International Journal of Information Security and Privacy
Volume 18 • Issue 1 • January-December 2024

Dynamic Adaptive Mechanism Design and Implementation in VSS for Large-Scale Unified Log Data Collection

Zhijie Fan
Department of R&D Center, Shanghai Chenrui Information Technology Company, Shanghai, China & School of Computer Science, Fudan University, Shanghai, China & Department of Information Security Technology, The Third Research Institute of the Ministry of Public Security, Shanghai, China

Bo Yang
Computer Network Information Center, Chinese Academy of Sciences, Beijing, China

Jing Peng
Sichuan Provincial Public Security Department, Chengdu, China

Bingsen Pei
School of Information Technology and Network Security, People's Public Security University of China, Beijing, China

Changsong Zheng
Sichuan Provincial Public Security Department, Chengdu, China & School of Computer Science and Engineering (School of Cybersecurity), University of Electronic Science and Technology of China, Chengdu, China

Xin Li
School of Information Technology and Network Security, People's Public Security University of China, Beijing, China

ABSTRACT

This paper studies the collection of large-scale log data of information system and puts forward a dynamic adaptive mechanism for large-scale unified log data collection. Furthermore, we design and implement our method for pan-government industry safety operation management platform. The data flow processing architecture based on message queue is adopted to realize the decoupling of log collection, log processing and log reporting. The traffic peak clipping technology of message queue is adopted to ensure the safety and reliability of log transmission. According to the characteristics of log traffic, a design mode supporting dynamic adjustment of consumption group is proposed to meet the high-performance requirements of the system. The whole system can meet the centralized analysis, security threat perception and intelligent analysis of various security data. Meanwhile, we analyzed and compared with the traditional open-source log collection technology, our proposed method and system has clear advantages.

KEYWORDS

Safe Operation Management Platform, Standardized Interface, Plug-In Technology, Message Queue, Streaming Architecture, Video Surveillance System (VSS)

INTRODUCTION

Rapid advances in cybertechnology have led to increased information warfare, with theft, breaches, and attacks on worldwide networks occurring one after another. Network security is a global issue that is intimately tied to, and as pressing as, the security of the world's political, economic, and cultural institutions. Breakdown, network paralysis, virus outbreak, data loss, and leakage are just a few of the severe consequences of a problem with a big data processing system. When the network's security encounters the aforementioned issues, major losses to the integrity of the national information

DOI: 10.4018/IJISP.349569

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

system, in particular for the pan-government industry, result. To protect big data, a data-centric security management system, a security operations and management platform, a professional security technology team, and a comprehensive security strategy are all required.

In a big data security system, the security operation management platform collects, analyzes, and commands security information. The platform system is capable of collecting huge amounts of data in real time, developing a data resolution model that is based on security strategy, centrally evaluating, recognizing, and issuing intelligent alerts and dynamic predictions. At the same time, by establishing the incident-handling procedure and working in conjunction with the coordinated handling mechanism, it is able to deal with unanticipated security issues swiftly (Fan et al., 2021; Liu et al., 2012; Zhang et al., 2016).

Because different systems use different information formats and accept a variety of transmission protocols, technical means must be sought to integrate, gather, and preprocess diverse heterogeneous data sources and integrate them into the safety operation management platform (Huang et al., 2021; Luo et al., 2018; Wang et al., 2019).

RELATED WORKS

Scholars have conducted a considerable amount of research on log collection, message caching middleware, data analysis, and so on. Log collection components include Logstash, Flume, and more. Logstash was developed by Jordan Sissel, an operations engineer at DreamHost, a web hosting service provider, in 2009, with a minimalist design and a free way of using it (Gu, 2020). Flume, a log harvesting component invented by Cloudera and donated to the Apache Foundation in 2009, is capable of harvesting, aggregating, and transmitting massive log messages (BalaAnand et al., 2019). The most popular message caching middlewares are ActiveMQ, RabbitMQ, and Kafka. ActiveMQ is a Java virtual machine–based messaging middleware introduced by the Apache Foundation; RabbitMQ is a message broker middleware that implements advanced message queuing that was introduced by Rabbit, an American microprocessor company; and Kafka originated from LinkedIn's data middleware, which was open-sourced in 2011 after refactoring its code (Georgios et al., 2018). Spark is a mainstream framework in batch data analytics applications that was developed by the AMPLab at the University of California, Berkeley, and is a batch computing engine developed specifically for large-scale data analytics scenarios (Al-Bana et al., 2022).

Log Data Collection and Transmission Technology

Two mainstream log collection and transmission technologies are Logstash and Flume. Logstash was developed before Flume was. Logstash works between the data source and storage and analysis tools and is capable of real-time processing of pipeline data (Guo et al., 2018). Logstash's design style is simple and standardized, and its data processing process can be divided into varying sources of data input. Flume was designed and developed by Cloudera, a distributed, highly stable massive log collection, aggregation, and transmission framework (Wang et al., 2021). Flume is deployed in the form of Agent, which is divided into Source, Channel and Sink. The function of the Source component is to collect log content in the form of Events and send it to the pipeline; the purpose of the Channel component is to solve the problem of a work rate mismatch between Source and Sink; and the Sink component scans the Event in the pipeline to dock the downstream data cache or data analysis (Zhao et al., 2020).

Log Data Cache Message Middleware Technology

In the history of log caching messaging middleware development, Kafka has dominated the messaging middleware. Kafka is a message queue based on a publish/subscribe pattern, because messaging middleware plays the role of asynchronous communication and traffic peak shaving (Mountasser et al., 2020). Around 2010, LinkedIn used ActiveMQ for data caching, but at that

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/dynamic-adaptive-mechanism-design-and-implementation-in-vss-for-large-scale-unified-log-data-collection/349569

Related Content

Information Technology Security Concerns in Global Financial Services Institutions: Do Socio-Economic Factors Differentiate Perceptions?
Princely Ifinedo (2009). *International Journal of Information Security and Privacy (pp. 68-83).*
www.irma-international.org/article/information-technology-security-concerns-global/34059

Aspect-Oriented Analysis of Security in Distributed Virtual Environment
Li Yang, Raimund K. Egeand Lin Luo (2009). *Handbook of Research on Information Security and Assurance (pp. 218-229).*
www.irma-international.org/chapter/aspect-oriented-analysis-security-distributed/20652

Ethical Boundary Between Network Security and Personal Privacy in the Era of Big Data
Tieguang Xu (2025). *International Journal of Information Security and Privacy (pp. 1-18).*
www.irma-international.org/article/ethical-boundary-between-network-security-and-personal-privacy-in-the-era-of-big-data/384489

Consumer Privacy Regulations: Considerations in the Age of Globalization and Big Data
Martha Davis (2021). *Research Anthology on Privatizing and Securing Data (pp. 1844-1860).*
www.irma-international.org/chapter/consumer-privacy-regulations/280259

An IIoT Temporal Data Anomaly Detection Method Combining Transformer and Adversarial Training
Yuan Tian, Wendong Wangand Jingyuan He (2024). *International Journal of Information Security and Privacy (pp. 1-28).*
www.irma-international.org/article/an-iiot-temporal-data-anomaly-detection-method-combining-transformer-and-adversarial-training/343306