PROCEEDINGS OF SCIENCE

A distributed framework for security operation center in the application of Institute of High Energy Physics

Jiarong Wang,[a,*] **Junyi Liu,**[a,b] **Qianran Sun,**[a,b] **Tian Yan,**[a] **Dehai An**[a] **and Fazhi Qi**[a,c]

[a] *Computing Center, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, P.R.China*

[b] *School of Nuclear Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, P.R.China*

[c] *Spallation Neutron Source Science Center, Dongguan 523803, P.R.China*

E-mail: wangjr@ihep.ac.cn

Security operations centers (SOCs) standardize what SOCs do. Most SOCs are designed as a centralized mode which serves for single organization. The single-organisation SOCs face structural challenges in addressing operational security scenarios where the incidents span muliple organisations, and where these organsiations habitually deliver services in a synergistic way. In this paper, we propose the distributed security operation center (DSOC) that provides the distributed working mechanism for multiple organizations over the wide area network by combining their security probes. The security probes of DSOC are deployed in the different organizations to collect data and the collected data is transferred over wide area network to the data analysis center of the DSOC. Especially, the data communication between security probes and data analysis center is encrypted to ensure the data security of every organization. The data analysis center adopts rule-based, AI-based and threat intelligence-based algorithms to detect cyber-attacks. The detection results are input into the automated response module in the DSOC. The automated response module is the client-server structure and the client are installed in the security probe. The server of the automated response sends commands across the wide area network to the target client of the security probe to block the attackers quickly, and meanwhile the communication between client and server in the response processes is encrypted. In addition, the threat intelligence component of DSOC can aggregation intelligence from the organizations and easily share this intelligence with all organizations based on the distributed security probes. The DSOC also builds the security situational awareness system that visualizes the cyber threats of every organization and set the permission to view the security situation by using access control for every organization. The DSOC has been applied to institute of high energy physics (IHEP) and deployed in several collaborative large scientific facilities and scientific data centers since 2021. The security protections are persistently provided to all organizations within the DSOC framework.

International Symposium on Grids and Clouds (ISGC) 2023,
19 - 31 March 2023
Academia Sinica Taipei, Taiwan

*Speaker

PoS(ISGC&HEPiX2023)026

1. Introduction

IHEP suffers frequent attacks by adversaries with different agendas. To minimize cybersecurity risks and improve the security operations of IHEP, we have created the Security Operation Center (SOC) in 2021 in IHEP [1], which is a centralized defense group in IHEP based on five data processing layers: data collection, data preprocessing, data storage, data analysis and data application. However, the current SOC setup is mainly used for ensuring the network security of one site and is insufficient to defend against cyber-attacks for several collaborative large scientific facilities and scientific data centers across the wide area network.

In this paper, a distributed SOC (DSOC) is proposed which can protect several sites from cyber threats and achieve coordinated operation among multiple sites across the wide area network. The proposed DSOC has three features:

- Multiprobe distributed framework: The proposed DSOC contains a data analysis center in IHEP and a few security probes in the sites. The data analysis center interacts with security probes in client-server mode and integrates data from several sites to achieve multi-site collaboration.
- Multi-function security probe: The security probe can support network traffic and security logs collection, and perform automatic response at the sites by combining data from the central data analysis center and security devices of sites. In addition, threat intelligence can be quickly shared with several sites by using security probe.
- Centralized data analysis: The data analysis center can centralize analysis of all site data and detect known and unknown attacks in the specific security scenarios by rule-based methods and machine learning algorithms.

2. Related work

Security Operations Centers (SOCs) have been created by many organizations or enterprises as effective cybersecurity monitoring solutions. D. Crooks presents a minimum viable security operation senter for the modern grid environment[2, 3]. This security operation senter reference design is based on 4 different stages: data sources and threat intelligence, data pipelines, storage and visualisation, alerting. Besides, Bidou presents the implementation of a completely integrated security operation center (SOC), called SOCBox in order to overcome the limitations of IDS[4, 5]. In addition, Ganame[6] presents several methods used to test the accuracy and the performance of the SOCBox. Miloslavskaya [7] presents security information center to empower information security management for intranets all organizations. Radu [8] proposes a few architectural considerations regarding frameworks and operating models that can be used when building a variably sized SOC. The SOC framework is proposed comprising log collection, analysis, incident response, reporting, personnel and continuous monitoring [9–12]. Settani[13] proposes a system architecture for a national SOC, defining the functional components and interfaces it comprises. Tafazzoli [14] customizes SOC for OpenStack environment to detect cloud specific attacks. The customized SOC receives and normalizes OpenStack alerts. The above mentioned SOC setup is mainly used for ensuring the network security of one organization or enterprise.

In order to defend against cyber-attacks for several collaborative organizations or enterprises across the wide area network, the distributed SOC is more appropriate. The Splunk can analyse and manage logs from multi-customer based on the Splunk agents and universal forwarders. In addition, the commercial NGSOC[15], T-sec[16], and ISOP[17] can provide the security operations for several organizations by using several probes. However, during the trial process of these products in IHEP, it was found that there were problems such as high deployment and maintenance costs, insufficient scalability, data security risks, and insufficient timeliness of response, which could not fully meet the needs of security operations in IHEP.

3. Key techniques of the distributed SOC

3.1 The distributed framework

The proposed distributed framework includes a server and a few clients. The client is the security probe deployed in every site and the server is the data analysis center deployed in IHEP. All the sites within the framework are highly collaborative and have mutual trust. To ensure data security, the transmission between the client and the server is protected. The framework is shown in figure 1.

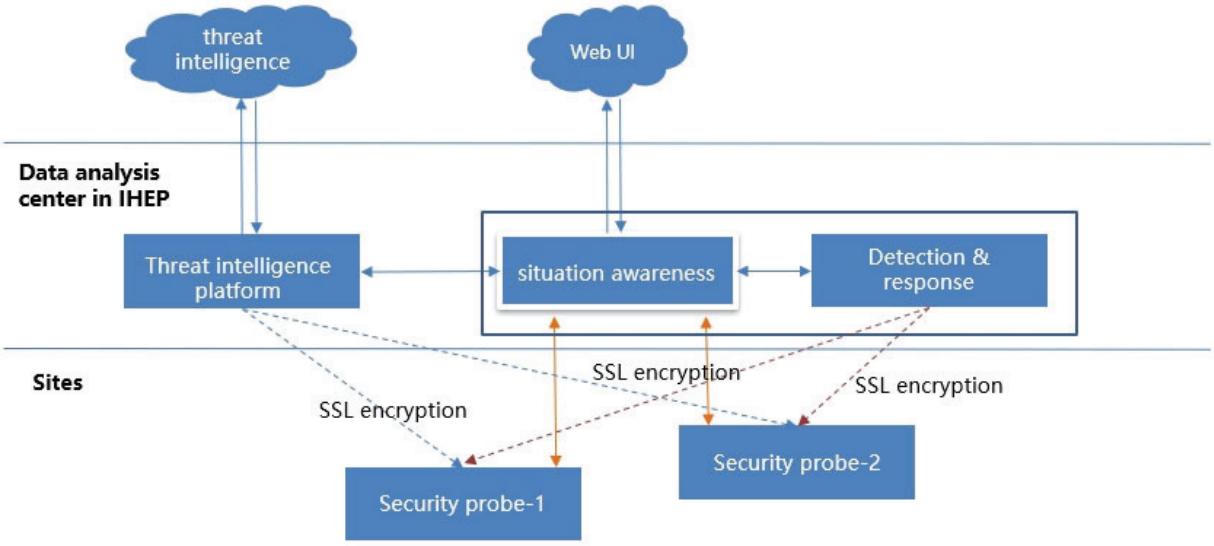

Figure 1: The distributed framework

The functions of security probe mainly include data collection and data application. The data application include automatic response and intelligence sharing. The collected data of the security probes are encrypted and transmitted to the server. The server then analyzes all site data and processes these data in the threat intelligence platform, the situation awareness system, the threat detection and the response services. The comparation between the original SOC and the DSOC based on the aforementioned five data processing sections is shown in figure 2. All data processing sections of the original SOC are deployed in IHEP, but parts of data processing sections of the DSOC are deployed in every site.

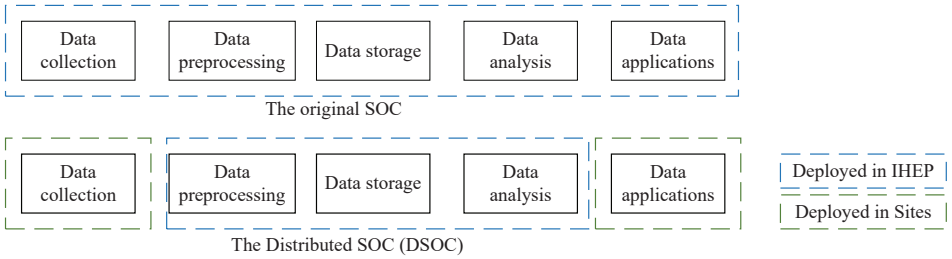

Figure 2: Comparation between the original SOC and the DSOC

The DSOC uses diffrent technologies to build five data processing sections as shown in figure 3. We will introduce these technologies in detail in the following.

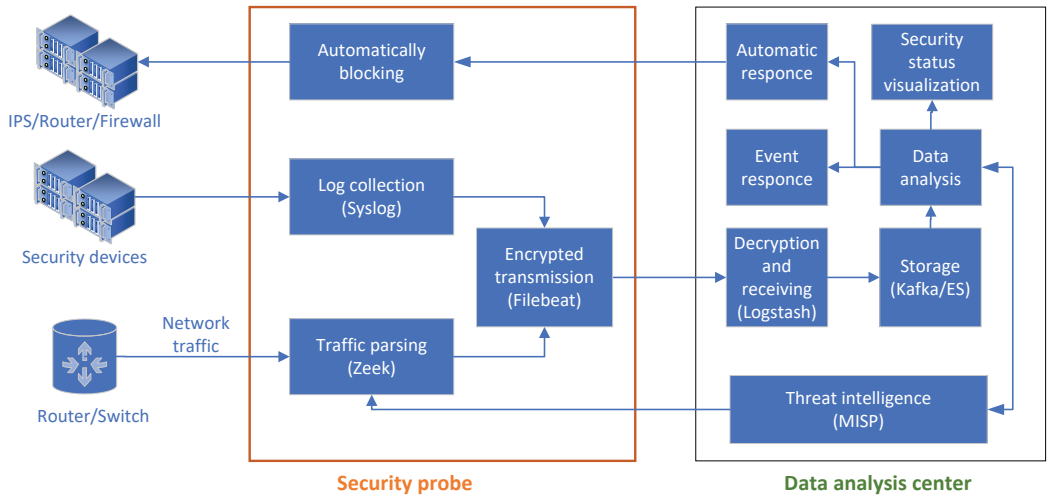

Figure 3: The technologies used in the DSOC

3.2 Data collection of security probe

The security probe collects security data of every site. On the one hand, we configure Rsyslog service to collect firewall logs, Intrusion Prevention System logs and other security devices logs. On the other hand, we set up Zeek [18] on the security probe to parse raw network packets. And then, the collected security device logs and network traffic data are sent to the data analysis center in IHEP from the probe by using Filebeat [19] which encrypts the transmission over wide area network to ensure the data security of sites.

3.3 Data preprocessing and storage

The data analysis center in IHEP firstly uses logstash to receive all sites data. The logstash [20] can decrypt the data and output into kafka [21]. And then we read the data stream from kafka and transform these data into structured form by using data correlation and data statistics methods. The structured data are stored into ElasticSearch [22]. The data preprocessing and storage process can be found in figure 4.

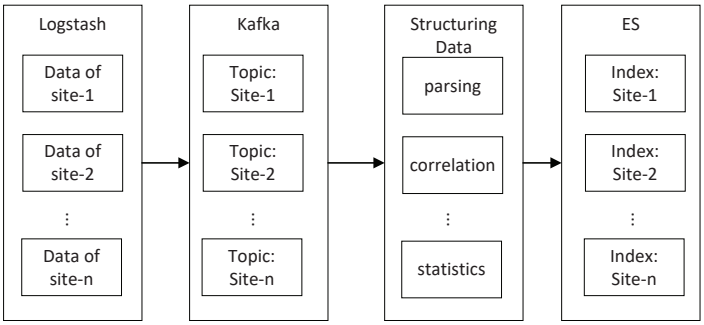

Figure 4: The data preprocessing and storage

3.4 Data analysis

There are three kinds of data analysis methods in the DSOC system: rule-based data analysis, threshold-based data analysis and AI-based data analysis.

The rule-based data analysis uses the known attack features to detect specific attack patterns. For example, the log4j2 vulnerability attack can be detected from the network traffic if the http protocol network packets contain string like 'jndi:ldap://' or 'jndi:rmi', and we use the feature string in the detection rule.

The threshold-based data analysis compares the statistics with the specific threshold to detect attacks. For example, the brute force attack detection for login servers counts the number of login failures, and if the number is larger than the threshold, the login IP address is detected as a malicious IP. The detection flow is shown in figure 5.

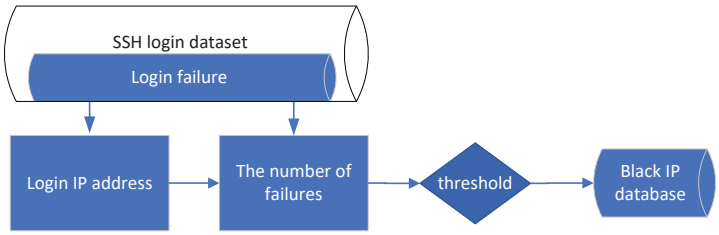

Figure 5: The brute force attack detection for login servers

The AI-based data analysis deploys anomaly detection algorithms and deep neural network to detect attacks. For example, we have developed two weakly-supervised method for malicious network traffic detection [23, 24]. These methods apply multiple instance learning to achieve the fine-grained classification of network traffic. The experiment results have demonstrated our detection methods are better than the state-of-the-art detection systems[23, 24].

3.5 Automatic response

The automatic response leverages agent and server components to handle the malicious IP addresses detected by the data analysis methods. The server, deployed in the data analysis center, generates an automatic response and sends the command to all distributed agents to block the event. The commands contain malicious IP and blocking operations. The agent of the automatic response

deployed in the security probe can receive the malicious IP from the server and then perform the blocking operations in the site by the API of IPS, firewall or router. In addition, we adopt dynamic blocking strategy which includes 6 blocking levels: 4 hours, 1 day, 1 week, 1 month, 1 year, indefinitely respectively.

3.6 Threat intelligence

The functions of the threat intelligence platform can be classified as four parts. The first part is threat intelligence collection which aggregates threat intelligence from two sources. One source is the malicious IP database which stores the detected and blocked IP addresses from all sites. The other source is the security related websites which post the malicious IP dataset periodically such as structured data feeds. The second part of the threat intelligence platform is threat intelligence storage which builds a graph database [25] to store the collected threat intelligence based on the STIX standard [26]. The third part is threat intelligence sharing in which security probes pull threat intelligence from the threat intelligence database. The last part is the threat intelligence application which uses the malicious IP for the threat detection and automatic blocking.

3.7 Security situation awareness

The security situation awareness can provide the overview of security status of every site by visualizing several statistics, such as the number of threats, the trends of threats, and the origins of threats. The used technologies include Django framework [27], Vue.js [28] and Echarts [29]. Meanwhile, the access control strategy is set up to ensure the data security so that one site can't view the security status of other sites.

4. Applications

The DSOC has been applied to institute of high energy physics (IHEP) and deployed in 4 collaborative large scientific facilities and 3 scientific data centers. The number of site data types collected and analyzed in the DSOC system is 9 as shown in figure 6.

For the authentication data of sites such as secure shell (SSH) logs, single sign-on (SSO) logs and mail system logs, the DSOC can send alert emails to the user after a detection of an unusual login to his account. For example, a user's email account was leaked because of phishing mail. And then the intruder logged into his account at mid-night. The DSOC detected the unusual login and sent an alert email to the user. This user read our notice in the morning and changed his password.

For the security alerts logs such as network intrusion detection system (NIDS), host intrusion detection system (HIDS), web application firewall (WAF) and intrusion prevention system (IPS), the DSOC can filter 99% false positives and perform automatic response. For example, the DSOC counts the number of backdoor attacks and compares with a threshold. If the number is larger than the threshold, the attack IP of backdoor attack is blocked automatically.

In addition, the DSOC also analyzes DNS logs of sites to detect DNS flood attack. The DSOC firstly counts the number of DNS queries. If the number is larger than the threshold, the DNS flood attack is detected and the attack IP is blocked automatically if the queries are from the external IP, but if the queries are from the internal IP, the DSOC sends alerts emails to security manager to handle.

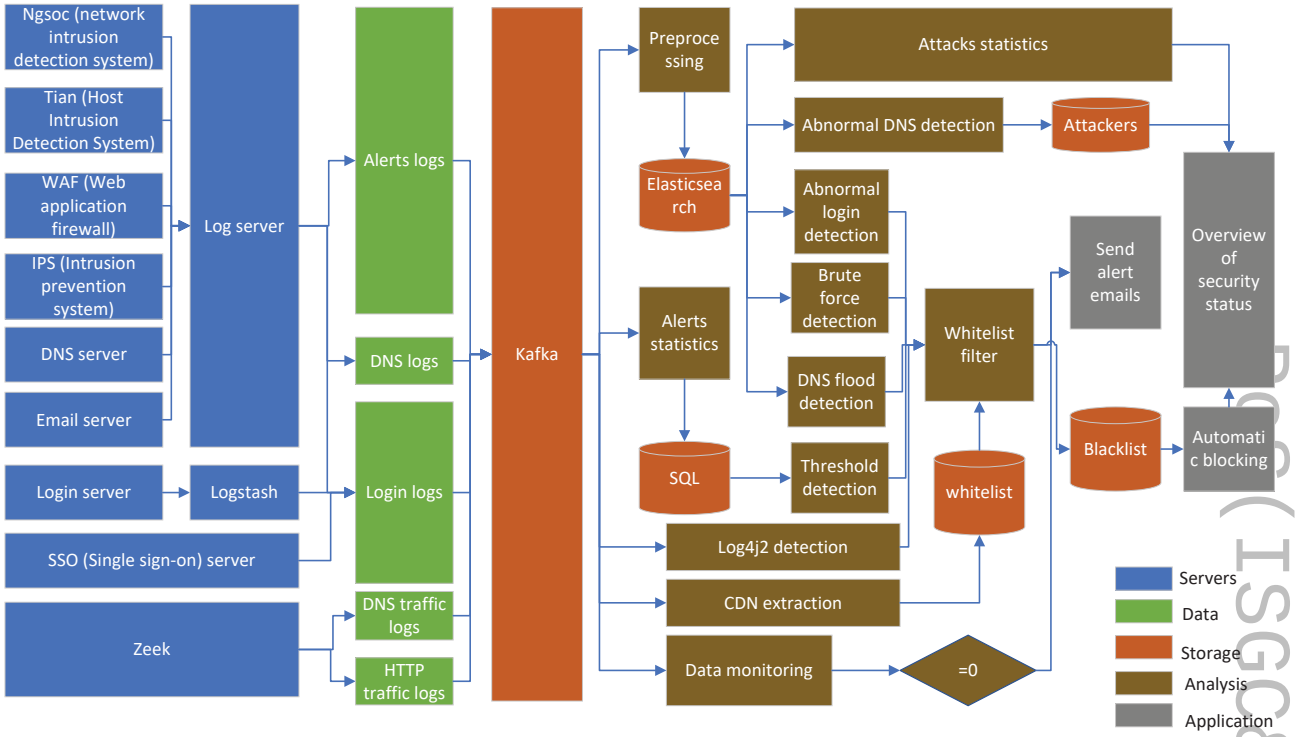

Figure 6: Data flow diagram for the DSOC in IHEP

The automatic blocking of sites has 4 steps in general. For example, the automatic blocking applied in China National Genomics Data Center (NGDC) includes: 1) security probe in NGDC sends the IPS, WAF and firewall logs; 2) data analysis center aggregates logs and extracts the malicious IP; 3) security probe pulls the malicious IP by the https protocol; 4) security probe operates the IPS in NGDC to block the malicious IP. The blocking strategy in site NGDC is set as blocking for 4 hours.

In total, the number of attack types identified by the DSOC is 12 and there are 6 blocking levels applied for different attack types. The figure 7 shows the attack types and the number of blocked IP addresses in 2022.

The DSOC has provided situation awareness for 7 sites. The figure 8 and 9 show the visualizations of the situational awareness in their operational environment in IHEP and in NGDC.

5. Conclusion

The paper presents the distributed SOC (DSOC) which can provide the centralized management of multiple sites. The DSOC is composed of the data analysis center in IHEP and multiprobe in sites which can achieve several sites security defense collaboration. The application of the DSOC in 7 sites has improved the security defense ability for multiple sites. The next goal would be to add more threat detection functions for the DSOC, such as network file leakage detection.

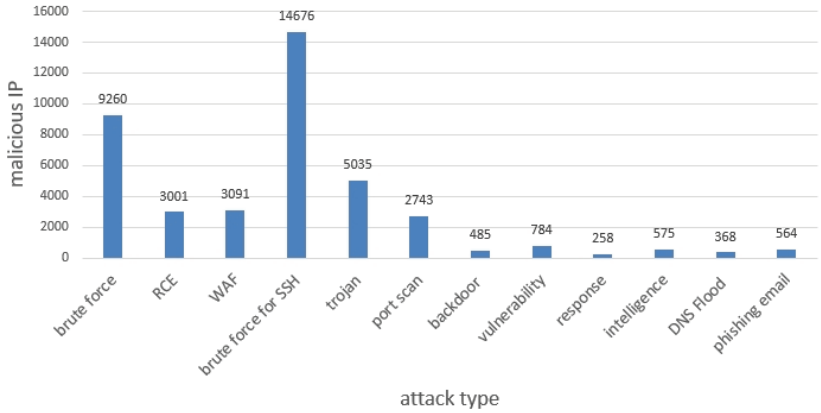

Figure 7: Attack type and the number of blocked IP addresses in 2022

Figure 8: Situation awareness in IHEP

Acknowledgments

This work is supported in part by the Xiejialin Project of Institute of High Energy Physics under Grant no. E25467U2 and the specialized project for cybersecurity and informatization in the 14th Five-Year Plan of CAS under grant no. WX145XQ12.

References

[1] J. Wang, T. Yan, D. An, et al. A comprehensive security operation center based on bigdata intelligent detection and threat intelligence. Proceedings of Science, 2021,378:028.

[2] D. Crooks, L. Valsan, K. Mohammad,et al. Operational security, threat intelligence & distributed computing: Thewlcg security operations center working group, EPJ Web Conferences,vol. 214, p. 15, May 2019.

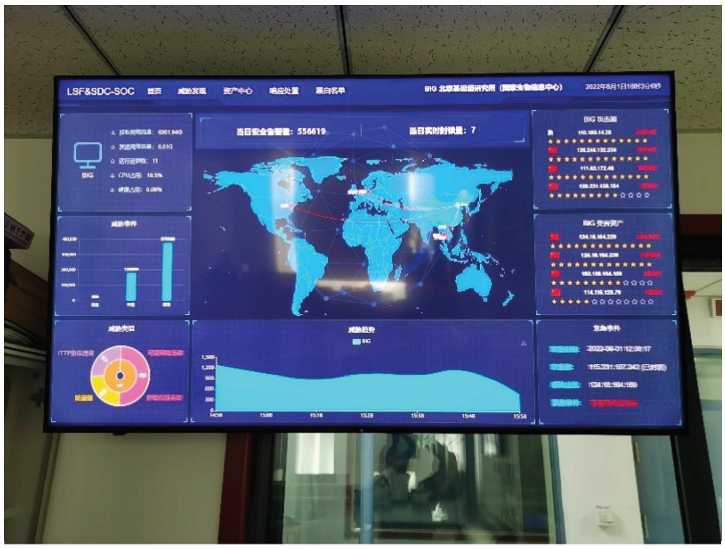

Figure 9: Situation awareness in NGDC

[3] D. Crooks and L.Valsan, Building a minimum viable security operationscentre for the modern grid environment, in Proc. Int. Symp. GridsClouds, Trieste, Italy, Nov. 2019, p. 10.

[4] R. Bidou, J. Bourgeois, and F. Spies, Towards a global security architecture for intrusion detection and reaction management, in Information Security Applications (Lecture Notes in Computer Science), vol. 2908. Berlin, Germany: Springer, 2004, pp. 111-123.

[5] J. Bourgeois, A. Ganame, I. Kotenko, and A. Ulanov, Software environment for simulation and evaluation of a security operation center, inInformation Fusion and Geographic Information Systems (Lecture Notesin Geoinformation and Cartography). Berlin, Germany: Springer, 2007,pp. 111-127.

[6] A. Ganame, J. Bourgeois, R. Bidou, and F. Spies, Evaluation of the intrusion detection capabilities and performance of a security operation center, in Proc. Int. Conf. Secur. Cryptogr., 2006, pp. 48-55.

[7] N. Miloslavskaya, Security intelligence centers for big data processing,in Proc. 5th Int. Conf. Future Internet Things Cloud Workshops(FiCloudW), Prague, Czech Republic, Aug. 2017, pp. 7-13.

[8] S. Radu, Comparative analysis of security operations centre architectures;Proposals and architectural considerations for frameworks andoperating models, in Innovative Security Solutions for InformationTechnology and Communications (Lecture Notes in Computer Science),vol. 10006. Cham, Switzerland: Springer, 2016, pp. 248-260.

[9] C. Onwubiko, Cyber security operations centre: Security monitoringfor protecting business and supporting cyber defense strategy, inProc. Int. Conf. Cyber Situational Awareness, Data Analytics Assessment(CyberSA), London, U.K., Jun. 2015, pp. 1-10.

[10] X. Hu and C. Xie, Security operation center design based on D-Sevidence theory, in Proc. Int. Conf. Mechatronics Autom., Luoyang,China, Jun. 2006, pp. 2302-2306.

[11] S. Yuan and C. Zou, The security operations center based on correlationanalysis, in Proc. IEEE 3rd Int. Conf. Commun. Softw. Netw., Xi'an,China, May 2011, pp. 334-337.

[12] E. G. Amoroso, Cyber attacks: Awareness,Netw. Secur., vol. 2011,no. 1, pp. 10-16, Jan. 2011.

[13] G. Settanni, F. Skopik, Y. Shovgenya, R. Fiedler, M. Carolan, D. Conroy,K. Boettinger, M. Gall, G. Brost, C. Ponchel, M. Haustein, H. Kaufmann,K. Theuerkauf, and P. Olli, A collaborative cyber incident managementsystem for European interconnected critical infrastructures, J. Inf. Secur.Appl., vol. 34, pp. 166-182, Jun. 2017.

[14] T. Tafazzoli and H. Gharaee Garakani, Security operation center implementation on Open-Stack, in Proc. 8th Int. Symp. Telecommun. (IST),Tehran, Iran, Sep. 2016, pp. 766-770.

[15] NGSOC, https://www.qianxin.com/product/detail/pid/110

[16] T-Sec, https://cloud.tencent.com/product/soc

[17] ISOP, https://www.nsfocus.com.cn/html/2019/209_1230/99.html

[18] The Zeek project, https://www.zeek.org/

[19] B. V. Elasticsearch. Filebeat, https://www.elastic.co/downloads/beats/filebeat

[20] B. V. Elasticsearch. Logstash, https://www.elastic.co/cn/downloads/logstash

[21] The Apache Kafka project, http://kafka.apache.org/

[22] B. V. Elasticsearch. Elasticsearch, https://www.elastic.co/downloads/elasticsearch

[23] J. Liu, Z. Li, J. Wang, et al. A weakly-supervised method for encrypted malicious traffic detection. Proceedings of Science, 2022,415:027.

[24] Z. Li, J. Liu, J. Wang, et al. Malicious Traffic Detection with Class Imbalanced Data Based on Coarse-grained Labels. Proceedings of Science, 2022, 415:030.

[25] The Neo4j, https://neo4j.com/.

[26] The Stix, http://stixproject.github.io/.

[27] The Django, https://www.djangoproject.com/.

[28] The Vue, https://v2.cn.vuejs.org/.

[29] The Echarts, https://echarts.apache.org/zh/index.html.