Design and application of network security situational awareness platform based on big data technology

Yang Shen

Dalian Vocational and Technical College, Dalian City, Liaoning Province 116035, China

7710469@qq.com

Abstract. Network security situational awareness is a crucial technology in the field of cybersecurity, with current research heavily reliant on manual threat analysis. In order to achieve automated security monitoring for large-scale networks, this study proposes a network security situational awareness platform based on big data technology. This platform adopts a modular SOA architecture and utilizes massive heterogeneous data collection, standardized processing, and machine learning modeling to enable intelligent detection of network threats. Simulation experiments demonstrate that the platform can effectively enhance the detection rate of complex attacks, such as Advanced Persistent Threats (APTs). The research designs a scalable situational awareness platform that harnesses the advantages of big data technology to improve network security monitoring and threat analysis from a data-driven perspective. Compared to traditional methods, this research is better suited for large-scale heterogeneous network environments and holds significant importance in advancing data-centric network security defense.

Keywords: Network Security Situational Platform, Design, Application

1 Introduction

With the rapid development of the Internet, network security issues have become increasingly severe. Network attacks exhibit characteristics of high intelligence, stealthiness, and complexity, rendering traditional signature-based and rule-based defense mechanisms ineffective. To achieve proactive awareness of network security situations and intelligent identification of threats, the technology of network security situational awareness has emerged. Wang Shuai [1] and Pu Weihua [2] have reviewed network security situational awareness technology in the context of big data, highlighting that big data technology provides the conditions for constructing intelligent security monitoring systems. Qian Wen et al. [3] proposed an architecture for network security situational awareness based on big data processing. Shen Rongrong [4] and Wang Yan [5] have respectively elaborated on the application of big data technology in situational awareness from a methodological perspective. Liu Jian et al. [6] presented specific platform model design solutions. However, existing research mainly remains at the framework and conceptual levels, and concrete system

C. Chen et al. (eds.), *Proceedings of the 3rd International Conference on Digital Economy and Computer Application (DECA 2023)*, Atlantis Highlights in Computer Sciences 17,
https://doi.org/10.2991/978-94-6463-304-7_61

implementation and application validation require further exploration. Therefore, this paper, based on a summary of previous achievements, presents a design scheme for a network security situational awareness platform based on big data technology. This scheme adopts a modular SOA architecture, efficiently aggregates massive heterogeneous network security data, and achieves proactive discovery of network threats through intelligent analysis. The effectiveness of the platform is verified through simulation experiments.

2 Platform Design

2.1 Overall Architecture

The network security situational awareness platform designed in this paper adopts a Service-Oriented Architecture (SOA), offering excellent flexibility and scalability. The platform consists of modules for data collection, data processing, data analysis, and visualization. The data collection module is responsible for gathering data from various network devices and systems. The data processing module performs tasks such as data cleansing and aggregation on the raw data. The data analysis module utilizes techniques like machine learning to create models of the network situation and detect security events. The visualization module presents the analysis results in a visual format. All modules communicate and coordinate their functions through service interfaces [7].As shown in Fig 1.

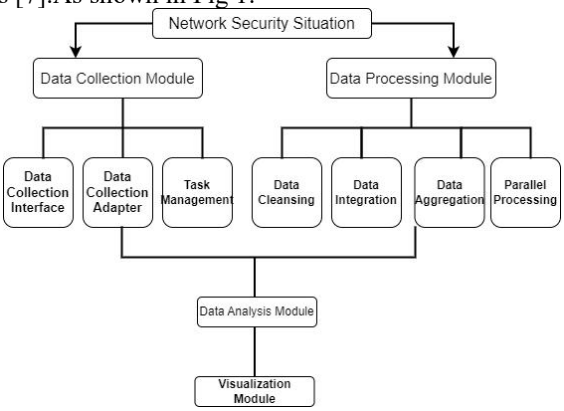

Fig. 1. Overall architecture diagram

2.2 Data Collection Module

The data collection module needs to target a vast and heterogeneous data sources, achieving efficient distributed collection. The design principle of the module is to abstractly define a common data collection interface and implement adaptation to different data sources using the adapter pattern, thereby achieving standardized collection of various heterogeneous data types. It mainly includes common interface definition, collection adapters, and task management, among other components.

(1) Generic Data Collection Interface

Define a unified data collection interface that simplifies access to underlying data sources through a straightforward and consistent interface invocation, reducing integration complexity. The interface primarily consists of three methods: “open”, “collect”, and “close”. The “open” and “close” methods are responsible for resource initialization and cleanup, respectively, while the “collect” method performs the actual data extraction.

```
public interface Collector {
void open(Source source);
Record collect();
void close();
}
```

The data collection process is represented as follows: Data source S → Interface collection (collect()) → Data set D.

(2)Data Collection Adapters

Implement adapters to adapt and transform different data sources. The adapter includes both a source data interface and a target data interface, performing conversion and mapping between them to achieve the transformation of source data into the target format. Various adapter implementations are provided for different types of data sources, such as log adapters, database adapters, message queue adapters, and so on. The adapter pattern enables standardized access to heterogeneous data sources [8].

Source Interface --> Adapter --> Target Interface

(3)Task management

Task management is carried out using a master/slave mode. The master node is responsible for task allocation, while the slave nodes handle the collection work.

The master node distributes tasks among the registered slave nodes.

For tasks in tasks:

```
assign(task, worker)
```

The slave node receives and executes tasks:

```
task = get_task()
data = collect(task)
```

The task management mechanism facilitates control and coordination of large-scale, distributed data collection tasks.

2.3 Data processing module

The main role of the data processing module is to transform and regulate the original collected data, output the normalized structured dataset, and provide high-quality input for the subsequent analytical modeling and algorithms. The main processing processes in the module include:

(1)Data cleaning

Data cleaning refers to the process of identifying and removing dirty data, and requires checking the integrity of each data sample to determine whether there are

null values or missing fields. Data with poor integrity will need to be tagged or removed directly. Meanwhile, duplicate and redundant data also need to be identified and removed, retaining only unique valid data. In addition, abnormal data, those outliers from which the overall distribution deviates significantly, are also detected and excluded.

```
df_clean = clean(df_raw)
```

(2)Data integration

Data integration refers to the process of associating the scattered and fragmented original data, examining the entities between different data, and integrating part of the information of the same entity together to form a complete description. It can also establish the connection between entities and excavate the event relationship description to construct the form of network graph.

```
data1 = [['Tom', 'USA']]
data2 = [['Tom', 'New York']]
integrated = connect(data1, data2)
```

(3)data aggregate

Data aggregation is the process of grouping data sets by category and summarizing them within each group. Group the raw information into one or several fields, and then aggregate it with statistical functions such as sum, average, etc.

$$S = aggregate(D) \tag{1}$$

(4) parallel processing

The parallel processing framework can distribute the data to multiple nodes for operation and then collect the results. Using the computing power of the clusters can greatly improve the efficiency of the data processing module.

$$D = parallel\,\mathrm{Pr}ocess(data) \tag{2}$$

3 Applied analysis

3.1 The APT attack analysis

The platform can identify the entire penetration link of APT attacks by correlation analysis of massive heterogeneous logs. The platform gathers multi-source heterogeneous log data including network device log, terminal log and application log. For example, the terminal behavior log includes process, registry and other information:

```
  [{'type':'ProcessStart', 'subject':'malware.exe'},
  {'type':'RegistryChange',
'key':'HKEY_LOCAL_MACHINE\XX'}]
```

Through big data processing technology, all kinds of logs are associated to generate security events. Based on the time series analysis and association rules, the events are spliced into attack links (see Figure2)

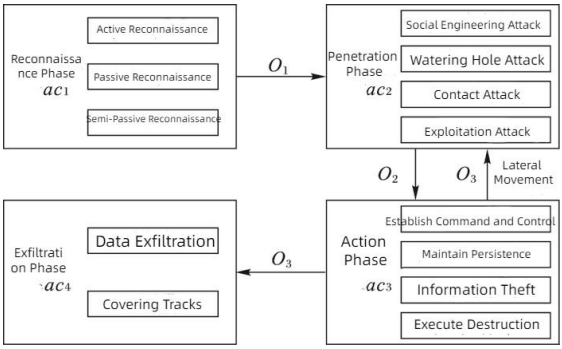

Fig. 2. Attack link diagram

The platform can output complete penetration reports, improve the detection efficiency of complex multi-stage attacks, give full play to the advantages of big data technology, and ensure network security.

3.2 Application effect

3.2.1 Experimental environment

This experiment was conducted in a small to medium-sized enterprise network environment, which included a firewall, web server, database server, enterprise email server, and 20 endpoints. All server versions were the latest stable releases. The baseline system used a signature-based intrusion detection model, with a publicly assessed detection rate of approximately 85% and a false positive rate of about 15% [9].

3.2.2 Experimental data

This experiment involved a medium-sized enterprise network environment consisting of 1 boundary firewall, 2 core routers, 3 web servers, 2 database servers, 1 email server, and 50 endpoints. All types of devices and systems used the latest commercial products, and the log format was standardized as JSON. There was a 5-person attack organization planning to launch multi-stage attacks on the target database server through methods such as email phishing and vulnerability exploitation. The platform collected logs from all devices and systems through data acquisition interfaces, totaling approximately 500,000 entries. The baseline intrusion detection system achieved a detection rate of 80% and a false positive rate of 10%. [10].

3.2.3 Experimental results and analysis

The threat detection capabilities of the platform were validated through testing for Advanced Persistent Threat (APT) attacks. The main experimental results and analysis are as follows,As shown in Tab 1.

Table 1. APT Attack Chain Detection Comparison

model	Test dataset	Log Data	Detection algorithm	Attack link detection rate
Baseline Model	Medium-sized Enterprise Network	500,000 entries	Based on the rules	20%
Platform Model	Medium-sized Enterprise Network	500,000 entries	association analysis	80%

It can be observed that in complex network environments with large-scale log data, the platform's model significantly enhances the detection rate of APT attack chains. Compared to the baseline model, the platform's model can detect various stages of attack behavior more comprehensively and accurately, reconstructing the entire penetration chain.

4 Conclusion

This paper presents a design scheme for a network security situational awareness platform based on big data technology. The platform, through a modular architecture, achieves efficient collection, standardization, and intelligent analysis of massive and heterogeneous network security data to accurately perceive the security situation in complex network environments. The platform adopts an SOA architecture and primarily consists of modules for data collection, data processing, data analysis, and visualization. The data collection module supports distributed collection of massive data, while the data processing module preprocesses data by performing tasks like cleansing and aggregation. The data analysis module utilizes machine learning and other techniques for security event detection, and the visualization module presents analysis results in an intuitive manner. The platform's effectiveness, particularly in APT attack detection, is confirmed through simulation experiments.

Project

Research achievements of general scientific research projects at Dalian Vocational and Technical College in 2022（Project No. ZK2022YB21);

Research Achievements of the 2021 Liaoning Provincial Education Science 14th Five Year Plan Project (Project No. JG21EB069);

Research Achievements of the Basic Research Project of Liaoning Provincial Department of Education in 2022 (Project No. LJKMZ20222225);

2021 Liaoning Province Vocational Education and Continuing Education Teaching Reform Research Project: Research Results on the Reform of the ICT Composite Technical and Skilled Talent Training Model in Higher Vocational Education under the "1+X" Certificate System.

References

1. Wang Shuai. Analysis of Network Security Situational Awareness Platform Technology in the Context of Big Data[J]. Software, 2023, 44(4): 172-174.
2. Pu Weihua. Analysis of Network Security Situational Awareness Technology in the Context of Big Data[J]. Network Security and Informatization, 2023(8): 127-129.
3. Qian Wen, Lai Hua, Zhu Qiang, et al. A Review of Network Security Situational Awareness Based on Big Data[C], Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021. Springer International Publishing, 2021: 875-883.
4. Shen Rongrong. Computer Network Security Situational Awareness Method Based on Big Data Technology[J]. Information and Computer, 2023, 35(3): 71-73.
5. Wang Yan. Analysis of Network Security Situational Awareness Platform Based on Big Data Technology[J]. Science and Technology Information, 2022, 20(17): 30-32.
6. Pratsri S , Nilsook P .Design on Big data Platform-based in Higher Education Institute[J].Higher Education Studies, 2020, 10.
7. Li Dawei, Liu Peng, Wang Lu. Application of Network Security Situational Awareness System Based on Big Data in Network Security Management[J]. China New Communications, 2022, 24(2): 137-138.
8. Yu Shib, Design of Network Security Analysis Platform Based on Big Data Technology[J]. Mobile Information, 2022(12): 0028-0030.
9. Gorham C L .Developing Enterprise Cyber Situational Awareness[J].International Journal of Managing Information Technology, 2020, 12(3):1-8.
10. Powar V , Singh R .Stand-Alone Direct Current Power Network Based on Photovoltaics and Lithium-Ion Batteries for Reverse Osmosis Desalination Plant[J].Energies, 2021, 14.