Asian Journal of Engineering and Applied Technology
ISSN 2249-068X Vol. 8 No. 1, 2019, pp. 44-49

Cloud Computing: An Exploration on Data Integrity and Its Techniques

R. K. Ramesh[1] and K. L. Neela[2]
[1]M. Phil Scholar, Department of Computer Science, E.G.S. Pillay Arts and Science College, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science, University College of Engineering, Tamil Nadu, India
E-Mail: ramesh26.vin@gmail.com, klneela@yahoo.com

***Abstract* - In the recent years, the cloud computing moves a step forward by providing more services for the individual users and organizations. Each and every technology and systems can be easily integrated with cloud environment. In future lot of new technology will come to play their role. But to make the new technology fulfill their role they surely need a small part of cloud computing. The cloud plays a vital role in managing the data of different users. Users may be individual or organization and they use any cloud deployment model. Cloud provides low cost unpredictable and untraceable storage space to store user's data. The main aim is to keep the data reliable, available, secure and sharable for the authorized users. But keeping the data secure is a big issue .In a public cloud any one can access the cloud easily. But the CSP or data owner has to prevent the unauthorized user's access to the cloud. CSP can easily modify or delete the data which are not used by the user for a long time so as to get the storage space .The data integrity problem can be overcome by some different techniques which are proposed in various models .We can surely say that this cloud computing technology will become popular than any other technology.**

Keywords: **CSP, Cloud Computing, Data Integrity, NIC, ALCOA, CCAF, BPMN, SEM, PKI, TTP**

I. INTRODUCTION

In a cloud computing environment the data plays a vital role. The key constraints of data are data storage, data security, data availability, and data integrity and data confidentiality. Data integrity means keeping the data accurate and errorfree.The data are stored in the cloud server. The cloud service provider (CSP) is responsible for managing the data. If some unauthorized user access the data, the data will be tampered and the tampered data is updated in the server .The result is data loss. This will affect the integrity of data. Data should be always kept secure. It is better to always keep a cop of data in client machine so as to avoid data loss. Data is important for each and every user. For example we are giving our useful information like National Identification Card (NIC) to bank and some other agencies. We actually don't know what they do with our data. The chances are there to misuse our data. Even though the technology is developed, keeping the data secured is a difficult task. Now adays data integrity is becoming a serious issue.

II. DATA INTEGRITY

Despite the fact that the next generation computing is fully depends on cloud computing, the integrity and security will make a major setback in enact it in the real time applications. Since the datacenter is used to stored data, the data security is a major issue [8]. Almost all the big and small organization uses cloud environment. Parallel development of both mobile and cloud computing make the users to store and access the data to the cloud environment by the handheld device like mobile. All the companies started to change their existing system to cloud environment.

Some small companies cannot afford the financial burden caused by migration. By using the cloud environment the companies need not to buy some high cost resources permanently. They can use the pay and use schema. The companies are using the cloud environment mainly for storage. Storing the data in cloud make them free from buying some high cost hardware for storage. The data integrity can be easily ensured through ALCOA [15]. This is manly used in pharmacy area. Here the datas are of hybrid. i.e. data is stored in paper and electronically. So that if any error occur in electronic format we can easily retrieve from the paper. The data is having some special constraints they are explained in Fig 1.

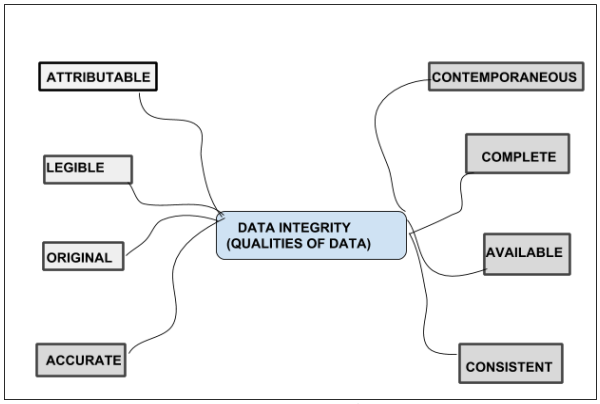

Fig. 1 Attributes of Data

Attributable (Person who acquire data)
Legible (One who read and understand data)
Contemporaneous (documented at activity)
Original (Observed first)
Accurate (No errors)
Complete (All data performed on sample)
Consistent (Sequence of element follow on)
Enduring (Not recorded on back of envelope)
Available (Can be accessed)

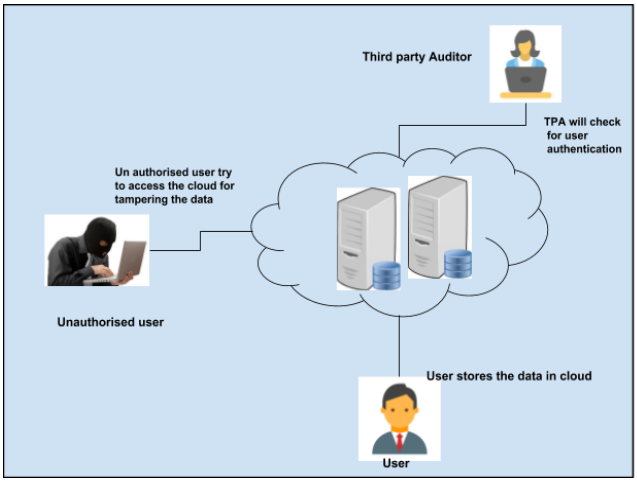

Fig. 2 Data storage in Cloud

Fig. 2 shows the data storage in cloud. Almost all the storage like S3, BLOB etc is managed by some cloud service providers [7]. It is the responsibility of the CSP to prevent the data from unauthorized access. Keeping the data untampered is a very difficult task. It is better to allow only authorized user to access the cloud so as to keep our data safe and secure. We can simply say data integrity means checking for correction or verifying if any changes occur for the original data. When the technology develops keeping the data safe become very hard. It is better to keep the data into the cloud in encrypted format using some algorithms like RSA, and AES.

The data integrity problem will usually occur in our real life. For example in colleges the Professor will correct the paper and give it to the students for correction. After recollecting the paper from student the Professor usually gives the paper to the class representative or class leader. The representative will write the marks in a sheet of paper and give it to professor.

Fig 3 shows high chance of data correction in class room. The chances are there for the student to approach the class representative and alter the mark.In this scenario we can take professor as an authorized user, Class representative as third-party and student as unauthorized user. If the professor keeps a copy of mark for his reference and give paper to the students for checking means the data tampering may be avoided. Fig 4 shows less chance of data correction in class room.

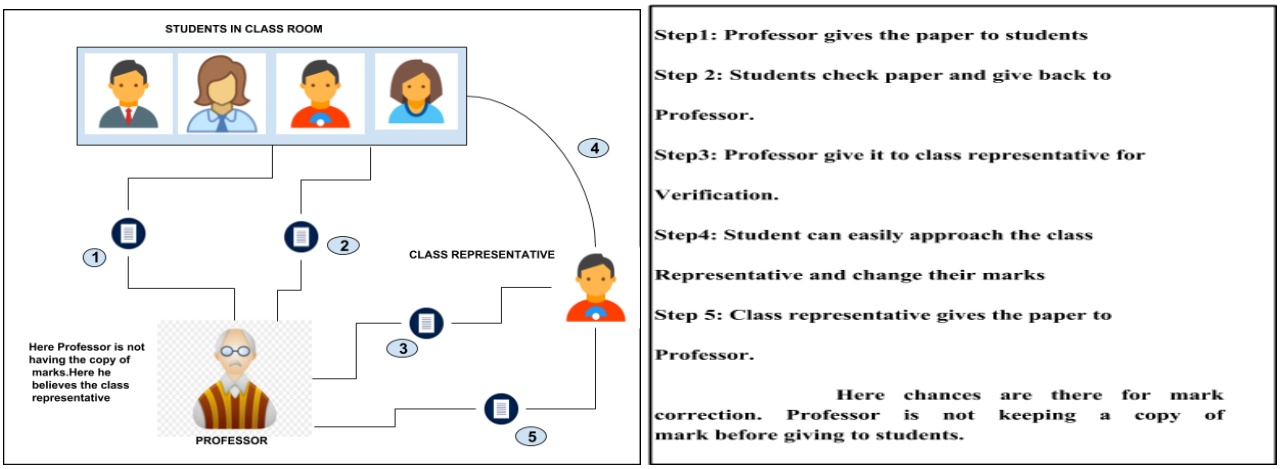

Fig. 3 High chance of data correction in class room

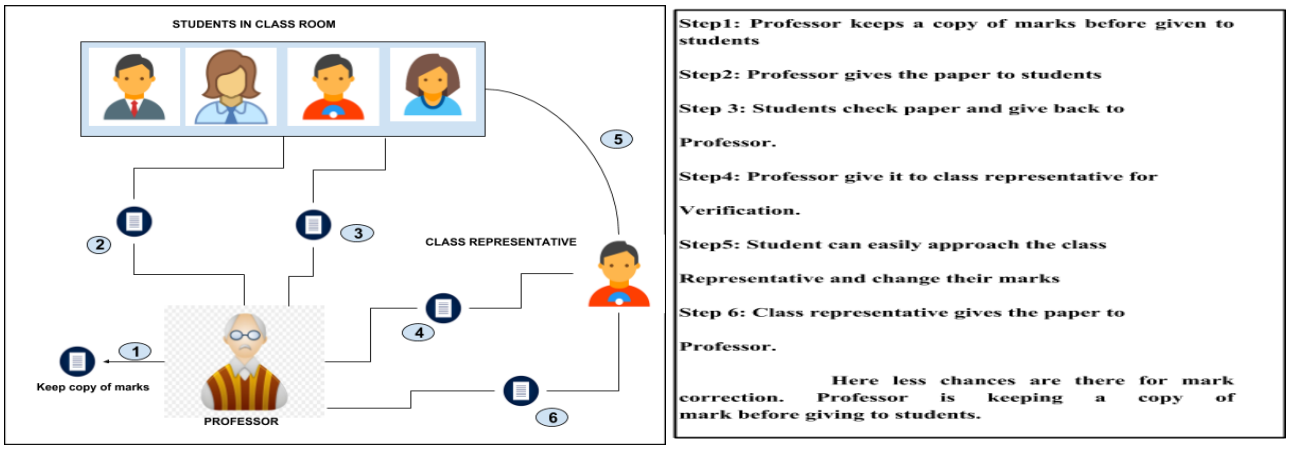

Fig. 4 Less chance of data correction in class room

In cloud environment the data are processed and stored with the help of third party data centers. The cloud security issues are faced both by the customer and the CSP. The company data is safe until it is in local storage. If the company uploaded the data to the cloud it is risk, because any one can access the data. The Cloud service supplier manages the data. To make the data more secure the cloud service supplier should use only employee with clear background check. The companies credential should be maintained by the employee. The entire cloud service provider assures the data security for all the organization and the individuals. Almost all the data owner needs there data should be secure and safe.ie trustworthy [1].

III. DATA INTEGRITY ISSUES

A. Data Loss or Manipulation

A company or an individual user uses huge amount of data. These data are stored in cloud and managed by the CSP. The user access the data when a need arise.ie daily or rarely. Sometimes the data are stored in some remote cloud which makes the data unsecured and unreliable .Since it is a public cloud the chances of unauthorized user access the cloud is high.CSP is responsible to keep the data from unauthorized access. Sometimes the data may be accessed by unauthorized user and they alter or modify the data. This may occur accidentally or purposefully. Sometimes data may be loss when functions like backup and restore occur. Since the data is stored in remote cloud, it is difficult for the user to control the data. [6].

B. Untrusted Remote Server Performing Computation

Since the computation is outsourced, it is very difficult to find whether the computation is executed with high integrity.
The cloud server will provide incorrect result to the user result because of less transparency in the computational details. [6].

C. Difficult in Data Recovery

In cloud environment sometimes data will be stored in remote cloud. The user will not aware of how and where the data is stored. They can only access their data. If any unauthorized user access the cloud and data is altered or loss means, it is difficult for the users to find and replace the data if backup of data is available. [6].

D. Data Integrity Compliance

The Compliance is one of the business goals of almost all big and small companies. The employees should not
be induced to avoid the accurate process. Employees should not be afraid of telling the truth if something went wrong. An open apology or an open communication can solve many problems. [23]

IV. RELATED WORK

Chang *et al.,* [12] proposed a cloud computing adaption frame work (CCAF) order to keep the data secure in a multilayered structure. Identity management, Firewall, access control and convergent encryption methods are used to provide three layer of security. Some experiments like ethical hacking is carried out in this paper. The experimental result shows that CCAF is highly secure than Business Process Modeling Notation (BPMN). With the help of the three layers viruses and Trojans were easily detected and blocked between 0.015 and 0.105 per second.SQL and MongoDB helped for SQL injection. No false alarm was detected by CCAF multi layered security. The study gives benefit for volume, velocity and veracity of big data service in the cloud.

Yong Yu *et al.,* [18] proposed a new IO based RDIC protocol with cryptographic key homomorphism method. The cloud server or a data storage server is enabled by the RDIC and it try to prove the verifier that it is honest to the data owner. Till now many RDIC protocol have been developed but it is mainly rely on public key infrastructure (PKI) expense. The main aim is to reduce the complexity. Here the ID based RDIC helps to secure against malicious server and a third party verifier with less knowledge privacy. The author clearly explains that a third party auditor (TPA) can check for data integrity instead of data owner. Two important constraints for security is proposed here i.e. soundness and perfect data privacy.

Mazhar Ali *et al.,* [19] explained about the various security issues in cloud computing environment. Also the author pointed out various security vulnerabilities in mobile computing. A detailed study about the cloud computing architectural framework is explained by the author. Cloud computing provides cost effective, manageable powerful resources on the internet

B. Mahalakshmi *et al.,* [20] has proposed that the data are stored in both public or private cloud i.e. hybrid cloud**...**Since public cloud is using the issue of data security will be high comparing to private cloud. To overcome the issue certain techniques to encrypt and decrypt data is proposed. Some important terms related to authentication security issue is discussed here. The access permission is given to the user by open authentication algorithm. The data is encrypted with convergence encryption technique. Data can be accessed by the authorized user in a secure way. The cloud server contains different types of data stored by different user (organization or individual). This affect the storage space and bandwidth. The data should not be accessed by the unauthorized user. A thorough check should be done for the duplication of data in the cloud environment. In some case a duplication of data will found in the cloud server.

Boyang Wang *et al.,* [21] detach the issue of revealing the identity of the data owner to the untrusted cloud server.

Here the author proposed a simple and efficient method to provide data integrity without revealing the identity of the data owner. A signature is generated by security mediator (SEM) and the data privacy is assured here. Here the system does not need extra storage overhead. The system is extending to work with multimodal SEM. Here the system modal contain data owner, user, cloud server, SEM. The paper clearly explains about the survey on few signatures like blind signature. The experimental result shows that computation, bandwidth, data privacy and identity privacy were minimized.

Deyan Chen *et al.,* [22] deeply discuss about the overall analysis of data security and privacy issues in the cloud environment throughout the life cycle. The author explained about the several problems faced by some cloud vendors like Amazon, Google doc in 2009.The paper gives a detailed explanation about data security issues. Some important points for solving the current solution for data integrity and security are explained in a comprehensive mode. Future work express in words about the need in improving the authorization and access control mechanisms in the cloud server environment.

Dimitrios Zissis *et al.,* allows making the cloud more secure and safe by providing various unique security requirements and by removing the potential data integrity threats and issues. To give some secure cloud environment the proposed system introduced a Trusted Third Party (TTP). The Public Key Infrastructure (PKI) is proposed in accordance with SSO and LDAP so as to provide authentication, integrity, security and confidentiality. In this system the client always fully trusted on third party. Here the author uses Public Key Infrastructure (PKI) so as to ensure confidentiality and integrity. TTP is responsible for addressing security issue in a multilevel distributed environment. Here PKI, Single-Sign-On, LDAP technologies are combined together to make a secure system. Using PKI, LDAP, SSO the security threats can be easily identified [5].

Yang Yang *et al.,* proposed a more efficient and privacy preserving auditing protocol with the help of auditing framework. Dynamic data operation get support from auditing protocol, which is more secure, reliable and efficient. With the help of bi-directional verification a new remote data auditing system is developed and it validates data storage security in cloud environment. By generating new authority credentials, the TPA becomes less credible. The role of CSP is to verify the user authentication and reject unauthorized user. Since there is possibility of computation mismatch between PC and CSP, which reduce the client communication, speed. A dynamical allocation strategy is used here. This paper added some extra methods to find and fix errors. The important blocks of users file will check for integrity by the data owner at low cost. In future the author will use more credential authentication and verification schemes. The System explained in this paper provides high security and computational load [17].

Bin feng *et al.,* proposed a system to protect the security of cloud storage, a scheme with large searchable one is introduced, which uses regular language encryption and DFA search function. The regular language search with privacy preserving and securely proved by decisional Diffie–Hellman (DDH) assumption helps to design a data storage system with more security. The system enables a regular language search with more flexible search pattern. A public key and regular language string is taken as input for the encryption algorithm. A cipher text generated is stored in cloud server. In order to get better performance analysis, the proposed system is simulated on laptop. The experimental result in the paper shows transmission bandwidth and computational overhead are comparatively less [2].

Salah H. Abbdal *et al.,* explained that the scrambled data is produced by combining iris features and original data. A two step process is there to check whether data is changed or not. The first process is installation process. Here shared key and metadata are generated by client. Next t step is verification process where the client provides a challenge for the server and the server had to respond. The server sends the original data and the signature. The client decrypts the meta data and check whether the data matches. If data is matched means the data is not changed [15].

V. METHODS TO PROVE DATA INTEGRITY

A. Proof of Data Possession (PDP)

As the technology improves day by day the use of cloud storage and services are also become an unavoidable one in the current system. As far as companies and organization concerns the rate of growth of remote storage system also emerged as a new storage system with lot of security issues. Proving the data integrity on remote server is a big issue in cloud environment. In PDP by keeping a copy of data .the client sends the data to the untrusted remote server for storage. The methods like MAC (Message Authentication Code), encryption scheme or some other methods are used to ensure data integrity [15]. Before sending the file to server the user fills the data with some metadata. The PDP method is used for encrypted and plain data.

The Fig 5 shows two step processes for PDP. In the first step, with the help of probabilistic key generation algorithm the public and secret key is generated by the client. This step is called Preprocess and store. In the second step the server responds to clients challenges by sending the file F`. The client verifies the file for correction. If no change in the file means the file is secured [4].

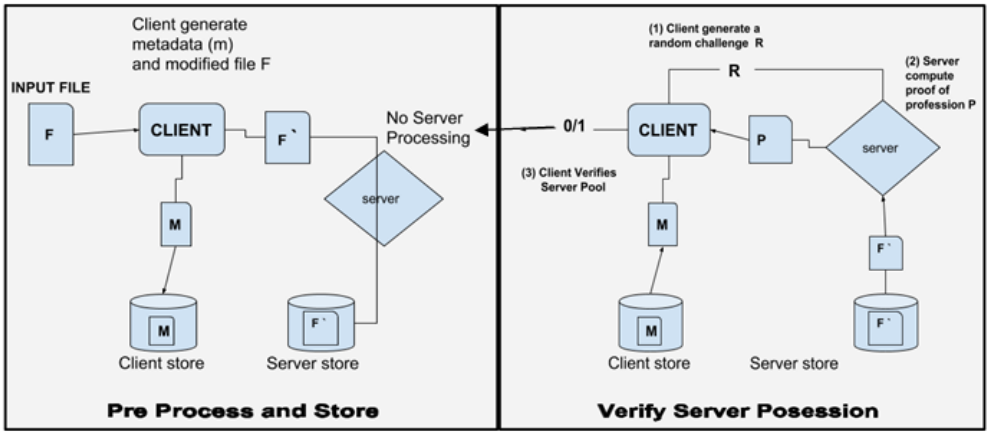

Fig. 5 Proof of Data Possession

Author Nesrine Kaaniche *et al.*, [14] with the help of Interactive Proof System (IPS) and a usage of GPS he proposed a determinic Proof of Data Possession scheme for providing tight security for data in the cloud. This paper removes the data owner's burden of periodic verification because it supports public verifiability. The message between storage server and clients are composed of small group elements due to constant communication complexity.

The result shows that the system used here is more efficient and secure. It prevents frauds and data leakage [14].

1. Scalable PDP

The scalable PDP fully depends on symmetric key operations in both verification and installation process. Since this type of PDP require less bulk encryption and of data and data expansion, the scalable PDP is better than PDP. In the random oracle model also this type of PDP proved it is secure. Also this PDP support dynamic operations like appending, modification and deletion. The system needs low cost and easily support dynamic outsourcing of data. [8]

2. Dynamic PDP

A Dynamic PDP (DPDP is an efficient and secure one which support provable updates o store data. A formal framework for DPDP is introduced .The security flaws in during hash aggregation are easily identified by the DPDP. A Dynamic PDP is most useful for data storage in cloud. It is very useful in insertion of new block, modification of a block and deleting a block. To organize the dictionary entry DPDP uses rank based information. It support efficient authentication on file operations. The Dynamic PDP is a group of some polynomial- time algorithms like Keygen, DPDP, Prepare Update DPDP, Prove DPDP, and GenChallenge DPDP [11].

B. Proof of Retrievability (POR)

POR is one of the schemes to ensure data integrity and availability in cloud environment. It is a protocol specially designed for cloud storage to make sure the client that the data is flawless. For getting better data retrievability POR is better than PDP. The uncorrupted data can be easily recovered by POR. Fig. 6 shows the process for POR. First step Setup phase, second one Verification phase [3].

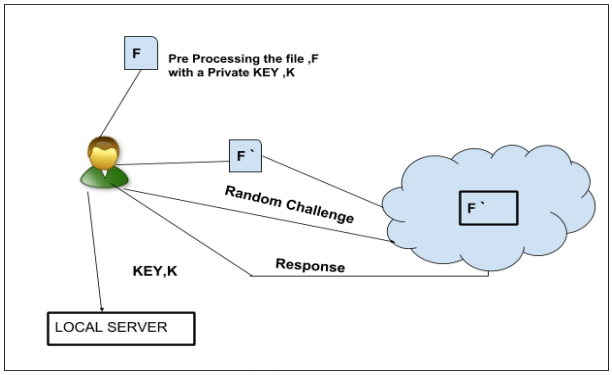

Fig. 6 Proof of Retrievability (POR)

1. Setup Phase

Some authentication code is generated by user by preprocessing the data file with the help of private key. With the authentication code the client sends the data file to the cloud sever and thereby removing that from the local server. Finally the user has the private key in the local server, and CSP has both data file and authentication code

2. Sequence of Verification Phase

Here the user wills generates some random challenges query and sends to the CSP. After receiving the random query the CSP response based on users file and authentication code. Finally the user will verify the CSP response using the private key and decide whether to accept or reject the CSP's response.

3. POR for Large Files

In this type of POR the cryptographic key is stored by the prover. The verifier also stores the key irrespective of the dimension and range of files. Here the pover access a tiny

portion of large file F. Verifier touches the part of file F, which is independent of length and contains few hundreds of block. This scheme encrypts the file F and includes a group of random valued blocks called as sentinels. Public auditability and data dynamics were achieved by this scheme [1].

4. Compact POR

Here two short homomorphic authenticator is being engaged. The first work with Pseudorandom functions so as to make the data retrieval more secure with standard model [10]. The second one work with BLS (Bobeh-Lynn-Scacham) signatures so as to make the retrievability more secure with random oracle model [9].

5. Dynamic Outsourced Proofs of Retrievability (DOPOR)

In this scheme the DOPOR makes the data dynamics more efficient from the corrupted code. It defends against malicious TPA. So as to ensure the logarithmic complexity DOPOR uses a special authenticated data structure called bc23Tree.To avoid the deletion attack , a hierarchical storage structure with uniform size level encoded data and encoded update operation is deployed [6].

VI. CONCLUSION

In this paper we explained about the basic cloud computing concepts. The cloud computing issue like data integrity and security is explained in a neat manner. Data integrity issue and its constraints were also explained. Some data integrity techniques like PDP, POR were explained with suitable diagram. A small survey is taken for data integrity. The various PDP likes scalable PDP, Dynamic PDP were also explained. We also put forward about POR types like POR for large files, Compact POR, and Dynamic Outsourced Proofs of Retrievability (DOPOR).

REFERENCES

[1] Juels and Burton S. Kaliski, "PORs: Proofs of Retrievability for Large Files", *CCS '07 Proceedings of the 14th ACM Conference on Computer and communications security*, 978-159593-703-2,USA

[2] Bin feng, Xinzhu Ma, Cheng Guo, Hui Shi, Zhangjie Fu, (MEMBER, IEEE), and Tie Qiu, (SENIOR MEMBER, IEEE), "An Efficient Protocol with bidirectional verification for storage security in Cloud computing, special section on emerging trends, issues and challenges in energy-efficient cloud computing", *Digital Object Identifier 10.1109/access.2016.2621005.*

[3] Choon Beng Tan, Mohd Hanafi Ahmad Hijazi, Yuto Lim and Abdullah Gani, "A survey on Proof of Retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issue", *Journal of Network and Computer Applications*, pp. 75-86, 2018.

[4] [Online] Available: http://cryptowiki.net/index.php?title=Proof_of_ data_possession.

[5] Dimitrios Zissis, Dimitrios Lekkas, "Addressing cloud computing security issues", *Future Generation Computer Systems*, pp. 583–592, 2012.

[6] Lu Rao, Tengfei Tu, Hua Zhang, Qiaoyan Wen, and Jia Xiao, dynamic Outsourced Proofs of Retrievability Enabling Auditing Migration for Remote Storage Security", *Wireless Communications and Mobile Computing*, 2018.

[7] Faheem Zafar, Abid Khan, Saif Ur Rehman Malik, Mansoor Ahmed,Adeel Anjum, Majid Iqbal Khan, Nadeem Javed, Masoom Alam, Fuzel Jamil, "A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends", *Computers and Security*, pp. 29–49, 2017.

[8] Giuseppe Ateniese, Roberto Di Pietro and Luigi V. Mancini, "Scalable and Efficient Provable Data Possession", in *the Proceedings of the 4th international conference on Security and privacy in communication networks*, Article No. 9 Istanbul, Turkey - September 22 - 25, 2008

[9] Gururaj Ramachandra, Mohsin Iftikhar and Farrukh Aslam Khan, "A Comprehensive Survey on Security in Cloud Computing", *The 3rd International Workshop on Cyber Security and Digital Investigation.* [Online] Available: www.sciencedirect.com, Science Direct Procedia Computer Science, Vol. 110, pp.465–472, 2017.

[10] Hovav Shacham and Brent Waters, "Compact Proofs of Retrievability", *ASIACRYPT 2008*, International Association for Cryptologic Research 2008, pp. 90–107, 2008.

[11] Jing Zou, Yunchuan Sun and Shixian Li, "Dynamic Provable Data Possession Based on Ranked Merkle Hash Tree", *International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*, 2016.

[12] Victor Chang, Yen-Hung Kuo, Muthu Ramachandran, "Cloud computing adoption framework: A security framework for business clouds", *in Future Generation Computer Systems*, pp. 24–41, 2016

[13] Nesrine Kaaniche, Ethmane El Moustaine and Maryline Laurent, "A Novel Zero-Knowledge Scheme for Proof of Data Possession in Cloud Storage Applications", *14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2014.

[14] [Online] Available: https://www.pharmout.net/data-integrity-alcoa/

[15] Salah H. Abbdal, Hai Jin, Deqing Zou and Ali A. Yassin, "Secure and Efficient Data Integrity Based on Iris Features in Cloud Computing", *7th International Conference on Security Technology*, 2015.

[16] M. A. Shah, R. Swaminathan and M. Baker, "Privacy-preserving audit and extraction of digital contents", *Cryptology ePrint Archive*, 2008.

[17] Yang Yang, Xianghan Zheng, Chunming Rong, Wenzhong Guo, "Efficient Regular Language Search for Secure Cloud Storage", *IEEE Transactions on Cloud Computing (Early Access)*, DOI: 10.1109/TCC.2018.2814594, 2018

[18] Y. Yu, M. Au, Ho. G. Ateniese, X. Huang, W. Susilo, Y. Dai, & G. Min, "Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage", *IEEE Transactions on Information Forensics and Security*, pp. 767-778, 2017.

[19] Mazhar Ali Samee U. Khan, Athanasios V. Vasilakos, "Security in cloud computing: opportunities and challenges", *Information Sciences*, pp. 357–383, 2015.

[20] B. Mahalakshmi and G. Suseendran, "Effectuation of Secure Authorized Deduplication in Hybrid Cloud", *Indian Journal of Science and Technology*, 2016.

[21] B. Wang, S. S. M. Chow, M. Li and H. Li, "Storing shared data on the cloud via security-mediator", In Proceedings of International Conference on Distributed Computing Systems, pp. 124–133, 2013 .

[22] D. Chen and H. Zhao, "Data security and privacy protection issues in cloud computing", *International Conference on Computer Science and Electronics Engineering*, pp. 647–651, 2012.

[23] [Online] Available: https://www.fda.gov/downloads/drugs/guidances/ ucm495891.