Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

Analytical Literature Survey on Existing Schemes to Improve Data Security and Robustness in Cloud Computing

N. Abirami
Student,
Department of Computer Science and Engineering
Sathyabama University
Chennai, India

S. Murugan
Professor,
Department of Computer Science and Engineering
Sathyabama University
Chennai, India

***Abstract*— In the recent years, cloud computing is largely adopted infrastructure by most of the corporate, health care organizations, educational institutions, and small growing software solution units for meeting their demand for hardware and software service offered by cloud infrastructure. Cloud provides an efficient way to store the documents on pay-per-use basis. In user perspective instead of storing the data in USB drive or hard-disk, the data is stored in a remote unknown location. Here security and robustness of the cloud storage system is of great concern. This paper gives the various issues of cloud computing systems, the existing solutions to avoid the issues like security and data robustness, performance analysis of the schemes and comparison of the schemes.**

Keywords — Data Security;robustness;cloud computing; cloud storage

I. INTRODUCTION

Cloud computing is an internet based computing where the user is able to store his document in a remote server and run his application software remotely. In desktop computing, the user stores all his documents, copies of software and database files and everything in a single computer, so there is a limitation in terms of storage space, required speed and hardware requirements. Cloud computing is a form of distributed computing which paves way for the user to store the documents and application instances among multiple computers, which are connected over the internet, so that the documents can be accessed or execute an application from any computer which is connected to the internet. Thus a cloud system is one in which multiple systems connected in a distributed environments run. With this type of environment security and robustness is of great concern for any user who stores the personal documents in the cloud storage system. There may be many reasons for failure of retrieving the documents back like server failure, server corruption, network failure, data interrupted by some unauthorized internal user in the cloud storage server or unauthorized person interrupting the network to access the data and so many. The main objective of the paper is to analyze various existing schemes to improve the data security and robustness in the cloud storage environment and the comparison of performance of schemes and detailed analysis of various methods incorporated.

II. VARIOUS ISSUES IN CLOUD COMPUTING

A. Reliability

Reliability can be better understood when it is said as 'Trust'. When the user A exchanges his data with user B, then user A believes and trusts that user B will behave as expected and he does not misuse the data sent user A. A service is said to be reliable when there is a certainty that the response will be received at the required time. Reliability in cloud environment depends upon the cloud deployment model. The various cloud deployment models includes public cloud, private cloud and hybrid cloud.

Public Cloud:

The security of the data is under the control of the cloud infrastructure provider. Security policies and techniques are provided by the owner of the cloud infrastructure provider. This type of deployment model has huge risks compared to private cloud since data is shared or stored in an external environment which may impose serious risk of unauthorized persons to access the data.

Private Cloud:

In the case of private cloud, the infrastructure is controlled by a private organization. Already practiced internal security schemes are imposed and particular security challenges are incorporated.

Hybrid Cloud:

Hybrid cloud is built by the combination of public and private clouds in order to achieve high customer satisfaction. Hybrid clouds are maintained internally by the organization. The most secured information are stored in the private cloud and public cloud is used to render cloud service's to user such as storage, running the users application. Both public and private clouds are connected but independent of each other.

B. Confidentiality

Confidentiality of information means that data is accessed only by the authorized person and when unauthorized person tries to interrupt the network or access the storage device to corrupt the data he is not allowed to do so in order to maintain the data privacy. Confidentiality in cloud environment is applicable to hardware and software confidentiality and the cloud service provider has the responsibility of ensuring data confidentiality and security to the user's data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

III. LITERATURE SURVEY

Shiuan-Tzuo et al [1] proposed an integrity check scheme for enhancing the data robustness in cloud storage services in the occurrence of corruption in server data and when cipher text is intruded by the unauthorized person. This integrity check scheme has homomorphism property. The researcher has tried to prove the security of the integrity check by creating appropriate parameters in order to retrieve the data successfully. They have used a threat model for proving the integrity check scheme considering two type of attacks namely storage failure and storage corruption. The integrity check scheme is proved by means of model security game. They have used bilinear map and a pseudo random function whose output is computed differently from the output of the random function, decentralized erasure code and threshold public key encryption. The researchers have computed communication and storage cost of the integrity check scheme. While storing user's data the user assigns identifier and defines verification key and sends it to the key server followed by dividing the file into multiple blocks, encrypts each block of data using the user's public key and the users defines an integrity unit for each of his cipher text and distributes cipher text-integrity tag pair to multiple storage server, which combines the data pairs and stores them in an encoded form. When the user wants to check the integrity he sends the request to some key server, upon receiving the request, the server queries storage server which consolidates these tuples and check their integrity using verification key. In case if the process fails, the key server splits the tuples into multiple groups and perform the integrity check, this process is repeated until all the corrupted tuples are identified. In the retrieval phase, the computed data are filtered and the key server defines decryption token using the available tuples. At the end of the process, the user receives the cipher text from the key server as response and decodes them to reconstruct the message.

Hsiao-Ying et al [2] proposed a scheme for forwarding the data to another user securely by ensuring data robustness, confidentiality by ensuring that the data owner has full control over the data to be forwarded to another user. This scheme uses decentralized erasure code so that the message is not under the security risk since it is encrypted and split into multiple blocks and stored in distributed storage server. This scheme tries to address the problem like maintenance of secret keys by the user, reduces communication traffic, and resolves the disadvantage of storage server of supporting other functionalities. The system model consists of highly distributed multiple key servers and storage server. In this scheme the operations encoding, encryption and data forwarding are integrated tightly. This scheme uses proxy re-encryption, which is understood as follows: When a user 'A' transfers his data user 'B' using his public key, creates re-encryption key and sends it to the proxy server. Here the proxy server is unaware of the original text message. The researcher has integrated encryption, re-encryption and the encoding techniques together, for improving storage robustness. This scheme reduces the communication cost and communication overhead. The re-encryption key is computed by the combination of User 'A' secret key and User 'B' public key. Failure system recovery is done by using additional storage server. The additional server obtains the encrypted units by querying remaining unaffected servers. This scheme uses multiplicative homomorphic encryption method which supports data encoding over encrypted message.

Yongjun Ren et al [3] have proposed a designated verifier provable data possession-DV-PDP scheme, which is used in a critical environment where the user cannot execute a check on the remote data. This system uses Elliptic-Curve cryptography (ECC) homomorphism authenticator. This system is an upgraded technology replacing bilinear computing, here the cloud storage server is independent from the verifier and it is stateless. Client uses independent verifier to ensure integrity of data stored in the cloud storage. The scheme is constructed as follows: Algorithm computes client's private-public key pair and the designated verifier private-public key pair. A challenge is generated by the client or verifier to check the data integrity by using the public parameter which is computed using the tag for the message block and the verifier public key. The cloud storage server proves the integrity of data and the metadata information as the response. The designated verifier receives the output from cloud storage server and checks whether it is true or false.

Bo Chen et al [4] proposed a Remote Data Checking Scheme-RDC. RDC provides a prevention tool and repair tool for the client, where the prevention tool is used to verify whether the data is corrupted and the repair tool is used to correct the data in the network code based distributed storage system. A verifier is used to check the data stored in the remote storage system and if any corruption is detected it retrieves the original message by making use of remaining uncorrupted servers. The retrieved data is then stored in a new server. Techniques used are network coding, replication and erasure coding. The proposed RDC scheme eradicated the additional attack in the network coding which includes replay and pollution attacks. To prevent replay attacks, the server stores the co-efficient of network coding. To eradicate pollution attacks, a repair verification tag is used, where the client checks whether the server performs combination of blocks correctly. The researcher proves this scheme using a mobile adversary model that is capable of corrupting the server data. This adversary corrupts any of the server with a time interval called as epoch. In the challenge phase, the adversary corrupts the servers and the client checks for correct data possession to the servers and detects any corruption. In the repair phase, the client computes the original data by containing the remaining healthy servers.

Guiseppe et al [5] proposed an interactive protocol called Proofs of storage (PoS) which enables the client to ensure trustworthy storage of his file with the server. By using any identification protocol as a base, the researchers have constructed a framework for computing which can be used for numerous times of verifications. Here the file size and the size of client's state and the communication complexity are independent. This system ensures whether the complete file sent by the client is stored as it is in the server, so that the client tags each segment of the file and stores it to the server. For verifying the storage, the client sends challenge to the server, the server returns the challenge vector and the file segment along with the tag.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

Shiuan-Tzuo et al [6] proposed a scheme for checking the integrity of messages stored in cloud storage. The owner of the message holds the full control to decide the authorized persons to check the integrity of data, hence this method prevents re-delegation from delegated verifiers. The user uploads delegation key, integrity verification tags and the data to the storage server. The cloud storage server uses the delegation key to transform the data tag which is in the form such that the delegated verifier can check using its private key. This scheme uses random oracle model.

H.Shacham et al [7] proposes two schemes for proofs of retrieval, the first scheme uses BLS signatures with very short client query and server response and this scheme allows public verifiability. The second scheme is constructed using pseudorandom functions and allows private verifiability, in this scheme the client's query is long but the server's response is extremely small compared to the first scheme. When homomorphic authenticators are used, the server response consists of combination of the data blocks and authenticators, so the response length is short. This scheme is proved using verifier and a prover; the verifier sends a request containing set which is formed with index and verifier co-efficient tuple. The prover sends the response pair containing data block and authenticator. The verifier checks whether the response is correctly formed. In the case of second scheme, BLS signatures are used for authentication. Public key is used for proof-of retrievability protocol by the verifier whereas private key is used for authenticator construction.

Yen Zhu et al [8] uses techniques such as fragment structure, index-hash table, random sampling to construct a dynamic cloud audit service for the data stored in cloud. In order to enhance the performance, probabilistic query and periodic verification is used. The dynamic audit service possess the property of being lightweight, supports dynamic operations, effective detection and timely detection in case of failure in integrity. The audit service does not store the original data offering public audit facility. Dynamic operations are implemented by using indexed-has table, which contains unique hash key for each record. Only the authenticated applications and the data owner holds the secret key of the data. In order to reduce the complexity of integrity checking each and every data, this scheme uses random sampling. Any data error is identified by the third party auditor and is recorded in the index hash table.

Bo Chen et al [9] proposed Remote Data Checking (RDC) scheme for improving robustness, in the case of insecure cloud storage provider. This scheme uses Cauchy matrices to determine the Reed-Solomon codes since they are compatible to dynamic operations.

N.Cao et al [10] has designed a LT Codes-based [11] secure and reliable cloud storage service in order to resolve the reliability issue. This scheme frees the data owner once he has uploaded his data in the data storage, he need not pay for data integrity checking process, which will be taken care by the third-party. This method is constructed as an improvement to erasure-coding and network coding-based systems in terms of reducing the computational cost compared to erasure coding and less storage cost compared to network coding. This method tries to reduce both the data owner's cost by avoiding the process of recreating the verification tags and data retrieval time by providing faster decoding at the time of retrieval by retrieving the corrupted data in its original form. This method uses Belief Propagation decoder [12] algorithm for decoding. It ensures data retrievability by checking the data decodability before outsourcing data in the cloud storage server provides security against replay attacks or pollution attack.

Wang et al [13] proposed a third-party auditor (TPA) to check the integrity of data, in order to reduce the overhead for the client to check the data integrity. In this method the TPA is designed in such a way that he performs multiple auditing tasks simultaneously. By improvising the existing proof of storage techniques this scheme also supports dynamic data operations. Block tag authentication is achieved by constructing Merkie Hash tree.

Yan Zhu et al [14] proposed the Cooperative Provable Data Possession (CPDP) scheme for hybrid cloud environment using Homomorphic Verifiable response and Hash Index Hierarchy. This scheme combines the existing private and public cloud storage servers to function cooperatively in order to store and manage user's data. This scheme provides better response in terms of bandwidth and time compared to other Provable Data Possession (PDP) schemes. They have proved that the computational overhead have reduces when the number of file blocks has increased.

Kevin et al [15] proposed HAIL (High-Availability and Integrity Layer) which provides mechanism for the cloud storage server to prove to their clients that their data are of high availability and reliability and secure over the attacks of mobile adversaries and Byzantine failures. This method was designed as an improvisation of Proofs of Retrievability (POR) methods. The researcher claims that the design of POR methods is single system based whereas HAIL is across distributed independent servers. HAIL uses Test-And-Redistribute strategy, which uses POR methods to detect data corruption followed by data repair mechanism which involves recovering the correct data blocks and setting it to the corrupted servers. HAIL proposes an Integrity-protected Error-Correcting code (IP-ECC) which is a combination of Pseudo-random functions (PRF), error correction codes and Universal Hash Functions (UHFs) into a single component. Message Authentication codes are produced by the combination of UHF and PRF. In order to avoid creeping-corruption attack, HAIL employs POR in each of the storage servers. The integrity is checked in two ways, by single-server approach or by cross-server redundancy.

IV. ANALYSIS OF VARIOUS SCHEMES

Table 1 shows the comparison of existing schemes for providing security in cloud computing. The table shows the name of researchers, name of the scheme, parameters used for performance analysis, environment, experimental setup, method used for proving the scheme and the results.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

TABLE 1. COMPARISON OF EXISTING SCHEMES

Researchers	Name Of the scheme	Performance analysis Parameters	Environment	Experimental setup	Method for Proving	Findings
Shiuan-Tzuo Shen, ,Hsiao-Ying Lin, and Wen-Guey Tzeng	An Effective Integrity Check Scheme for Secure Erasure Code-Based Storage Systems	Storage cost, Computational cost, Communication cost	Hybrid cloud environment	NA	Theorem.	Result shows success rate of data retrieval under server failure and corruption.
H.-Y. Lin and W.-G. Tzeng,	A secure decentralized erasure code for distributed networked storage	Storage cost, Computational cost, probability of successful retrieval	Distributed and de-centralized	NA	Theorem	Result shows the successful retrieval of data at low storage and computation cost.
H.-Y. Lin and W.-G. Tzeng	A secure erasure code-based cloud storage system with secure data forwarding	Storage cost, Correctness, Computational cost, probability of successful retrieval	Distributed and de-centralized	NA	Theorem	Result shows the successful retrieval of data at low storage and computation cost.
Yongjun Ren,Jiang Xu,Jin Wang,Jeong-Uk Kim	Designated-Verifier provable data possession in public cloud storage	Communication cost, computation cost	Public cloud environment	NA	Theorem	Their approach produces less communication cost compared to RSA and an efficient computation cost.
B. Chen, R. Curtmola, G. Ateniese, and R. Burns	Remote data checking for network coding-based distributed storage systems	Server Storage space, communication cost, computation cost	Distributed	Intel Core 2 Duo system, 1.333 GHz frontside bus, 4GB RAM,360 GB hard-disk, Ubuntu 9.10, OpenSSL version 1.0.0	Experiment	The result shows that the computational cost is linear to the file size and varies by no: of file blocks, rather than by no: of servers. Computational costs of these components are decreasing with no: of file blocks for fixed file size. Encoding cost increases with no: of file blocks for fixed file size
S.-T. Shen and W.-G. Tzeng	Delegable provable data possession for remote data in the clouds	Computation cost, Storage cost, Communication cost	Distributed	NA	Theorem	The result shows that when binary co-efficient is used it reduces computational cost.
Hovav Schacham and Brent Waters	Compact proofs of retrievability	Security parameter	Distributed	NA	Theorem	The result shows the successful retrieval of message when a adversary challenges the storage server.
Yan Zhu, Gail-Joon Ahn,Hongxin Hu,Stephen S. Yau,Ho G. An, and Chang-Jun Hu	Dynamic audit services for outsourced storages in clouds	Computation cost, Communication cost	Public clouds	Amazon S3,local IBM server with 2.16 GHz two Intel Core 2 processors running Windows Server 2003, C libraries	Experiments	The results shows that the audit service of our scheme shows minimal overhead, minimum computational and communication cost
Bo Chen, Reza Curtmola	Robust dynamic remote data checking for public clouds	Computation and I/O cost	Public clouds	NA	Examples	The result shows reduced client-server communication overhead.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

N. Cao, S. Yu, Z. Yang, W. Lou, and Y. T. Hou	Lt codes-based secure and reliable cloud storage service	Retrieval time, Storage cost, Computation cost and communication cost.	Distributed environment	C, Linux Server with Intel Xeon Processor 2.93GHz	Experiments	The result shows that the communication and storage cost are linear with the erasure-code technique, but the computational cost is much lesser than the erasure code. This method produces less storage cost and lesser retrieval time compared to network-code based systems.
Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li	Enabling public auditability and data dynamics for storage security in cloud computing	Public auditability, Communication cost, Computational time	Distributed, Public Cloud environment	C, IntelCore Processor of 2 GHz,768 MB RAM, 8 MB buffer, Pairing-Based Cryptography (PBC) library version 0.4.18 and the crypto library of OpenSSL version 0.9.8h	Experiments	The results shows that the communication cost has linear growth with size of the file block.
Yan Zhu1,Huaixi Wang, Zexing Hu, Gail-Joon Ahn, Hongxin Hu and Stephen S. Yau	Efficient Provable Data Possession for Hybrid Clouds.	Computation and communication cost	Hybrid Cloud environment	C++, Intel Core 2 processor with 2.16 GHz	Experiments	The result shows that the overall communication overhead has not changed significantly compared to existing PDP schemes. The computational overhead is decreases with increase in the number of file sectors.
Kevin D. Bowers, Ari Juels and Alina Oprea	HAIL:A High-availability and Integrity Layer for Cloud Storage	Retrieval time, Storage cost	Cloud Environment	C++ ,Intel Core 2 processor 2.16 GHz RSA BSAFE C library, Jera	Experiments	The result shows reduced storage cost due to the use of dispersal code and high availability of data against mobile adversaries.

From Table 1. we infer that storage cost, computational cost, communication cost, I/O cost, success rate of retrieval and retrieval time are the parameters used for performance analysis in the existing schemes. Erasure-code, replication, network-codes based system, public auditing, third-party auditing, proofs-of-retrieval protocols are commonly used techniques by the researchers for ensuring security, reliability and robustness of cloud storage system.

V. CONCLUSION

Security and robustness are an important issue in cloud computing today. In this paper we have analyzed almost thirteen existing schemes and have studied the performance analysis parameters, tools and software used in the environment setup for simulation. We have come to a conclusion that most of the schemes have tried to achieve maximum success of data retrievability in case of attackers, server failure and server corruption with minimum storage, computation and communication cost.

VI. REFERENCES

[1] Shiuan-Tzuo Shen,Hsiao-Ying Lin, and Wen-Guey Tzeng, "An Effective Integrity Check Scheme for Secure Erasure Code-Based Storage Systems,"IEEE Trans. on Reliability.,vol. 64,no.3,Sep. 2015.

[2] H.-Y. Lin and W.-G. Tzeng, "A secure decentralized erasure code for distributed networked storage," IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 11, pp. 1586–1594, Nov. 2010.

[3] H.-Y. Lin and W.-G. Tzeng, "A secure erasure code-based cloud storage system with secure data forwarding," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 6, pp. 995–1003, Jun. 2012.

[4] Yongjun Ren,Jiang Xu,Jin Wang,Jeong-Uk Kim,"Designated-Verifier provable data possession in public cloud storage",International journal of security and applications.,vol. 7,no.6,pp.11-20,2013.

[5] B. Chen, R. Curtmola, G. Ateniese, and R. Burns, "Remote data checking for network coding-based distributed storage systems," in Proc. 2nd ACM Workshop Cloud Computing Security (CCSW'10), 2010, pp. 31–42.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

[6] S.-T. Shen and W.-G. Tzeng, "Delegable provable data possession for remote data in the clouds," in Proc. 13th Int. Conf. Information and Commun. Security (ICICS'11), 2011, pp. 93–111

[7] H. Shacham and B. Waters, "Compact proofs of retrievability," J.Cryptol., vol. 26, no. 3, pp. 442–483, 2013.

[8] Y. Zhu, G.-J. Ahn, H. Hu, S. S. Yau, H. G. An, and C.-J. Hu, "Dynamic audit services for outsourced storages in clouds," IEEE Trans. Services Comput., vol. 6, no. 2, pp. 227–238, 2013.

[9] B. Chen and R. Curtmola, "Robust dynamic remote data checking for public clouds," in Proc. 19th ACM Conf. Computer and Commun. Security (CCS'12), 2012, pp. 1043–1045.

[10] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. T. Hou, "Lt codes-based secure and reliable cloud storage service," in Proc. 31st M. Luby, "Lt codes," in Proc. of FoCS, 2002, pp. 271–280.

IEEE Int. Conf. Computer Commun. (INFOCOM'12), 2012, pp. 693–701.

[11] M. Luby, "Lt codes," in Proc. of FoCS, 2002, pp. 271–280.

[12] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman,"Efficient erasure correcting codes," ITIT, no. 2, pp. 569–584, 2001.

[13] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing,"IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847–859, May 2011.

[14] Yan Zhu,Huaixi Wang, Zexing Hu, Gail-Joon Ahn, Hongxin Hu and Stephen S. Yau," Efficient Provable Data Possession for Hybrid Clouds ",CCS'10, October 4–8, 2010, Chicago, Illinois, USA. ACM 978-1-4503-0244-9/10/10.

[15] Kevin D. Bowers,Ari Juels,Alina Oprea,"Hail:A High-availability and Integrity Layer for Cloud Storage", CCS'09, November 9–13, 2009, Chicago, Illinois, USA