# Log File Authentication and Storage on Blockchain Network

Dávid János Fehér[1], Barnabás Sándor[2]
[1]Keleti Károly Faculty of Business and Management, Óbuda University, Budapest, Hungary
[2]Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University
david.janos.feher@gmail.com; sandor.barnabas@gmail.com

***Abstract*—Authentication is the issue most talked about nowadays. Comparing with Blockchain we want to research the justification for the blockchain technology to be used in enterprise log analysis and the appearance of blockchain technology to develop and make remarkable change in enterprise log analysis.**

***Keyword: blockchain; siem; authentication, cyber security***

## I. INTRODUCTION

The main problem in the enterprise environmentregarding to security is the fast and efficient incident response. Most of the significant companies have a dedicated Security Operations Center and spend a significant amount of money to operate it. It's essential to guarantee the dependability of the information they handle in the Security Information Event Management solution and ensure its safe long-term preservation. The blockchain can help to provide the uncorrupted Confidentiality and Integrity elements of the CIA criteria.

## II. BACKGROUND

Nowadays, most of the high-end companies use Security Information and Event Management (SIEM) solutions to make their security more efficient. This is a complex system to make the information more readable and usable. This log management provides real-time monitoring opinion, compliance adequacy and makes it more pleasant to pass compliance and get the required certifications like ISO27001, or to pass other audits. System log analysis is not only in the company's best interest but, in the financial sector or at telecommunication service providers it is compulsory! This analysis is executed by checks using regular audits to make sure that the data safety is guaranteed, while everything is entirely operational.

The automatized and managed log analysis is not worth conduct for the sake of auditors, but it also gives a comprehensive picture to the organization about the operation of IT systems from either a financial or technical aspect. Log analysis can give answers to such questions as „How much of the systems are used?" or „What upgrades will the system need when the number of users increases?" It helps to make deductions of the possible causes of an error not only after but during its occurrence. By getting all the information from logs, these errors and system failures can be avoidable in the future. So far security is the main reason for log analysis. The better use of a system could avoid most attacks. During the incident, the primary cause of gathering evidence is to resolve the event, but additional information may be needed in case of a legal procedure. In such cases, it is essential to document everything the evidence states, including the vulnerable system. Evidence should be collected according to procedures that meet the requirements of legal organizations and law enforcement agencies. It is necessary to consult with the local authorities on what this is about when it comes to generating and guarding evidence. If any information is transmitted to someone the transfer must be documented by official personal signatures. For all evidence, a detailed log has to be ordered and kept. This includes identification data such as locations, serial numbers, model numbers, MAC and IP addresses, datesas well as other useful data or information, such as a log of the evidence gathered and proof of evidence storage.

Gathering evidence comes with many challenges. It is advised to start saving data from the affected systems early on. An initial screensaver can help in finding the source and the problem. It is important to save the device status before incident managers and administrators or others accidentally change the machine status during the scan. Users and administrators should be informed about the steps they need to take to ensure that the evidence is safely preserved. [1] [2] [3] [4]

### *A. Growing tendencies*

There has been a growing interest in log analysis systems in recent years due to the growing trend in the number and severity of cyber attack sequences. It is expected that electronic data will become more and more targeted in the future, as we are continuously storing more and more quantities and quality in this form, and we are doing more and more activities online. Recent threats, which have been growing ever since, have caused enormous financial damage and are expected to keep growing. Many organizations, corporations, or private individuals have become victims to these systems, as there are environments where they can not provide support for systems that are regularly upgraded or applications that run only in a previous system environment; typically old target applications such as cashier software used in some stores and in healthcare. [3]

Since it is not possible to continuously update these systems, they may also have other solutions such as appropriate network protection and support for log analysis

978-1-5386-6841-2/18/$31.00 ©2018 IEEE 

systems. Already, with this enthusiasm, many companies have joined the interest base, but after almost every big-scale attack, the interest in these systems further increased. [4]

### *B. Legal Changes*

In most sectors, including the finances, banking, health and industrial sectors, appropriate, tracked and documented IT operations are required to comply with internal and external audits but are also a prerequisite for effective operation. The statutory and legal requirements give leeway in resolving compliance. Using the log analysis continuously for IT activities to be monitored to the standards (SOX, PCI DSS, HIPAA, FISMA, NERC CIP) requires or recommends these systems to comply, but also helps a great deal, because it can determine and reduce the risks.

Evolving electronic information security regulations and laws increasingly focus on adequately documented and stored log files and incident management. According to The EU General Data Protection Regulation[5] paragraph 85 containing the point, the company must have up to 72 hours tospend and generate a report on the incident. The general principle is where the legislator can expect an automated solution. For both this and subsequent investigations, it is recommended to operate the proper log analysis system on our network, thus increasing the need for building and operating the appropriate log analysis system. [5] [6] [7]

### *C. Compliance regulations*

There are many requirements defined in the following regulations: PCI, HIPAA, SOX, GLBA.

From the log management point of view, the most important is the logging of all relevant and necessary events and securely storing them for a specified interval.

### *D. PCI DSS*

The 5 big payment card industry (VISA, MasterCard, AMEX, JCB, Discovery) created together a security standard for the handling procedures of the payment card data. Payment card industry data security standards (PCI DSS) is a standard known and used worldwide. There is a complicated audit procedure for every company that handles, stores and forwards payment card information. They need to pass it every year to get the permission to use these data. Technically it is not hard to comply it, but guaranteeing the generating of these logs and managing the created logs is the challenging part. *[8] [9] [10]*

**Log management**

The first step to sufficiently integrate a SIEM solution is the well-managed log management. It is essential to get every piece of information from everywhere around the network or else blind spots will turn up, which can reduce the efficiency of the SIEM and the level ofsecurity.

Parts of the infrastructure, operation systems, application logs and transaction logs are real treasure chests of information concerning safe and consistent operation. However, using this information can cause serious problems for most companies because information can come from various sources in various formats which does not help their unified handling, research and analysis.

Every organization has different expectations of log managing regarding standard compliance, incident response time reducing and better investigation efficiency. Depending on the need every organization can have a personalized logging plan and a specific log retention time and log retention area sizes based on the loggable and useful data and the sources. The budget is an essential part of the designing.

Primary sources for the SIEM are the network or host-based intrusion prevention detection systems, firewalls, proxies, routers, and other network devices. The most significant information is the correlated information found on this data. [11] [12] [13] [14]

*Log collection*

Devices nowadays use push or pull log collection method. It is important to check and design the log management with the right properties.

- Push log collection is the simplest way and easy to set up at the SIEM. The log source device is sending the log information to the log collector or the SIEM solution directly immediately or as soon as possible. Forwarding this information in TCP connection is important to avoid data loss.
- Pull log collection is the method when the SIEM solution connects to the source device and collects the information. It is not as secure as the push method. The time gap between two collection periods can cause a risk.

The real environment is more complicated so essential to handle the mixed log collection methods. [15]

*Log Formats*

Log format and syntax define the formatting, transport, storage, and analysis of the log contents. Users and designers of automated event analysis tools strive to maintain a consistent syntax for ease of processing. Popular formats include W3C Extended Log File Format (ELF), Apache Access Log, Cisco SDEE / CIDEE , ArcSight Common Event Format (CEF), Syslog. Syntax defines the text and its report. The most typical components of the log are the date, log source, generating system name, an application that created it and the story that it describes. [16]

"Flat text file" is a standard text file that contains information that can be easily understood by people as well. Some separation needed in a file is, for example, an exact comma or another character so that the information is legible and can be processed properly. This type of storage is not very popular because it not designed for large companies. Its performance is not satisfactory for this solution, reading and writing from a text is much slower than other methods. There is little point in using the Flat text file format for data storage, but it makes it easy for external applications to access these data. If the logs are stored in a text file, it is not difficult to write the code to open the file and transfer the information it

contains to another application. Another advantage is that this text file is accessible for people to read and makes analyzers' work easier when looking for a file. It is easy to open in any text editor, and we can easily find the information we are looking for without opening the management console.

The binary file format uses a unique format to store binary information, which is only used by SIEM specifically. The specified target device is the only one that can read and write these confidential files. The Logs are usually stored in a database, typically in standard formats such as Oracle, MySQL, Microsoft SQL, or other accessible databases in the company. With this method, we can easily modify our data interactively as part of the database application. The performance is excellent, depending on the database type, but most of them are optimized for SIEM. Using databases is a great way to store logs, but it can cause some errors depending on the log analysis solution. [16]

*Syslog standard log recording method*

The open source product syslog-ng has been on the market for a long time. This software provides an excellent foundation for creating a complete logging framework for the demands of the market. The syslog project aims to create a mechanism for information systems, especially networks and machine rooms. It realizes the events that occur during its operation. The system to be created will forward log entries from different sources to a central location in hierarchical order. As a result of the project, a complete logging infrastructure was established, based on the open industry standards of the Internet technology, allowing connections to other (such as monitoring, analysis, archiving) systems. This provides the best system for the user's needs, as it can flexibly configure for manageable infrastructure and it is easily adaptable to any monitoring system.

The Syslog is the most used method for security auditing and system management and other log user activity. The most popular network devices such as printers, routers and firewalls support this method. There are reporting level configurations opinions to make the amount of data personalized to the requirements. [17]

The main benefit of this method is the encrypted storage and the reliable information transport. The update Syslog servers work with TCP connection, so this grants protection against the data loss based on the wrong protocol. Moreover, it can ensure encrypted storage for the messages, which could be useful to regulatory compliance. [18] [19]

*Artificial intelligence in log analysis*

Every large corporation seeks to analyze an increasing amount of data with fewer and fewer trained professionals to increase the company's expenses. The most popular topic is to include artificial intelligence and cognitive computing capacities in log analysis to reach the analytical team's reliance on false alarms and further facilitate users' work.

Watson, which can be used by IBM Qradar, is a pioneer in this category. Watson is continuously researching IT resources available on the Internet, such as reports, recommendations, blogs, and correlating them to structured data to gain a greater insight into the incidents. We can entrust some of the exploration tasks to Watson to collect information about an attack. This not only facilitates the work of explorers today, but it can also enhance its efficiency as it can process much larger information sets than analysts have time for. [20] [21] [22]

**Blockchain**

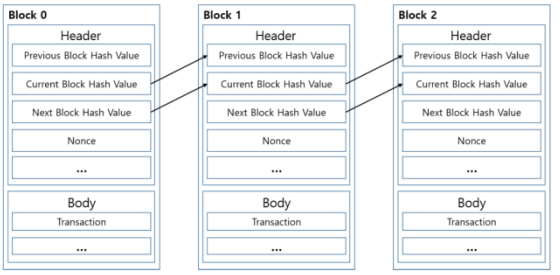
Blockchain technology is known as a shared, distributed ledger that facilitates the process of recording transactions and tracking assets in a business network. Virtually in real time anything of value can be tracked and transferred ina blockchain network, decreasing risk and cutting costs for all included.An essential disadvantage of the Blockchain data storage is that a blockchain is not fitting for storing vast amounts of data by design. It can store simple activities and some arbitrary data, but it is indeed not suitable for storing a highvolume of data.

Blockchain technology is not just a unique technique, but contains Cryptography, mathematics, Algorithm and economic model, combining peer-to-peer networks and using distributed consensus algorithm to solve traditional distributed database synchronizing problem, it is an integrated infrastructure construction.

Every block carries a hash (a digital fingerprint), a timestamped identifier of current valid actions, and the hash of the past block. The previous block hash links the blocks together and prevents any block from being altered, or a block inserted between two existing blocks. In this way, each subsequent block strengthens the verification of the previous block and hence the entire blockchain. The method renders the blockchain tamper-evident, lending to the key attribute of immutability. [23] [24] [25] [26]

*Blockchain Working Scheme*

First of all sending node records of new data and broadcasting to the network. (2) After receiving the node checks in the message about data it received and, if the message was correct,it will be stored in a block. (3) All received nodes act as proof of work (PoW) or proof of stake (PoS) algorithm to the block in the network. (4) The block will store up into the chain after executing consensus algorithm and every node in the network admits this block and will continuously.[Figure 1] [23]

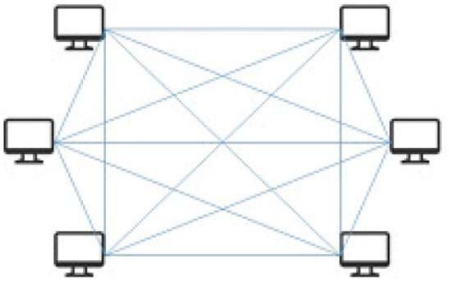



*1. Figure: Blockchain connection structure [27]*

*Blockchain types [23]*

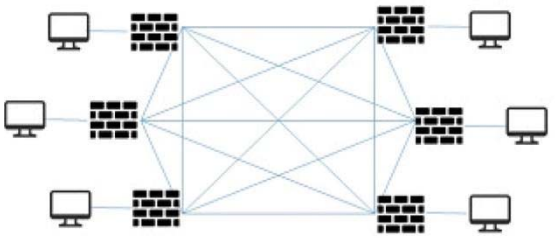
Public blockchain: The transactions are transparent; therefore anybody can verify and audit it. The users can

participate in the process of getting consensus; according to Bitcoin and Ethereum these are both Public Blockchains. [Figure 2]



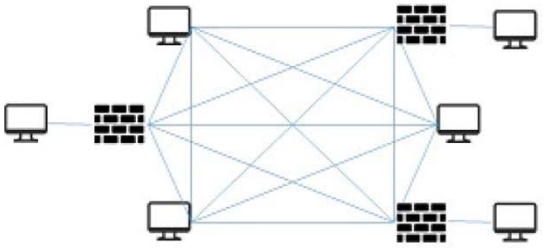
*2. Figure: Public blockchain*

Private blockchain**:** Node is restricted, and can participate in this blockchain, and has harsh authority management on data access. [Figure 3]



*3. Figure: Private blockchain*

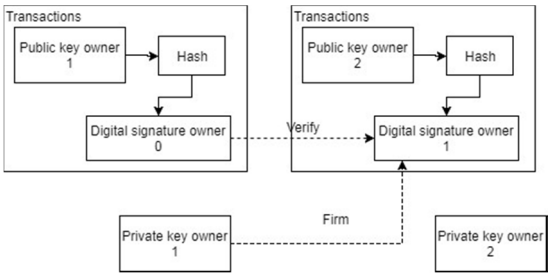
Consortium blockchain: The node having authority can be chosen in advance, usually having partnerships like business to business, the data in blockchain can be open or private or it can be seen as Partly Decentralized. Hyperledger and R3CEV, for example, are both consortium blockchains. [Figure 4]



*4. Figure: Consortium blockchain*

*Blockchain Security*

Satoshi Nakamoto described a chain of digital signatures, where each partner makes a transfer, each transfer uses an asymmetric encryption system and information integrity. Every owner has a public and private key. The public key is stored inside a transaction block, but to keep its integrity this, in turn, stops by a hash function that together with the private key as confirmation method of the parent owner of the block guarantees that the transaction is made in a P2P way. [Figure 5] [28]




*5. Figure: Blockchain transaction diagram*

## III. CONCLUSION

Depending on the security requirement the blockchain based log management could be a great opportunity. Partial solutions are already available on the market like the Tamper Protection but the new technologies may contain unknown potential. The complexity of the Big Data and other new trends like the Cloud Computing already made significant changes in the word of information security. With the technology evolving, new, better and improved solutions can decrease the dangers of log management and improvement of their reliability. The future is the hashgraph technology which can replace the blockchain's current position and this new technology can lead to more efficient solutions. The main problem of Blockchain is wasting of resources if it is handling too much data. Hashgraph is the answer to the blockchain loss problem with higher efficiency rate so it may reduce the disadvantages of blockchain usage. [29] [30] [31]

## REFERENCES

[1] Computer Security Incident Handling Guide // www.nist.gov. - 2012 augusztus. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf Visited: May 2, 2018

[2] ERNW - EventLog Manipulation: https://www.ernw.de/download/EventManipulation.pdf Visited: May 2, 2018

[3] ISTR 22: Extraordinary Attacks, High-Dollar Heists, Electoral Disruption // symantec.com - 2017 - https://www.symantec.com/security-center/threat-report Visited: May 1, 2018

[4] Hart, Catherine V. "Security information and event management system employing security business objects and workflows." U.S. Patent No. 9,069,930. 30 Jun. 2015.

[5] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=ENVisited: May 8, 2018

[6] Nyrén, Olof, Magnus Stenbeck, and Henrik Grönberg. "The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research." European journal of epidemiology29.4 (2014): 227-230.

[7] Voss, W. Gregory. "European union data privacy law reform: General data protection regulation, privacy shield, and the right to delisting." (2017).

[8] PCI Security Standards Council:https://www.pcisecuritystandards.org/documents/PCI_DSS_v3-2.pdf Visited: May 4, 2018

[9] Williams, Barry L. Information Security Policy Development for Compliance: ISO/IEC 27001, NIST SP 800-53, HIPAA Standard, PCI DSS V2. 0, and AUP V5. 0. Auerbach Publications, 2016.

[10] SANS Institute - Successful SIEM and Log Management Strategies for Audit and Compliance: https://www.sans.org/reading-

room/whitepapers/logging/successful-siem-log-management-strategies-audit-compliance-33528 Visited: May 9, 2018

[11] Wright, Clifford C. "Behavioral-based host intrusion prevention system." U.S. Patent No. 8,776,218. 8 Jul. 2014.

[12] Lin, Wei-Chao, Shih-Wen Ke, and Chih-Fong Tsai. "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors." Knowledge-based systems 78 (2015): 13-21.

[13] Kamiya, Kazunori, et al. "The method of detecting malware-infected hosts analyzing firewall and proxy logs." Information and Telecommunication Technologies (APSITT), 2015 10th Asia-Pacific Symposium on. IEE, 2015.

[14] John, Pramod, et al. "Monitoring network traffic by using event log information." U.S. Patent No. 9,584,522. 28 Feb. 2017.

[15] Sawada, Shigenori, et al. "Master device, slave device, information processing device, event log collecting system, control method of master device, control method of slave device and control program." U.S. Patent Application No. 15/381,128.

[16] Miller, David R., et al. Security Information and Event Management (SIEM) Implementation (Network Pro Library). McGraw Hill, 2010.

[17] Gerhards, Rainer. The syslog protocol. No. RFC 5424. 2009.https://tools.ietf.org/html/rfc5424)

[18] Anastopoulos, Vasileios, and SokratisKatsikas. "Design of a Log Management Infrastructure Using Meta-Network Analysis." International Conference on Trust and Privacy in Digital Business. Springer, Cham, 2016.

[19] Maki, Nobuhiko. "Monitoring apparatus, monitoring system, log management method, and computer program." U.S. Patent No. 9,413,915. 9 Aug. 2016.

[20] Lee, Sangdo, and Yongtae Shin. "The Direction of Information Security Control Analysis Using Artificial Intelligence." Advances in Computer Science and Ubiquitous Computing. Springer, Singapore, 2017. 872-877.

[21] Atzmon, Asaf. "Smart Cyber Security Solution for Driving in the Future." ATZ worldwide 119.12 (2017): 38-43.

[22] Andrecht, Torsten, et al. "Cognitive Security." IT-Prüfung, Sicherheitsaudit und Datenschutzmodell. Springer Vieweg, Wiesbaden, 2017. 145-169.

[23] Lin, Iuon-Chang, and Tzu-Chun Liao. "A Survey of Blockchain Security Issues and Challenges." IJ Network Security 19.5 (2017): 653-659.

[24] Juan Benet: IPFS - Content Addressed, Versioned, P2P File System (DRAFT 3) https://github.com/ipfs/papers/raw/master/ipfs-cap2pfs/ipfs-p2p-file-system.pdf Downloaded: May 13, 2018

[25] Tapscott, Don, and Alex Tapscott. Blockchain revolution: how the technology behind bitcoin is changing money, business, and the world. Penguin, 2016.

[26] Swan, Melanie. Blockchain: Blueprint for a new economy. " O'Reilly Media, Inc.", 2015

[27] Park, Jin Ho, and Jong Hyuk Park. "Blockchain security in cloud computing: Use cases, challenges, and solutions." Symmetry 9.8 (2017): 164. Figure 1. Blockchain connection structure Accessed: Mar 27, 2018

[28] Monsalve, Fabián, Octavio José Salcedo Parra, and Roberto AlbeiroPava Díaz. "Blockchain: 3.0 the Technological Solution to Face Corruption." (2017).

[29] Baird, Leemon, Mance Harmon, and Paul Madsen. "Hedera: A Governing Council & Public Hashgraph Network." (2018).

[30] Baird, Leemon. "The swirldshashgraph consensus algorithm: Fair, fast, byzantine fault tolerance." Swirlds, Inc. Technical Report SWIRLDS-TR-2016 1 (2016).

[31] Shu Yun Lim, M. L Mat Kiah, Tan Fong Ang: Security Issues and Future Challenges of Cloud Service Authentication. Acta Polytechnica Hungarica Vol. 14, No. 2, 2017

[32] Dániel Csubák, Katalin Szücs, Péter Vörös, Attila Kiss: Big Data Testbed for Network Attack Detection. Acta Polytechnica Hungarica Vol. 13, No. 2, 2016