IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Secure distributed cloud service using trusted third party

Hari Kumar P.
haridavidwil@gmail.com
SRM Easwari Engineering College, Chennai, Tamil Nadu

Chandra Sekar S.
csekar28jun97@gmail.com
SRM Easwari Engineering College, Chennai, Tamil Nadu

Dinesh Kumar V.
dineshvssv@gmail.com
SRM Easwari Engineering College, Chennai, Tamil Nadu

Gopikrishna G.
gopignanasekar98@gmail.com
SRM Easwari Engineering College, Chennai, Tamil Nadu

ABSTRACT

Cloud storage is a system with a distributed data center that takes advantage of virtualization technology and provides an interface for data storage. It makes servers or data centers able to work together for conveniently sharing and accessing resources. Enterprise cloud user demand that there is a secure supply chain and that every step in that supply chain can be verified in real-time and when things go wrong it is more possible to figure out what went wrong and that there is someone who can be held accountable. Before storing data to an untrusted cloud server, some measures should be adopted to guarantee the security of data. However, the communication overhead will increase when users transmit files encrypted by traditional encryption schemes. Remote data integrity checking enables a data storage server to prove to a verifier that it is actually storing a data owner's data honestly. We use Hadoop file system to provide high throughput access to user data. If the semi-honest cloud server does not delete the data honestly and returns an incorrect deletion result, the misbehavior of the cloud server can be detected by the data owner with an overwhelming probability. The user cannot deny after requiring the cloud server to delete the data. This implies that the proposed scheme can support traceability.

Keywords— *Data integrity, Cloud storage, Cloud security, Hadoop file system*

1. INTRODUCTION

Cloud computing provides computing services such as servers, storage, databases, networking, software, analytics, intelligence and more over the Internet. Typically, the user pays only for cloud services they use, helping lower user's operating costs and run user's infrastructure more efficiently. Cloud computing provides a straightforward way to access servers, storage, databases and a set of application services over the Internet. Cloud computing has three main division that is commonly referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Cloud computing is changing businesses in the way they store their data or how they protect their secure information. Cloud computing is benefiting all businesses in every sector in the world.

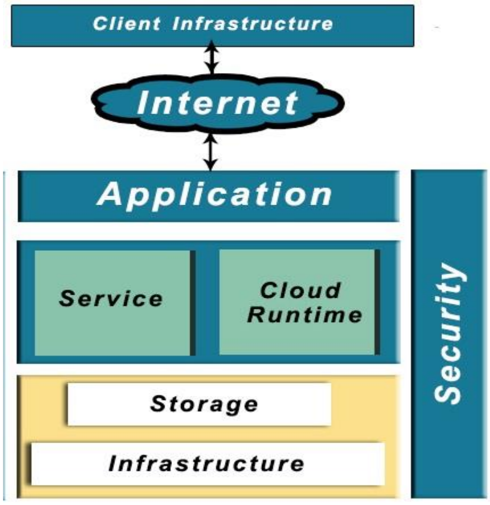

Fig. 1: Cloud computing

© 2019, www.IJARIIT.com All Rights Reserved

Smart businesses are often looking for the most innovative ways to improve and accomplish their business objectives. When it comes to cloud computing, more businesses realizing the benefits this technology can provide them and are beginning to seek more cloud computing options to manage their business activities. With the great number of future technology trends in cloud computing, companies in each and every sector are benefited by the opportunities offered by cloud technology. A huge aspect affecting the future improvement of cloud computing is the amount of storage cloud computing will offer companies and individual users. Many businesses are adopting cloud technology as a part of doing business because of their growth.

Cisco estimates the storage capacity of the cloud will increase exponentially this year alone. With this additional storage, more businesses will be able to store massive data sets and perform analytics using cloud computing. Being able to perform analytics on this enormous amount of data will help companies to gain valuable insights into customer behaviour, strategic financial investments and human systems. Storage demand is increasing nowadays and most IT organizations are under pressure to lower the cost of their

IT infrastructure through the use of shared cloud computing resources. It is predicted that storage software will overtake the storage hardware by 2020.

A major concern of cloud computing is security. Usually, the data in the cloud should be stored in encrypted form. Cloud Security Alliance (CSA) defines the boundary between the responsibilities of the service provider and user. Cloud Security Alliance stack model defines the boundaries between each service models. To avoid security breach data protection mechanisms like auditing, access control, authentication and authorization are incorporated. To avoid data being compromised cloud uses encryption. Encryption is capable to protect unauthorized access but it does not prevent from data loss. One of the key concern of cloud computing is that applications should be migrated easily from one cloud provider to another. Since most of the businesses are now becoming dependent on cloud services are given by the third party so it is necessary for a cloud system to be reliable and robust.

2. LITERATURE SURVEY

Yong Yuet.al [1] proposed Remote Data Integrity Checking (RDIC) which proves honesty of cloud server to, verifier. The approach uses an identity-based signature scheme consists of four polynomial-time and probabilistic algorithms. The approach also uses bilinear paring to maps of pair of group elements to another group element. A probabilistic algorithm TagGen is run by the data owner, it takes system parameter, the secret key and the file to be stored. It outputs the tags for each file block which is stored in the cloud along with the file.

2.1 Results Achieved

Table 1: Summarize of the time cost for a 1 MB file

Setup	Extract	TagGen: off-line	TagGen: on-line	Challenge	GenProof
4.8 ms	0.1 ms	241.9 seconds	20.3 seconds	351 ns per challenge	1.3 ms per challenge

The major achievement is verifier need to challenge 460 blocks in order to achieve the probability of server misbehaviour detection of at least 99%. Verifier takes 3.0 seconds to verify a response and the server takes 0.7 seconds to generate a response when challenging 460 blocks. The major issues are time cost for tag generation for a 1 MB file for on-line generation takes 20.3 seconds and off-line generation takes 241.9 seconds.

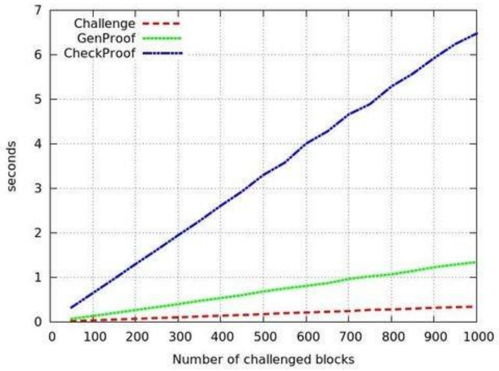

Fig. 2: An increasing number of challenges for a fixed size of the file

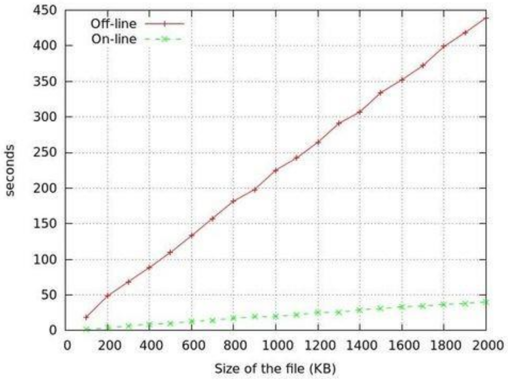

Fig. 3: The tag generation time for the increased size of files

Jingwei Li et.a [3] proposed secure auditing and data deduplication in the cloud using two entities namely SecCloud and SecCloud+. Where SecCloud introduces an auditing entity with maintenance of a Map Re-duce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in the cloud. It is supported on both block level and sector level. In addition, SecCoud also enables secure deduplication. In order to prevent the leakage of such side channel information, they use proof of ownership protocol between clients and cloud server, which allows clients to prove to cloud servers that they exactly own the target data.

The second entity, SecCloud+ enables the guarantee of file confidentiality, the property of deterministic encryption in convergent encryption. The challenge of deduplication on encrypted is the prevention of dictionary attack. The issues in this system are that 92% of the blocks to be challenged to achieve high confidence (i e: 99%) of detecting any small fraction of corruption. The Time cost for the response from cloud storage server increases as the growth of the number of blocks for the challenge.

Lizhen Cui et.al [5] proposed a genetic algorithm, where the parents make the fittest gene. Their approach uses a tripartite graph in which the mapping is between user tasks, data sets and data nodes. A genetic algorithm is used for optimization of storing replica. The algorithm's performance improved than the random strategy used in the Hadoop Distributed File System (HDFS). The total data transmission time is small as a number of replicas grows. The issue is that the algorithm stores more than one replica usually 2 or 3 which require some additional storage for the same data.

Zhihua Xia et.al [6] proposed a privacy-preserving and copy deterrence context-based image retrieval scheme in cloud computing. The secure KNN algorithm is used to encrypt the visual features. They use locality sensitive hashing to improve the search efficiently. The image features are secure against ciphertext-only attack model and the image contents are secure against chosen-plaintext attack model. The results showed that the extraction accuracies are larger than 95% when 30% bits on 5 lower bit plane are flipped.

Zhihua Xia et.al [7] proposed a secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. The approach used greedy depth-first search algorithm which is a recursive procedure in a tree. Greedy depth-first search algorithm is used to obtain better efficiency than a linear search. To update a leaf node, the data owner needs to update log n nodes. It involves an encryption operation for index vector at each node, which takes O(m^2) time, the time complexity of update operation is thus O(m^2 log n).

It is a secure and efficient search scheme which supports accurate multi-keyword ranked search and the dynamic deletion and insertion of documents.

Table 2: Storage consumption of index tree

Size of Dictionary	1000	2000	3000	4000	5000
BDMRS (MB)	73	146	220	293	367
EDMRS (MB)	95	168	241	315	388

3. PROPOSED SYSTEM

In our proposed system, we have three components namely key distributor, third-party auditor and cloud server. We have three modules in our proposed system and their overview as follows. In the first module, the client gives its identity and file name to the key distributor. The key generator provides the user session key using the user's identity. In the second module, the user sends the file and user key to the third party.

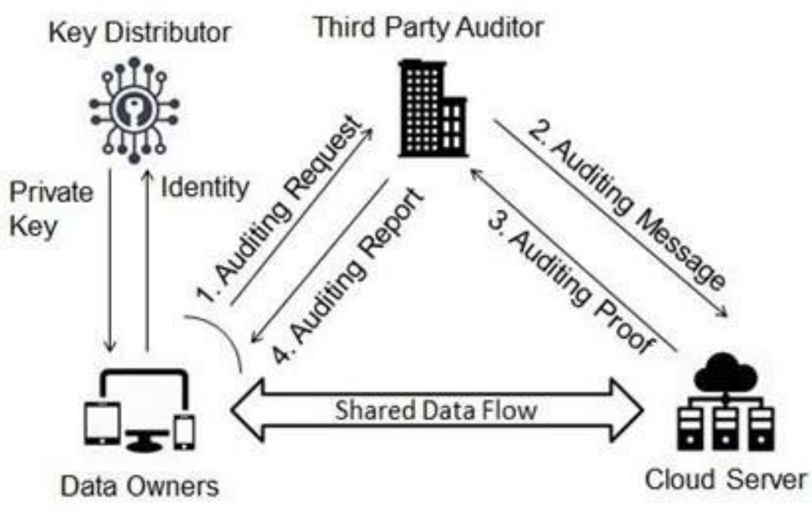

Fig. 4: The architecture of our proposed system

The third-party auditor validates the user and forward the request to the cloud server. If the file already exists then the user is verified and allowed to edit the file else the file is stored.

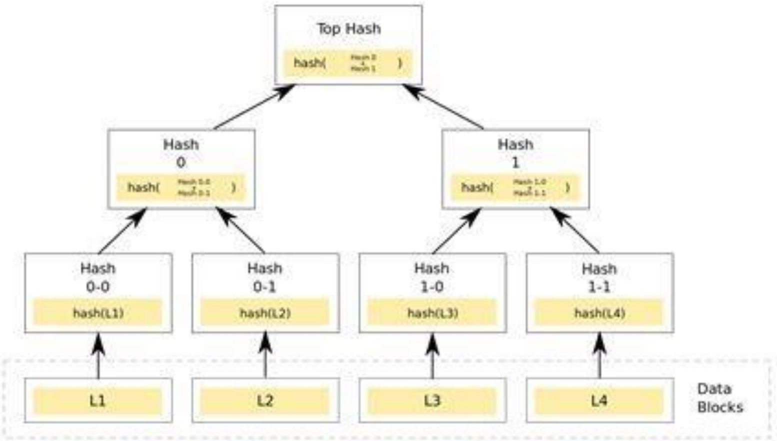

Fig. 5: Merkle tree

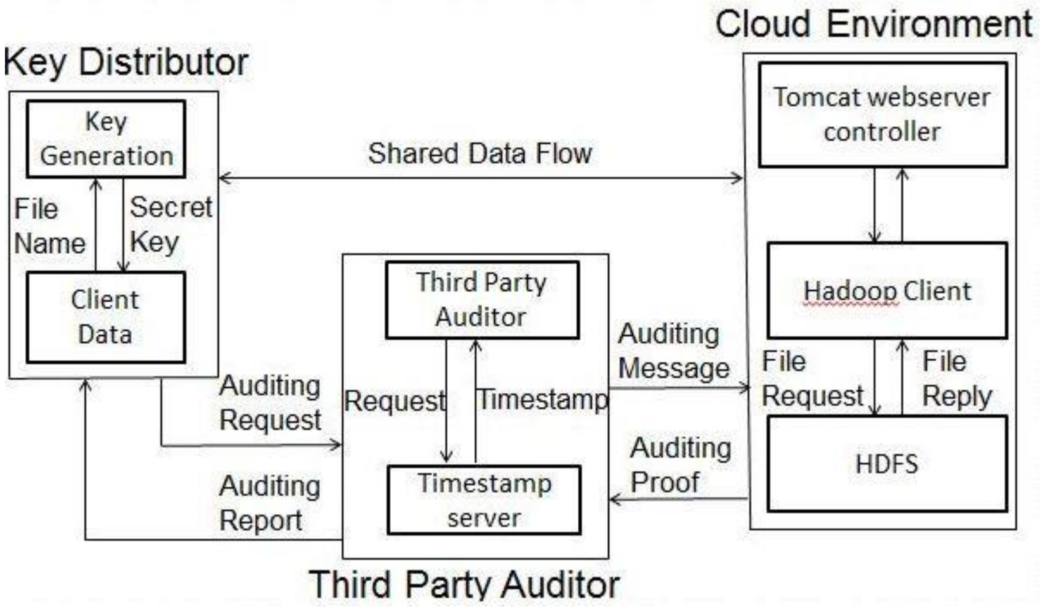

Fig. 6: Implementation of our proposed system

We divide our work into three modules namely- key management using a key distributor, simulation of the cloud environment and third-party auditing (Integrity and Deletion). Our system takes sensitive data as input and provides audit record and retrieval of data. To begin with the service, the user needs to create an account in the cloud system. Successful login allows the user to start using our cloud service. The user can upload their files to the cloud server using upload files option in the user's workspace. Secure Hashing Algorithm (SHA) is used to generate a hash value using the uploaded file name and user's id. This obtained hash value is used as mapping for the file in the cloud storage. Encryption is carried out through the Data Encryption Standard (DES) algorithm. The encrypted file is divided into blocks and stored in the Hadoop File System (HDFS). Namenode have all the metadata information and the mappings of the file. The file is decrypted and retrieved whenever user request for the file. The Third Party Auditor (TPA) checks the integrity of the file and the report is given to the user.

4. PERFORMANCE ANALYSIS

Our scheme allows the server to prove possession of selected blocks of File F. It divides the file into blocks of fixed size with a replication factor set to 2. The storing and retrieval of file occur in constant time.

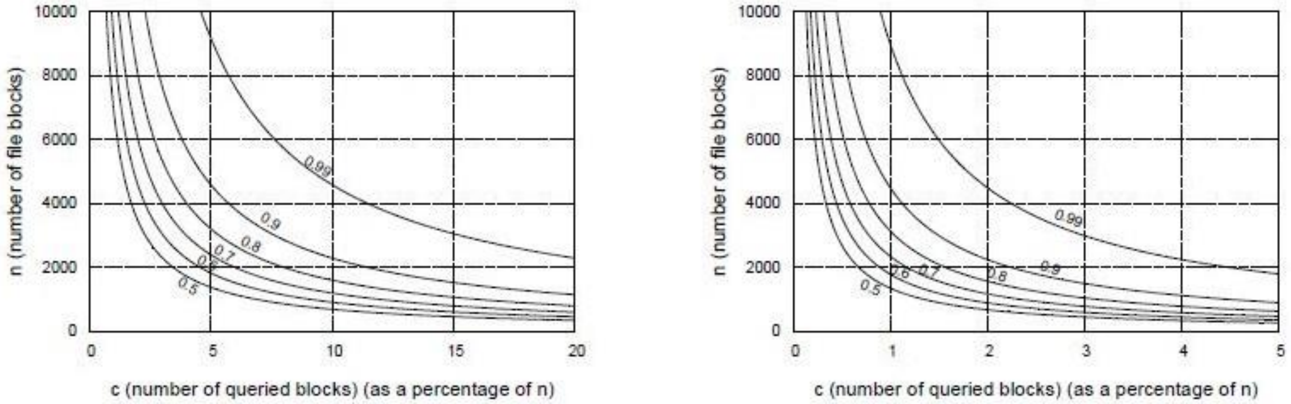

Fig. 7: The probability of server misbehaviour detection

© 2019, www.IJARIIT.com All Rights Reserved

5. CONCLUSION AND FUTURE WORK

We focused on the problem of verifying if an untrusted server stores a client's data. We introduced a model for provable data possession, in which it is desirable to minimize the file block access, the computation on the server, and the client-server communication. Our solutions for this model: They incur a low overhead at the server and require a small, constant amount of communication per challenge. Weprovided construction of this primitive and showed that it achieves soundness and perfect data privacy. The numerical analysis and the implementation demonstrated that the proposed protocol is efficient and practical. In future, we can improve the performance using the optimisation algorithm.

6. REFERENCES

[1] Yong Yu, Man Ho Au, Giuseppe Ateniese, Xinyi Huang, Willy Susilo, Yuanshun Dai, and Geyong Min, (2017), "Identity-Based Remote Data Integrity Checking With Perfect Data Privacy Preserving for Cloud Storage", IEEE Transactions on Information Forensics and Security, Volume: 12, Issue: 4, PP: 767-778, April 2017.

[2] Jia Yu, Kui Ren, Cong Wang and Vijay Varadharajan, (2015), "Enabling Cloud Storage Auditing with Key-Exposure Resistance", IEEE Transactions on Information Forensics and Security, Volume: 10, Issue: 6, PP: 1167-1179, February 2015.

[3] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, (2016), "Secure Auditing and Deduplicating Data in Cloud", IEEE Transactions on Computers, Volume: 65, Issue: 8, PP: 2386-2396, August 2016.

[4] Kim-Kwang Raymond Choo, Omer F. Rana, and Muttukrishnan Rajarajan, (2017), "Cloud Security Engineering: Theory, Practice and Future Research", IEEE Transactions on Cloud Computing, Volume: 5, Issue: 3, PP: 372-374, July-September 2017.

[5] Lizhen Cui, Junhua Zhang, Lingxi Yue, Yuliang Shi, Hui Li and Dong Yuan, (2018) "A Genetic Algorithm Based Data Replica Placement Strategy for Scientific Applications in Clouds", IEEE Transactions on Services Computing, Volume: 11, Issue: 4, PP: 727-739, July-August 2018.

[6] Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun and Kui Ren, (2016) "A Privacy-Preserving and Copy-Deterrence Content-Based Image Retrieval Scheme in Cloud Computing", IEEE Transactions on Information Forensics and Security, Volume: 11, Issue: 11, PP: 2594-2608, November 2016.

[7] Zhihua Xia, Xinhui Wang, Xingming Sun and Qian Wang, (2016) "A Secure and Dynamic Multi- Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Volume: 27, Issue: 2, PP: 340-352, February 2016.

[8] Wan-Chi Chang and Pi-Chung Wang, (2019) "Write-Aware Replica Placement for Cloud Computing", IEEE Journal on Selected Areas in Communications, Volume: 37, Issue: 3, PP: 656-667, March 2019.

[9] [9] Colette Langos and Mark Giancaspro, (2015) "Does Cloud Storage Lend Itself to Cyberbullying?", IEEE Cloud Computing, Volume: 2, Issue: 5, PP: 70-74, Sept-Oct 2015.

[10] [10] Dongmahn Seo, Suhyun Kim and Gyuwon Song, "Mutual exclusion method in client-side aggregation of cloud storage", IEEE Consumer Electronics Society, Volume: 63, Issue: 2, PP: 185-190, August 2017.