BLOCK CHAIN BASED DATA LOGGING AND INTEGRITY MANAGEMENT SYSTEM FOR CLOUD FORENSICS

Jun Hak Park, Jun Young Park, Eui Nam Huh

Department of Computer Science and Engineering,
Kyung Hee University, Yongin-si, South Korea

ABSTRACT

Along with the increasing use of cloud services, security threats are also increasing and attack methods are becoming more diverse. However, there are still few measures and policies to deal with security incidents in the cloud environment. Although many solutions have been proposed through research on digital forensics for responding to security incidents, but it is still difficult to prove the integrity of evidence collection and storage in the cloud environment. To solve these problems, in this paper, we propose a blockchain based data logging and integrity management system for cloud forensics. In addition, compare the performance of the proposed system with the other blockchain based cryptocurrency.

KEYWORDS

Cloud Computing, Cloud Forensics, Block chain, Data Integrity

1. INTRODUCTION

Cloud computing is a technology that provides physical resources to users through virtualization technology. Profit of enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand profit, cloud computing market is getting bigger. Because of these characteristics, the number of users using cloud computing is also increased. However, with the growing cloud computing market, security threats began to grow. Many security solutions for the cloud environment are being researched, it is difficult to apply the existing digital forensic methods because of virtualization technology [1]. When the cloud environment is classified according to the service model, access to some system layers is limited in Software-as-a-Service(SaaS) and Platform-as-a-Service(PaaS) environments, access to that layer is controlled by Cloud Service Provider(CSP). So the log data generated in the inaccessible layer needs to be provided to the CSP through agreements[2]. In traditional digital forensics, investigators have full control over the evidence. However, in a cloud environment, the data centers are distributed globally, Cloud Service Customers(CSC) share physical resources, volatile data that disappears when CSC shut down the instance, virtual network, load balancing and auto scaling for providing seamless service environment. Therefore, it is necessary not only to record data for cloud forensics before a security incident for investigation but also to ensure the

Natarajan Meghanathan et al. (Eds) : ICCSEA, WiMoA, SPPR, GridCom, CSIA - 2017
pp. 149– 159, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71112

integrity of the log data because it is difficult for the investigator to collect the data directly and collect the data from the remote site.

There are several methods for ensuring the integrity of data, one of which is a blockchain. A technique called blockchain or distributed ledger is being studied as a method for ensuring integrity since the previous block affects the value of the next block. Since all blocks are connected like chain, it is possible to verify the integrity of all past blocks simply by verifying the hash value of the immediately preceding block. In this paper, we describe the need for data logging system for cloud forensics and propose a blockchain based data logging and management system for cloud forensics. The paper is structured as follows. Section 2 review about cloud forensics, blockchain and related works for ensuring data integrity. Section 3 describes the proposed system. Section 4 compare the performance of the proposed system with the some blockchain based cryptocurrency. Finally in Section 5 describes conclusions and suggest future research directions.

2. RELATED WORKS

2.1. Cloud Forensics

Cloud forensic is a branch of forensic science encompassing the recovery and investigation of material found in cloud environment, often in relation to computer crime[3]. According to NIST[4], computer forensics consists of four steps: Collection involves the process of physical acquisition of data. Examination is the process of combing through the data for items of interest. Analysis is the application of the interesting items to the investigative question. Reporting describes the output of analysis. The difference between cloud forensics and traditional digital forensics is the collection and identification steps. Because outsourcing resource is one of characteristics in the cloud computing. For the more improve forensic investigation procedure, such as the storage and transportation of data stored in the cloud server is added. Because this is need to guarantee the reliability of such data confidentiality and integrity of data forensic investigation. The problem of applying the collection and identification method of digital forensics to the cloud environment is that the cloud environment is an outsourcing resource to use the desired service or resource from the CSP and it is difficult to know the actual location of the data because of the virtualization technology applied. Therefore, there is a need for a reliable identification and collection method of data that takes into account the characteristics of the environment.

2.2 Cloud Forensic Challenges

Unlike legacy systems that own all of the computing resources, in the cloud computing environment, the CSP provides infrastructure, platform, application. The CSC utilizes the services provided. This structural difference causes many issues in cloud forensic, such as the storage of data and storage locations, and access the data.

The first reason why difficult to apply forensic technology in cloud computing is that data processing is dispersed in large scale of computing resources. Second, in traditional computer forensics, investigators have full control over the evidence. However, it is very hard in cloud environment. Third, there is a lack of reliable evidence as it is difficult to collect evidence due to the multi-tenant features. Fourth, when VM shut down, it is difficult to preserve volatile data.

Fifth, chains of custody might clearly depict how the evidence was collected, analyzed, and preserved. Sixth, investigators are completely dependent on CSPs for acquiring evidence[5]. In order to solve the problem, it might be saved for prevent loss of volatile data (e.g. snapshot). Moreover, collected evidence might be ensured that the integrity has not been manipulated during the process.

2.3 Previous Research on Data Management Methods in a Cloud Computing Environment

In the cloud environment, data centers are scattered around each country, so there is a possibility that users feel that one data but it is distributed among several physical machines actually. Chun, Byung-Gon et al.[6], propose a method to manage data through a replica set that replicates all data in order to prevent data loss in a distributed node environment and to minimize damage. Although this method has the advantage of solving the data loss problem, there is a disadvantage that the data is managed through the replica, which causes a large maintenance cost. Moreover, since the cloud environment has a service model that pay-as-you-go, it is difficult to apply the method as it is.

Nepal, Surya et al. propose a service that guarantees the integrity of data in cloud storage service[7]. The system provides a way to prevent data tampering in a cloud environment by adding an integrity service provider in a scenario where a cloud service user uses a storage service to upload / download data to / from the cloud, called Data integrity as a Service(DIaaS). This service consists of a Key Management Service (KMS) that manages key values, a Trust Management Service (TMS) that ensures trust, and an Integrity Management Service (IMS) that manages integrity of data. In addition, they propose a model that can guarantee the integrity of data by categorizing the CSP and IMS into four cases, which are trust and untrust respectively.

Edorado Gaetani et al.[8] propose a block chain based data management method for cloud federation environment based on the European SUNFISH project, to solve security problems such as data management method and data integrity in the cloud federation environment. Intrinsic goal of cloud federation is sharing services among members by creating regulated, secured inter-cloud interactions. In order to define possible threats in the cloud federation environment and solve the problem of performance degradation due to the application of block chain technology, they devised a two-layer blockchain based database structure. First layer ensures adequate performance by lightweight distributed consensus protocol, second layer ensures strong integrity guarantees by PoW based blockchan methods.

2.4. Block Chains

A block chain is a distributed ledger technique in which a plurality of peers manage and store data by mutually agreed rules. The nodes (peers) that want to manage the data participate in the P2P network and each node can verify the integrity of the block. Each peer can create a block, where the block of the first successful peer propagates to all peers, and if all the peers agree that the block is justified, the block is added to all peers. If the new block is properly created, it means that the verification of the previous block is also completed. Therefore, the longer the block length, the higher the reliability of the entire block. Verification of the integrity of a block can also verify that all past blocks are correct by comparing the hash value. However, this does not guarantee that the block is completely trustworthy, and that it has been acknowledged that it has done a lot of work proofing. Therefore, the more peers participating, the safer it is.

New blocks are created using the Proof of Work (PoW) or Proof of Stake (PoS) method. The PoW method is a task to find a hash value that satisfies a certain condition, and it is operated by adjusting the degree of difficulty for an average of 10 minutes in case of Bitcoin[9]. The PoS method is a method for saving the cost and maintenance cost of hardware equipment and is a concept to solve the problem of PoW method in the field of cryptocurrency[10]. Recently, cryptocurrency has been developed that combines both methods properly due to system maintenance cost and security problems. In addition, research is underway to apply not only cryptocurrency but also the fields that need to guarantee the integrity of data. For example, the blockchain based digital content distribution system[11], using blockchain for medical data access management[12], a framework for preventing double-financing[13], blockchains and smart contratcs for the Internet of Things[14] are researched.

3. LOGGING SYSTEM FOR CLOUD FORENSICS BASED ON BLOCK CHAIN

As mentioned above, the most important consideration for cloud forensics is “how to collect the data?” In cloud computing environment, CSP need to collect and store their own data, in which case there is a possibility of data manipulation and loss, so that the integrity of the data needs to be guaranteed. Therefore, in this section, we propose a system structure that can guarantee the integrity by blockchain technology while CSP collect data itself.

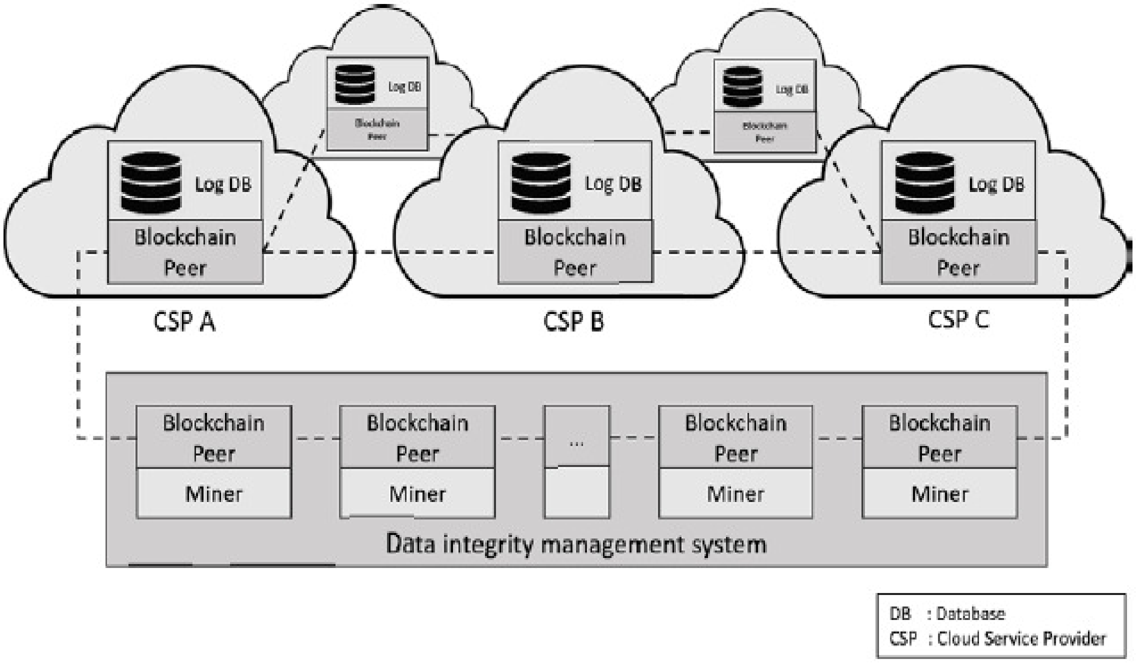

Figure 1. Blockchain based data integrity management system

Figure 1 shows the structure of the blockchain based data integrity management system. In this paper, the data of CSP is stored by itself, but the procedure for verifying the integrity of the data is performed through blockchain. Data that requires storage for cloud forensics is determined through agreement between CSP and CSC. At this time, the CSC should consider the additional cost incurred to store the data in the cloud environment. The collected log data is converted into a hash value through a hash function. These data are used to create a hash tree and construct a block. In the case of permission-less blockchain such as Bitcoin, all peers participating in the network can perform mining to create new blocks. However, this is not suitable for proposed systembecause not all CSP peers can be trusted. Also, if all CSPs participate in mining by PoW method, the proposed system is very inefficient because it needs to consume more computing power than mining power of CSPs. Therefore, only the data integrity management system performs making block and the each block consists of the hash value of the CSP data. One block can contain data of one CSP and the data of CSP participating in the system are stored in order. The generation period of the block is determined by the agreement of the CSPs participating in the system, and it is determined in consideration of the processing performance.

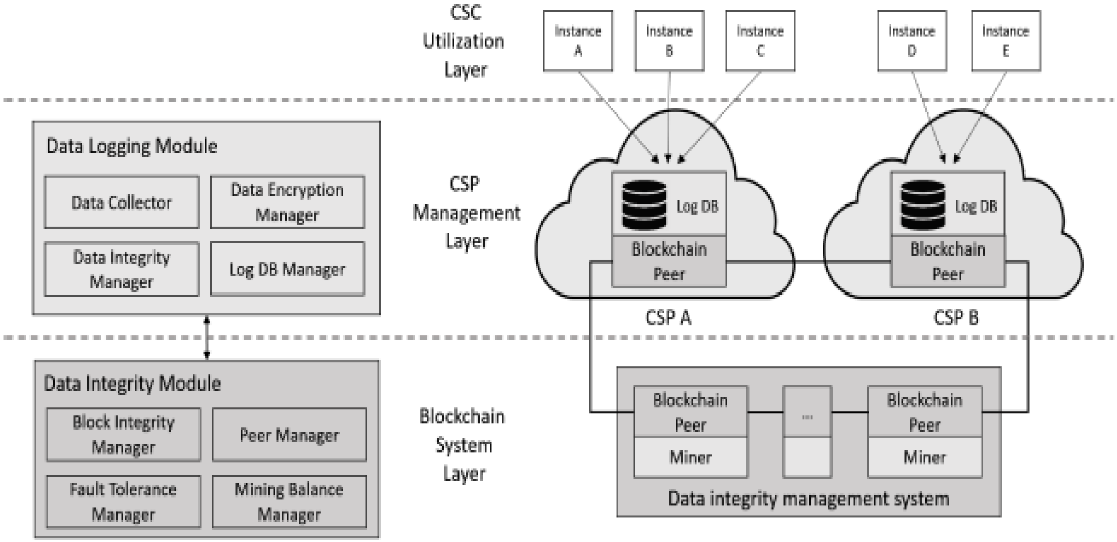

Figure 2. Overview and detailed functions of the proposed system

Figure 2 shows the overall flow and detailed functions of the proposed system. In CSC Utilization Layer, the data of instance used by CSC is stored in the CSP Management Layer, and the Blockchain System Layer manages the integrity of stored data. The Data Integrity Module consist of four functions: Block Integrity Manager, Peer Manager, Fault Tolerance Manager, and Mining Balance Manager. Details of each function are as follows.

3.1 Block Integrity Manger

The Block Integrity Manger performs integrity check on data received from CSP when new block is created. Integrity verification is the process of checking the hash value to see if the encrypted data has not changed and verifying that the data is being sent by that CSP.

3.2 Peer Manger

The Peer Manager monitors the number of CSPs participating in the network and adjusts the number of peers so that the Byzantine Generals Problem does not occur. When a block is created and propagated, it performs a task of maintaining a minimum number of 3n + 1 peers so that n malicious CSP nodes do not interfere with the block generation process. In addition, the size of the block is adjusted to obtain the optimum performance considering the number of peers. It also manages the peers of the CSPs that want to join or leave the network.

3.3 Fault Tolerance Manger

The Fault Tolerance Manger performs tasks such as building a block by solving a fault situation such as branching or consensus falling into a deadlock when stacking blocks.

3.4 Mining Balance Manager

The Mining Balance Manager performs the task of adjusting the cycle of generating block time. If the period is not constant, size of the data in each block may be unbalanced, which may complicate the integrity verification of the data when doing cloud forensics. Therefore, by adjusting the minimum time and maximum time range in which a new block is created, it is possible to stack data of a proper size into one block.

The Data Logging module in CSP Management Layer consist of four functions: Data Collector, Data Encryption Manager, Data Integrity Manager, Log DB Manager Details of each function are as follows.

3.5 Data Collector

The Data Collector performs the task of collecting the service data or log that the CSC requested to be collected. It is recommended that you use a public tool that can be used for cloud forensics when collecting, for example snort to store network packet data.

3.6 Data Encryption Manger

Because data in the cloud environment may be related to the privacy of the CSC, the Data Encryption Manager performs encryption of the data collected by the data collector. Encrypted data is recommended to be encrypted using the CSC's public key.

3.7 Data Integrity Manager

The Data Integrity Manager manages the integrity of data collection and storage. It is difficult to trust CSP's integrity of data in an environment provided by CSP itself. This means that CSP manages the data that is stored before it can be used as evidence for the cloud forensic investigator by this procedure.

3.8 Log DB Manager

The The Log DB Manager performs the task of storing the collected data. The stored data is transmitted to the Data Integrity Management System after performing a hash operation.

4. PERFORMANCE CALCULATION

In this section, we compare transactions per second(tps) with the other cryptocurrency mechanism with our proposed system. One of the reasons why mining-based permission-less cryptocurrency are not used in other area is that the number of transactions per second is too small. The tps formula of cryptocurrency is as follows.

$$tps = \frac{Blocksize}{Blocktime \times Size\ \ of\ \ the\ \ Transaction} \qquad (1)$$

VISA, a credit card company, handles 100,000 transactions per minute in 2016[15]. Compared to that, the tps of the cryptocurrency is too low. For example, 6.41tps for Bitcoin, 15.65tps for Ethereum, 26.67tps for Zcash (unshielded), and 6.67tps for Zcash (shielded). The contents are shown in Table 1.

Table 1. Comparison of TPS of the proposed system with other cryptocurrency

	Blocksize (MB)	Blocktime (sec)	Transaction size (bytes)	tps
Bitcoin	1	600	260	6.41
Ethereum	4.7	14.3	21000	15.65
Zcash(unshielded)	2	150	500	26.67
Zcash(shielded)	2	150	2000	6.67
Proposed System (tps : per CSP)	3.2	600	32	166.67

The proposed system, assuming that uses a permission blockchain such as Hyperledger[16], the manager can revise the chaincode to set the rule. In the proposed system, when the data of one CSP is stored in a cycle of 10 minutes, the transaction of the blockchain system in 10 CSP environment can be thought to occur once a minute. To assume the size of the log data, previous research about security data logging system for cloud forensics proposed by Zawoad et al.[17] each log uses SHA-256 hash function. Therefore, it is assumed that our proposed system also uses that method. Assuming that the log is generated once per second, about 600 logs are created because one block is stored every 10 minutes in one CSP. The hash value of each log can be defined as transaction. If the size of encrypted log is 100byte, size of one block can be 3.2MB including Hyperledger's block header in order to save this log in hash tree. Based on this situation, we calculate about 1667 tps of the proposed method and about 167 tps per CSP because there are 10 CSPs. The graph comparing TPS with other cryptocurrency is shown in Figure 3.

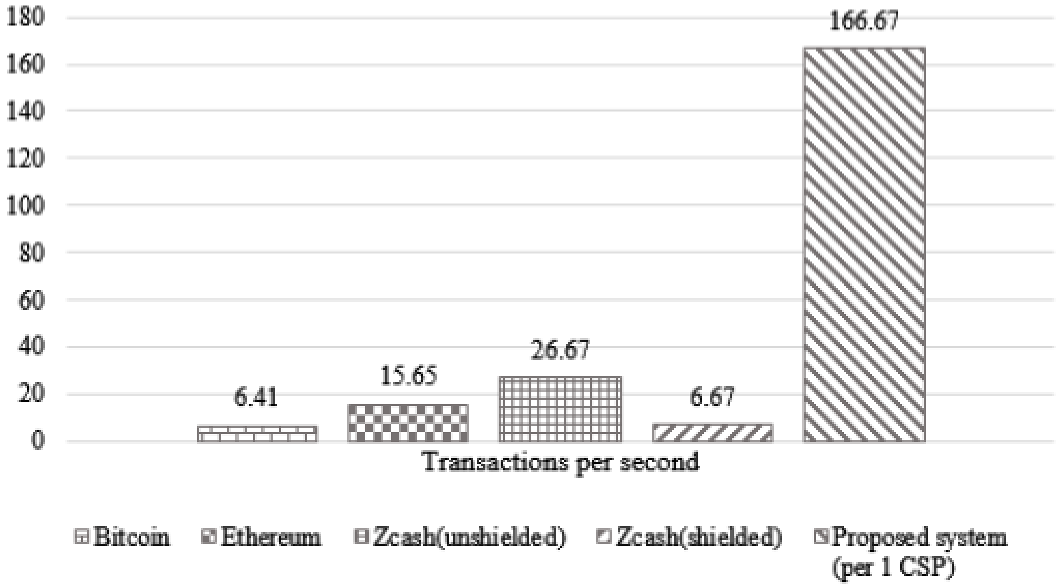

Figure 3. Compare transaction per second with other cryptocurrency

In Figure 4, we check tps according to the number of CSP participating in the proposed system. The horizontal axis represents the number of CSP. In the proposed system, the blocktime and the transaction size of the block chain are assumed to be the same, so the overall TPS shows an increasing tendency. The tps per CSP shows a certain range depending on the number of CSP. This is because the hash tree is organized in a binary tree. If the number of logs is less than the available number of hash tree, the height of the hash tree is expanded. It can be seen that the tps per CSP decreases as the number of logs and the size of the hash tree become equal.

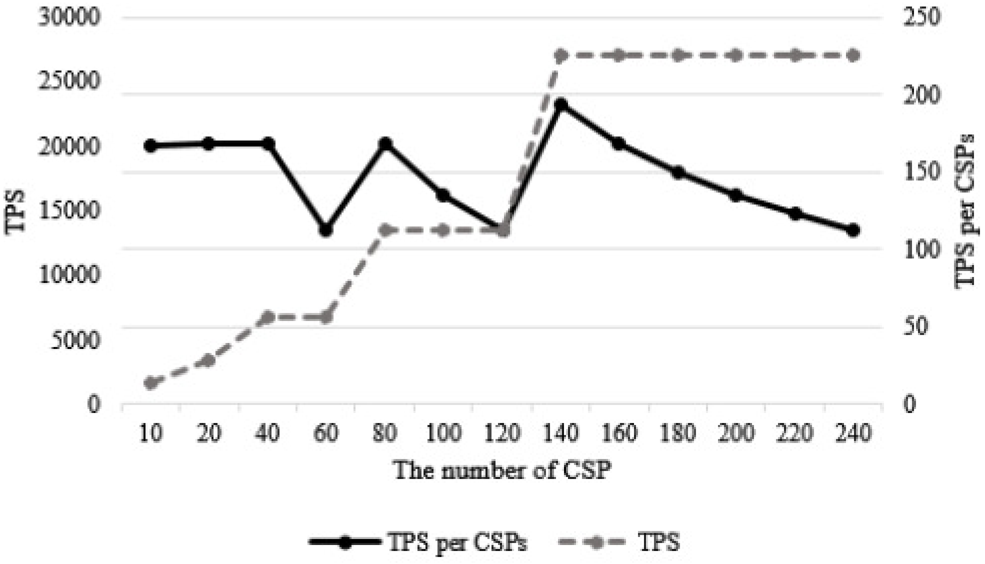

Figure 4. Comparison of tps according to the number of CSP

Permission blockchain such as Hyperledger cannot guarantee the accuracy of the tps because the user can arbitrarily control the size of the block and the block generation period, and the performance may vary depending on how the environment is configured. In addition, the proposed system does not consider the time required for consensus and the time required to process each transaction, so actual performance is expected to be lower. However, the existing PoW-based permission-less blockchain network has a low processing speed because untrusted persons are allowed to participate in the network while maintaining reliability. Thus, the processing speed of a permission block chain can be expected to be faster.

5. CONCLUSIONS

In this paper, we investigate the reason for logging in cloud environment for cloud forensics and propose the permission blockchain based data integrity management system. The proposed system is able to guarantee the integrity of data while processing more transactions than existing permission-less based blockchains. However, there is a limitation that the performance evaluation of the present system can not perform the actual evaluation merely by comparing the calculated result values by calculating the expected data size. The proposed system can be used as one of the methodologies for coping with security incidents in the cloud environment. As future work we collect network data with snort and perform simulation to calculate accurate tps by using Hyperledger. The reason for choosing network data is that cloud environment has a complex network environment due to the virtual network configuration, and there are many incidents that exploit its vulnerabilities. We will also perform a performance evaluation comparing the time required for various consensus algorithms for comparison between permission blockchains.

REFERENCES

[1] K. Kent, S. Chevalier, T. Grance, and H. Dang, “Guide to integrating forensic techniques into incident response,” NIST Special Publication, pp. 80-86, 2006.

[2] Josiah Dykstra, Alan T. Sherman, “Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques”, Digital Investigation 9, 2012

[3] J. Dykstra and A. T. Sherman, “Understanding issues in cloud forensics: two hypothetical case studies,” in Proceedings of the Conference on Digital Forensics, Security and Law, 2011, p. 45.

[4] K. Kent, S. Chevalier, T. Grance, and H. Dang, “Guide to integrating forensic techniques into incident response,” NIST Special Publication, pp. 80-86, 2006.

[5] S. Zawoad and R. Hasan, “Digital forensics in the cloud,” DTIC Document, 2013.

[6] Chun, Byung-Gon, et al. "Efficient Replica Maintenance for Distributed Storage Systems." NSDI. Vol. 6. 2006.

[7] Nepal, Surya, et al. "DIaaS: Data integrity as a service in the cloud." Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011.

[8] Gaetani, Edoardo, et al. "Blockchain-Based Database to Ensure Data Integrity in Cloud Computing Environments." ITASEC. 2017.

[9] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008): 28.

[10] King, Sunny, and Scott Nadal. "Ppcoin: Peer-to-peer crypto-currency with proof-of-stake." self-published paper, August 19 (2012).

[11] Kishigami, Junichi, et al. "The blockchain-based digital content distribution system." Big Data and Cloud Computing (BDCloud), 2015 IEEE Fifth International Conference on. IEEE, 2015.

[12] Azaria, Asaph, et al. "Medrec: Using blockchain for medical data access and permission management." Open and Big Data (OBD), International Conference on. IEEE, 2016.

[13] Oudejans, Joris, and Zekeriya Erkin. "DecReg: A Framework for Preventing Double-Financing using Blockchain Technology." Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts. ACM, 2017.

[14] Christidis, Konstantinos, and Michael Devetsikiotis. "Blockchains and smart contracts for the internet of things." IEEE Access 4 (2016): 2292-2303.

[15] Jan Vermeulen, “VisaNet – handling 100,000 transactions per minute” [Online]. Available: https://mybroadband.co.za/news/security/190348-visanet-handling-100000-transactions-per-minute.html. 2016.12.

[16] Cachin, Christian. "Architecture of the Hyperledger blockchain fabric." Workshop on Distributed Cryptocurrencies and Consensus Ledgers. 2016.

[17] Zawoad, Shams, Amit Kumar Dutta, and Ragib Hasan. "SecLaaS: secure logging-as-a-service for cloud forensics." Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013.

AUTHORS

Jun Hak Park

Jun Hak Park received B.S. degree in Department of Computer Science and Engineering from Kyung-Hee University, South Korea, in 2016. He is pursuing his M.S. degree in the Department of Computer Science and Engineering at Kyung-Hee University, South Korea. His research interests include Cloud Computing, Cloud Forensics, and Blockchain.

Jun Young Park

Jun-Young Park received his B.Eng. degree in Computer Engineering from Hannam University, Korea, in 2010, and a Master’s degree in Computer Engineering from the Kyung Hee University, Korea in 2012, He is currently working toward a Ph.D. degree in the Department of Computer Science and Engineering at Kyung Hee University, Korea. His research interests include cloud computing, mobile cloud computing, cloud computing security, security-as-a-service.

Eui-Nam Huh

Eui-Nam Huh earned a B.S. degree from Busan National University in Korea, a Master's degree in Computer Science from the University of Texas, USA in 1995, and a Ph.D. degree from the Ohio University, USA in 2002. He is the director of Real-time Mobile Cloud Research Center. He is a chair of Cloud/BigData Special Technical Committee for Telecommunications Technology Association(TTA), and a Korean national standards body of ITUT SG13 and ISO/IEC SC38. He was also an Assistant Professor at Sahmyook University and Seoul Women's University, South Korea. He is now a Professor in the Department of Computer Science and Engineering, Kyung Hee University, South Korea. His research interests include Cloud Computing, Screen Contents Coding(Cloud Streaming), Internet of Things, Distributed Real-Time Systems, Security, and Big Data