International Journal of Informatics and Information Systems ISSN 2579-7069
Vol. 5, No. 3, September 2022, pp. 108-113

Integrating Big Data Technologies to Strengthen Network Security Awareness

Taqwa Hariguna*

[1.] Department of Information System, Universitas Amikom Purwokerto, Indonesia
[1] taqwa@amikompurwokerto.ac.id
* corresponding author

(Received: April 23, 2022; Revised: May 15, 2022; Accepted: July 12, 2022; Available online: September 30, 2022)

Abstract

The Network Security Alarm Warning System (NSAWS) is a sophisticated, real-time, large-scale database management system designed to enhance cybersecurity by analyzing user identity data and access rights within the amassed information to detect potential threats and issue timely alarm notifications. This paper explores methods to bolster network attack prevention in a Big Data framework, starting with an overview of prevalent internet technologies in our country and their current applications. It then elaborates on the architecture, deployment strategies, and operational methodologies of the NSAWS, emphasizing the integration of fundamental security infrastructures like cloud computing platforms and firewalls. Following this, the traditional NSAWS is analyzed, and a simulation platform is tested to evaluate its performance in identifying and alerting network security threats. The results indicate that the NSAWS platform demonstrates exceptional accuracy and stability in its warning capabilities, effectively safeguarding network security by swiftly addressing potential vulnerabilities and threats. This paper underscores the importance of leveraging advanced technologies and robust security frameworks to fortify network defenses against evolving cyber threats, highlighting the NSAWS as a vital tool in maintaining and enhancing network security in an increasingly digital and interconnected world.

Keywords: Big Date Technology, Network Security, Security Early Warning, Early Warning Reminder

1. Introduction

With the widespread application of computer networks, various forms of malicious network attacks have continuously emerged, evolving towards large-scale, automated, and coordinated patterns, thereby increasing their harmful impact. To maximize network security, people commonly use firewalls, antivirus software, intrusion detection systems, and other security measures. However, these tools can only address certain aspects of network security needs. Users also desire systems that can predict future network attacks based on current security conditions, identify the attackers' ultimate intentions, and provide early warnings of impending attacks.

In recent years, the rapid development of cloud computing, the Internet, and the Internet of Things (IoT) has led to a significant increase in the number of data collection terminals. Numerous scholars have made significant strides in enhancing network security through Big Data (BD) processing. However, the rapid growth of network BD presents severe challenges to the storage and computing capabilities of traditional hardware. According to a report by the International Data Corporation (IDC), the volume of data was estimated to reach 1.8 zettabytes by 2020. The characteristics of network BD are summarized by the 5Vs: Volume, Variety, Veracity, Velocity, and Value. Network BD encompasses large-scale, diverse (including structured, semi-structured, and unstructured data), and often unexpected and emergent data, making it challenging for researchers to evaluate and predict its changing states.

This paper aims to overcome the limitations of traditional Network Security Situational Awareness (NSSA) models by introducing a NSAW reminder model tailored to address network danger scenarios. The proposed model leverages a variety of distributed algorithms and incorporates the high-dimensional, voluminous, and complex nature of network big data. By focusing on these advanced methodologies, the NSAW reminder model seeks to provide a more robust and proactive approach to network security, enhancing the ability to predict, identify, and mitigate potential threats effectively.

2. Literature Review

2.1. NSAW

The Network Security Alarm Warning (NSAW) system primarily operates by identifying indicators of intrusion through abnormal network traffic, unusual network behaviors, virus threats, and similar anomalies. Once these signs are detected, NSAW employs pre-configured attack models to analyze the intrusion process and predict the attacker's potential next steps. This analysis includes evaluating the impact of the attack on the network and assessing the severity of the threat posed by the attacker's actions. The ultimate goal is to anticipate and preemptively mitigate potential risks before the attacker can escalate their activities and compromise the system further.

By leveraging advanced detection algorithms and real-time monitoring capabilities, NSAW aims to implement proactive defense measures. These measures may include isolating compromised network segments, blocking malicious traffic, and implementing other preventive actions to neutralize threats promptly. The proactive approach of NSAW not only helps in safeguarding critical network assets but also minimizes potential downtime and operational disruptions caused by cyber-attacks. This proactive defense strategy is crucial in today's cybersecurity landscape, where threats are increasingly sophisticated and continuously evolving, necessitating agile and preemptive responses to protect sensitive information and maintain operational continuity.

In essence, NSAW serves as a vital component of comprehensive cybersecurity frameworks, offering proactive threat detection and response capabilities. By continuously monitoring network activities and analyzing potential threats, NSAW enables organizations to stay ahead of cyber adversaries, mitigate risks effectively, and uphold the integrity and availability of their network infrastructure and services.

3. NSAW and Reminder Based on BD Technology

3.1. NSAWS Architecture

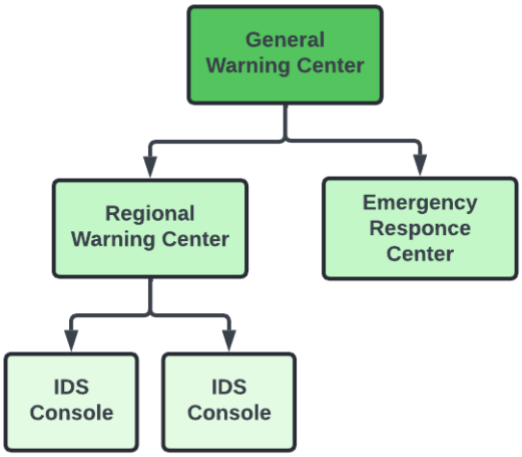

Figure 1. Distributed NSAW architecture

The Network Security Alarm Warning System (NSAWS) is a crucial decision support system designed to assess threats and issue early warnings for network attacks based on open information sources. Various architectures have been proposed to achieve the goals of NSAWS. This paper advocates for a distributed early warning architecture utilizing a multi-level structure, as depicted in Figure 1. In this architecture, risk assessments trigger early warnings that are aggregated at an Advanced Early Warning Center for comprehensive threat evaluation across distributed data. In case of an attack, the system alerts an emergency call center beyond the NSAWS scope. The interconnected early warning centers ensure reliable, accurate, and timely notifications by facilitating effective information exchange.

3.2. Architecture of the System

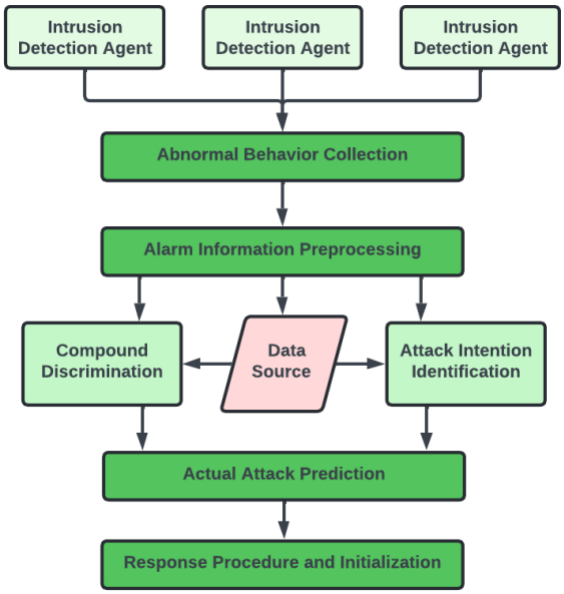

Figure 2. System architecture

The process of intrusion exhibits significant complexity, diversity, and distribution across modern network environments. Attackers often pursue their objectives through composite attacks, which involve orchestrating multiple simple attack methods rather than relying on a single approach. Traditional intrusion detection systems typically focus on detecting and alerting simple attacks, potentially overlooking the orchestrated nature of composite attacks. To effectively combat these sophisticated threats, the Network Security Alarm Warning (NSAW) system emphasizes the need for a more comprehensive approach. In response to the evolving threat landscape, the prototype NSAW system adopts a distributed architecture tailored to detect and respond to composite attacks. This architecture, illustrated in Figure 2, integrates multiple detection points and analytical capabilities distributed throughout the network. By decentralizing detection and response mechanisms, the system enhances its ability to correlate diverse indicators of compromise and detect patterns indicative of composite attacks. This proactive stance enables the NSAW system to provide early warnings and prompt responses, mitigating the impact of advanced cyber threats before they can compromise network integrity or operational continuity.

The distributed NSAW architecture not only improves the system's ability to detect composite attacks but also facilitates comprehensive threat assessment and mitigation strategies. By leveraging distributed resources and advanced analytics, the system can effectively analyze complex attack vectors and predict attacker behavior more accurately. This approach empowers organizations to bolster their cybersecurity posture by preemptively identifying and neutralizing sophisticated threats, thereby safeguarding critical assets and ensuring uninterrupted business operations in an increasingly hostile digital landscape.

3.3. Functional Components of the System

3.3.1. Collection of Anomalous Behavior

Real-time monitoring and detection of abnormal network behaviors are crucial for the Network Security Alarm Warning System (NSAWS), ensuring timely identification of attacker intentions and potential threats. This component plays a pivotal role in NSAWS by receiving a multitude of low-level alarm data generated by intrusion detection agents. These alarms are gathered and processed through sensor agents deployed across various network segments to detect security policy violations and signs of impending attacks. This module serves as the primary source of input data for the NSAWS, enabling comprehensive threat assessment and proactive defense measures.

3.3.2. Alarm Information Pre-processing Mechanism

In scenarios involving complex composite attacks, individual attack steps can trigger multiple alarms from intrusion detection systems. This influx of alarm data may include false positives or redundant information. To accurately discern

the attacker's intentions and predict subsequent actions, the system employs an alarm information pre-processing mechanism. This mechanism filters and consolidates alarm data, ensuring that only relevant and accurate information is retained. By refining the alarm data, the system enhances its ability to derive actionable insights and facilitate effective threat response strategies.

3.3.3. Identification of Composite Attacks

As attackers execute composite attacks, disparate intrusion detection systems across the network generate alarm information. Post-processing of these alarms involves distilling essential data points necessary for identifying the most probable attack scenarios. Utilizing predefined hidden Markov models tailored to composite attack patterns, the system calculates and evaluates alarm sequences. The hidden Markov model with the highest probability signifies the most likely composite attack scenario, enabling the system to prioritize response efforts and mitigate potential risks effectively.

3.3.4. Attack Prediction Module

Critical to the early warning system, the attack prediction module anticipates attacker behavior based on analyzed alarm data and identified attack intentions. This module focuses on predicting two key aspects of an attacker's course of action: firstly, assessing the attacker's likely intentions for the next attack step, and secondly, forecasting the specific attack methodologies they may employ to achieve their objectives. By forecasting both intent and method, the module provides essential intelligence for preemptive security measures, thereby bolstering the system's capacity to mitigate emerging threats and safeguard network integrity.

4. Test of NSAW Reminder

4.1. Overview of the Simulation Platform

Table 1. Main steps and procedures of the experiment

Step	Experimental process
Step 1	Generate training datasets and model data activities during offline learning to create templates for abnormal activities.
Step 2	Preprocess incoming network data flow; extract feature attributes of each record using novel data dimension reduction methods to facilitate subsequent analysis.
Step 3	Use templates generated in Step 1 to match and recognize active features in preprocessed data; classify each data record based on attribute features.
Step 4	Apply a novel unsupervised clustering method to analyze data records with unclear characteristics or those not belonging to known types; store newly identified exceptions in the exception library.
Step 5	Implement online learning for the new abnormal activity library; update template library with new exceptions; assign and update weights for each template using error feedback learning to effectively reduce false detection rates.

To validate the efficacy of the proposed NSAW and reminder system based on Big Data (BD) technology, a dedicated simulation platform was designed and constructed. The platform leverages Spark, which offers a simpler programming model compared to Hadoop MapReduce, ensuring a stable and efficient experimental environment. The experiment focuses on detecting and analyzing 25 different types of abnormal activities within the dataset, each characterized by distinct attribute features. Through analysis, corresponding templates for abnormal behavior are established to enhance the system's capability to identify specific anomalies accurately.

The simulation platform follows a structured experimental process divided into five main steps, as outlined in Table 1. Firstly, training datasets are generated and offline learning models are developed to create templates for various abnormal activities. Secondly, incoming network data undergoes preprocessing where feature attributes of each data record are extracted using a novel dimension reduction method. Subsequently, the system utilizes these templates to

match and classify data records based on their attribute features. The platform employs a novel unsupervised clustering method in the third step to analyze data records that exhibit unclear or unfamiliar characteristics, updating the exception library with newly discovered anomalies through online learning. This iterative process ensures the model continuously improves its detection accuracy and reduces false positives through error feedback learning.

4.2. Analysis of Simulation Results

In addressing BD challenges, particularly with massive datasets that exceed the processing capabilities of single machines within feasible time frames, the NSAW system introduces new demands. These demands include real-time capabilities to handle large-scale network data effectively. The experiments conducted on the open-source BD platform involve partitioning data into logical streams based on time slices, processing data blocks distributed across cluster nodes, and ultimately performing batch processing to analyze each time slice comprehensively.

Table 2. Misdetection rate

Data Volume	Misdetection Rate (%)	Stability and Detection Rate
Low	High	Poor stability, inconsistent detection rates with traditional methods
Medium	Moderate	Improved stability with optimized algorithms and adaptive strategies
High	Low	High detection accuracy as data processing capacity increases

Table 2 illustrates the misdetection rates observed during the experiments under varying data volumes. Traditional methods exhibit poor stability and inconsistent detection rates across different anomaly types due to their limited scalability. In contrast, the model proposed in this study enhances stability through optimized dynamic time warping algorithms and adaptive weight adjustment strategies. As data processing capacity increases, the model demonstrates improved stability and higher overall detection accuracy. This trend highlights the effectiveness of the proposed approach in handling large-scale network data efficiently while maintaining robust detection performance across diverse anomaly types. By leveraging Spark's capabilities and implementing advanced data processing strategies, the NSAW and reminder system proves instrumental in advancing network situational awareness, ensuring timely and accurate detection of anomalies critical to maintaining network security in dynamic and evolving environments.

5. Conclusion

In the era of Big Data (BD), extracting meaningful insights and actionable knowledge from vast datasets is paramount. This challenge is particularly pronounced in the realm of Network Security Situation Awareness (NSSA), where traditional algorithms and models face significant hurdles such as managing large volumes of data, dealing with complexity, high dimensionality, redundancy, and noise. Addressing these challenges necessitates the parallelization or enhancement of existing algorithms and the development of novel models tailored to BD environments. The primary focus of current research revolves around integrating BD processing technologies into NSSA models to establish a cohesive framework capable of effectively managing network big data.

Efforts are concentrated on refining each stage of NSSA, aiming to enhance accuracy, performance, efficiency, and overall effectiveness. This involves optimizing data preprocessing techniques to handle diverse and dynamic network behaviors, implementing advanced anomaly detection algorithms capable of identifying complex attack patterns, and leveraging distributed computing frameworks like Spark for real-time data analysis and decision-making. The goal is to bolster NSSA capabilities in detecting and mitigating cybersecurity threats swiftly and accurately amidst the rapid proliferation of data across interconnected networks.

Central to these advancements is the development of comprehensive models that can assimilate disparate data sources, including structured, semi-structured, and unstructured data types. By integrating advanced analytics and machine learning approaches, researchers seek to fortify NSSA frameworks against evolving cyber threats, ensuring robust defense mechanisms that adapt to the intricate and constantly evolving nature of network attacks. Ultimately, the ongoing research in BD-driven NSSA aims to establish a standardized approach that enhances network security posture by leveraging the full potential of BD technologies to anticipate, prevent, and respond to emerging threats proactively.

References

[1] Cui Ping. Design of laboratory security early-warning system based on wireless network monitoring technology. Modern electronics technology, 2019,042 (012): 37-39, 44.

[2] Lu Guosheng, Tian Lin, Chen Junhao. Early warning and quantitative analysis of power network security risk. Modern scientific instruments, 2019,000 (001): 101-103145.

[3] Wang Z. Research on information security early warning and decision support system based on risk control. Boletin Tecnico/Technical Bulletin, 2017, 55 (20): 581-590.

[4] Li C Y, Zheng L. Analysis of Tai Chi Ideological and Political Course in University Based on BDand Graph Neural Networks. Scientific Programming, 2021, 2021 (1): 1-9.

[5] Yi M, Xu X, Xu L. An Intelligent Communication Warning Vulnerability Detection Algorithm Based on IoT Technology. IEEE Access, 2019, 7 (99): 164803-164814.

[6] Yuan T, Zhang Y X, Ma S Y, et al. Combining the BDanalysis and the threat intelligence technologies for the classified protection model. Cluster Computing, 2017, 20 (2): 1-12.

[7] ShengliZhou, XinWang, ZeruiYang. Monitoring and Early Warning of New Cyber-Telecom Crime Platform Based on BERT Migration Learning. China Communications: the English version, 2020 (3): 140-148.

[8] YAN, Yan, YANG, et al. Disaster reduction stick equipment: A method for monitoring and early warning of pipeline-landslide hazards. Journal of Mountain Science, 2019, v.16 (12): 4-17.

[9] Einy S, Oz C, Navaei Y D. The Anomaly- and Signature-Based IDS for Network Security Using Hybrid Inference Systems. Mathematical Problems in Engineering, 2021, 2021 (9): 1- 10.

[10] Xue Y, Zhu L, Wang W, et al. Research on fault analysis and positioning technology of distribution network based on BD. Journal of Physics: Conference Series, 2019, 1176 / 062029.

[11] Huang J C, Ko P C, Fong C M, et al. Statistical Modeling and Simulation of Online Shopping Customer Loyalty Based on Machine Learning and BDAnalysis. Security and Communication Networks, 2021, 2021 (3): 1-12.

[12] Yi L, Niu D, Wang H, et al. Assessment Analysis and Forecasting for Security Early Warning of Energy Consumption Carbon Emissions in Hebei Province, China. Energies, 2017, 10 (3): 391.