Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 8, No. 4, November, Pp. 773-780

Anomaly detection analysis with graph-based cyber threat hunting scheme

Didit Hari Kuncoro Raharjo[1*], Muhammad Salman[1]
Universitas Indonesia, Indonesia[1]

Article Info
Keywords:
Deep Learning; Artificial Intelligence; Visual Impairment; Visual Aids; Assistive Technology

Article history:
Received: June 29, 2023
Accepted: September11, 2023
Published: November 30, 2023

Cite:
D. H. K. Raharjo and M. Salman, "Anomaly Detection Analysis with Graph-Based Cyber Threat Hunting Scheme", KINETIK, vol. 8, no. 4`, Nov. 2023.
https://doi.org/10.22219/kinetik.v8i4.1773

*Corresponding author.
Didit Hari Kuncoro Raharjo
E-mail address:
ditharikuncoro@gmail.com

Abstract
As advanced persistence threats become more prevalent and cyber-attacks become more severe, cyber defense analysts will be required to exert greater effort to protect their systems. A continuous defense mechanism is needed to ensure no incidents occur in the system, one of which is cyber threat hunting. To prove that cyber threat hunting is important, this research simulated a cyber-attack that has successfully entered the system but was not detected by the IDS device even though it already has relatively updated rules. Based on the simulation result, this research designed a data correlation model implemented in a graph visualization with enrichment on-demand features to help analysts conduct cyber threat hunting with graph visualization to detect cyber-attacks. The data correlation model developed in this research can overcome this gap and increase the percentage of detection that was originally undetected / 0% by IDS, to be detected by more than 45% and can even be assessed to be 100% detected based on the anomaly pattern that was successfully found.

1. Introduction

Cyber-attacks have become more and more prevalent, especially the form of attack that has no information related to attack patterns such as TTP or what is often called a 0-day attack [1]. Even for attacks with advanced levels (using high technical attack methods and the ability to survive in the system), which are currently better known as APT, the pattern / TTP can only be known sometime after launching and, on average it is only known within 90 days [2, 3]. This condition forces the blue team to race against time to ensure that there are no cyber-attacks, either normal level or APT or 0-day attacks on the system. Blue teams must change their mindset from "we're fine until proven otherwise" to "we're owned until proven otherwise" [4]. Blue teams must actively analyze the data traffic entering the system and metadata from endpoints to obtain information and prove whether there is abnormal activity based on what has been determined (baseline). However, the analysis is not easy to do, considering the number of logs and evidence that the blue team must analyze has a large number, and based on research by Cisco in 2012, only 56% of alerts per day can be handled by analysts [5]. On the other hand, the concept of attacking and defending in the cyber world is often analogous in the form of team colors, where the Red Team conducts tests and attacks on the system. In contrast, the Blue Team defends and monitors the system's security. The red team will continue to look for security holes to get into the system, while the blue team will continue to try to analyze attack patterns and methods launched by the red team to defend and secure the system owned [6, 7]. It would be normal and necessary if the blue team tried to have and even create rules (applied to the security perimeter) as comprehensively and as much as possible to detect attack patterns. Indirectly, the syllogism formed is that the more complete the attack pattern detection rules are, the higher the chance of detection percentage that can be done. This means that the approach of creating complex rules is expected to minimize the chance of being exposed to cyber-attacks from APT. However, the limit of the resource performance capability of the security perimeter to perform detection based on the rules that have been applied is a challenge that the blue team must face. Research conducted by Raharjo et al. [8, 9] showed a significant decrease in detection performance of the IDS Suricata device, along with the increase in the number of rules applied. Their research shows a decrease in performance due to the failure of detection rules that can occur with an average value of 19.64%. Even more, when the total number of rules reaches 1,000,000, IDS Suricata fails to detect all types of rules (0%). It means that attacks that are already known and have been defined in the rules but are not successfully detected due to performance failures from detection devices pose just as much of a threat to the system as previously undetected attacks due to unknown information, such as 0-day attacks or APT.

That potential risk suggests that an ongoing defense mechanism is required to ensure that no incidents occur in the system, one of the approaches is to conduct cyber threat hunting activities. Cyber threat hunting is a form of analysis by the blue team to hunt for indications or signals of cyber incidents or attacks on system logs [3, 10]. Threat hunting is carried out according to the new paradigm of cyber defense that thinks paranoia has occurred in cyber incidents, so it

continues striving to prove the initial paranoid hypothesis. Cyber threat hunting can be done by checking proactively or reactively. When done proactively, the blue team will focus on long-tail analysis, which is an analysis method that focuses on data that is not detected by the parameters or rules of the security perimeter because it has not been defined in the rules or there is still no information related to the characteristics of the anomaly [11]. Blue teams intensively analyze anomalies through the long-tail analysis method so that they can update the parameters or rules of the security perimeter based on the anomaly patterns found from the analysis results [10, 12-14]. While reactively, threat hunting is carried out by looking for evidence of an IoC that has been recorded in all logs based on information obtained from the CTI report [5, 15, 16]. One approach that can be taken in conducting the analysis is to utilize graphical visualization charts to perform correlation analysis of various logs [17, 18]. Graph visualization is expected to help the blue team perform correlation analysis or long-tail analysis in the context of cyber threat hunting from logs considered suspicious by analysts.

Several previous studies have discussed research related to correlation analysis using graphs. Diederichsen et al.[19] have created a graph schema using Neo4j from Zeek IDS logs that can generate graph schemas in near real-time based on models defined on network logs. Schindler [20] detected APT using graph visualization by creating abstraction layers and mapping adapted cyber kill chain models based on machine learning. Djanali et al. [21] used clustered graphs from honeypot logs to create IDS signatures. Research related to cyber threat hunting schemes has been conducted by Milajerdi et al. [14] showed a hunting scheme called Poirot, which uses graph correlation analysis of OS kernel logs, and continued by Wei et al. [13] improved Poirot by introducing DeepHunter which is claimed to be more precise. These studies show that the approach using graphs can be used in analyzing the correlation between log data [13, 14, 19, 20, 22, 23]. The results of the correlation analysis between data logs can help conduct threat-hunting activities from indications of cyber threats on the system owned to prove whether a cyber incident has occurred or not [10, 12-14]. One of the evidence parameters is obtained from CTI, which is high-value information containing attack parameters in the form of IoC [15, 16].

With the potential risk of cyber incidents requiring cyber defense mechanisms to detect in greater depth, previous research has shown that graph-based correlation models have helped analysts perform anomaly detection or cyber threat hunting. However, the graphs compiled in earlier studies used all the log data obtained, which requires enormous resources to visualize. Even then, some studies only focus on one of the network logs or host/OS kernel logs, only one of which uses logs from both the network and host sides. To cover these deficiencies, this study proposed graph visualization using a long-tail analysis scheme to reduce logs (both network and host) that do not need to be visualized and enhanced capabilities for on-demand information enrichment based on IoC to help analysts filter out anomaly information and concentrate on logs that may become potential cybersecurity incidents.

2. Research Method

2.1 Design of Cyber Threat Hunting Scheme

This research simulated a cyber-threat hunting scheme to develop a graph visualization model for anomaly analysis. The schema adapted the data flow analysis process using the HeteMSD framework created by Angkang Ju et al. Al, which was used to perform big data analysis in conducting targeted cyber-attack detection using a variety of heterogeneous data sources [24], as illustrated in Figure 1, which was divided into 4 phases:

a. Data Processing
   It is a phase of collecting and processing parsing log data from NIDS Suricata and Zeek, HIDS Wazuh, Winlogbeat, and Sysmon contained on the SIEM server (Elastic Stack) so that it can be processed for further analysis. Elastic Stack is a platform capable of performing big data processes, including collecting, processing, and using [25].
b. Data Analysis
   It is a long-tail analysis phase by filtering data from events considered dangerous by IDS rules. In a security scheme, log data that falls into the alert category (based on rules) will be followed up as a cyber incident response. However, in the cyber threat hunting scheme, analysts will search for anomalies from event not detected by IDS rules. The threat-hunting process can be run reactively or proactively. Reactively indicates that the analyst will look for evidence of a cyber incident on the system (in the SIEM) based on the IoC parameters provided by CTI. While proactively looking for anomalies that have the potential to become 0-day attacks, which are unlikely to be available from CTI. In this study, the proactive analysis mechanism uses a log correlation model using graph visualization.
c. Data Correlation
   This is the supporting phase of long-tail analysis, where a correlation model is developed from the logs. The data from the long-tail analysis will be modeled using the concept of a graph to create a connection between data fields/nodes. The correlation modeling results are implemented in the Maltego application to visualize the correlation model with graphs. Maltego is an application more often used by red teams to carry out information-gathering activities from targets through open-source intelligence techniques and then find interrelationships between data and visualize these connections [26]. This research utilized Maltego's capabilities to analyze and enrich the value of nodes (IoCs) by making API calls from Maltego to the CTI repository. With this scheme, the enrichment process is on-demand, making it more effective without having to download and store all IoC databases on the system, which

is likely to affect resources during storage and processing, considering that the IoC sources from CTI can amount to millions of data (There are more than 110 billion total IoCs from CTI IntelligenceX (https://intelx.io/) and more than 89 million total IoCs from CTI AlienVault (https://otx.alienvault.com); Accessed on June 29, 2023).

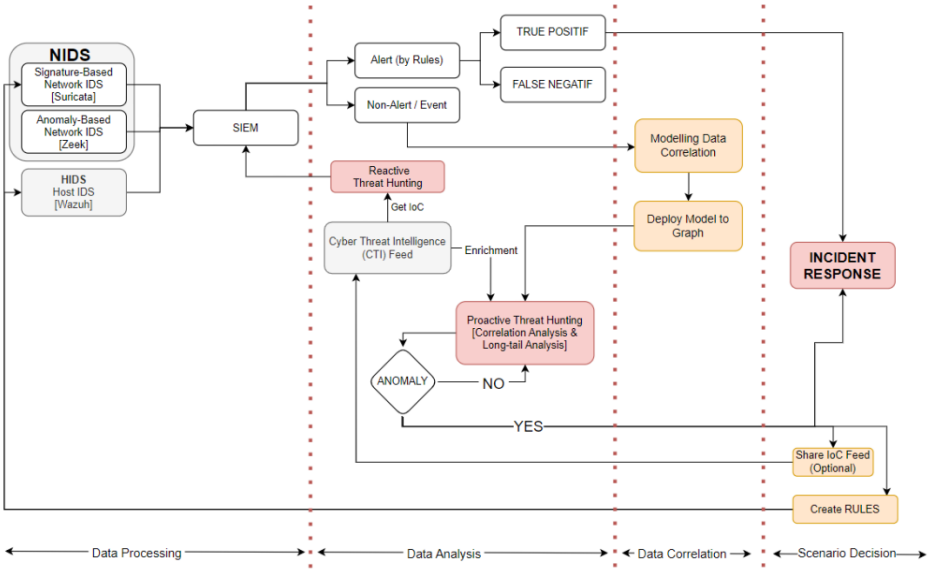

Figure 1. Design of Cyber Threat Hunting Scheme

d. *Scenario Decision*

When a system is identified with an incident, an incident response must be carried out immediately to counteract or minimize the impact of threats on the system. In incident response activities, several stages and procedures are carried out, including triage, containment, system backup, forensics, eradication, and lessons learned [27]. In the scheme in Figure 1, the lesson learned is to create new rules for the security perimeter based on the findings and share them (optionally) with CTI so that other entities can consume them.

2.2 Design of Cyber Threat Hunting Simulation Scenario

The scenario begins with a cyber-attack from the attacker host to the client host. From the network side, all data traffic will be duplicated (mirroring) or activated promiscuously on the virtualization system so NIDS can inspect. Client hosts are also installed with HIDS and log collectors. All logs from NIDS, HIDS, and log collectors will be sent to the SIEM device for data processing. Host management will access the SIEM to perform data processing and run graph correlation modeling to perform analysis. Attack simulation will use the CobaltStrike tool, an attack simulation tool for red team operation. CobaltStrike is widely used by threat actors in creating APTs, with the number of uses increasing by 161% from 2019 to 2020 [28]. Social engineering techniques such as spear-phishing are still a trend used by APTs to steal data, such as Skypot APT, Ghostnet, APT-29, and APT-37 [29]. Based on this fact, this research will design a simple attack simulation (not as complex as APT) involving spear-phishing, remote shell, and data theft techniques, as illustrated in Figure 2 below:

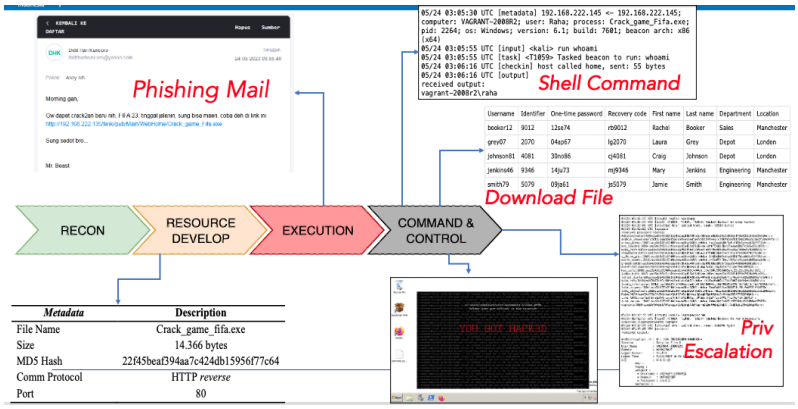

Figure 2. Illustration of Cyber Threat Hunting Simulation Attack Scenario

2.3 Design of Correlation Model

Without information related to the timeline of cyber-attacks carried out by attackers, the blue team must be able to analyze and reconstruct whether a cyber-incident occurred or not based on the data and logs owned. The logs used in this research included network logs (Suricata, Zeek) and host logs (Winlogbeat, Sysmon). Before implementing graph visualization, data correlation modeling must be done first (it is a data correlation process from the cyber threat hunting scheme flow in Figure 1). The focus of network logs whose correlation analyzed are logs generated from Zeek but not detected as alerts by Suricata (filter based on the community_id field). At the same time, the focus of the host logs that were analyzed for correlation was Winlogbeat logs that had been integrated with Sysmon but were not detected by the Wazuh alert log. In network logs, each log has a unique session_id data, a link from each other node, such as HTTP (Domain URL, Original URL), IP (source IP, Destination IP), DNS accessed (DNS Query), and Files transferred over the network (MD5 hash value), and SSL (TLS Client JA3, TLS Server JA3). In the host log, correlation is performed based on event_id data, the link of each other nodes, such as the User who is running the activity, PID, and PPID (for conditions where an event runs on the trigger of another PID), DataImage (refers to the name of the process or application being run), FileName (refers to whether there is a file name generated), and CommandLine (refers to whether there is a terminal command being executed). Each link formed between nodes will affect the weight of the nodes (represented in the form of node size), both incoming and outgoing. The correlation model that has been created is illustrated in Figure 3, which connects the nodes involved:

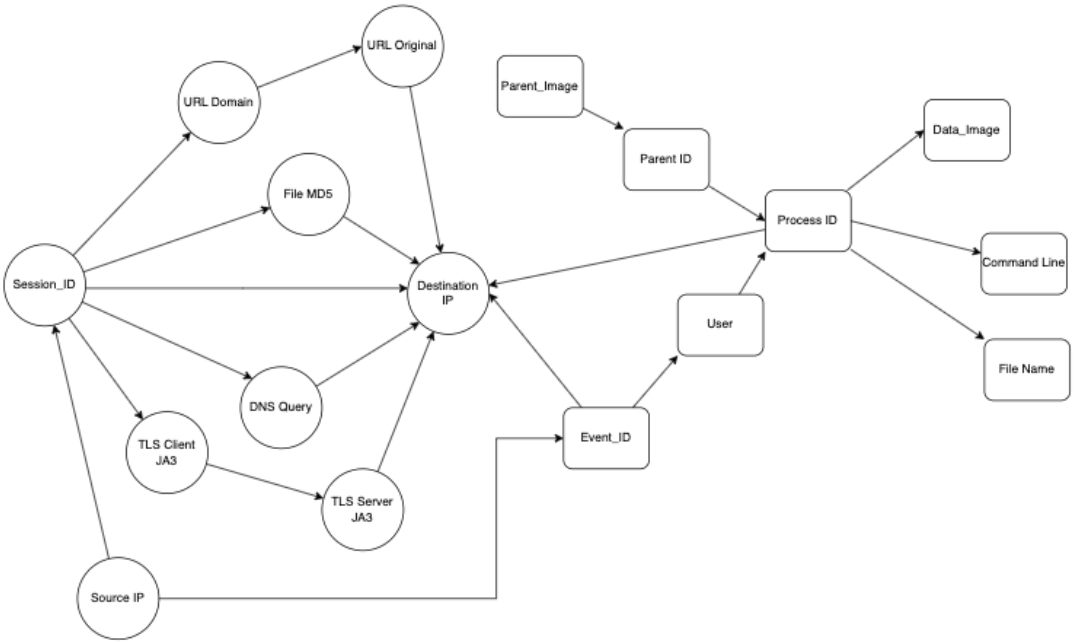

Figure 3. Correlation Model Design

Figure 3 shows the correlation integration of host and network logs. It can be seen that DestinationIP (the IP of the host that is the object of monitoring) is the connecting node of the two model designs. The integration of the models is expected to make it easier for analysts to understand the information chain between the network and the host. The design of the correlation model that has been made will be the basis for mapping nodes and configuring relations in the implementation of graph visualization in the Maltego application. The implementation process will be done in stages to make the analysis more focused and structured. If all data is directly implemented simultaneously, the graph visualization formed will be very large and difficult to map the correlation, as well as consume high enough resources from the computer to perform the visualization process. A flowchart diagram is created to facilitate this process. Figure 4 shows the looping process of graph implementation based on data types, which will continue until the visualized graph network shape meets analysts' expectations in conducting the cyber threat-hunting process. That is why the on-demand scheme is used, because if all logs are enriched at once, then in addition to burdening device resources, it also makes it difficult for analysts because the visualization network will be more complicated. Analysts will hunt for cyber security incident indicators, enrich information through CTI, but are limited to logs that are considered suspicious. Analysts will carry out an elimination process on data (in the form of nodes) that are considered normal and not affiliated with threats from CTI results.

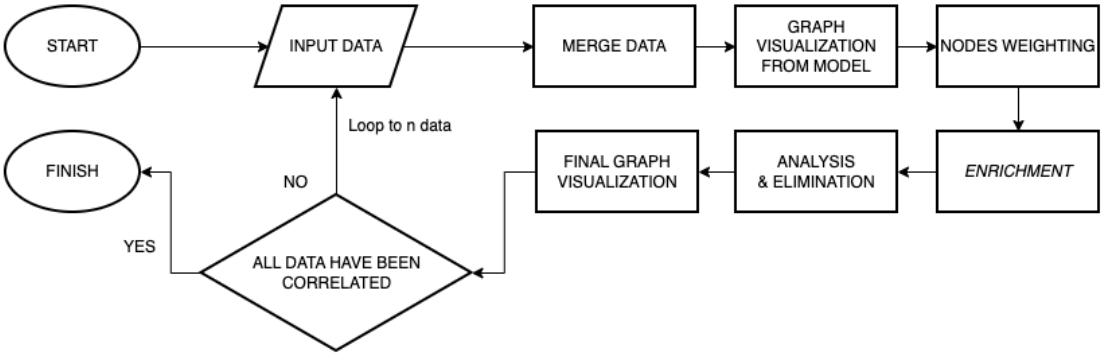

Figure 4. Graph Visualization Implementation Flowchart

3. Result and Discussion

By running the flowchart in Figure 4 based on data input in the form of logs based on the correlation model (Figure 3), the graph visualization is obtained as illustrated in Figure 5 below:

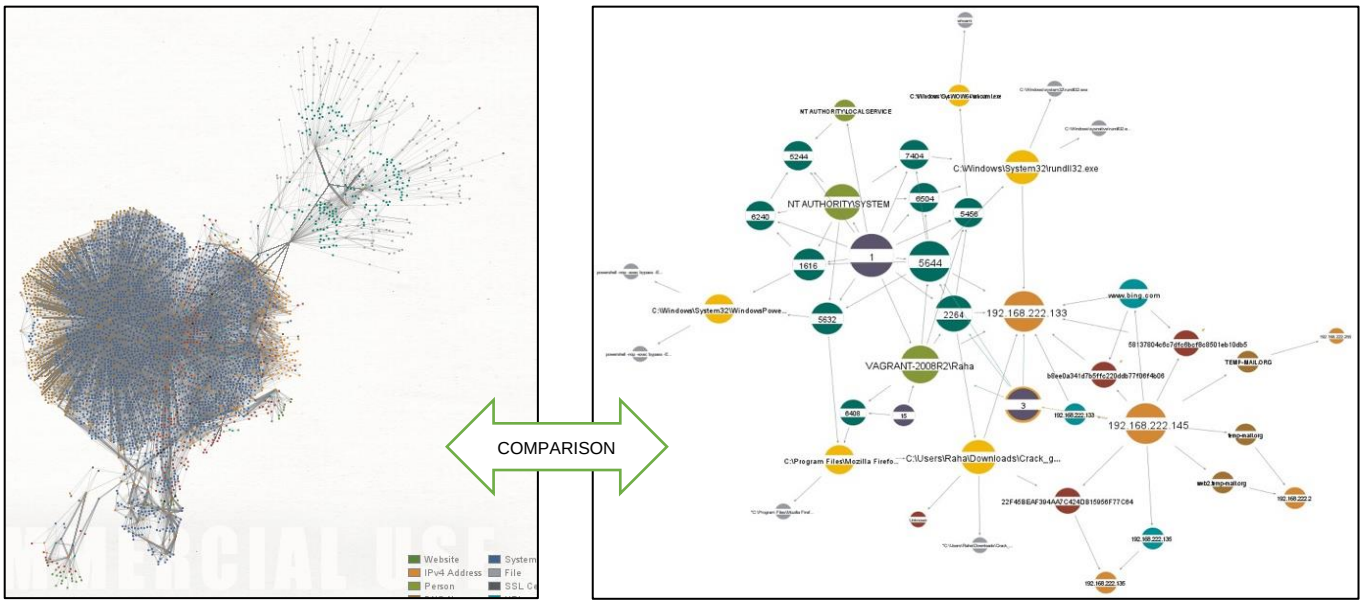

Figure 5. Comparison between Graph Visualization Implementation, without (Left) vs. With (Right) Flowchart Figure 4

Figure 5 clearly shows that the analyst needs to implement the visualization gradually according to the flowchart diagram. In the left part of Figure 5, it will be difficult for analysts to sort out and understand the shape of the graph compared to the right part. It can be seen that there is a significant elimination process of the nodes assessed by the analyst (based on the knowledge of the analyst and the help of IoC enrichment information from threat intelligence), resulting in a more structured graph visualization.

There will be a considerable gap compared to the detection results by IDS on simulation attack scenario. IDS did not generate any alert notifications of cyber incidents related to the CnC activity within the attack timeline. The five IDS alert notifications that occurred during the attack timeline (2 from Wazuh and 3 from Suricata) did not indicate a causal relationship to the CnC activity, which was only in the form of name resolution, DNS Query, and detection of Kali Linux devices (which are often used as the red team OS). The comparison between the CnC activity, alert detection by IDS (Suricata and Wazuh), and detection from cyber threat hunting using graph visualization will be shown in Table 1:

Table 1. Comparison of CnC Activity, IDS Detection and Detection from Graph Analysis

No	CnC Activity	Suricata Alert	Wazuh Alert	Detection from Graph Analysis
1	[Crack_game_Fifa.exe] initial beacon	X	X	Identified from IP relation 192.168.222.133 and eventID 3
2	run: whoami	X	X	Identified from PID and shell relation

No	CnC Activity	Suricata Alert	Wazuh Alert	Detection from Graph Analysis
3	Task beacon explore process list	X	X	Identified from the HTTP relation. Encoded URL to BING domain (with a suspicious IP address), assumed to be CnC activity.
4	Task beacon explore file folder	X	X	Identified from HTTP relation. Encoded URL to BING domain (with a suspicious IP address), assumed to be CnC activity.
5	download C:\Users\Raha\Documents\username-password-recovery-code.csv	X	X	Identified from HTTP relation. Encoded URL to BING domain (with a suspicious IP address), assumed to be CnC activity.
6	Task Beacon to run windows/beacon_http/reverse_http(192.168.222.133:80) via ms15-051	X	X	Identified from the PID, ParentPID, user, and shell relation. Privilege escalation from user Raha to System.
7	dump hashes	X	X	Identified from the connection of rundll32.exe to IP 192.168.222.133 by user SYSTEM.
8	run mimikatz's sekurlsa::logonpasswords command	X	X	Identified from the connection of rundll32.exe to IP 192.168.222.133 by user SYSTEM
9	run: wget "http://192.168.222.135/twiki/pub/Main/WebHome/pict.jpg" -OutFIle "C:\Windows\Temp\pict.jpg"	X	X	Identified from the relation PID, user, and shell. Decode PowerShell command
10	import: /home/kali/Downloads/Cobalt Strike4.1/x.ps1	X	X	Identified from the connection rundll32.exe to IP 192.168.222.133 by user SYSTEM
11	run: Set-WallPaper -Image "C:\Windows\Temp\pict.jpg"	X	X	Identified from the relation PID, user, and shell. Decode the Powershell command

From Table 1, five forms of attack (numbers 1, 2, 6, 9, and 11) can be identified in the graph visualization. As for the six forms of attack (numbers 3,4,5,7,8,10 in gray), these attacks are identified but not clearly visible in the log/graph due to encoding constraints (rundll32.exe and URL www.bing.com) which cannot be decoded in this study. However, from the pattern formed, the identification assumption is likely correct. In real life, the blue team will not have timeline data from attacks by attackers/threat actors (CNC activity). Instead, the blue team is required to be able to analyze, review, understand, and anticipate the conditions of the incident or attack that occurred. Referring to the case of this attack simulation, it can be stated that the IDS security perimeter fails to detect seizures, even though it is equipped with relatively updated rules (community / free version; updated on May 2023). If there is no in-depth analysis or cyber threat hunting, then the IDS do not detect this attack, and the blue team will assume the system is in good condition, even though there has been an intrusion, data theft, and changes to the system.

4. Conclusion

This research succeeded in developing a correlation model to perform graph visualization that can help blue team analysts to parse and recreate the existence of cyber incidents in the system in the simulation conducted, where in the previous simulation, cyber-attacks were not detected by IDS, both Suricata and Wazuh. With the long-tail analysis scheme, the visualized graph focuses on logs not detected (not triggered alerts) by IDS devices from both the network and host sides. By involving enrichment and elimination processes, the final graph visualization can clearly illustrate that a cyber-incident has occurred in the system. From the results of the graph visualization analysis, the opening point of the incident in the simulation is when the host downloads and runs the executable file and connects to suspicious IP, which is disguised through the domain www.bing.com. From this starting point, it is indicated that there is a CnC process carried out by the attacker via the HTTP protocol with a total of 514 variants of access to www.bing.com which contains the execution of javascript files, including remote commands, privilege escalation, and data download. After the cyber threat hunting process has successfully identified a cyber-incident, the next step is to carry out an incident response, including the containment of the affected host and eradication of threat evidence. The next step is to create rules based on the identified TTP / IoC that can be applied to perimeter security. The correlation model from this research can be further developed by involving other types of data to be more comprehensive and follow the needs of the blue team

analyst team, who needs an overview of the relationship of the data owned. The simulation of cyber-attacks carried out in this research is still simple or not as detailed and structured as APT, so it is necessary to test graph visualization analysis on other forms of cyber-attacks, as well as an effort to improve and learn.

Notation

The following is a description of the notation used:

API	: Application Programming Interface
APT	: Advance Persistence Thread
CnC	: Command and Control
CTI	: Cyber Threat Intelligence
DNS	: Domain Name System
HeteMSD	: Heterogeneous Multi-Source Data
HIDS	: Host Intrusion Detection System
HTTP	: Hypertext Transfer Protocol
IDS	: Intrusion Detection System
IoC	: Indicator of Compromise
IP	: Internet Protocol
JA3	: A method for creating SSL/TLS client fingerprints, stands for the creator John Althouse, Jeff Atkinson, Josh Atkins
MD5	: Message Digest Algorithm (Hash Function)
NIDS	: Network Intrusion Detection System
OS	: Operating System
PID	: Process Identifier
PPID	: Parent Process Identifier
SIEM	: Security Information and Event Management
SSL	: Secure Socket Layer
TLS	: Transport Layer Security
TTP	: Tactics, Techniques, and Procedures
URL	: Uniform Resource Locator

References

[1] J. Song, H. Takakura, and Y. Kwon, "A generalized feature extraction scheme to detect 0-day attacks via IDS alerts," in *2008 International Symposium on Applications and the Internet*, 2008: IEEE, pp. 55-61. https://doi.org/10.1109/SAINT.2008.85

[2] M. Li, W. Huang, Y. Wang, W. Fan, and J. Li, "The study of APT attack stage model," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 2016: IEEE, pp. 1-5. https://doi.org/10.1109/ICIS.2016.7550947

[3] V. Palacin, *Practical Threat Intelligence and Data-Driven Threat Hunting: A Hands-on Guide to Threat Hunting with the ATT&CK"! Framework and Open Source Tools*. Packt Publishing Limited, 2021.

[4] R. Bhargava. "How threat hunting enhances cybersecurity -- GCN.".

[5] C. Pace, "The threat intelligence handbook: A practical guide for security teams to unlocking the power of intelligence," *Annapolis, CyberEdge Group,* 2018.

[6] M. Richter, K. Schwarz, and R. Creutzburg, "Conception and Implementation of Professional Laboratory Exercises in the field of ICS/SCADA Security Part II: Red Teaming and Blue Teaming," *Electronic imaging,* vol. 2021, no. 3, pp. 74-1-74-13, 2021. https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-074

[7] J. Rehberger, "Cybersecurity Attacks–Red Team Strategies," ed: Packt Publishing, 2020.

[8] D. H. K. Raharjo, A. Nurmala, R. D. Pambudi, and R. F. Sari, "Performance Evaluation of Intrusion Detection System Performance for Traffic Anomaly Detection Based on Active IP Reputation Rules," in *2022 3rd International Conference on Electrical Engineering and Informatics (ICon EEI)*, 2022: IEEE, pp. 75-79. https://doi.org/10.1109/IConEEI55709.2022.9972298

[9] D. H. K. Raharjo and S. Muhammad, "ANALYZING SURICATA ALERT DETECTION PERFORMANCE ISSUES BASED ON ACTIVE INDICATOR OF COMPROMISE RULES," *Jurnal Teknik Informatika (Jutif),* vol. 4, no. 3, pp. 601-610, 06/26 2023. https://doi.org/10.52436/1.jutif.2023.4.3.1013

[10] P. Gao *et al.*, "Enabling efficient cyber threat hunting with cyber threat intelligence," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2021: IEEE, pp. 193-204. https://doi.org/10.1109/ICDE51399.2021.00024

[11] F. Skopik, M. Wurzenberger, and M. Landauer, "The Seven Golden Principles of Effective Anomaly-Based Intrusion Detection," *IEEE Security & Privacy,* vol. 19, no. 05, pp. 36-45, 2021. https://doi.ieeecomputersociety.org/10.1109/MSEC.2021.3090444

[12] A. Oktadika, C. Lim, and K. Erlangga, "Hunting Cyber Threats in the Enterprise Using Network Defense Log," in *2021 9th International Conference on Information and Communication Technology (ICoICT)*, 2021: IEEE, pp. 528-533. https://doi.org/10.1109/ICoICT52021.2021.9527434

[13] R. Wei, L. Cai, A. Yu, and D. Meng, "DeepHunter: A Graph Neural Network Based Approach for Robust Cyber Threat Hunting," *arXiv preprint arXiv:2104.09806,* 2021. https://doi.org/10.1007/978-3-030-90019-9_1

[14] S. M. Milajerdi, B. Eshete, R. Gjomemo, and V. Venkatakrishnan, "Poirot: Aligning attack behavior with kernel audit records for cyber threat hunting," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 1795-1812, doi: https://doi.org/10.1145/3319535.3363217.

[15] M. Bertovič, "Utilization of Threat Intelligence in Information Security," Computing and Information Center, Czech Technical University in Prague., 2017.

[16] H. Almohannadi, I. Awan, J. Al Hamar, A. Cullen, J. P. Disso, and L. Armitage, "Cyber threat intelligence from honeypot data using elasticsearch," in *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, 2018: IEEE, pp. 900-906. https://doi.org/10.1109/AINA.2018.00132
[17] T. Freeman *et al.*, "Graphia: A platform for the graph-based visualisation and analysis of complex data," *bioRxiv,* 2020. https://doi.org/10.1371/journal.pcbi.1010310
[18] I. Robinson, J. Webber, and E. Eifrem, *Graph databases: new opportunities for connected data*. " O'Reilly Media, Inc.", 2015.
[19] L. Diederichsen, K.-K. R. Choo, and N.-A. Le-Khac, "A graph database-based approach to analyze network log files," 2019: Springer, pp. 53-73. https://doi.org/10.1007/978-3-030-36938-5_4
[20] T. Schindler, "Anomaly detection in log data using graph databases and machine learning to defend advanced persistent threats," *arXiv preprint arXiv:1802.00259,* 2018. https://doi.org/10.48550/arXiv.1802.00259
[21] S. Djanali, A. P. BASKORO, H. Studiawan, R. Anggoro, and T. Henning, "CORO: GRAPH-BASED AUTOMATIC INTRUSION DETECTION SYSTEM SIGNATURE GENERATOR FOR E-VOTING PROTECTION," *Journal of Theoretical & Applied Information Technology,* vol. 81, no. 3, 2015.
[22] P. Neise, "Graph-based event correlation for network security defense," The George Washington University, 2018.
[23] R. Tan, "Zeek Log Recon with Network Graph," *SANS Whitepaper,* 2020.
[24] A. Ju, Y. Guo, Z. Ye, T. Li, and J. Ma, "Hetemsd: A big data analytics framework for targeted cyber-attacks detection using heterogeneous multisource data," *Security and Communication Networks,* vol. 2019, 2019. https://doi.org/10.1155/2019/5483918
[25] A. Talaş, F. Pop, and G. Neagu, "Elastic stack in action for smart cities: Making sense of big data," in *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2017: IEEE, pp. 469-476. https://doi.org/10.1109/ICCP.2017.8117049
[26] K. Schwarz and R. Creutzburg, "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools-Part 3: Maltego," *Electronic Imaging,* vol. 2021, no. 3, pp. 45-1-45-23, 2021. https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-045
[27] P. Kral, "The incident handlers handbook," *Sans Institute,* 2011.
[28] S. Larson and D. Blackford. "Cobalt Strike: Favorite Tool from APT to Crimeware | Proofpoint US." @proofpoint.
[29] H. J. Hejase, H. F. Fayyad-Kazan, and I. Moukadem, "Advanced persistent threats (apt): an awareness review," *Journal of Economics and Economic Education Research,* vol. 21, no. 6, pp. 1-8, 2020.