Exploration and Practice of Big Data Analysis for Broadcasting Network Security Logs (Postprint)

Authors: Zhu Yizhong

Date: 2023-10-08T00:00:00+00:00

Abstract

To address increasingly sophisticated network attacks, analyzing broadcast network and device logs combined with cloud-provided threat intelligence can effectively detect Advanced Persistent Threat (APT) activities within the network, rapidly identify compromised hosts, and generate timely alerts. This approach enables multi-dimensional, rapid, and automated correlation analysis of massive datasets for threat discovery and situational awareness of anomalous behaviors, while providing visualization aids and evidence retention for rapid analysis, characterization, and response.

Full Text

Abstract

To address increasingly sophisticated cyber attacks, this paper explores the analysis of broadcasting network and equipment logs combined with cloud-based threat intelligence. This approach effectively identifies Advanced Persistent Threat (APT) attacks within networks, rapidly detects compromised hosts, and generates timely alerts. The system enables multi-dimensional, rapid, and automated correlation analysis of massive datasets for threat detection and situational awareness of anomalous behavior, providing visual aids and evidence preservation for rapid analysis, characterization, and response.

Keywords: cybersecurity; logs; big data; APT attacks

Introduction

Traditional monitoring of broadcasting networks has achieved considerable maturity in conventional DVB services and is steadily improving in new media applications. However, as network scale and complexity continue to grow, data

traffic has surged dramatically, intensifying the cybersecurity arms race. Organizations now face emerging security challenges: complex network environments leave security teams without clear visibility, allowing attackers to access sensitive data areas undetected, rendering expensive security defenses potentially ineffective. Conventional security technologies prove inadequate against APTs, exhibiting severe limitations in threat detection, discovery, response, and traceability.

To address these challenges, organizations urgently require new technological approaches to maintain comprehensive security situational awareness and optimize security operations, thereby controlling cybersecurity risks within acceptable ranges. Following the enactment of China' s Cybersecurity Law, which mandates log retention, organizations now have a critical opportunity to enhance their security capabilities by supervising log preservation and collection across locations and conducting unified analysis.

Logs—textual records of every operational state in our networks and systems—can be understood as digital footprints of activities in the virtual world. Log types include system logs, application logs, operational behavior logs, and database logs, each containing timestamps, device identifiers, users, and action descriptions. Centralized log collection and analysis enable: (1) monitoring of overall network operations, problem diagnosis, root cause analysis, and error correction; (2) rapid assessment of end-user service quality across applications; and (3) identification of security risks and intrusion attacks for timely intervention and threat mitigation.

System Architecture

The cybersecurity log big data analysis system is a specialized platform for security risk and intrusion analysis based on extensive network and system logs. The system comprises four primary modules: data collection, correlation analysis, situational awareness, and visualization.

The **data collection module** normalizes logs from thousands of devices across the network. These devices may include dozens of vendors and product types —firewalls, bastion hosts, intrusion detection systems, and more than a dozen other categories. Even firewalls from the same vendor may have different log formats due to different production years. Recognizing these formats and storing effective information is the foundation of the entire big data analysis system.

The **correlation analysis engine** serves as the system' s engine. Big data is like a gold mine awaiting development; mere storage creates no value. An excellent analysis engine employing advanced search technology can rapidly detect anomalous behavior and hacker intrusions.

The **situational awareness module** provides threat intelligence-based augmentation, feeding known criminal tactics and characteristics from the internet into the system to give the correlation analysis engine heightened sensitivity.

The **visualization module** must present information that goes beyond simple IP addresses and incomprehensible alerts for non-professionals. It should display intrusion sources and pathways while correlating and showing associated assets, departments, and personnel names.

[Figure 1: see original paper] System Architecture Design Diagram

The ultimate objective is a threat intelligence and log-based system capable of rapid, automated security big data analysis, comprehensive threat and anomaly monitoring, rapid response and remediation, and in-depth security incident investigation. Through visualization technology, the system presents the organization's overall security posture, enabling security managers and operations personnel to maintain real-time situational awareness, proactively identify threats, and ensure smooth business operations.

Implementation and Practice

Following initial system development, we partnered with Zhejiang Hua Shu, a technologically leading broadcasting network company, to conduct exploration and practice. Through extensive research on broadcasting network companies' broadcast and transmission processes nationwide and deep collaboration with system developers, we enhanced three key capabilities and resolved a broadcasting-specific log collection challenge.

1. Ensuring Comprehensive Data Analysis Through "All-Dimensional, Three-Dimensional" Collection

Our practice revealed that network physical entry points are not the only attack vectors. Therefore, relying solely on inline security devices and their log statistics is insufficient for detecting advanced threats. We implemented comprehensive collection across three dimensions: (1) **Full terminal collection** from PCs, smart devices, enterprise users, applications, and internet-facing services; (2) **Full traffic data collection** from aggregation layer switches, core layer switches, and access layer switches; and (3) **Full log collection** encompassing system-level logs, data logs, security device logs, and alert logs.

Second, we fully leveraged cloud security resources, which form the core of the new security defense architecture. Threat intelligence and big data utilization provide massive data support for unknown threat discovery and APT detection in real network environments, enabling comprehensive feature tracking of attackers and continuous discovery of unknown threats to ensure detection accuracy.

A broadcasting-specific challenge emerged: DVB live broadcast networks consist of multiple independent small LANs with potentially overlapping IP addresses. The challenge was to collect logs without breaking existing isolation, changing original IP addresses, or compromising network independence. We addressed this by connecting these LANs through firewalls, maintaining isolation while

enabling log transmission to collection servers with automatic source address translation to a planned new address range.

2. Threat Intelligence Data Classification, Grading, and Consolidated Management

Threat intelligence is evidence-based knowledge about known or emerging threats that provides precise decision-making references for security personnel. The challenge lies in automating and operationalizing this intelligence to drive attack behavior analysis and traceability queries.

Our system design first classifies and grades threat intelligence, then standardizes its format to enable direct matching with local big data search keywords. This foundation supports network attack analysis and traceability. The big data search technology stores indexes as multiple shards and replicas in a distributed system, improving retrieval performance while ensuring threat data reliability and accuracy. Its default in-memory indexing enables near-real-time queries for recently ingested data, while terabyte-scale disk-stored data can be queried within seconds, enabling full data collision and local risk analysis for threat intelligence-driven attack analysis and traceability.

3. Improving System Efficiency Through Big Data Processing Technology

The system' s correlation analysis core technology is big data search-based warning monitoring and analysis, where enhanced data discovery capability is critical. Traditional approaches using SQL and relational databases cannot meet current performance requirements for data processing and correlation at our data volumes. Therefore, we integrated big data search technology seamlessly into the system.

Innovatively adopting search engine technology as the core for local data storage and retrieval, with JSON format for input/output, we achieved dramatically improved retrieval performance and terabyte-scale rapid search capabilities while significantly reducing interface development effort compared to traditional architectures. This technology provides the architectural foundation for large-scale log data mining in provincial broadcasting networks, enabling rapid storage, extraction, and analysis that satisfies requirements for real-time, high-efficiency, concurrent, and reliable data processing and analysis, while improving warning detection time efficiency.

[Figure 2: see original paper] Typical Prefecture-Level Front-End Deployment Topology

The system supports remote configuration and permission management through the provincial company control center, features low deployment costs, and offers excellent scalability, significantly reducing technical difficulty for branch office deployment. This avoids the security risks and hardware costs of traditional

multi-stage deployment approaches while reducing resource consumption and operational costs.

Results and Evaluation

The system has operated stably for nearly one year. Previously, provincial security engineers had minimal visibility into local conditions and could only passively respond to security incidents. Post-implementation, the system provides a complete province-wide solution of "pre-warning + in-process protection + post-incident service." Through horizontally scalable big data capabilities and analysis, it achieves extensive multi-dimensional massive data collection, processing, and presentation. By integrating comprehensive detection results with rapid response and in-depth investigation, it forms a closed-loop security incident handling process, dramatically improving security operations effectiveness and ensuring smooth business operations.

References

[1] Ren Kai, Deng Wu, Yu Yan. Research on Network Log Analysis System Based on Big Data Technology[J]. Modern Electronics Technique, 2016, 39(2): 39-41, 44.

[2] Chen Shimin. Big Data Analysis and High-Speed Data Updates[J]. Journal of Computer Research and Development, 2015, 52(2): 333-342.

[3] Rossi Dario, Traverso Stefano, Finamore Alessandro, Mellia Marco, Khatouni Ali, Safari Munafo Maurizio. Methodology and application to carrier-grade NAT[J]. Computer Networks, 2016, 107(Oct.9 Pt.1): 20-35.

[4] Li Tianfeng, Yao Xin, Wang Jinsong. Research on Real-Time Cloud Monitoring Platform for Large-Scale Network Anomalous Traffic[J]. Information Network Security, 2014(9): 1-5.

[5] Deka Ganesh Chandra, Walczak Steven. Special Issue on Bigdata Analytics in Practice[J]. Journal of Organizational and End User Computing, 2017, 29(4): vi-viii.

[6] Gao Jing, Duan Hui. Research on JSON Data Transmission Efficiency[J]. Computer Engineering and Design, 2011, 32(7): 2267-2270.

(Author' s Affiliation: Zhejiang Radio and Television Development Corporation)

Note: Figure translations are in progress. See original paper for figures.