Available Online at http://www.journalajst.com

ISSN: 0976-3376

ASIAN JOURNAL OF
SCIENCE AND TECHNOLOGY

Asian Journal of Science and Technology
Vol. 10, Issue, 04, pp.9642-9646, April, 2019

RESEARCH ARTICLE

AN IN-DEPTH SURVEY ON INTEGRITY AUDITING AND INFORMATION HIDING IN CLOUD

[1]Roma Joseph and [2, *]Samadhan Sonavane

[1]Computer Science and Engineering, Sandip University, Nashik, India
[2]Professor, Department of Computer Science and Engineering, Sandip University, Nashik, India

ARTICLE INFO

Article History:
Received 27th January, 2019
Received in revised form
24th February, 2019
Accepted 28th March, 2019
Published online 30th April, 2019

Key words:
Cloud Computing,
Sensitive information hiding,
Data auditing.

Corresponding author:
Samadhan Sonavane

ABSTRACT

Due to the enormous size and complexities of cloud there is been always a threat to the data from the external and internal entities. Internal entities are generally are accountable for the data theft, this leads to the third party to audit the data integrity. Whereas the external entities are exploits the sensitive information stored in the cloud through the intelligent query fired on the cloud. There are many methodologies and tools are available to achieve the any one of them out of these two major issues and very few are there to tackle both sensitive information hiding and data integrity together. This research paper introduces the idea of handling both the techniques by analyzing the other past works in depth.

***Citation: Roma Joseph and Samadhan Sonavane,* 2019.** "An in-depth survey on integrity auditing and information hiding in cloud", *Asian Journal of Science and Technology*, 10, (04), 9642-9646.

INTRODUCTION

The term cloud originated through chance as the network flow charts and diagrams during the earlier stages of development of the internet, predominantly represented the internet as a big cloud. And that just stuck. So basically, the basic understanding is that the cloud is almost shorthand for the internet. It couldn't be further away from the truth. The cloud must have started off as a service which required an internet connection to work, but it has since evolved to be one of the most influential of the technologies in the current times. Most of the cloud's capabilities require an internet connection, but referring to the cloud as the internet itself would be a misnomer. The cloud is fundamentally an advanced form of a server that provides an almost limitless potential. Data can be stored on to the cloud with the added benefit of being able to access it anywhere in the world through any device connected to the internet, unlike your regular local storage. And that is one of most basic utilizations of cloud. To be able to store the data on a server and access it on any device without being limited to your local storage or the hard drive is a an immensely useful feature. A great deal of data processing and handling transpires unbeknownst to the user, to be able to achieve such basic functionality. One of the most widely used and yet, almost invisible to the end-user is IaaS which stands for Infrastructure as a Service, wherein cloud based companies provide a foundation for other businesses to integrate their own services on their cloud servers.

Just like the popular and well-known content streaming service deploys its content on a backbone provided by a cloud computing service company. The business industry is the biggest consumer of the cloud. Unlike an average consumer, they utilize the cloud for drastically different purposes. The businesses utilize the cloud for innovative applications, for example, Software as a Service or SaaS enables them to use an application which resides and computes in the cloud, offloading resources and time as they're handled by the cloud, which eliminates installing and executing the same application on powerful and expensive machines every time they planned to implement it. This technique is extremely beneficial to the organization by minimizing overhead expenses and provides added benefits of being available anywhere and anytime ubiquitously. The cloud also implements PaaS, which refers to Platform as a Service, where companies can design and deploy, their own applications and utilize the cloud's computing capabilities to execute those applications. This particular service enables a larger degree of control over the scalability and maintenance of the product thereby facilitating the organization to focus their resources on more pressing issues. The Cloud has unknowingly transformed into an inseparable part of our life's, it has made our life convenient and reduced the consumption of time, money and energy that would've been squandered if the cloud servers didn't exist. It has an immense impact on a person's day to day life without being too intrusive. Cloud storage facilities provide a very reliable and dependable storage, unlike local storage which are more prone to crashes and data corruption. Since the advent of the age of internet, data security has been of paramount importance. The internet was conceived in late 1960's, specifically designed to enable fast communication for the researchers and military. Therefore, data hiding was an immediate requirement even at the inception of the internet we know today. Even today, data confidentiality is a major security concern in the field of cloud computing. As the cloud is a ubiquitous service which is good for personal convenience, but that comes with the understanding that if you can access your data from everywhere, so can anyone else like an attacker or thief. Therefore, the biggest assets of cloud computing is also its security loophole that can be exploited by someone with the tools to do so. A number of methods have been proposed to ameliorate this effect, data hiding is one the most widely used and easily applicable in many situations. Various techniques of data hiding are available such as encryption of the data, as anyone without the correct encryption key cannot decipher your data even if they have acquired your data. But this method is time consuming and requires a lot of computing power to encrypt and decrypt data every time you want to access it. On a slower machine of the client, it can prove to a nightmare.

Therefore, steganography is rightful alternative, it is different from encryption as it does not completely change your data into a stream of scrambled elements that would take considerable processing power to compute. Steganography masks the data present, it does it in a way that the masking is not visible to the naked eye and can also pass off as legitimate data to the intruder. This process is not CPU intensive and provides comparable levels of security than encryption, and in essence is a better form of Data hiding in the Cloud. Storing large amounts of data on the cloud is subject to a slew of tampering, the cloud administrator may move it around, dishonest cloud providers, intruders trying to tamper with your data. All of these scenarios will reduce the data integrity of the database in the cloud. If the data gets corrupted through time, it would be of no use after a certain amount of time and is a pressing concern for the sensitive data stored on the cloud. Therefore, regular auditing of the cloud data must be performed, to keep the databases in a healthy condition. Data Auditing looks for errors and data corruption on the cloud servers and attempts to correct them before a catastrophic data loss takes place. As the process of data auditing is a complex and required to be performed regularly to avoid losing the database integrity, the cloud companies outsource the Data Auditing process, independent companies then exchange keys as most of the cloud data is cryptographically encrypted, then they perform it on their behalf. Data Auditing is a very essential part of the Cloud maintenance as failing to perform one can lead to irreparable losses, data corruption and loss of integrity, which are extremely undesirable as a part of the functioning of the Cloud.

In cryptographic it is the technique for generating keys by using algorithms. The generated key is utilized to encrypt and decrypt information, data, and communication. To generate key web tool or key generator software is utilized to generate keys in an alphanumeric sequence which inform an installer software that the end user who installs software is the owner of the license. Two distinguish types of algorithm are utilized to generate two types of Keys. First one is a Symmetric Encryption Algorithm and the second one is the Public Key Encryption Algorithm. Symmetric Encryption Algorithm generates keys for DES and AES, this utilizes the identical cryptographic keys for both plaintext encryption and ciphertext decryption. Symmetric is mostly used in an organization like the military, big financial corporations, and governments for their communication. The public key encryption method is more practical then symmetric encryption algorithms, in public key system different keys are utilized for both encryption and decryption. A private decryption key is an address to each receiver and they need to publish a public key known as the encryption key. Authentication assurance of public key is also requiring for avoiding spoofing by the adversary. In public key encryption, the deduction of plaintext from cipher text is very difficult due to the complexity of the encryption algorithm. Maximum latest software has another way of validation than a product key to make sure that the software is legally certified and no longer pirated. A key generator may additionally allow the person to install in the software however confirmation above on the internet would then prevent the software from the operation. But, crackers and hackers utilize extra than the key generator which will illegally utilize software program. Some keygens are prepared with parody servers that cut off the conversation among the software program and the real servers, giving it with the confirmation respond waiting for from the real servers, thereby tricking the software into questioning that it has been confirmed. To protect unauthorized access to websites, database, and computer, some digital privacy action is applied that is called Data Security. It also prevents data corruption and it is a necessary step for every big or small software companies. It is also referred to as Computer security or information security. Data erasure, backups and data masking are some of the technologies related to Data security. The main technology measure of data security is encryption, where important data are hiding from unauthorized users. Most used encountered technique for data security is authentication. Authentication provides the user biometric data, password, code or data in some other forms to confirm user identity before login to the system is granted. Cloud computing is defined as the application and services like servers, platforms for software development, storage above the internet.

The different data centers across the world offered these services which together called as the “cloud”. A general argument from critics is that cloud computing can't prevail as it means that companies have to lose manage in their data, inclusive of an email issuer that saves records in multiple locations across the world. A huge regulated industry, like a bank, is probably required to keep data inside the USA at the same time as this isn't always an insurmountable issue, it demonstrates the kind of problem that some agencies might also have with cloud computing. So the data safety in cloud computing is a very important and superior to safety for the corporate data center. There are various technologies, policy, and controls are established to safe data, associated infrastructure, and applications related to cloud computing which all combined referred to as "cloud security". When an organization chooses to save data on the common public cloud, it's losing its potential to have physical entry to the servers hosting its information, which increases the threat of insider attacks to the data security in the cloud. There is numerous security risk are linked with the cloud data services, some are traditional like DOS attacks, network eavesdropping and illegal invasion and some are particular cloud computing risk such as abuse of cloud services, virtualization vulnerabilities, channel attacks etc. So, the cloud facility providing companies take some major step to control the data security threats like, confidentiality of data means to prevent illegal users to access data contents, access controllability measure control access for information i.e. outsourced by data owner to the cloud and data integrity which define as integrality and correctness of data stored in cloud. In this paper, section 2 is dedicated for literature review of past work and Finally Section 3 concludes this paper.

LITERATURE REVIEW

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

Ren *et al.* (2012), proposed a methodical public auditing protocol for outsourced data with the unique dynamic formation in the cloud. The performance of the presented structure is much better than the state of the art. To establish mutual trust amongst data owners (Dos) and cloud service providers (CSPs), verification of global and sampling is presented. New Dynamic formation provides the data dynamics efficiently. Their protocol also overcomes the common basic challenges of cloud auditing such as lazy update, batch auditing, and block less verification. Experimental outcome and analytical analysis proved that the presented protocol provide the required efficiency in practice.

Ateniese *et al.* (2007) presents two efficient PDP (provable data possession) schemes. Their model required to minimize the server computation, file block access, and communication amongst client and server. They implemented their E-PDP scheme (which incurs constant overhead on the server) and protocols for remote data to examine and compare their performance. Experimental results proved that the probabilistic possession assurance makes it practical to confirm ownership of the big data sets.

Shen *et al.* (2017) developed the latest cryptographic structure called as PORs (proof of retrievability). In POR the users authorized to search out what prover contain data object F and or a file. Greater exactly, a successfully accomplished POR confirm a verifier that the protocol interface is given by prover and through which F is completely retrieved by the verifier. But the prover refused to free F even after successfully cooperating in a POR. The author discloses that a POR can be efficient sufficient to offer everyday checks of document retrievability. Therefore, as a well-known tool, a POR can supplement and make stronger any of a style of archiving architectures, which include those that involve information dispersion.

Wang *et al.* (2013) proposed a system for security of storage data in cloud computing named as privacy-preserving public auditing system. They use the random masking and homomorphic linear authenticator to confirm that any information of data content that is saved in cloud server is not learned by TPA during the auditing process. This reduces the cloud user burden from the expensive and tedious auditing task, it also relieves users fear for their data leakage.

Worku *et al.* (2014) proposed a scheme for public auditing that is more reliable and has superior performance then privacy-preserving public auditing method. The proposed public auditing method contains a TPA (third party auditor), whose work is to perform auditing of data on the user's behalf. The TPA handles data auditing for many users simultaneously. They minimize the bilinear mapping to improve their system performance. The analysis of their system on the ground of performance and security they proved that their system is more efficient than the previous one.

Shen *et al.* (2017) proposed a cloud storage auditing scheme for group users, which is lightweight and it lessens the calculation load on the user side. They include TPM (Third Party Medium) which perform time taking operations on user's behalf. The TPM generates authenticators for users and approves the integrity of data on user's behalf. They use simple operations to blind data auditing and uploading phase, this protects data privacy averse to the TPM. The users need not to conduct decryption operation while utilizing cloud data. They also set the authorization expiration time for TPM to ensure that TPM performs their operation within given time.

Ateniese *et al.* (2008) present a comprehensive design of a lightweight and safe PDP scheme which depend on symmetric key cryptographic. Their scheme doesn't support third-party verification due to its dependency on symmetric key cryptographic.

Wang *et al.* (2011) explored the complication of providing concurrently data dynamics and general auditability for faraway data probity in cloud computing. Their structure is designed to meet these two major goals. They improved the storage models existing proof by handling the classic structure of Merkle Hash Tree for confirmation of block tag. They also explore the method of bilinear aggregate signature to favor the handling of several auditing tasks efficiently and they extend their main outcome into a numerous user setting, where several auditing tasks performed by TPA simultaneously. The analysis of performance and security of their structure prove that the suggested structure is safe and highly efficient.

Yu *et al.* (2016) present the design of cloud storage auditing. In their paradigm key-upgrade operations aren't completed by the client, however, it is completed by a verified party. The verified party holds a secret encrypted key of the client for cloud storage auditing and updates it underneath the encrypted state every time. The secret encrypted key is downloaded by the client from the verified party and the client decrypts it when he needs to upload files to the cloud. The TPA plays the role of the verifier party who is responsible for key updates and they also examine the integrity of files which are stored in the cloud by the clients.

Yu *et al.* (2018) present the IRIBS (Intrusion Resilient Identity Based Signature) scheme which based on the framework of the FSIBS scheme, in which to binary tree framework is utilized to join time periods. To revive secret key in the single time period they utilize homomorphism framework in the key upgrades. This structure increases the efficiency of the proposed scheme. The author presents the generic framework of IRIBS scheme. Their generic framework needed the separable framework in FSIBS scheme amongst key materials used by users for updating and for real signing. They also give indirect safety proof to their IRIBS scheme.

Yang *et al.* (2016) proposed a unique public auditing scheme in cloud storage for shared data, which favor identity traceability and privacy. They also present the numerous security needs for public auditing scheme. In their scheme, the identity of group members is unknown to the TPA. If any dispute happened only the group manager has the right to open the identity of the fraudulent member. They also proved the security and efficiency of their scheme by concrete implementations.

Wang *et al.* (2015) present Panda, a unique public auditing technique for the shared information nobility with systematic user abrogation in the cloud. They implemented the plan of proxy re-signatures in their system due to which the blocks are resigned by the cloud and revoked user signed the block with re-signing key. This increases the user revocation efficiency and existing user resources for communication and calculation are easily saved. More so their scheme not only reinforces task of batch auditing and multiple auditing simultaneously but also reinforce data sharing between numerous users efficiently.

Wang, (2015) designed the identity-based distributed provable data possession (ID-DPDP) protocol which depends on bilinear pairings and it is provable safe under the speculation that the computational Diffie-Hellman (CDH) difficulty is tough. The author also formalizes the security and system model of ID-DPDP. The presented model also realize the delegated verification, public verification, and private verification depend on client authorization.

Yu *et al.* (2017) proposed the new protocol for RDIC (remote data integrity checking) which is identity based, by utilizing key-homomorphism cryptographic primitive. This lessens the complexity of the system and cost for managing and establishing a public key verification structure in PKI-depend RDIC scheme. They standardize Id-depend RDIC and its safety model, which includes safety against a harmful cloud server and no information policy against a TPV. The proposed protocol leak no stored knowledge to the third party verify throughout the RDIC procedure. They proved that the suggested protocol is efficient, practical and safe in the actual word implementation.

Zhang *et al.* (2018) design a unique identity depend on cloud storage auditing scheme for shared information, that supports actual systematic user revocation. The user revocation does not dependent on the gross amount of file blocks which revoked user possessed in the cloud. To achieve this they utilize unique key generation strategy, in which group's identity information replaced the group's public key for a whole lifetime. The two components are utilized to generate the group's private key, one component is fixed from the time it issued and another one is changed with user revocation. A unique update method for private key is also presented to favor user revocation.

Li *et al.* (2016) proposed See Cloud and See Cloud+ to achieve data integrity and deduplication in the cloud. In See Cloud before uploading, the client is able to create data tags and it also introduced verification of possession protocol to enable safe deduplication, that prevent side channel knowledge leakage throughout data deduplication. User computation during auditing and file uploading phases are also reduced in See Cloud. Seecloud+ is the improved structure that based on the reality that the user wants to encrypt their data earlier than uploading and permit auditing and safe deduplication in already encrypted data.

Ateniese *et al.* (2005) introduced the concept of sanitizable signatures that have numerous safety features for emerging and current applications. This concept permits the modification in verified semi-trusted censors in a manageable way without communicating with the real signer.

Ateniese *et al.* (2005) present the formal safety model for chameleon hash functions. This model contains accurate description of the properties of information hiding and key exposures. They decided that the single trapdoor scheme is not enough for the framework of a chameleon but double trapdoor mechanism is needed. Their outcome contains three frameworks of schemes that satisfied the safety model fully, one based on pairings and two on RSA.

Li *et al.* (2017) introduced the framework of fuzzy identity depend on information auditing protocol that used the biometrics as fuzzy identity. In the proposed structure, the private key binds with one identity and it can be utilized to check the accuracy of the response produced with a second identity. Safety of the model was also revealed by the author in the selective Id model.

CONCLUSION

This paper analyzed all the past works on sensitive information hiding and Data auditing technique in the cloud. And thereby come to a conclusion that most of the methodologies are dealing with one of the methodologies and many suffer from time and space complexities. So this paper decides to deal with the concept of bilinear pairing for data integrity which not only checks the integrity of the data and also retain the original data. And also this research paper works on correlation based sensitive information hiding scheme, which will be reflected in our coming edition of research articles.

REFERENCES

Ateniese G. and B. de Medeiros, “On the key exposure problem in chameleon hashes,” in Security in Communication Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 165–179.

Ateniese G., D. H. Chou, B. de Medeiros, and G. Tsudik, “Sanitizable signatures,” in Proceedings of the 10th European Conference on Research in Computer Security, ser. ESORICS’05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 159–177.

Ateniese G., R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable data possession at untrusted stores,” in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS ’07, 2007, pp. 598–609.

Ateniese G., R. D. Pietro, L. V. Mancini, and G. Tsudik, “Scalable and efficient provable data possession,” in Proceedings of the 4th international conference on Security and privacy in communication networks, 2008, pp. 1–10.

Li J., J. Li, D. Xie, and Z. Cai, “Secure auditing and deduplication data in cloud,” IEEE Transactions on Computers, vol. 65, no. 8, pp. 2386–2396, Aug 2016.

Li Y., Y. Yu, G. Min, W. Susilo, J. Ni, and K. K. R. Choo, “Fuzzy identity-based data integrity auditing for reliable cloud storage systems,” IEEE Transactions on Dependable and Secure Computing, 2017. [Online]. Available: DOI:10.1109/TDSC.2017.2662216.

Ren K., C. Wang, and Q. Wang, “Security challenges for the public cloud,” IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, Jan 2012.

Shen J., J. Shen, X. Chen, X. Huang, and W. Susilo, "An efficient public auditing protocol with novel dynamic structure for cloud data," IEEE Transactions on Information Forensics and Security, vol. 12, no. 10, pp. 2402–2415, 2017.

Shen W., J. Yu, H. Xia, H. Zhang, X. Lu, and R. Hao, "Light-weight and privacy-preserving secure cloud auditing scheme for group users via the third party medium," *Journal of Network and Computer Applications*, vol. 82, pp. 56–64, 2017.

Wang B., B. Li, and H. Li, "Panda: Public auditing for shared data with efficient user revocation in the cloud," IEEE Transactions on Services Computing, vol. 8, no. 1, pp. 92–106, Jan.-Feb. 2015.

Wang C., S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," IEEE Transactions on Computers, vol. 62, no. 2, pp. 362–375, 2013.

Wang Q., C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 5, pp. 847–859, May 2011.

Wang, H. "Identity-based distributed provable data possession in multicloud storage," IEEE Transactions on Services Computing, vol. 8, no. 2, pp. 328–340, 2015.

Worku S. G., C. Xu, J. Zhao, and X. He, "Secure and efficient privacy-preserving public auditing scheme for cloud storage," *Comput. Electr. Eng.,* vol. 40, no. 5, pp. 1703–1713, Jul. 2014.

Yang G., J. Yu, W. Shen, Q. Su, Z. Fu, and R. Hao, "Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability," *J. Syst. Softw.,* vol. 113, no. C, pp. 130–139, Mar. 2016.

Yu J., K. Ren, and C. Wang, "Enabling cloud storage auditing with verifiable outsourcing of key updates," IEEE Transactions on Information Forensics and Security, vol. 11, no. 6, pp. 1362–1375, June 2016.

Yu J., R. Hao, H. Xia, H. Zhang, X. Cheng, and F. Kong, "Intrusion-resilient identity-based signatures: Concrete scheme in the standard model and generic construction," *Information Sciences,* vol. 442-443, pp. 158 – 172, 2018.

Yu Y., M. H. Au, G. Ateniese, X. Huang, W. Susilo, Y. Dai, and G. Min, "Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage," IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, pp. 767–778, April 2017.

Zhang Y., J. Yu, R. Hao, C. Wang, and K. Ren, "Enabling efficient user revocation in identity-based cloud storage auditing for shared big data," IEEE Transactions on Dependable and Secure Computing, 2018. [Online]. Available: DOI:10.1109/TDSC.2018.2829880.

****