WIPP - Towards a Delay-Sensitive Provenance Framework for Operational Technology Networks

Christopher Porras
Department of Computer Science
University of Texas at El Paso
Email: ceporras@miners.utep.edu

Deepak Tosh
Department of Computer Science
University of Texas at El Paso
Email: dktosh@utep.edu

2024 Resilience Week (RWS) | 979-8-3503-8898-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/RWS62797.2024.10799471

***Abstract*—In the rapidly advancing age of cybersecurity, detecting and mitigating advanced cyber threats is a critical challenge. This work-in-progress paper (WIPP) explores the use of low-level system log auditing for effective threat detection, focusing on the innovative use of provenance-based techniques. We focus our attention on three advanced systems —ShadeWatcher, SPADE, and UNICORN, which leverage data provenance to enhance cyber-threat detection capabilities. By analyzing their methods and integrating their strengths, we propose a delay-sensitive audit tracking system suitable for critical infrastructure networks. Our system aims to detect anomalous activities with high precision and reduce false alarms within a guaranteed time bound, thus providing robust protection against cyber threats.**

***Index Terms*—Industrial control systems; operational technology; provenance; cyber-threat detection;**

I. Introduction

In the field of critical infrastructures, such as power grids, water supply systems, and transportation networks, maintaining the security and reliability of control systems is essential. These systems, which form the basis of modern societies, are increasingly susceptible to cyber threats. The growing complexity of cyber threats necessitates advanced techniques for threat detection and response. Traditional approaches often fall short due to the high false alarm rates and the need for extensive expertise. In the case of power grids, for instance, the inability of traditional Intrusion Detection Systems (IDS) to accurately filter out benign from malicious activities often results in a deluge of false positives. This not only overwhelms the monitoring teams but also diminishes their ability to respond to actual threats swiftly. Such high false alarm rates are indicative of a fundamental limitation in traditional systems: they are often rule-based and lack the sophistication to adapt to the nuanced and dynamic nature of cyber threats, leading to potential delays in identifying and addressing real security breaches. Moreover, the 2015 cyberattack on Ukraine's power grid [1] starkly illustrates how the absence of specialized knowledge can cripple the effectiveness of conventional security methods. Advanced cyber threats often manifest through subtle, complex patterns that require a deep understanding of both cybersecurity and the operational specifics of the targeted systems. However, the scarcity of personnel with such dual expertise means that critical indicators of attacks can go unnoticed until irreparable damage occurs. This reliance on specialized knowledge creates a bottleneck in the timely and effective response to threats. This underscores the pressing need for more automated, adaptable, and context-aware security solutions, such as provenance-based methods [2], [3], which can enhance threat detection and mitigation capabilities without the prohibitive requirement for extensive human expertise.

In critical infrastructure networks, provenance plays a vital role in threat mitigation by providing a granular and contextual view of system activities [4]. By examining the lineage of data flows and system operations, security systems can more accurately pinpoint the source and nature of anomalies, distinguishing between benign operational variances and actual malicious activities. This capability allows for a more proactive approach to threat detection, as potential issues can be identified and addressed before they escalate into significant breaches. However, implementing provenance-based methods in critical infrastructure networks is challenging due to the sheer scale and complexity of these systems [5]. The vast amounts of data generated, coupled with the intricate and often proprietary nature of control systems, make comprehensive provenance tracking and analysis a daunting task. Additionally, ensuring that provenance data itself is secure and tamper-proof adds another layer of complexity to the deployment of these methods in critical infrastructures.

Current practices in logging and auditing within industrial networks often involve collecting system logs and event data to monitor and analyze activities across the network [6], [7]. These practices are essential for understanding system behavior, diagnosing problems, and responding to incidents. However, traditional logging and auditing techniques typically focus on capturing snapshots of system states or sequences of events without fully contextualizing the relationships and causality between them. This can lead to delayed or incomplete threat detection, as crucial context and the interconnectedness of events are missing. Delay-sensitive provenance addresses this need by ensuring that comprehensive and context-rich data is available even in environments where immediate or continuous data transmission is not feasible. In industrial networks, where real-time data processing might be constrained due to bandwidth, computational, or operational limitations, delay-sensitive provenance allows for the

† This work is support by National Science Foundation, Award# 2239609.

collection and analysis of detailed event histories at a later time. This approach enhances the ability to understand and respond to security threats effectively, providing a critical advantage in maintaining the integrity and reliability of critical infrastructure systems. This paper explores the implementation of delay-sensitive provenance-based methods for low-level system log auditing to enhance threat detection and mitigation in operational technology networks within critical infrastructure systems.

Low-level system log auditing [8] involves the thorough collection and analysis of log data generated by various control systems within a critical infrastructure's network. These logs provide an overview of system activities, including user actions, system events, and interactions between different system components. By analyzing these interactions, we can identify anomalous behaviors indicative of cyber threats. The primary objective of utilizing a provenance-based approach is to improve the accuracy and granularity of threat detection. Unlike traditional methods that utilize predefined signatures or rules, provenance analysis explores the relationships and interactions within the system to detect anomalies, providing a dynamic and adaptable defense mechanism. By understanding the provenance of each data point or action, security measures can more accurately identify and respond to anomalies that may indicate a threat. This approach is particularly beneficial in control systems, where the integrity and traceability of data are crucial.

In the context of critical infrastructure, where the reliability and security of industrial control systems are paramount [9], [10], incorporating low-level system audit logs and provenance-based approaches into the security framework of control systems is essential for bolstering their resilience against cyber threats. These methods enhance the ability to detect and respond to anomalous activities, traceability, and accountability, and ensure compliance with security regulations. As cyber threats continue to evolve, the integration of these advanced techniques will be crucial in safeguarding the critical systems that support modern society.

In this paper, we aim to create a flexible provenance platform that is adaptable in operational technology (OT) networks for tracking and analyzing low-level network activities from supervisory servers. By establishing a network of nodes that audit low-level system calls and transmitting this data to a collection/analysis server, we can visualize and examine the data in detail. Leveraging insights from prior works [11], [12], [13], we propose a delay-sensitive tracking system that analyzes logs from each control system within the network to detect anomalous activities in a time threshold. By continuously updating and refining our analysis, we can provide accurate and reliable threat assessments even in limited time bounds. Our system's objective is to integrate dynamic threat detection with high-precision anomaly identification and to establish metrics that provide timely responses. This will enable operators to act swiftly and confidently in mitigating threats, thus maintaining the integrity and resilience of critical infrastructure networks. While still under development, our work-in-progress paper (WIPP) promises significant advancements in both online and offline threat detection and the overall security of vital OT environment.

The paper is organized as follows. We provide some background on three important provenance frameworks from past literature in Section II. The Section III provides a high level design on the delay aware provenance system for OT environment with the help of different tools for log analysis and visualization. In Section IV, we present some preliminary results to validate the provenance data collection process of different network activities. Finally, the Section V and VI provide concluding remarks and future research directions to be taken to further this work.

II. Background

To contextualize our proposed architecture, we provide background on three advanced systems that leverage provenance for threat detection: ShadeWatcher [11], SPADE [12], and UNICORN [13]. Each system offers unique insights and methodologies that inform our proposed solution.

A. ShadeWatcher: Recommendation Guided Cyber Threat Analysis

This work utilizes recommendation systems to improve cyber threat analysis by establishing a mapping between the concepts of threat detection in cybersecurity and Deep recommendation in information retrieval. The system treats system entity interactions as user-item interactions in a recommendation context [11]. By doing so, It can predict the preferences of a system entity on its interactive entities, thereby identifying cyber threats with high precision and recall rates. It employs a recommendation model built upon graph neural networks (GNNs) to exploit high-order connectivity by recursively propagating information from neighboring entities. This allows the system to capture both first-order and high-order information from audit records, which is crucial for understanding the semantics of system entity relationships and detecting threats effectively.

Furthermore, it dynamically updates its models with analyst's feedback on detection signals, integrating false recommendations as additional supervision to improve its detection capabilities. This semi-supervised approach enables the system to learn from both unlabeled benign system entity interactions and labeled analyst feedback on false alarms, enhancing its generalization, and reducing false alarm rates. Its innovative use of recommendation systems provides an end-to-end solution for detecting threats without prior knowledge of attacks and produces fine-grained detection signals that highlight key indicators of an attack. The system's effectiveness and efficiency are demonstrated through systematic evaluation against both real-life and simulated cyber-attack scenarios, showing high accuracy and low false alarm rates.

B. SPADE: Open-Source Provenance Management

SPADE (Support for Provenance Auditing in Distributed Environments) [12] represents a robust open-source infrastructure tailored for comprehensive management, querying,

and auditing of data provenance in distributed computing settings. Rooted in a graph-based data model aligned with the Open Provenance Model, SPADE captures provenance metadata through vertices and directed edges, facilitating the representation of data lineage and dependencies across distributed systems. At its core lies a versatile provenance kernel designed to manage the storage, querying, and distribution of provenance data, supporting a spectrum of query operations essential for auditing and verifying data integrity. Data provenance collection in distributed environments presents numerous challenges, all of which the SPADE system endeavors to tackle comprehensively. Chief among these challenges is the sheer volume of data generated and processed within distributed systems, necessitating efficient methods for collecting and managing extensive provenance information. Additionally, ensuring the security and privacy of provenance data [14] becomes paramount, given its potential sensitivity regarding data sources, processes, and user information traversing multiple trust domains.

Moreover, the complexity of relationships between data elements, processes, and agents across diverse locations and administrative domains poses a significant hurdle, requiring nuanced approaches to accurately capture and represent these intricate dependencies. Another critical issue is the heterogeneity inherent in distributed systems, where varying technologies, protocols, and data formats further complicate the uniform collection of provenance data. This diversity not only affects data collection but also impacts the system's performance, introducing potential latency and overhead concerns, particularly under heavy operational loads.

The SPADE addresses these multifaceted challenges by employing a robust, graph-based data model designed to manage provenance information flexibly and effectively. This model supports complex relationships and dependencies, adheres to the Open Provenance Model (OPM) [15], and decouples provenance collection, storage, and querying processes. SPADE's architecture accommodates various query operators and storage backends, enhancing flexibility and scalability. Moreover, the system incorporates a filtering mechanism to optimize metadata management and improve query precision, ensuring efficient handling of large volumes of provenance data over time. These features collectively empower SPADE to serve as a powerful tool for tracking, managing, and querying data lineage in distributed environments, thereby facilitating informed decision-making and analysis across diverse application domains.

C. UNICORN: Anomaly-Based APT Detection

The main goal of UNICORN [13] is to effectively detect Advanced Persistent Threats (APTs) by leveraging data provenance analysis. It is designed to identify stealthy anomalous activities within a system without relying on pre-defined attack signatures. It achieves this by efficiently analyzing provenance graphs over time and space, focusing on the entire span of system operations. The system learns and adapts to typical changes in normal system behavior, allowing it to distinguish these from anomalies caused by attackers. It employs a graph sketching technique to summarize long-running system activities, making it adept at detecting slow-acting attacks. Key components include a real-time updating streaming graph histogram, a fixed-size graph sketch for preserving histogram similarities, and an evolutionary model that captures normal system behavior changes.

D. Limitations of Existing Systems

Despite their strengths, these systems fall short in addressing a critical need for delay-sensitive threat detection, which is essential for rapid response in dynamic and critical environments.

ShadeWatcher: ShadeWatcher's design is centered around achieving high accuracy and fine-grained detection signals without a specific focus on the rapid identification and response to emerging threats. This lack of emphasis on timeliness is a notable gap, especially in scenarios where immediate detection and response are critical.

SPADE: Despite its robust management capabilities, it is primarily designed for detailed provenance auditing and querying rather than the prompt detection of threats. The system's architecture can handle large volumes of data, but this capability comes at the expense of potential latency. This makes SPADE less effective for scenarios requiring immediate threat identification and mitigation.

UNICORN: While UNICORN excels in identifying sophisticated, prolonged anomalies, its emphasis on long-term behavior analysis may delay the detection of threats that require swift response times. This approach may not be sufficient for environments where rapid detection and immediate action are crucial to maintaining system security and integrity.

Addressing the Gap with Delay-Sensitive Provenance: These systems highlight important advancements in threat detection through provenance analysis, yet they do not fully address the need for rapid and precise threat detection within short time frames. Our delay-sensitive provenance system addresses this gap by focusing on swift analysis and response to threats. By prioritizing timely detection, our approach enables immediate identification of anomalous activities, allowing for quick and decisive actions to mitigate potential threats. This capability is crucial for enhancing the security and resilience of critical infrastructure networks against imminent attacks, an aspect that is currently underrepresented in systems like ShadeWatcher, SPADE, and UNICORN.

III. Towards a Delay-Aware Provenance System for OT Networks

In this Section, we outline an overview of our environment setup for the system in Figure 1. The architecture of the proposed audit tracking system is designed to ensure robust monitoring and detection of potentially malicious activities within critical infrastructure networks. The system utilizes elements of the ELK stack, comprising of Elasticsearch, Logstash, and Kibana, integrated with Auditbeat to collect, analyze, and visualize system-level logs.

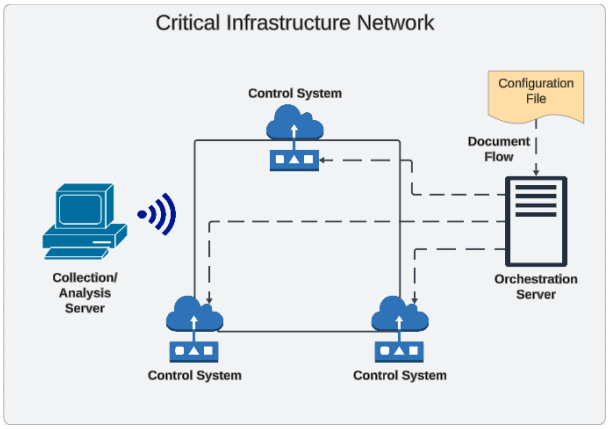

Fig. 1. Overview of Provenance Environment for OT

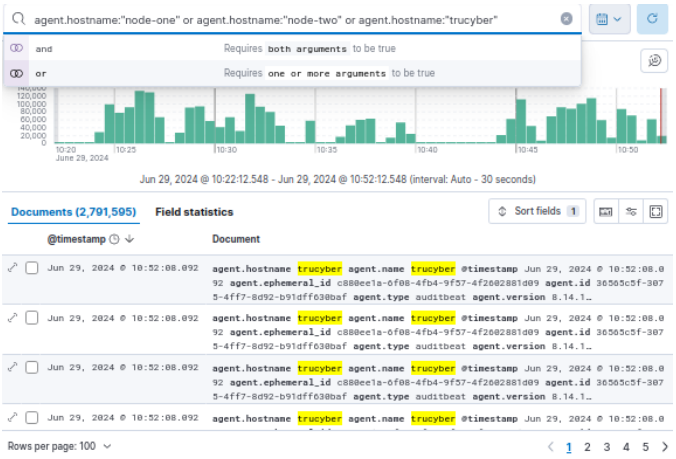

Fig. 2. 30 Minute Event Capture For Preliminary Testing Environment

A. Preliminary Testing Environment

In Figure 1, we have established the foundation for our preliminary provenance collection environment consisting of three supervisory control systems servers, a collection/analysis server, and an orchestration server. Our control systems perform critical functions within the critical infrastructure network. All the control systems are connected and can communicate with one another within the network, but each operates independently and is able to audit low-level system data logs. As the auditing process occurs, each control system sends these low-level system data logs to an off-location collection/analysis server. There the server can provide log aggregation and custom fine-grained analysis.

We have included an orchestration server whose primary purpose is to provide information on the auditing configuration for the network. This configuration file will be sent to the orchestration server, and it will distribute this configuration to the appropriate control systems in the network. The configuration files on each node define security rules and tracking markers tailored to the specific functionalities of the control systems. These rules establish baselines for "normal" system behavior. Deviations from these baselines generate flags that are tagged in the logs with keywords identifying potentially malicious activities.

B. Network Setup and Node Configuration

In the preliminary setup, we have three Virtual Machines (VMs) running Ubuntu 20.04 that have been deployed. These VMs act as our control systems that would be in critical infrastructure networks. One of these VMs acts as both a control system and the collection/analysis server in this preliminary setup. The VMs have been configured to reside on the same LAN network, allowing seamless communication and data transmission between nodes. In our setup, we have it acting as if it were connected to the LAN network. In the future we aim for our collection/analysis server to not be connected via LAN and act as an off-location entity that the control systems can still interact with.

Each VM has the Elasticsearch and Kibana services available from the ELK Stack. These services assist in facilitating the collection and analysis of system logs on the collection/analysis server. Additionally, the Auditbeat service is utilized on each server to interact with Linux's Auditd framework. This built in Linux auditing service enables the collection of system-level logs based on predefined specifications. Auditbeat's integration ensures that logs are systematically collected and sent to Elasticsearch on the collection/analysis server for further analysis.

C. Log Analysis and Visualization

The collection/analysis server, equipped with the ELK stack, serves as the foundation for log analysis and visualization processes. Our current focus is on time-bound log monitoring, where logs are reviewed during specific periods rather than continuously. This approach is particularly crucial for critical infrastructure systems, where timely detection and response to cyber threats are essential.

Using Kibana, we can perform granular searches within the logs, filtering data by system identifiers such as name or IP address, or by keywords tagged in the logs. We can also customize Kibana's dashboard views to get better visualizations of the activity on the network. These capabilities allow for efficient sifting through log data to prioritize and investigate potentially malicious activities. Although we have not yet implemented a detection model, our system is designed to support future enhancements that incorporate advanced threat detection and mitigation techniques.

IV. Testing and Evaluation

Our system is capable of monitoring a distributed network of nodes over specified temporal intervals, effectively auditing both standard operational activities and potential malicious behaviors. The infrastructure supports sophisticated querying of accumulated log data using diverse field names and specific tags, facilitating a nuanced and hierarchical analysis of security events.

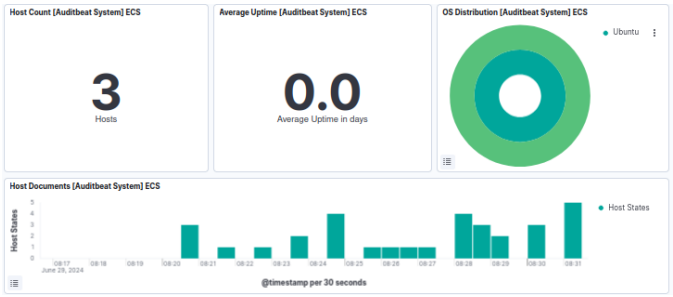

Fig. 3. Kibana Dashboard With Network Host Information

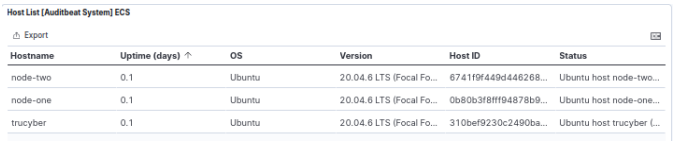

Fig. 4. Kibana Dashboard With List of Hosts in the Network

A. *Kibana Discover/Dashboard Capabilities*

As illustrated in Figure 2, a preliminary simulation is conducted within our test environment to demonstrate the functionality of our system. Figure 2 showcases a screenshot of the Kibana Discover dashboard, which displays the comprehensive low-level system logs collected from the network nodes. In this simulation, we executed a query utilizing the *agent.hostname* field within our system logs, targeting specific nodes across the network. By querying each node based on its system hostname, we successfully retrieved all relevant logs for those nodes. This capability is particularly beneficial in scenarios where the system is busy with voluminous logs, yet a focused analysis on a single node is required. This system not only streamlines the process of log management and threat detection but also enhances our ability to perform targeted audits on individual network components, thereby significantly improving the efficiency and accuracy of cybersecurity operations.

Figures 3, 4, and 5 illustrate the versatile capabilities of the Kibana Dashboards in providing detailed insights into network host information. These dashboards can be customized to create targeted visualizations that address specific auditing and monitoring needs within the network. Figure 3 offers a comprehensive overview, including:

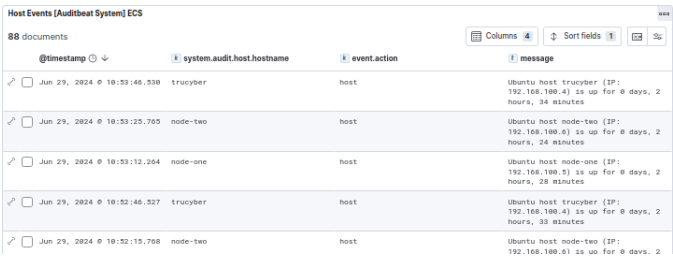

Fig. 5. Kibana Dashboard With List of Host Events in the Network

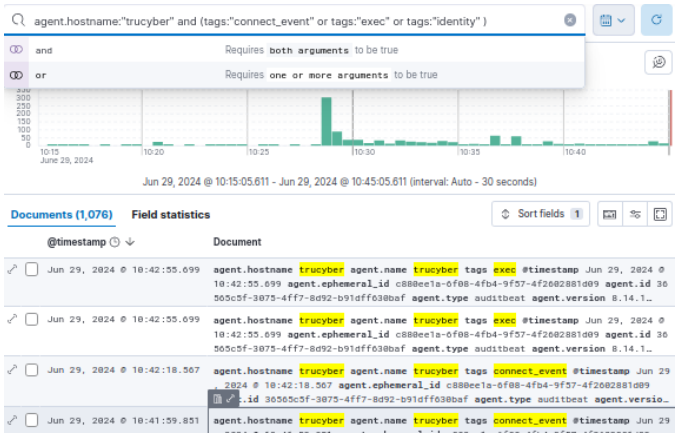

Fig. 6. TRUCYBER Auditbeat Rule Queries

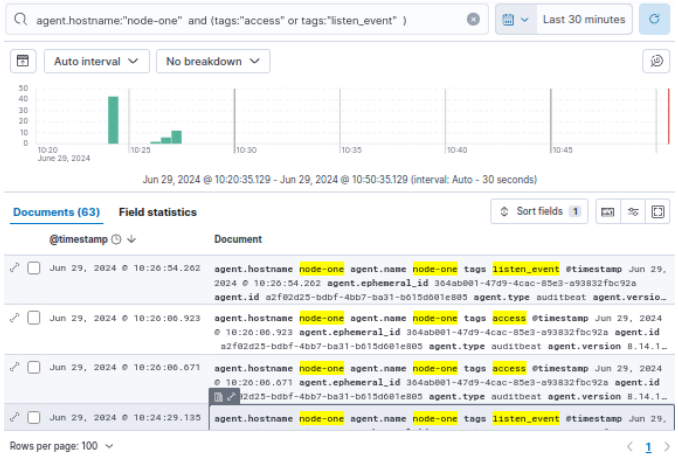

Fig. 7. Node One Auditbeat Rule Queries

- The total number of hosts within the network.
- The average uptime for each host.
- The distribution of different operating systems used across the network.
- A time series visualization representing host-related documents over a given interval.

These detailed breakdowns offer a robust understanding of the types of nodes within the network and their operational activity. This comprehensive view is crucial for evaluating network health and performance and identifying patterns that may suggest either normal or abnormal behaviors. Figure 4 presents a detailed table listing all nodes within the network. This table serves as a tabular representation of the visualizations shown in Figure 3, offering a comprehensive view of the complete data for each host. In contrast, Figure 5 delves deeper by providing a focused log of host-related events for each node. This detailed log analysis is essential for refining our detection models. The raw data from these logs enable us to extract critical metrics and uncover patterns that are crucial for improving our threat detection and response strategies.

B. Auditbeat Rules Capabilities

In our experimental setup, we evaluated the capability to incorporate predefined rules within the Auditbeat configuration file to monitor potentially malicious events. Below, we delineate the specific rules integrated for this testing phase:

RULE SET FOR MALICIOUS EVENT MONITORING

- **Execution Events:**
  - `-a always,exit -F arch=b64 -S execve,execveat -k exec`
- **Identity Modifications:**
  - `-w /etc/group -p wa -k identity`
  - `-w /etc/passwd -p wa -k identity`
- **Unauthorized Access Attempts:**
  - `-a always,exit -F arch=b64 -S open,creat,truncate,ftruncate,openat, open_by_handle_at -F exit=-EACCES -k access`
  - `-a always,exit -F arch=b64 -S open,creat,truncate,ftruncate,openat, open_by_handle_at -F exit=-EPERM -k access`
- **Network Events:**
  - `-a always,exit -F arch=b64 -S socket -k socket_event`
  - `-a always,exit -F arch=b64 -S bind -k bind_event`
  - `-a always,exit -F arch=b64 -S connect -k connect_event`
  - `-a always,exit -F arch=b64 -S listen -k listen_event`
  - `-a always,exit -F arch=b64 -S accept -k accept_event`
  - `-a always,exit -F arch=b64 -S sendto -k sendto_event`
  - `-a always,exit -F arch=b64 -S recvfrom -k recvfrom_event`
  - `-a always,exit -F arch=b64 -S sendmsg -k sendmsg_event`
  - `-a always,exit -F arch=b64 -S recvmsg -k recvmsg_event`

These rules were implemented to facilitate the monitoring of security threats within each network node. Subsequently, we executed testing scripts on each node to simulate a variety of malicious activities over a 30-minute duration. During this interval, we utilized the central Kibana interface to observe the logged malicious events. This was achieved by querying specific hosts and keywords defined in the aforementioned rules.

In Kibana, to effectively locate the keywords associated with the implemented rules, it was necessary to specify the *tags* field. Figures 6, and 7 illustrate that each network node successfully captured the events as per the rules, accurately tagged them, and transmitted the logs to the central Kibana host, designated as TRUCYBER in this study. This demonstrated our system's capability to perform fine-grained analysis and enable targeted detection of malicious activities across the network.

Limitations. The current implementation of our delay-sensitive provenance framework for OT networks has some limitations in reporting system performance, particularly regarding delay and detection sensitivity. While our preliminary results demonstrate the feasibility of provenance data collection, we acknowledge the need for more comprehensive performance metrics. Future work will focus on developing and implementing precise measurements for system latency and detection accuracy. We plan to conduct extensive testing to quantify the delay between the occurrence of an anomalous event and its detection by our system. Additionally, we will establish sensitivity thresholds to optimize the balance between threat detection and false alarm rates. To address the delay problem, we are exploring several strategies, including optimizing our data processing algorithms, implementing parallel processing techniques, and investigating the integration of machine learning models to enhance our detection processes. Machine learning approaches could potentially improve the system's ability to identify patterns and anomalies more quickly and accurately, thereby reducing overall detection time. These enhancements will be crucial in ensuring that our framework can provide timely and accurate threat detection in critical OT environments where rapid response is essential.

V. CONCLUDING REMARKS

In the face of ever-evolving cyber threats, particularly in critical infrastructure systems, this research delves into the application of low-level system log auditing through provenance-based techniques to enhance threat detection capabilities. Focusing on three pioneering systems—ShadeWatcher, SPADE, and UNICORN, we examine their methodologies for leveraging data provenance to identify and mitigate threats with precision. Our proposed system, inspired by these advanced models, aims to create a delay-sensitive audit tracking framework tailored for critical infrastructure networks. By continuously auditing low-level system logs and analyzing them through a dynamic, time-bound approach, we enhance our ability to detect anomalous activities swiftly and accurately. This system is designed to provide high precision in identifying threats and minimizing false alarms, thereby ensuring robust protection for essential services such as power grids, water supply systems, and transportation networks.

In our preliminary testing environment, we've established a network of nodes capable of auditing system activities and transmitting logs to a central collection/analysis server. This setup allows for the granular analysis and visualization of network activities using tools like Elasticsearch and Kibana. By integrating proven methodologies from ShadeWatcher, SPADE, and UNICORN, we aim to not only monitor and detect potential cyber threats in real-time but also adapt to evolving threats, thus promising significant advancements in the security and resilience of critical infrastructure systems.

VI. Future Research Directions

Our research will continue to advance by integrating more sophisticated detection models and enhancing the system's ability to respond to cyber-attacks in real-time. Several key areas will be the focus of our future developments:

A. Time-Sensitive Response Mechanisms

Incorporating a delay-sensitive response mechanism is a priority. We plan to develop a machine learning model capable of computing a certainty metric, which will help determine whether an attack has occurred based on audited logs. Our goal is to guarantee a response to anomalous threats within 5-10 minutes of detection, providing a preliminary assessment of the threat's likelihood. Continuous monitoring beyond this initial period could enhance the certainty of our assessments, refining our response strategy over time.

B. Integration of Provenance-Graph Techniques

We aim to incorporate the advanced provenance-graph techniques demonstrated in ShadeWatcher and UNICORN to enhance our system's detection capabilities. These techniques have proven valuable for tracing the lineage of data and system activities, providing insights into potential security threats. By running ShadeWatcher on our virtual machines and utilizing Auditbeat logs to construct provenance graphs, we have observed significant potential for real-time threat detection. Our objective is to integrate these models into a delay-sensitive detection system that leverages machine learning to recognize and respond to new anomalies in real-time.

C. Utilization of Comprehensive Log Data

Leveraging the extensive log data collected through the ELK stack (Elasticsearch, Logstash, and Kibana) [16] and Auditbeat will be critical. These tools have been instrumental in managing our auditing nodes network and simplifying the analysis of low-level system logs. They enable detailed, granular investigations and visualizations of system activities. Although our focus may shift towards provenance-graph detection for a more sophisticated approach to malicious threat detection, the ELK stack and Auditbeat will continue to play supportive roles in our overall system architecture.

D. Scalability and Network Expansion

To achieve scalability, we need to expand our network capabilities. The SPADE project provides valuable insights into managing data provenance in distributed environments, and its findings will be crucial as we add more nodes to our network. This expansion will ensure that our detection system remains robust and effective as it scales to handle larger and more complex environments.

E. Enhancing Machine Learning for Real-Time Analysis

Future work will focus on refining machine learning algorithms to improve their ability to analyze and respond to threats in real-time. This will involve training models on comprehensive datasets and logs, enabling them to identify patterns and anomalies that indicate potential security breaches. The ultimate goal is to develop a highly responsive and accurate detection system that can adapt to evolving threats.

These future directions underscore our commitment to advancing the field of real-time threat detection and response. By integrating these sophisticated techniques and expanding our system's capabilities, we aim to develop a robust, scalable solution capable of effectively countering cyber threats in increasingly complex environments.

References

[1] "Cyber-Attack Against Ukrainian Critical Infrastructure," https://www.cisa.gov/news-events/ics-alerts/ir-alert-h-16-056-01, 2021, [Online; accessed July 20, 2021].

[2] W. U. Hassan, A. Bates, and D. Marino, "Tactical provenance analysis for endpoint detection and response systems," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 1172–1189.

[3] M. M. Alam and W. Wang, "A comprehensive survey on data provenance: State-of-the-art approaches and their deployments for iot security enforcement," *Journal of Computer Security*, vol. 29, no. 4, pp. 423–446, 2021.

[4] R. Ahmad, A. Modi, E. Jung, C. de Senne Garcia, H. Irshad, and A. Gehani, "Analyzing discrepancies in whole-network provenance," in *International Conference on Critical Infrastructure Protection*. Springer, 2023, pp. 93–114.

[5] J. E. Rubio, C. Alcaraz, R. Roman, and J. Lopez, "Current cyber-defense trends in industrial control systems," *Computers & Security*, vol. 87, p. 101561, 2019.

[6] M. A. Inam, Y. Chen, A. Goyal, J. Liu, J. Mink, N. Michael, S. Gaur, A. Bates, and W. U. Hassan, "Sok: History is a vast early warning system: Auditing the provenance of system intrusions," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 2620–2638.

[7] Y. Xie, D. Feng, Z. Tan, and J. Zhou, "Unifying intrusion detection and forensic analysis via provenance awareness," *Future Generation Computer Systems*, vol. 61, pp. 26–36, 2016.

[8] J. Zeng, Z. L. Chua, Y. Chen, K. Ji, Z. Liang, and J. Mao, "Watson: Abstracting behaviors from audit logs via aggregation of contextual semantics." in *NDSS*, 2021.

[9] E. Esiner, U. Tefek, D. Mashima, B. Chen, Z. Kalbarczyk, and D. M. Nicol, "Message authentication and provenance verification for industrial control systems," *ACM Transactions on Cyber-Physical Systems*, vol. 7, no. 4, pp. 1–28, 2023.

[10] A. Al Farooq, J. Marquard, K. George, and T. Moyer, "Detecting safety and security faults in plc systems with data provenance," in *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, 2019, pp. 1–6.

[11] J. Zengy, X. Wang, J. Liu, Y. Chen, Z. Liang, T.-S. Chua, and Z. L. Chua, "Shadewatcher: Recommendation-guided cyber threat analysis using system audit records," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 489–506.

[12] A. Gehani and D. Tariq, "Spade: Support for provenance auditing in distributed environments," in *ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing*. Springer, 2012, pp. 101–120.

[13] X. Han, T. Pasquier, A. Bates, J. Mickens, and M. Seltzer, "Unicorn: Runtime provenance-based detector for advanced persistent threats," *arXiv preprint arXiv:2001.01525*, 2020.

[14] B. Pan, N. Stakhanova, and S. Ray, "Data provenance in security and privacy," *ACM Computing Surveys*, vol. 55, no. 14s, pp. 1–35, 2023.

[15] L. Moreau, J. Freire, J. Futrelle, R. E. McGrath, J. Myers, and P. Paulson, "The open provenance model: An overview," in *International provenance and annotation workshop*. Springer, 2008, pp. 323–326.

[16] A. F. Rochim, M. A. Aziz, and A. Fauzi, "Design log management system of computer network devices infrastructures based on elk stack," in *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*. IEEE, 2019, pp. 338–342.