International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
www.ijresm.com | ISSN (Online): 2581-5792

Effective and Enhanced Data Security with Encrypting, Hashing and Chunking Using MHT and Data Retrieval System in Cloud

M. Dhivya[1], P. Bharathi[2], B. Saranya[3], S. Sharannya[4]
[1]*Assistant Professor, Dept. of Computer Science & Engineering, Panimalar Inst. of Technology, Chennai, India*
[2,3,4]*Student, Department of Computer Science and Engineering, Panimalar Inst. of Technology, Chennai, India*

***Abstract*: In the Existing system, NFV (Network Function Virtualisation) is incorporated which is a network security pattern based on security deployment constraints. In the Proposed system, an identity-based dynamic data auditing scheme which supports dynamic data operations, including modification, insertion and deletion. As far as we know, there's no different identity-based statistics auditing scheme which helps dynamic operations. In particular, to achieve efficient dynamic operations, we use Merkle hash tree data structure for block tag authentication, which helps update data with integrity assurance. The Analyses of security and performance reveal that the proposed scheme is efficient and secure. In the modification process, we are implementing a system to recover the crashed files automatically. For that we are having TPA Third Party Auditing to monitor the file crashes TPA will intimate the user that filer crashed. After that we are using the concept Erasure code, it will recover the file from replica.**

***Keywords*: Data Security, Encrypting, Hashing, Chunking, MHT, Data Retrieval System, Cloud.**

1. Introduction

The development of the information society, big data has become an important strategic resource, influencing the development of different fields. In certain area, a large amount of data is produced every day, stored by various institutions, serving as an intangible social wealth. If these institutions share data with each other, we can maximize the value of data through fully data mining and analyzing, thus improving knowledge discovery and promoting the development of smart products. To achieve this goal, a big data platform should be established for data storage to be shared among various institutions. Many institutions no longer keep data locally and instead they store the data on the platform with an easy and convenient approach of accessing and updating. However, the convenience comes with challenges. Considering various threats such as malicious attackers tampering data or hardware damages, data integrity becomes a major concern of all users. A service should be provided for users to ensure that the data is correctly and accurately stored in the remote platform. Is there any manner to check the integrity of data, mainly privacy information in a large facts platform? especially privacy data in a big data platform? Remote data auditing is an effective way and can solve this problem by means of remote data integrity verification without downloading all storage data. The existing remote data auditing schemes mainly use public key infrastructure (PKI) and identity-based cryptography. However, PKI has the drawback of overspending on key management, so identity-based cryptography is more commonly used. After analyzing the existing schemes, we found that there is none of identity-based remote data auditing schemes which can support dynamic auditing. Moreover, these schemes cannot be extended to dynamic auditing scheme directly. In those schemes, tag technology is linked to the index of data block. For example, if an information block m2 is inserted after m1, for every block whose index is behind, its tag needs to be recomputed, which leads to massive computation cost. So these schemes cannot be prolonged to dynamic auditing scheme directly. Based on the scheme, we advise an identity-based totally dynamic records auditing scheme for big data storage.

2. Literature survey

Z. J. Xu says that Massive Internet invasions implemented through the distributed platform fabricated by rapid diffusion of malwares, has emerge as a great issue in community safety. Based on the Cloud SEC architecture, both statistics distribution and challenge scheduling overlays may be simultaneously applied in a loosely coupled fashion, which can efficaciously retrieve facts sources from heterogeneous network security facilities, and harness disbursed series of computational assets to process facts-in depth tasks.

S. T. Zargar says that with the developing recognition of cloud computing, the exploitation of viable vulnerabilities grows at the identical pace; the distributed nature of the cloud makes it an attractive goal for ability intruders. Despite protection issues delaying its adoption, cloud computing has already emerged as an unstoppable force; thus, protection mechanisms to ensure its steady adoption are an immediate want. Here, we recognition on intrusion detection and prevention systems(IDPSs)to guard towards the intruders. In this paper, we suggest a Distributed, Collaborative, and Data-

driven Intrusion Detection and Prevention system (DCDIDP). Its purpose is to utilize their sources within the cloud and offer a holistic IDPS for all cloud provider vendors which collaborate with other peers in a disbursed way at different architectural stages to reply to attacks.

A. R. Khakpoura says that to lessen the complexity and value in deploying and dealing with firewalls, companies have started out to out supply the firewall provider to their Internet Service Providers (ISPs), together with AT&T, which offer cloud-primarily based firewall service. Such firewalling model saves groups in managing, deploying, and upgrading firewalls. The modern-day firewall provider outsourcing version requires companies fully trust their ISPs and provide ISPs their firewall regulations. However, organizations typically need to keep their firewall rules confidential. In this paper, we advise the first privacy retaining firewall outsourcing technique where agencies outsource their firewall services to ISPs without revealing their firewall rules to the ISPs. By no way that we declare our scheme is perfect; however, this attempt represents the first step closer to privacy preserving outsourcing of firewall offerings.

Saman Taghavi Zargar, says that DDoS flooding attacks are commonly express attempts to disrupt valid users get entry to services. Attackers typically gain access to a big range of computers by exploiting their vulnerabilities to setup attack armies (i.e., Botnets). Once an attack navy has been setup, an attacker can invoke a coordinated, massive-scale attack in opposition to one or greater targets. Developing a comprehensive protection mechanism towards identified and expected DDoS flooding attacks is a desired goal of the intrusion detection and prevention research network. However, the improvement of the sort of mechanism calls for a complete know-how of the hassle and the techniques that have been used to date in preventing, detecting, and responding to various DDoS flooding assaults. In this paper, we explore the scope of the DDoS flooding attack trouble and tries to fight it. We categorize the DDoS flooding assaults and classify present counter measures primarily based on wherein and once they prevent, detect, and respond to the DDoS flooding attacks. Moreover, we highlight the need for a comprehensive allotted and collaborative protection approach. Our primary goal for this paintings is to stimulate the studies community into growing creative, effective, efficient, and complete prevention, detection, and reaction mechanisms that address the DDoS flooding trouble before, for the duration of and after a real assault.

3. Existing system

Identity-primarily based remote records auditing schemes can verify data integrity and provide a simple identity authentication and management for a couple of users with the subsequent disadvantages of expanded ready time, occurrence of Congestion, Unreliable, Less rate of records transmission and additionally less effective.

4. Proposed system

We are deploying this Application in Cloud. Data is made & Encrypted, Split and stored in Cloud. Replica is created for data backup. Top Hash Key is stored in Separate Cloud as well as in the Local Backup. We apply MHT algorithm for data splitting. In our implementation we are using cloud for data storage. Because in existing system automatic data recover is drawback. If user loss any file in their cloud storage user have to give recover option and cloud service provider will recover the file and provide that to the user. But in this paper we are implementing a system that will recover the crashed files automatically. For that we are having TPA Third Party Auditing to monitor the file crashes TPA will intimate the user that filer crashed. after that we are using the concept Erasure code, it will recover the file from replica server. So through this we are providing security to the user and his data.

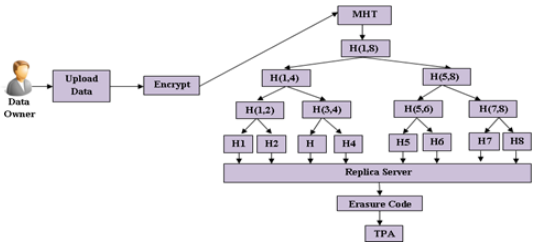

Fig. 1. Architectural diagram

5. Related works

A. *Cloud computing*

Cloud computing is a fast-developing era that has installed itself in the next generation of IT enterprise and business. Cloud computing promises reliable software, hardware, and IaaS delivered over the Internet and remote data centers. Cloud offerings have end up an effective structure to carry out complex huge-scale computing duties and span more than a few IT features from garage and computation to database and application services. The want to store, process, and analyze big amounts of datasets has pushed many businesses and individuals to adopt cloud computing. A massive variety of scientific packages for massive experiments are currently deployed in the cloud and may hold to increase due to the lack of available computing centers in neighborhood servers, decreased capital costs, and increasing volume of facts produced and ate up by using the experiments. In addition, cloud provider companies have begun to integrate frameworks for parallel information processing in their offerings to assist users get admission to cloud sources and deploy their programs .Cloud computing “is a version for permitting ubiquitous, convenient, and on-demand network get entry to some of configured computing sources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management attempt or provider issuer interaction”. Cloud computing has some of favorable components to address the rapid growth of economies and

technological barriers.

B. Hashing algorithm

The hashing algorithm used is Merkle Hash Tree algorithm. In the subject cryptography a hash tree otherwise is a tree where each leaf node is specified with the cryptographic hash of a data block. Each non-leaf node is named with the cryptographic hash in the labels of their child nodes. Hash trees paves way for efficient and secure verification and authentication of the contents of large data structures. Hash trees are a deducement of the hash lists and hash chains.

Representing that a leaf node is a segment of a given binary hash tree needs computing a number of hashes proportional to the logarithm of the number of leaf nodes of the tree. This variation with hash lists, where the number is proportional to the number of leaf nodes.

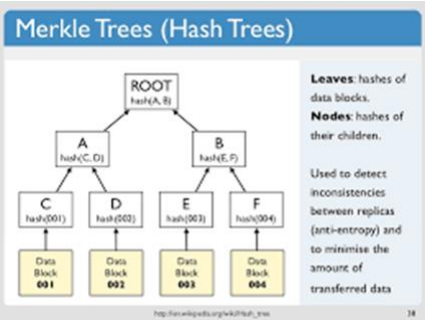

Fig. 2. Merkle Trees

Hash trees are used to verify any sort of data stored, handled and transferred in and between computers. They can aid to ensure that data blocks received from other peers in a peer-to-peer network are received undamaged and unaltered, and it is also used to check that the other peers do send fake blocks.

Hash trees are useful in hash-based cryptography. Hash trees are also used in the IPFS and ZFS file systems, Dat protocol, Git and Mercurial distributed revision control systems, the Tahoe-LAFS backup system the Bitcoin and Ethereum peer-to-peer networks; the Certificate Transparency framework; and few number of NoSQL systems such as Apache Cassandra, and Dynamo. Improvisation are made to use hash trees in reliable computing systems.

The initial Bitcoin implementation of Merkle trees was by Satoshi Nakamoto applies the compression step of the hash function to a great extent, which is mitigated by using Fast Merkle Trees.

The structure of the tree allows efficient mapping of enormous amount of data and minimal changes made to the data can be easily found out. If we want to identify where data change has occurred, then we can check if data is consistent with root hash and we will not have to traverse the whole structure but instead a small. The root hash is used as the fingerprint for the entire data.

C. Splitting and replication of data

Data separation or splitting is a way to secure sensitive and trusted data from unauthorized access and this is done by encrypting the data and storing different sections of a file on different servers. When the split data is accessed, the parts are retrieved, combined and decrypted for further use. An unauthorized person should know the locations of the servers containing the sections, to be able to get access to each server, to know what data to combine, and how to decrypt it.

Data splitting can be made more enhanced by systematically retrieving and reorganizing the parts, and then splitting the data in a number of ways among different servers, and using a different encryption key. Thus, even if an unauthorized user makes progress by obtaining split data, the data will have been reorganized before the hacker manages to obtain all the necessary components. By rearranging the data often enough, a network administrator can stay ahead of even the most expert hacker.

In this Module User interface is created so that the data owner will upload the data to the server. The main objective of this module is to store and share the data by uploading the file to the remote machine. Data is Hashed and applied XOR functionality and then finally stored in the main server. Data owner will upload their data to the cloud server and request for a particular file is send to cloud server. Both the upload and the file request are handled by the main Cloud Server. During the file request is processed main server will communicate with the data owner and the files are retrieved only after the approval given the data owner.

When the parity bits are added, the data will be given to Trusted Parity auditor. The Trusted Parity Auditor will generate the signature by the means of change and response method. The data will be audited and checked in the module, if any changes occurs it will provide the intimation of what the changes are. The data are stored in the respective data servers and the keys are stored in the key servers. Parity bit is added to the data to make the data so secured. Then by adding the parity bits to the data, the data will be changed. These chunked data is stored in the other servers as replica so as to retrieve the data when the main storage data is or not available.

6. Future enhancement

The business side for cloud computing requires a better understanding of costs as compared to an organization's solution. The main feature is that cloud must reduce capital and operational expenses without sacrificing user functionalities such as availability. The outstanding model for cloud functionality is a hardware related approach that fondle the commodity architectures in use by the world's popular Internet and Software as a service provider.

This can be accomplished by less budget servers and disks merged with intelligent management software, it provides true cloud-based economies of scale and efficiency. When moved to the cloud, the organization's intellectual property is sold to a third party. Infact the smallest entry point can create an opening for unauthorized access and theft of valuable data. Authentication and access controls are more critical in a public cloud wherein the cluster attacks aimed towards a hypervisor can compromise the needs of multiple customers. The cloud provider must offer a broad set of security solutions enabling an information focused approach to securing critical interfaces between services and end users, private and public services, as well as virtual and physical cloud infrastructures.

7. Conclusion

Thus the paper infers that now a day's people were using cloud in huge number of ways, but the question is proper security is there or not? We can't able to provide security to the entire cloud server but we can able to secure our data. Here, we provide the security system using MHT by splitting the data on cloud server and if it any case our data was crashed on cloud TPA will send notification to the user through Mail and also crashed file will automatically recovered using Replica server.

References

[1] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," IEEE Trans. Med. Imag., vol. 13, no. 4, pp. 601–609, Dec. 1994.
[2] J. Xu, J. Yan, L. He, P. Su, and D. Feng, "Cloud SEC: A cloud architecture for composing collaborative security services," in Proc. IEEE 2nd Int. Conf. Cloud Comput. Technol. Sci., 2010, pp. 703–711.
[3] J. S. T. Zargar, H. Takabi, and J. B. Joshi, "DCDIDP: A distributed, collaborative, and data-driven intrusion detection and prevention framework for cloud computing environments," in Proc. 7th Int. Conf. Collaborative Comput.: Netw. Appl. Work sharing, 2011, pp. 332–341.
[4] A. R. Khakpour and A. X. Liu, "First step toward cloud-based fire-walling," in Proc. IEEE 31st Symp. Reliable Distrib. Syst., 2012, pp. 41–50.
[5] Z. A. Qazi, C. C. Tu, L. Chiang, R. Miao, V. Sekar, and M. Yu, "SIMPLE-fying middle box policy enforcement using SDN," in Proc. ACM SIGCOMM Conf. SIGCOMM, 2013, pp. 27–38.
[6] A. Gember, A. Krishnamurthy, S. S. John, R. Grandl, X. Gao, A. Anand, T. Benson, V. Sekar, and A. Akella, "Stratos: A network-aware orchestration layer for virtual middle boxes in clouds."
[7] S. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDS) flooding attacks," IEEE Common. Surveys Tuts., vol. 15, no. 4, pp. 2046– 2069, Oct.–Dec. 2013.
[8] B. Hedlund, "What is a distributed firewall?" VMware, 2013.
[9] Y. Jarraya, T. Madi, and M. Debbabi, "A survey and a layered taxonomy of software-defined networking," IEEE Common. Surveys Tuts., vol. 16, no. 4, pp. 1955–1980, Oct.–Dec. 2014.
[10] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," IEEE Commun. Mag., vol. 53, no. 2, pp. 90–97, Feb. 2015.