International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 4 Apr. 2024, pp: 681-686 www.ijaem.net ISSN: 2395-5252

Secure Data Logging and Processing with Blockchain

First Author[#1], Second Author[*2], Third Author[#3]
[#]First-Third Department, First-Third UniversityAddress Including Country Name
[*]Second Company Address Including Country Name

Date of Submission: 13-04-2024 Date of Acceptance: 26-04-2024

Abstract— The book Secure Data Logging and Processing with Blockchain offers a way to use blockchain technology into data management systems to improve security and transparency. Because they frequently lack strong security safeguards, traditional data logging and processing techniques are susceptible to manipulation and illegal access. In order to securely record and timestamp data exchanges, this research makes use of blockchain's decentralized and immutable ledger. To ensure transparency and do away with the need for middlemen, smart contracts are used to automate data processing processes and enforce predetermined regulations.This project intends to offer a transparent, tamper-proof data management solution that improves data security, integrity, and auditability by incorporating blockchain technology. With the help of the suggested method, businesses may be sure that their data is safe and trustworthy at every stage of its existence.

Keywords- Secure data logging, Data processing, Blockchain technology, Data security.

I. INTRODUCTION

Log analysis can help security management by enhancing malware detection, which is crucial to reducing the harm that malware outbreaks can cause. Machine learning techniques for log analysis are drawing interest because they simplify sophisticated techniques for detecting fraudulent traffic and lessen the workload for analysts.within a SOC (security operations center). Since malware authors frequently recycle their already-existing attack methods, most newly created malicious traffic shares features with previously created malicious traffic, despite the fact that malware producers are always improving their attack tactics. Thus, by capturing certain features, machine learning techniques are likely to identify malicious communications. Originally developed by Cisco IOS software, NetFlow is now the de facto standard for gathering traffic data on IP networks. A large amount of legitimate and malevolent traffic that does not show up in labeled training data may be included in unlabeled training data. We can therefore produce a more accurate classifier if the classifier is able to learn some properties from unlabeled training data in addition to labeled training data. In particular, ISP NetFlow includes a lot of malicious and lawful traffic that does not exist in many corporate networks because an ISP connects to numerous businesses. This information would be helpful for developing a flexible and extremely accurate classifier.

II. LITERATURE SURVEY

1.Paper Name: Inferring Workflows with Job Dependencies from Distributed Processing Systems Logs
Author: Gladys E. Carrillo, Cristina L. Abad
Abstract: Evaluation of new enhancements to distributed processing platforms such as Spark and Hadoop is one of the issues we examine. Workloads released by organizations with substantial data clusters, such as Google and Facebook, are a popular method for assessing these systems. Under practical workloads, these assessments aim to illustrate the advantages of making enhancements to important framework elements, such as the job scheduler. The information on dependencies among the jobs is usually absent from reported workloads, though. This is cause for concern because detecting dependencies may cause a large error in the speedup that results from a given enhancement. We address the need for task dependency information in this position paper and demonstrate how workflow mining techniques may be applied to extract dependencies from job traces that are missing. task dependency information is a crucial component of distributed processing framework evaluations. In an attempt to demonstrate the methodology's viability, we find that the suggested approach can identify workflows in Google traces.

2.Paper Name: Secure Log Storage Using Blockchain and Cloud Infrastructure
Author: Dr. Manish Kumar1,Ashish Kumar Singh2
Abstract: Information technology has taken over our world. It plays a vital and critical role in our daily lives, hence its safe and secure operation is crucial. A cyberattack occurs every minute, involving

thousands of victims. globalized. The means by which attackers choose to target their victim are always becoming more complex and devious. To ensure that attack traces, such as system logs and related data, cannot be found again, they take all necessary precautions to erase them from the victim systems. In an attempt to avoid being discovered, attackers purposefully dispersed their actions across an extended duration. Maintaining secure log records for a prolonged amount of time is necessary in order to comprehend and recognize such complicated attacks. Long-term log record preservation is a difficult problem, though. The integrity of the log files and the logging procedure should also be guaranteed by the system. We propose in this study a secure log storage using Blockchain on Cloud platform to overcome these problems. Audit logs will be more difficult to tamper with thanks to blockchain technology. Proof of log tampering and non-repudiation is provided. Deep analysis is supported and the system is made scalable overall by the cloud platform.

3.paper Name: Towards Automated Log Parsing for Large-Scale Log Data Analysis
Author: Pinjia He, Jieming Zhu, Shilin He, Jian Li, and Michael R. Lyu
Abstract: For dependability assurance, logs are widely used in system management since they are often the only data available that capture unique system runtime behaviours in production. Given that log volumes are constantly increasing and that structured input data, such as matrices, is required, developers (and operators) are interested in leveraging data mining techniques to expedite their analysis. Research on log parsing begins as a result, with the goal of organising disorganised log messages into structured events.Unfortunately, as there are no open-source implementations of these log parsers or benchmarks for performance comparison, developers are unlikely to be aware of the drawbacks and effectiveness of current log parsers when putting them to use in practice. When they have to modify or reintroduce one, it is tedious and unnecessary.To evaluate the efficacy of the state-of-the-art log parsers, we first offer a characterization analysis of five real-world datasets including over ten million log messages. Overall, these parsers are fairly accurate, although their robustness varies depending on the dataset, it is determined. As logs get big (e.g., 200 million log messages), which is the usual in practice, these parsers become ineffective to analyse such data on a single machine. To get beyond these limitations, we develop and implement a parallel log parser, termed POP, on top of Spark, a platform for managing enormous volumes of data. There have been numerous research conducted to evaluate POP on both synthetic and real-world datasets. According to the evaluation's results, POP can accurately, effectively, and efficiently handle increasingly challenging log mining tasks.

4. Paper Name: Securing Big Data in the Age of AI
Author name: Murat Kantarcioglu,Fahad Shaon.
Abstract: To develop sophisticated machine learning, AI, and data analytics models, businesses are gathering ever-increasing volumes of data. Moreover, the unstructured data (text, images, and videos) required to create these models may be present. Accordingly, these kinds of data could be kept in a variety of DMSs, from relational databases to more recent NoSQL databases designed specifically to hold unstructured data. Additionally, more and more data scientists are processing data with numerous existing libraries by using programming languages like Python, R, and others. Under certain circumstances, the NoSQL system on the stored data will automatically run the developed code. The aforementioned advancements highlight the necessity for a comprehensive data security and privacy resolution that can consistently safeguard information housed in various data management systems and implement security protocols, even in cases where sensitive data is processed by a data scientist's intricate program. We lay forth our plan in this paper to create a massive data protection solution of that kind. In particular, our suggested SECUREDL system enables organizations to: 1) implement policies that restrict access to sensitive data; 2) automatically maintain audit logs required for data governance and regulatory compliance; 3) sanitize and redact sensitive data as needed based on the requirements of AI models and data sensitivity; 4) identify potentially unauthorized or unusual access to sensitive data; and 5) automatically create attribute-based access control policies based on data sensitivity and data type.

5.Paper Name: Kratos: A secure, authenticated and publicly verifiable system for educational data using the blockchain
Author: Dr. Velislava Hillman,Varunram Ganesh
Abstract: Institutions and students are placing a lot of emphasis on data ownership as a result of the growing interest in educational data mining (EDM) and learning analytics (LA) as ways to use big data to improve education and the science of learning. In order to improve the quality of teaching and learning, EDM and LA can offer valuable information. However, it is now imperative to protect student autonomy over data and data privacy. Our study presents Kratos, an unchangeable and publically

verifiable data management system that permits both EDM and LA. It also protects student privacy by providing an interface for data governance and student engagement in school procedures.While putting student agency ahead of their data, the system seeks to establish data interoperability, which makes EDM and LA easier to implement and provides incentives to educational stakeholders (policy makers, educators, tech developers, etc.). With our technology, schools and students have complete access to data that is otherwise dispersed across several vendors and systems, as well as an unchangeable log. An existing nonvirtual agreement [1] between education technology (edutech) manufacturers and schools served as the model for the smart contracts that specify the fundamental rules of the system.

6.Paper Name: SILU: Strategy Involving Large-scale Unlabeled Logs for Improving Malware Detector
Author: Taishi Nishiyama , Atsutoshi Kumagai, Kazunori Kamiya
Abstract: This paper presents a novel approach to semi-supervised learning dubbed SILU. Taking full advantage of unlabeled data, it increases detection performance without requiring more manually labelled data. SILU automatically uses combined labelled and unlabeled training data to enhance labelled training data through the screening process. As a result, a classifier is created. As opposed to previous semi-supervised learning methods used in cyber security, which use test data as unlabeled training data, SILU can employ separate datasets for test data and unlabeled training, negating the need for retraining whenever test data changes. This facilitates the SOC's efficient suppression of the detection time. Furthermore, while SILU makes use of certain supervised learning strategies, it does not require a specific method. As a result, SILU can be applied to any kind of supervised learning approach classifier. Furthermore, SILU can prevent test data's classification performance from declining by using screening. We assessed SILU with two kinds of real-world data: NetFlow from a large ISP and proxy logs from a large organisation. We have shown that SILU consistently enhances detection performance for supervised learning techniques when tested with various classifier types. Furthermore, SILU works better than existing semi-supervised techniques. All things considered, SILU outperforms traditional supervised learning techniques and can be added to current methods with little overhead. Our analysis also demonstrates that the use of unlabeled training data from ISP NetFlow is more effective than the use of simply labelled proxy logs from the same organisation. These findings imply that when many organisations, such as SOCs and ISPs, work together and share unlabeled data, SILU can increase its detection capability.

7.Paper name: Securing Logs of a System - An IoTA Tangle Use Case
Author: Mohan Bhandary,Manish Parmar
Abstract: The proliferation of the Internet and our reliance on technology have led to an accelerated growth in the cyber technology sector. The most important thing we have now is cyber-security, which helps shield systems from online threats. Blockchain technology, or distributed ledger technology, has several benefits and features that make it a promising tool for cybersecurity and risk mitigation. With the help of these technologies, resource tracking and recording are made easier by eliminating the requirement for any one central, reliable party or organisation. Distributed ledger technology, or DLT, is the main topic of this paper's discussion on cybersecurity. It centers on the features and operation of the recently developed distributed ledger technology, IOTA.The application of IOTA Tangle technology, which offers a means of system log security, is another topic of this article. In the investigation phase of cyber forensics, logs are the most significant piece of evidence, hence it is imperative to securely store and preserve the logs. The application of IOTA for system log security is proven later in this paper.

8.Paper name: Securing Car Data and Analytics using Blockchain
Author: Gokay Saldamli, Kavitha Karunakaran, Vidya K. Vijaykumar
Abstract: The automotive business is rapidly evolving and changing thanks to manufacturers' partnerships with the technology sector. Since there is a high demand from related entities like insurance companies, vehicle buyers/sellers, and government authorities, the present trend of connected automobiles depends on obtaining various types of vehicular data. As things like duplicate or fabricated vehicle data records, tampered safety checks, and manipulated driving histories arise, the current methods of data collection—manual or unsupervised—pose challenges to trust, legitimacy, and accuracy. Therefore, a robust tool that can safeguard vehicle data, record modifications for auditing purposes, and ultimately increase system trust is required. We suggest applying blockchain technology to these requirements.Manufacturers' collaborations with the technology industry are causing the automotive industry to change and evolve quickly. Getting different kinds of vehicle data is

essential to the current trend of connected cars since relevant institutions such as insurance companies, car buyers/sellers, and government agencies have significant demands. The present methods of data gathering, whether manual or unsupervised, put confidence, legitimacy, and accuracy at risk due to things like duplicate or falsified vehicle data records, tampered safety checks, and manipulated driving histories. Because of this, a strong tool is needed that can protect vehicle data, log changes for audits, and eventually boost system confidence. For these requirements, we propose to use blockchain technology.

9.Paper name: A More Efficient ORAM for Secure Computation
Author: Bo Zhao, Zhihong Chen, Hai Lin, Lin Chen
Abstract: When scaling a distributed ORAM to a two-party secure computation, the amount of pseudo-random generator (PRG) calls in the distributed point function (DPF) evaluation and generation—which are O(logn) and O(n), respectively—determines the majority of the overhead, where n is the number of data blocks. Our suggested distributed ORAM approach reduces the PRG calls to O(log(zn/)) for generation and O(2log(zn/)) for evaluation, where z is the length of shares of outputs in DPF and is the secure parameter. From a technical standpoint, we first expand the early termination optimization of Function Secret Sharing (FSS), which is applicable to functions with small output groups, to the ORAM context for secure computation.Next, we create a plan called etoram to take advantage of the great efficiency attained by early termination. Every data block in etoram has an access counter added to it. During a write operation, a random string is hidden to update the data block quietly and the access counter is updated privately using the DPF's outputs. In actuality, depending on various access rates, z ranges in sequential mode from 1 to logn and in random mode from 3 to logn depending on the total number of accesses.

10.Paper name: Secure Opportunistic Contextual Logging for Wearable Healthcare Sensing Devices
Author: Muhammad Siddiqi, Syed Taha Ali, Vijay Sivaraman
Abstract: Medical uses of wearable technology are growing in popularity, including the ongoing home and hospital monitoring of patients with chronic illnesses.
Patients, physicians, and insurers are among the stakeholders who have a vested interest in guaranteeing not just the integrity of the data but also the verifiability of the context, meaning that the information gathered can be linked to the correct time and location. In earlier studies, these topics were examined separately and usually with the use of cryptographic methods. We present a novel approach in this research that makes use of the abundance of wireless devices nearby the transaction to generate witness recordings that guarantee data integrity and tamper resistance within the constraints of time and place.The first thing we've done is create a secure logging architecture that uses Bloom filters to compress witness recordings and hashchains them to the data so that forensic verification can be completed quickly and accurately. We also quantify the impact of these configuration choices on verification accuracy. These configuration settings influence our scheme's performance in terms of storage, processing, and transmission efficiency. Our final contribution uses real trace data from a multi-story building that simulates a hospital setting to implement our method, show that it is feasible, and quantify its effectiveness through simulation.

III. PROPOSED METHODOLOGY

It is suggested that a decentralized blockchain network be put in place in order to securely log and process data transactions as part of the Secure Data Logging and Processing using Blockchain approach. To maintain data integrity and immutability, data logging is accomplished using consensus methods and cryptographic hashing. Automating data processing operations and implementing pre-established rules are possible with smart contracts. Data security is improved by the use of encryption methods and access controls. To further preserve the integrity of the blockchain ledger, routine audits and transparency initiatives are put in place. With a variety of applications, the approach seeks to offer a stable and dependable foundation for safe data processing and logging.

A. APPLICATION

Blockchain-based secure data logging and processing is being used in a number of businesses where data security and integrity are top priorities. This technology is beneficial to supply chain management, healthcare, finance, and the Internet of Things, among other industries. Blockchain technology can be utilized in banking to securely log and audit transactions. It can protect patient records' integrity in the medical field. Blockchain technology facilitates traceable and transparent logistics for supply chain management. In real time, data can be safely logged and processed by IoT devices. All things considered, the use of blockchain in Secure Data Logging and Processing offers a reliable foundation for handling sensitive data in a variety of contexts.

B. SYSTEM ARCHITECTURE

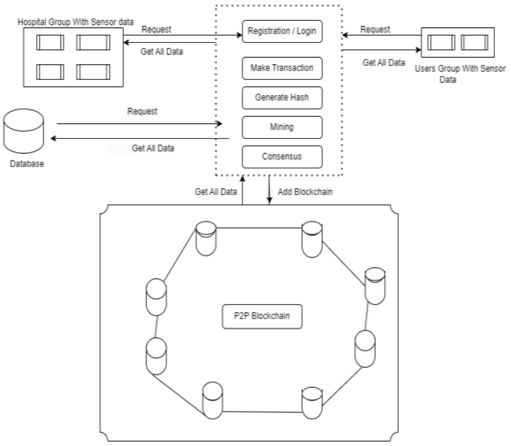

Fig. 1 System Architecture

IV. CONCLUSION AND FEATURE SCOPE

In order to produce a flexible and advanced automatic log analysis tool that accurately determines if a host is infected with malware or not, we invented SILU. This allows us to completely utilize valuable information from both labeled and unlabeled network traffic logs. The primary benefits of SILU over other semi-supervised learning techniques in cyber security are its ability to prevent classification performance degradation, work as an add-on to any kind of supervised-learning-based classifier, and eliminate the need for retraining each time test data changes during the detection phase. Through assessments, we highlighted the impact of learning with unlabeled training data and proved that SILU outperforms both traditional supervised learning methods and current semi-supervised learning techniques for cyber security.

Future developments in Secure Data Logging and Processing with Blockchain are expected to be substantial. New security protocols will be created to provide strong defenses against constantly changing cyberthreats. A greater number of industries—including government and supply chain—are expected to utilize blockchain technology more frequently as its advantages in data management become increasingly apparent. The smooth interchange of data and collaboration between various blockchain platforms will be facilitated by efforts to increase interoperability. Furthermore, it is anticipated that blockchain systems will advance to provide transparent and auditable procedures, which will better support regulatory compliance. These developments will spur innovation in data processing and logging, opening up new avenues for both individuals and enterprises.

REFERENCES

[1]. K. Bartos and M. Sofka, “Optimized invariant representation of network traffic for detecting unseen malware variants,” In Proceedings of the 25th USENIX Security Symposium, pp. 807–822, 2016.

[2]. M. Antonakakis et al., “From throw-away traffic to bots: detecting the rise of dga-based malware,” In Proceedings of the 21th USENIX Security Symposium, pp. 491–506, 2012.

[3]. J. Jang, D. Brumley, and S. Venkataraman, “BitShred: feature hashing malware for scalable triage and semantic analysis,” In Proceedings of the 18th ACM Conference on Computer and Communications (CCS), pp. 309–320, 2011.

[4]. B. Claise, “Cisco systems NetFlow services export version 9,” https://tools.ietf.org/html/rfc3954, 2004.

[5]. L. Shi, D. Lin, C. V. Fang, and Y. Zhai, “A hybrid learning from multi-behavior for malicious domain detection on enterprise network,” In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 987–996, 2015.

[6]. I. Santos, J. Nieves, and P. G. Bringas, “Semi-supervised learning for unknown malware detection,” In Proceedings of the 8th International Symposium on Distributed Computing and Artificial Intelligence (DCAI), pp. 415–422, 2011.

[7]. G. Osada, K. Omote, T. Nishide. “Network intrusion detection based on semi supervised variational auto-encoder,” In Proceedings of the 22nd European Symposium on Research in Computer Security (ESORICS), pp. 344-3612017.

[8]. K. Zhang, C. Li, Y. Wang, X. Zhu, and H. Wang, “Collaborative support vector machine for malware detection,” In Proceedings of the International Conference on Computational Science (ICCS), pp. 1682– 1691, 2017.

[9]. D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scho¨lkopf, “Learning with local and

global consistency," In Advances in Neural Information Processing Systems (NIPS), pp. 321–328, 2004.

[10]. G. Stringhini, Y. Shen, Y. Han, and X. Zhang, "Marmite: spreading malicious file reputation through download graphs," In Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC), pp. 91–102, 2017.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper before submission to the conference. Failure to remove template text from your paper may result in your paper not being published.