Secure Log Storage Using Blockchain and Cloud Infrastructure

Dr. Manish Kumar[1]
Department of Computer Applications
M S Ramaiah Institute of Technology
Bangalore, India
manishkumarjsr@yahoo.com

Ashish Kumar Singh[2]
Department of Computer Applications
M S Ramaiah Institute of Technology
Bangalore, India
ashish.msrit10@gmail.com

Dr. T V Suresh Kumar[3]
Department of Computer Applications
M S Ramaiah Institute of Technology
Bangalore, India

***Abstract* — We are living in the era of information technology. Our day to day life is heavily dependent on it and hence it's safe and secure functioning become very crucial and important. Every minute, there are thousands of cyber-attacks taking place around the world. Attackers are continuously evolving sophisticated and stealthy techniques to target the victim. They takes all the precautionary measures to remove attack traces such as system logs and related information on the victim systems so that they cannot be traced back. Attackers intentionally spread their activities over longer period of time to evade detection. To understand and identify such complex attack, it is required to maintain a secure log records over extended time period. However, preserving the log records for longer time is challenging issue. The system should also ensure the integrity of log files and logging process. In order to overcome these issues, in this paper we are proposing a secure log storage using Blockchain on Cloud platform. Blockchain technology will help to create tamper-proof audit logs. It provides proof of log manipulation and non-repudiation. Cloud platform makes the overall system scalable and support for deep analysis.**

Keywords—Blockchain, Secure Log, Cloud Computing.

I. INTRODUCTION

Network monitoring and log auditing is one of the most important process to keep any organization safe and secure. Network security devices can provide the adequate security but as the attackers keep changing their attacking strategies, security devices also has to be tuned accordingly.

An effective network monitoring and auditing help the administrator to identify the complex attack pattern and attacker's strategy. Based on the analysis of network log report, administrator develop the patches to be deployed in security devices or re-configure the devices accordingly.

Active Network Monitoring and Log Analysis helps to discover many loop-holes in networks which is open intentionally by attackers or unintentionally because of misconfiguration, bugs in the Software, Hardware and Firmware.

Log report gives the detailed information about any access or change for the network resources. However, it has many problems and challenges against advanced persistent threats (APTs).

A major challenge for network monitoring system is, limited window time. Most of the modern network monitoring system focus on real-time detection and its support a short time frame in which attack can be detected. The time frame or duration for which the state of particular network connection is maintained by the security monitoring devices are very small. Knowing this weakness, attackers now a days, cleverly spread their actions over a longer period of time to evade the detection. Thus, it's become important to shift the focus from real-time detection and have the option of archive log analysis.

Though preservation of logs for long time period is important for identification of advanced persistent threats, it also pose some challenges. Log generated by devices may be very small in size, however accumulatively the size will become very large with time. Situation becomes very worrisome, when the network is very big and the log has to be collected from each and every devices. It will become mammoth size data. Storing and analysis of such large size of log data is real challenge. Huge infrastructure and investment required in collecting, storing, sorting, or indexing a large quantity of log messages.

To reduce the cost and effort, cloud based centralized log server could be the best option. Centralized log management using cloud helps the organization to auto-scale the infrastructure based on their need. Once the log data is collected and stored centrally, Big Data System can be used for analysis and identification of various types of threats and attack patterns[5][6].

Big Data system perfectly suits the requirement of storing such a mammoth size of log data. It's not only offer the cost effective and scalable solution for log storage but it also provide a supporting platform for various types of analytics. Normally, it's really very time consuming and tedious to do the log analysis pertaining to specific issue. However, functionality like pattern matching and rule-based approach makes the administrator works easy. The big data analytics tools has great ability to correlate various data of long time span and identify the suspicious activities. It's really a great help to detect the Advanced Persistent Threats or attacks.

The objective of this paper is to explore how to preserve the log in a centralized log server for the longer time period using Blockchain techniques. The paper highlight importance of log preservation and what the benefits are of centralize log analysis

and Blockchain. Further section discuss the advantage of Blockchain and how it is used in the proposed architecture followed by general discussion and conclusion.

II. Literature Review

The effective way of Log storage and analysis has attracted the attention of many researchers. Many scholars has done remarkable work in this direction. Some of the researchers have focused on secure log storage whereas some of them have given priority for analysis and scalability issues.

Holt[2] has discussed about storing the log in cryptographic secure manner so that it cannot be modified or any attempt for modification can be detected. He describes how the process of log creation can be separated from log verification process. The main focus of his research is on the security aspects of the log.

Indrajit Ray et al. described the algorithms for log file preparation and the algorithms necessary for uploading the log data to the cloud[3]. They described a novel scheme for transforming and transmitting the log records to the cloud storage.

Andrew Sutton and Reza Samavi[6] has presented the work on "Blockchain Enabled Privacy Audit Logs". It's a Linked Data based technique for a tamper-proof privacy audit log. It's also help the log auditors for integrity and authenticity verification of the logs.

Researchers have actively contributed on secure storage and analysis of logs, however scalability of the overall system has slightly lost the attention. As the network size increases and number of devices increase, the size of logs also gets exaggerated. It pose a scalability issue for the overall system. Scalability is one of the objective which is addressed in this research work. We are proposing the approach mixed of secure log and cloud based platform. Blockchain is introduced for integrity and security features whereas cloud platform is used for scalability.

III. Importance of Secure Log Preservation

Its normal phenomenon, that whenever an attacker break into system, they do not want to leave any trace for their source or their activities. It may give leads to investigator to catch the attackers. Since the log file contains all the activities carried out by attackers, it's become the most important target for the attack. It's observed that attackers often delete the logs or damage the log files before closing their attack session.

Log records also play a significant role in cybercrime investigation and digital forensic analysis. Regulatory compliance such Payment Card Industry Data Security Standard (PCI DSS) or Health Insurance Portability and Accountability Act (HIPAA) normally mandate that log record should be preserved in forensically sound manner. Only non-tampered log can be considered in a court of law as an evidence for the resolution of any dispute. Hence the preservation of log in a secure manner is not only required for the sound analysis of attack but it's also a legal obligation for the organization [3][4].

IV. Benefits of Centralize Log Analysis

As mentioned in the earlier section, most of the network monitoring devices has short time frame to detect the attack. The attacks which is crafted for slow exploitation and spread their attacking activities for long time spam may not be detectable. Knowing this weakness of the security system, attackers are intentionally crafting such attacks and evading detection. The only way to solve this problem is preservation of logs for the longer time period and do the archived log data analysis.

A combined approach of big data analysis, deep packet analysis, long term behavior analysis and pattern correlation gives a great capabilities to identify and understand the complex attack patterns. Even though attacker spread their activities across several days and months, every event generates a significant log. With the help of good analytical tools, the series of event can be correlated and the complete pattern of exploit can be identified, provided the log data for the complete time span of the attack is preserved in secured manner. The correlation of events for the complete timescale and from the multiple source is crucial for identifying the sophisticated attacks and to protect from the Advanced Persistent Threat.

V. Blockchain

We have proposed the use of Blockchain to make the logging storage safe and secure. Since the storage of logs is from the same network and sub-network components, it can be considered as a private Blockchain network.

The motive behind using the Blockchain is to make the logging system temper proof. It is easy to detect any manipulation and alteration, if the logs are stored using Blockchain. Log stored using Blcokchain cannot be replaced or forged. Storing the log in centralized system using Blockchain also ensure its availability, easy monitoring and analysis.

The centralized log server can be updated only with the consensus of all participating network and sub-network and through a valid approval. The whole process is based on Blockchain consensus algorithm. In the proposed system of private Blockchain, administrator can give relevant accessibility to the participating nodes to do the transaction with central log server. Each participating node can maintain their own copy of ledger, which can be verified with the central log server for its authenticity and integrity.

The whole idea can be implemented with a Blcokchain and Cloud based infrastructure. In next section we briefly describe an immutable log solution that stores the logs with Blockchain.

VI. Proposed System Architecture

The overall architecture of the cloud based log management system using Blockchain is shown in Fig. 1. There are four major functional components in the system.

i.) Log Generators Devices: These are the devices that generate logs and generally stores it in local log server.

ii.) Local Log Server: The local log server collects the logs from various log generator devices connected in network or sub-network. The log data is further processed and normalized in the standard format to be

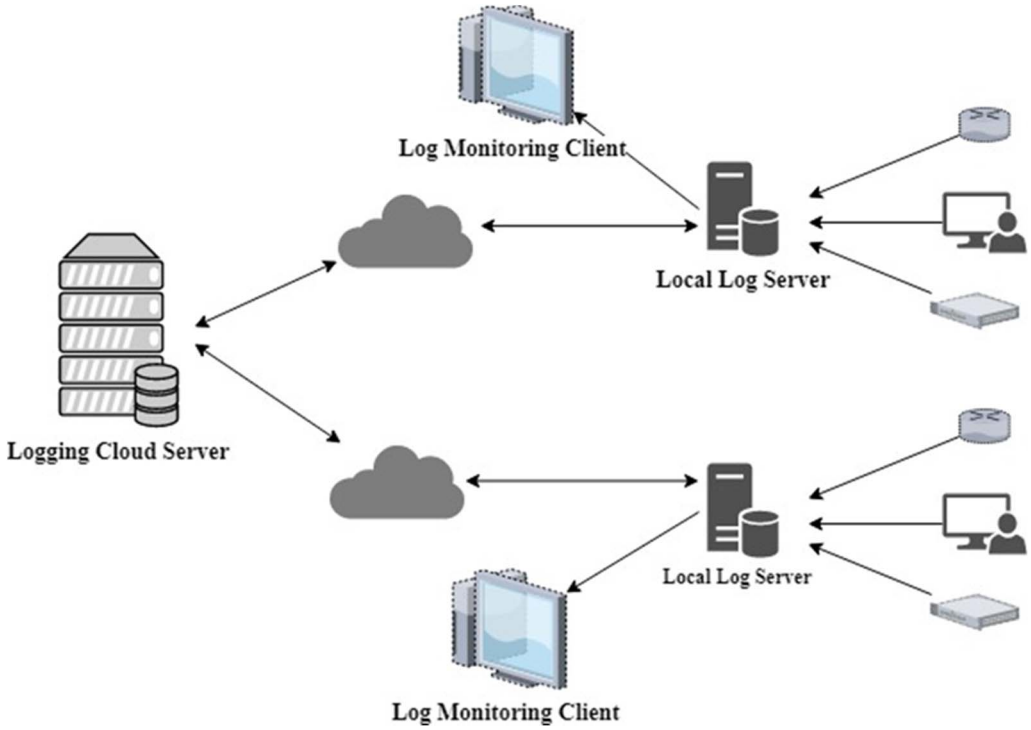

Fig. 1:- System Architecture For Cloud-Based Secure Logging Using Blockchain

pushed in cloud server for long term storage and easy analysis. The local log server upload the log to centralized server in batches, based on schedule or as and when specific threshold value reach. The threshold value can be set in term of log size or in terms of number of log records.

iii.) Log Monitoring Client: These are local hosts which is used by administrator to monitor and analyze the log data from local server as well as from cloud server.

iv.) Logging Cloud Server: The logging cloud server is based on Blockchain techniques. It collects the log from local log server belonging to different Sub-networks and stores it using Blockchain technique for long term archival and analysis.

A. Storage of Logs Using Blockchain

Blockchain makes a tamper-proof immutable chain of records. As per the Blockchain proof of concept, once the block is written (i.e. log record in this case) cannot be modified or change [1]. The complete explanation of how Blockchain work is out of the scope of this paper. However, a simplified approach is explain here.

The Blockchain software makes use of cryptographic algorithm to calculate cryptographic hashes for each block including the checksum of the previous block. A hash function takes inputs of block and generate a unique fixed size alphanumeric string, which is commonly called as 'message digest', 'checksum', 'hash value' or 'digital fingerprint'. Since, previous block's checksum is included for the calculation for new block, it makes the complete chain of block immutable. The whole process makes use of encryption algorithm using public and private key. Each logger have a signing key, which is require to write the logs in cloud log server [1]. In our proposed system the logs are classified in two categories i.e. regular log and the checkpoints.

The normal log generated by the system and devices are stored using Blockchain in local server. As discussed in the previous section each log (Log_i) is chained with the previous block using hash value. The input of the hash function is log entry ($LogInfo_i$) concatenated with the hash value of previous block (h_{i-1}) as shown in Equation 1, where S_k is the private key used by individual logging systems. The whole scheme prevents the log records from intermediate addition, deletion and modifications.

$$Log_i = (LogInfo_i, h_i) \quad where\ h_i = HMAC\ (S_k\ , (LogInfo_{i-1} | h_{i-1}) \tag{1}$$

$$Chekpoint_i = \left(\left(LogInfo_{j\ldots k}, h_{j\ldots k}\right) | Chekpoint_{i-1}\right) where\ h_j = HMAC\ \left(S_k\ , \left(LogInfo_{j-1} | h_{j-1}\right)\right) \tag{2}$$

The overall system is designed to archive the log records for longer time duration. It is obvious that with time, the number

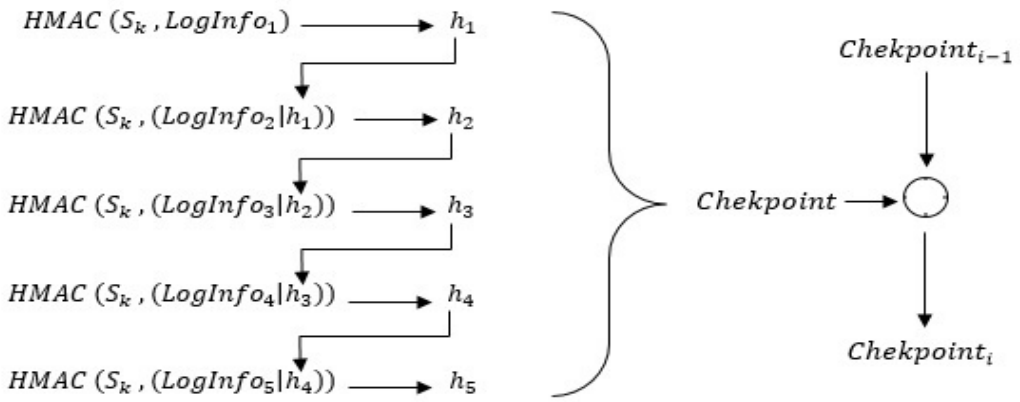

Fig. 2:- Blockchain of Logs and Checkpoint Generation

and volume of log records will multitude and will piles up. In such situation it will be challenging to search and verify the integrity of individual logs from the local server and with cloud server. To solve this issue, idea of checkpoint log is introduced.

The checkpoint logs are special log generated for the set of 'n' logs. A single checkpoint log guarantee the integrity of 'n' logs. The checkpoint log are also generated and stored similar to normal log records using Blockchain.

The administrator can configure the system to generate a checkpoint $Checkpoint_j$ for every certain number of logs. Checkpoints can also be generated based on fix time duration or based on various other threshold values fixed by administrator. Checkpoint generation and arrangement is shown in Fig. 2. Each Checkpoint ($Checkpoint_i$) is chained with the previous Checkpoint ($Checkpoint_{i-1}$) and stored using Blockchain techniques. Such arrangement makes easy for bulk log verifications and integrity identifications.

Input: Log_i

Output: Valid/Invalid

LogValidationProcedure *(Log_i)*
Search $Chekpoint_i$ ***where*** $Log_i \in Chekpoint_i$
CloudHash = ***GetCloudHash***($Chekpoint_i$)
LocalHash = ***GetLocalHash***($Chekpoint_i$)
If *(CloudHash = = LocalHash)*
Return Valid
Else
Return Invalid
End If
End LogValidationProcedure

Fig. 3:- Log Verification and Validation Algorithm

B. Log Verification and Validation

The validation of logs are generally done with the help of Checkpoints. If any log is suspected for tamper, the relevant checkpoint will be identified. The hash value of respective checkpoint will be retrieved (since the hash value of checkpoint is already stored in the next checkpoint, so re-calculation of checkpoint hash do not require) and hash value of the same checkpoint will be retrieved from cloud server. Since the logs and checkpoints are stored in cloud server using Blockchain, its integrity undoubtingly can be trusted. If the hash value of the checkpoints from local server and from cloud server matches, it assure that logs are untampered and intact. The algorithm used to perform the validation is depicted in Fig. 3.

VII. CONCLUSION

Logging plays vital role in keeping the organization safe and secure. However, maintaining logs securely for long time periods is difficult and expensive in terms of the resources needed. The emerging technology like Cloud computing and Blockchain gives economical and trust worthy solution. The architecture proposed in the paper not only gives a safe log storage platform for long time but also gives the capability for in-depth analysis of archived log record to better understand the complex attack pattern.

REFERENCES

[1] Cucurull J., Puiggalí J. (2016) Distributed Immutabilization of Secure Logs. In: Barthe G., Markatos E., Samarati P. (eds) Security and Trust Management. STM 2016. Lecture Notes in Computer Science, vol 9871. Springer, Cham

[2] Holt JE. Logcrypt: forward security and public verification for secure audit logs. Proceedings of the 4th Australasian workshops on grid computing and e□research (ACSW '06), Tasmania, Australia, 2006; 203–211.

[3] I. Ray, K. Belyaev, M. Strizhov, D. Mulamba and M. Rajaram, "Secure Logging as a Service—Delegating Log Management to the Cloud," in IEEE Systems Journal, vol. 7, no. 2, pp. 323-334, June 2013.

[4] J. Buric and D. Delija, "Challenges in network forensics," 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2015, pp.1382-1386.

[5] N. Virvilis, B. Vanautgaerden and O. S. Serrano, "Changing the game: The art of deceiving sophisticated attackers," *2014 6th International Conference On Cyber Conflict (CyCon 2014)*, Tallinn, 2014, pp. 87-97.

[6] Sutton A., Samavi R. (2017) Blockchain Enabled Privacy Audit Logs. In: d'Amato C. et al. (eds) The Semantic Web – ISWC 2017. ISWC 2017. Lecture Notes in Computer Science, vol 10587. Springer, Cham.