully identified by detailed comparison with existing rules in the rule base. Detected anomalies are immediately displayed in a specific list, enabling security personnel to quickly view and respond. This function adds value to the system and provides users with a simple and effective way to deal with potential security threats.

Integrated algorithms with disabled interfaces and background boards provide excellent analysis capabilities. Thanks to sophisticated algorithms, powerful connections can detect patterns and patterns, embedded in large amounts of data, and it is interested in identifying subtle and complex security threats. With the help of big data technology and machine learning models, the processing and analysis of massive data can be more efficient and accurate. The buttons used for identification reflect the system's focus on the user's design. It provides automated and simplified application of security protocol standards that not only improve operational efficiency but also incorporate cybersecurity best practices and ensure timely identification and resolution of potential threats. With the help of rapid response and in-depth analysis, the system can help create a safer and more robust online environment to meet the multiple dynamic challenges posed by current cybersecurity threats.

These tests demonstrate the robustness of the system and its applicability to the practical and scientific needs of the field of cyber topography to ensure that it meets the requirements of a complex and ever-changing cybersecurity environment.

4. Conclusion

With the support of big data analysis technology, the computer network log analysis system designed in this article not only achieves multiple important functions but also significantly improves the system's automation and intelligent processing capabilities. The system can continuously monitor the security logs of computer systems and deeply explore and analyze hidden association rules through quantitative association algorithms. This process not only provides accurate data support for subsequent rule analysis but also effectively strengthens the system's warning function and overall security. Secondly, after each operation is completed, the system will automatically generate relevant records and store these record results in a specific log repository through data conversion technology, ensuring the integrity and traceability of the data. Finally, the system extracts logs and corresponding rules from the log library and rule library one by one, analyzes and compares them, quickly identifies abnormal logs, and visually displays this

information to users for timely action.

Although the system has relatively improved in terms of functional implementation, there is still room for further improvement. At present, the accuracy of the system has not been fully validated, which may affect its reliability in practical applications. In addition, existing algorithms still need to be optimized in terms of processing efficiency and complexity to better meet the real-time analysis requirements of massive logs in big data environments. Future research directions should include improving algorithms and conducting comprehensive testing of system performance to ensure that the system can fully utilize its effectiveness in a wider range of application scenarios.

References

[1] Andalib A, Babamir S M. Anomaly detection of policies in distributed firewalls using data log analysis. Journal of supercomputing, 2023, 79 (17): 19473-19514.

[2] Muratov S Y, Muravyov S B. Framework architecture of a secure big data lake. Procedia Computer Science, 2023, 229: 39-46. DOI: 10. 1016/j. procs. 2023. 12. 005.

[3] Zouhri A, Ezzahout A, Chakouk S, et al. A Numerical Analysis Based Internet of Things (IoT) and Big Data Analytics to Minimize Energy Consumption in Smart Buildings[J].Journal of Automation, Mobile Robotics and Intelligent Systems, 2024, 18(2):46-56. DOI:10.14313/jamris/2-2024/12.

[4] Qiu X, Fan P, Xie B .The Application of Software Engineering Technology in the Era of Big Data[J].Journal of Electronics and Information Science, 2023.DOI:10.23977/jeis.2023.080205.

[5] Wang M, Lu S, Xiao S,et al.An Unsupervised Gradient-Based Approach for Real-Time Log Analysis From Distributed Systems[J].International Journal of Cooperative Information Systems, 2024, 33(02).DOI:10. 1142/ S0218843023500181.

[6] Tim Fürmann, Lechl M, Meer H D, et al. From computer systems to power systems: using stochastic network calculus for flexibility analysis in power systems[J].Energy Informatics, 2023, 6(Suppl 1).DOI:10.1186/ s42162-023 -00286-z.

[7] Tian X, Zhiyuan W U, Cao J,et al.ILIDViz:An incremental learning-based visual analysis system for network anomaly detection[J].Virtual Reality & Intelligent Hardware, 2023, 5(6):471-489.

[8] Rakib M H, Hossain S, Jahan M, et al. A Blockchain-Enabled Scalable Network Log Management System[J]. Journal of computer sciences, 2022.

[9] Yang C T, Chan Y W, Liu J C,et al. Cyberattacks detection and analysis in a network log system using XGBoost with ELK stack[J].Soft Computing, 2022, 26(11):5143-5157. DOI:10.1007/s00500-022-06954-8.

[10] Quan X .A Neural Network Classifier Intrusion Detection Vulnerability System Based on HFE-CAMMLP Intrusion Detection Model and Weblog Analysis[J].2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), 2022:718-721.DOI:10.1109/ICFTIC57696.2022.10075290.