ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Data Integrity Check in Cloud Computing using Hash Function

Shephali Singh[1], Puneet Sharma[2], Dr. Deepak Arora[3]
Department of Computer Science & Engineering,
Amity University, Lucknow, UP, India

***Abstract-*** Storage, server, networking, database and more are the services delivered by cloud computing over the internet. Cloud Computing is globally scaled that provides virtualization, scalability, fine grained data access control and performance benefit. Despite all this benefits, there are various issues of security and challenges for cloud. The challenges are that data is no longer in full control of owner, compromised cloud server and access control desired by owner. Customer needs a platform that stores its information and returns it without any changes or modification. Most of the cloud computing environment guarantees safety of data by various advancedencryptions. The information stored is encrypted at server side and is decrypted at client side. This makes data secure so that no third party can access or understand the data. The concern is unfaithful behaviour of the cloud itself. The data stored on cloud may be possible altered or modified without the knowledge of client.There should be a mechanism that verifies the stored and the data being retrieved is same. This paper investigates such mechanism.
This research paper points out how data integrity of information stored in cloud can be checked. In this technique after encryption of data, hash is created using hash function. At customer side after decrypting the data each hash is compared with other hash set to verify uniqueness of the data. This verifies is the data has been altered or it is same as the original data stored by the client.

**Keywords-** Client, Cloud Computing, Data, Encryption and Integrity.

## I. INTRODUCTION

Various computing resources are provided by cloud computing paradigm as a service to the customers. Cloud computing has gained a lot of popularity in both IT industries and academic world. Cloud computing has a paramount advantage of resources, storage, dynamic scalability and virtualization of the infrastructure and services. Models of cloud computing architecture consist ofPaaS (Platform as a Service), SaaS (Storage as a service) and IaaS (Infrastructure as a Service). Pay-per-use system is provided by these models giving users the authorization to access its application and services. Some of such clouds areMicrosoft Azure, Amazon's EC2 and Salesforce CRM. Storing of data and computing services are few of the typical services provided by the cloud computing paradigm. Moving business applications and services into the cloud, customers can save the capital investment of implementation and maintenance of their applications.

A third party or technology has to be used and trusted by organisations for safeguarding their data and information when using cloud services[28]. The main issue is that the client and customer is aware that third party auditor can behave unfaithfully because they are out of customer and clients reach. Thus it arises several danger to data like loss or modification of important data located in the clouds. Very confidential and sensitive data are stored in the cloud by individuals or industries that cannot be open to unauthorized access by any third party. Loss of any such data can lead to business at risk and loss. To protect these sensitive data from unauthorized access different methods can be used. Generally, before storing data in cloud it is encrypted using various algorithms [1]. Though the toughest encryption techniques also do not guarantee 100% security as hackers and unauthorized people can compromise its integrity. Therefore, a process needs to be developed to check integrity of the data to verify that integrity of data has not been compromised and data modification has not occurred. This will provide satisfaction and confidence to the customer of cloud that their data is preserved and protected by the cloud.

A third party auditors could be brought into role to verify data integrity stored in the cloud. A challenge response system is used by the third party auditors for the data stored in the cloud. Challenge response authentication is done after a fixed intervals of time for checking the integrity of data. But for successful execution of this method customer should develop a trust in the third party auditors too. From the customer's view point TPA is same as cloud service providers since both have the access to customer's data stored in cloud. The third party auditor itself could behave unfaithfully or can be compromised leading to data integrity loss. Instead of safekeeping the customer's data, third party auditor itself could act as the insecure channel leading to data leakage. This vulnerable link of third party auditor for integrity check should not be used due to security reasons. Instead of TPA integrity of data can be checked by the customers themselves.

This paper deals with the uses of a hash function which is calculated by the customer. The customer calculates the hash value of the content stored in cloud before the encryption and stores the hash locallyin a secure hash repository and then uploads the file to the cloud for storage. The data is then encrypted by the cloud and is stored. To check the integrity of a data file stored in cloud, the customer takes that data contents and calculate hash value which is matched with earlier calculated hash value. Change in the two hash values, will depict that data integrity has been compromised since the hash values doesn't matches

which means data has been modified. This is a less complicated scheme and is more secure than thethird party auditor scheme. It provides a simple and efficient way to check integrity of data.

## II. LITERATURE SURVEY

Researcher Gennaro et al. [5] has shown literary working of secure computation outsourcing. He has shown feasibility of input and output privacy maintenance as well as correctness and soundness of the result. However, it is not practically sound method due to high computing complexity. Later Atallah et al. did a list of works [1] [3] [4] [8] for secure outsourcing computation. The set of techniques used were string matching, linear algebra, sorting etc. However, these mechanism were not very efficient in securing input and output information and did not confirmed correctness of result, which is the main concern in secure computation on cloud. Atallah et al. [3] [4] gave two protocols later that were used for secure sequence outsourcing and algebraic computation outsourcing. Since these two protocols used burdened cryptography algorithm like homomorphic encryption [11]. Thus were not very successful for large problem set due to huge complexity. Based on above concept, Hohenberger [2] defined secure outsourcing protocol of modular exponentiation, which was taken to be a public key cryptography method that was very expensive.

Latest, Atallah [8] et al. presented a safe protocol based on secret key concept for secure outsourcing that used matrix multiplication [12]. This mechanism performed very well due to assumption of only one server and computation effectiveness. The only lacking point is lot of overhead because of message passing. Concluding, these methods are still not efficient enough for secure LP outsourcing computation.

Safe multiparty working was given by Yao [6]. Two or more parties are allowed to execute functions for obtaining result along with preserving their input from each other. Result is computed along with hiding the input from both the parties individually. Basic SMC is efficient, Du and Atallah et al. gave customized solution under SMC context for problem such as scientific computation, sequence comparison, statistical analysis etc. [13]. Though applying these concepts directly to the cloud is problematic. The computational power of customer and cloud was not similar which could not be handled effectively, which is avoided in the given design by shifting all computational load to cloud only. The other problem is security asymmetry because no party alone knows all problem input, leading to difficulty in result validation.

In SMC, Li and Atallah [14] gave a solution to the participating parties to apply additive split of constraint matrix along with few cryptographic techniques that are executed in each step of simplex algorithm. The above method does show practical performance for big size problem and does not guarantee optimal solution. With the same method, Toft [15] gave a secret key sharing secure simplex algorithm that had less complexity than other protocols. In [16], Vaidya formulated a new improved simplex algorithm that worked on safe scalar product and protocol comparisons. Lately, Catrina et al. [17] gave a safe multiparty LP using fixed point arithmetic. Some other works are of Du [18] and Vaidya [19] who studies distinguished approaches of matrix based transformation to look inti privacy preserving linear programming. Later, Bednarz et al. [20] proved Du's and Vaidya's approach infeasible and proposed to use permutation matrices. Recently, Mangasarian gave two privacy protecting methods of linear programming on vertical [21] and horizontal divided [22] constraints matrix. Although,many techniques were proposed but all had computation asymmetry issues.

Cloud computing is not a very trustworthy platform. It may behave unfaithfully during computation which may lead to incorrect computation of result without the knowledge of the customer. Detecting this is not an easy task that when the data is being modified in cloud which results in incorrect computation outsourcing. Verifiable computation delegation has found huge interest in theoretical computer science communities where weak customers can find out the correctness of computational result with the help of powerful but not trusted servers with the use of very less resources. Some of latest results and outcome is specified by Glodwasser et al. [23]. Golle et al. [24], to defeat the untrusted servers presented the idea to append pre-calculated results along with the computation. In [25], Szada et al. further worked on ringer scheme to defeat servers that cannot be trusted. In [26] Du. Et al. gave a mechanism for grid computing to find the cheating done in outsourcing computation. Based on Merkle tree the servers gives a commitment on the result computed by it. This commitment is then used by the customer followed by a sampling technique to do result verification.

The above schemes looks into data and the computed result by it that is prohibited in cloud computing for security and safety of data. Thus it becomes tough task to provide result verification as well as input/output privacy. The introduction of concept of duality of LP problem efficiently performs the result validation, appending some overhead on customer server as well as cloud server.

Cong Wang [9], recently, gave an efficient and feasible method for securely outsourcing linear programming that will protect input/output and also find cheating servers. The data that are outsourced contains very insecure information such as medical history, financial details, research related works etc. To protect these important data and ensure its confidentiality it is encrypted before being outsourced to the cloud for computation. On the other hand, the details of computation is not transparent to user so there exists chances of cloud to behave unfaithfully and produce wrong output. Fully homomorphism encryption (FHE) scheme, has been shown feasible in theory for secure computation outsourcing.

Computation outsourcing involves two different parts, cloud customer and cloud server. Cloud customer outsources LP problems to cloud server for computation. The cloud server has huge computing resources such as memory, storage and processing power. The customer sends its LP problem to CS after encrypting it with a secret key[27]. The CS then computes the solution with the help of public LP solver running on the cloud and also produces a correctness proof. The customer on receiving the result verifies the result with the appended proof and then decrypt the result.

## III. EARLIER WORK

Both academic and IT world still consider cloud computing as new and immature technology. Many researches have looked into security and safety domain of cloud computing and still researches are going on [28]. Cloud computing's one of the latestresearch topic related to data privacy preserving is data integrity check by third party auditor. Reservations are always made by customer in trusting third party cloud service provider. The important concern is that, third party auditor services and existing system can be integrated to check the integrity data stored in cloud.

For maintaining the security of cloud data the method of auditing service externally that will check that data stored in a cloud is same or not can be used [29]. In this technique random masking technique along with the public key based authenticator is applied to achieve the aim of privacy preserving auditing. This method ensures that no extra overhead is created for the customer sincedata is not saved locally for third party auditors. And it is guaranteed, after combining no vulnerability in the existing security system will occur. This leads to a very efficient, secure and high performance privacy preservation method. [10]

The followingmethod uses extraction protocol by third party auditor to ensure data integrity of customer [30]. This method does not involve any type of encryption of data using symmetric keys by the customer. This is because there are chances of keys being lost by the customer itself which may lead to loss of data. Generation of any secret keys or hashing of data or encryption is not required by customer in this privacy preserving technique. This is one of the major advantages of this method. The customer can retrieve data whenever required as per use.

## IV. PROPOSED SYSTEM

The concept of third party auditors does not remove the trust problem for checking ofdata integrity of the cloud. This is more like not solving the problem and finding alternatives to depend on one party instead of another party. If cloud platform cannot be trusted by customer, in the same way it cannot trust the third party auditors.

The third party auditor is given the access to data and the key to decrypt it which is not a very reliable method. Though the third party auditor provide the assurance of security but they cannot be trusted since their main agenda is money making, not the safety of customer data.

Complexity increases with the introduction of the third party as a new member is involved. Now there are three involvement customer, cloud and the third party auditors. Thus customer has to involve with the cloud service provider and the auditor. This will increase communication overhead and complexity.Customer will send audit request, the auditor will communicate with the cloud and reply to the client. This communication between customer and auditor/auditor and cloud service provider needs additionalchannels thus exhausting network bandwidth and creating more overhead.

Along with cloud service charges, customer will have to pay additional charges to third party which will be a waste if proper security is not provided by the TPA. Auditors have access to customer's data which can be made available to any unauthorized people. This will lead to concern for customers to keep the data safe from the third party. This trust problem can be handles by removing the role of third party auditor. Instead of TPA, customer involvement will be made to check the integrity of data. Hash function can be used for integrity check. The hash function will be calculated at customer side to avoid any trust problem.

### *A. Hash Function to check data integrity*

There should be a simple method for integrity check of data that could beeffectively and efficiently implemented for users or customer. The trust problem between customer and TPA can be solved if user perform the integrity check themselves rather than using third party services such as third party auditing. This can be done by calculating the hash value of the data by customer itself and saving these computed hash values locally in secure hash repository created by the customer. Figure 1 depicts this scheme.

This paper depicts the use of property of hash for checking the precomputed and recomputed hash values for checking data integrity. The method lets the user compute the hash value/hash digest of file and then upload the file for storage to the cloud. The data on the cloud side is encrypted and stored. The customer then stores the calculated hash value in a secure localhash repository. The file can be retrieved from the cloud whenever the customer wants to check the integrity of data. The data is decrypted at the cloud side and returned to the customer. The hash value of the data is computed and is matched with the previously calculated hash values that is stored in the hash repository. If both the hash values are same, it indicates that the integrity of data has been maintained. This leads to the customer satisfaction that the data has not been modified and is safe and secure. The hash value of any message doesn't changes if the message is same, thus any change in hash value will depict that the message has been altered. If the previously created hash value and the recomputed hash values matches, it shows that the data has not been modified and its integrity is intact.

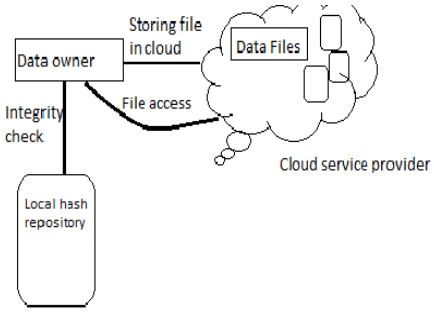


Figure 1. Data integrity check in cloud environment using hash function

### *B. Case Study*

This paper mostly focusses on data integrity of information stored in cloud. Though many security measures are applied by the cloud service providers but those techniques do not guarantee data integrity. To check the data here we are implementing hash function concept to check correctness of user's data. The data is correct or not can be checked by the user itself using hash value.Here we are using a software named HashCalc to calculate hash value of any file, text or string. It can calculate hash value of files having size upto 15 GB. It can calculate different hash function values such as MD4, MD5, SHA1, SHA256, MD2, CRC32 etc. We take a file and upload it in this software and calculate SHA1, SHA256 and SHA384. The resultant hash values are as follows:

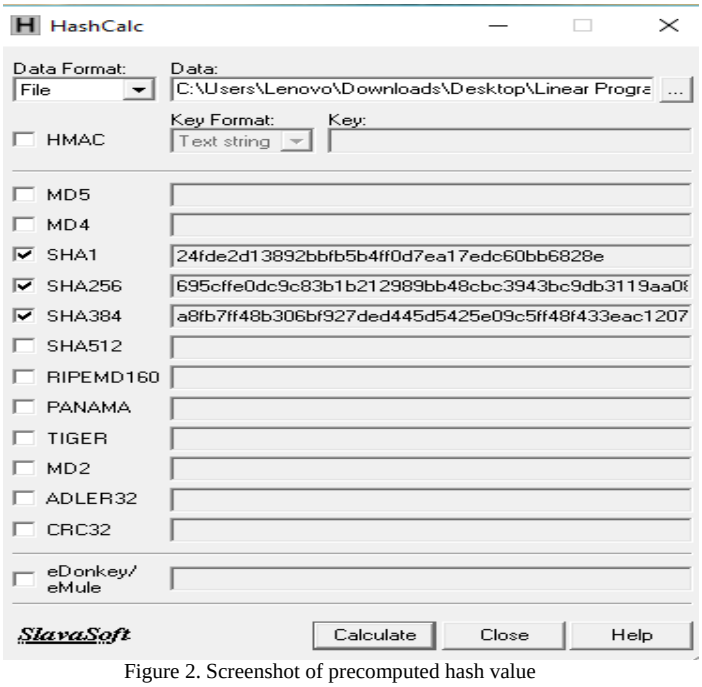


Figure 2. Screenshot of precomputed hash value

Now we make changes in the same file and modify some data in the file. Again, the same hash values of the file are calculated using the software. This time there are changes in the hash value of the file. The following hash values are:

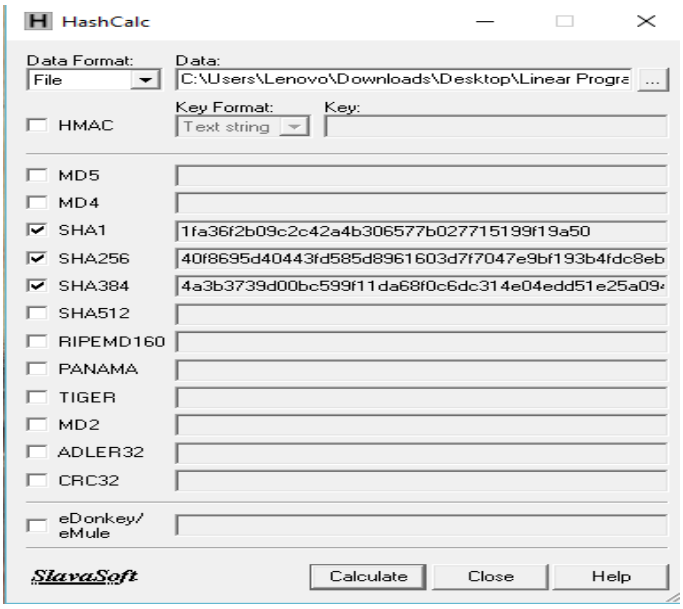


Figure 3. Screenshot of recomputed hash value

The change in hash value suggests that the data has been modified and data integrity has been compromised. This same method can be used in cloud computing environment to check the data integrity of files stored in the cloud. This method is suggested to be applied on the customer side, the hash values calculated can be saved in secure hash repositories which can be later checked by the customer at the time of retrieving the file from the cloud.

## V. CONCLUSION

The primary challenging issues in cloud computing has been the security of data and trust problem. This paper tries to find the security issues of cloud computing and deals with data integrity check of the customers data. The method discussed in this paper, calculates the hash value of customers data on customer side, which checks the integrity of data and as the removes the role of third party auditors. A local hash repository is used to store the computed hash value safely and securely. The files of customer can be downloaded at any time from the cloud and computed hash value can be matched with the hash value stored in repositories.This can be very efficient and effective for small scale to check the faithfulness and authenticity of cloud and already developed SLA. This method saves money by human involvement rather than being dependent on technology that involves money and is not trustworthy. Cloud computing still needs to be developed to provide a more secure and safe platform for the users to use. The method used and discussed in this paper is very basic and easy with the involvement of customer itself.

## VI. REFERENCES

[1] M. J. Atallah, K.N. Pantazopoulos, J.R. Rice, and E.H. Spafford, "Secure Outsourcing of scientific computations", Adv. Comput., vol.54,pp. 216-272, 2001.

[2] S. Hohenberger and A. Lysyanskaya, "How to securely outsource cryptographic computation", in Proc. 2nd Int. Conf. Theory Cryptography, 2005, pp. 264-282.
[3] M.J. Atallah and J. Li., "Secure outsourcing of sequence comparisons," in Int. I. Inf. Sec., vol. 4, no. 4, pp. 277-287, 2005.
[4] D. Benjamin and M.J. Atallah, "Private and cheating-free outsourcing of algebraic computation," in Proc. Int. Conf. Privacy, Secur., Trust, 2008, pp. 240-245.
[5] R. Gennaro, C. Gentry, and B. Parno, "Non-interactive verifiable computing: Outsourcing computation to untrusted workers," in Proc. 30thAnnu. Conf. Adv. Cryptol., Aug. 2010, pp. 465-482.
[6] A. C.-C. Yao, "Protocols for secure computations (extended abstract)," in Proc. of FOCS, 1982, pp. 160–164.
[7] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc of STOC, 2009, pp. 169–178.
[8] M. Atallah and K. Frikken, "Securely outsourcing linear algebra computations," in Proc. of ASIACCS, 2010, pp. 48–59.
[9] Cong Wang, Kui Ren, and Jia Wang, "Secure and Practical Outsourcing of Linear Programming in Cloud Computing", IEEE, 2011, pp. 820-821.
[10] D. Luenberger and Y. Ye, Linear and Nonlinear Programming, 3rd ed. Springer, 2008.
[11] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proc. 17th Int. Conf. Theory Appl. Cryptographic Tech., 1999, pp. 223–238.
[12] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11, pp. 612–613, 1979.
[13] W.Du and M. J. Atallah, "Secure multi-party computation problems and their applications: A review and open problems," in Proc. New Secur. Paradigms Workshop, 2001, pp. 13–22.
[14] J. Li and M. J. Atallah, "Secure and private collaborative linear programming," in Proc. Int. Conf. Collaborative Comput., 2006, pp. 1–8.
[15] T. Toft, "Solving linear programs using multiparty computation," in Proc. 13th Int. Conf. Financial Cryptography Data Security, 2009, pp. 90–107.
[16] J. Vaidya, "A secure revised simplex algorithm for privacy-preserving linear programming," in Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl., 2009, pp. 347–354.
[17] O. Catrina and S. De Hoogh, "Secure multiparty linear programming using fixed-point arithmetic," in Proc. 15th Eur. Conf. Res. Comput. Security, 2010, pp. 134–150.
[18] W. Du, "A study of several specific secure two -party computation problems," Ph.D. dissertation, Comput. Sci. Dept., Purdue Univ., West Lafayette, IN, USA, 2001.
[19] J. Vaidya, "Privacy-preserving linear programming," in Proc. 24th ACM Symp. Appl. Comput., 2009, pp. 2002–2007.
[20] A. Bednarz, N. Bean, and M. Roughan, "Hiccups on the road to privacy-preserving linear programming," in Proc. ACM Workshop Privacy Electron. Soc., 2009, pp. 117–120.
[21] O. L. Mangasarian, "Privacy-preserving linear programming," Optim. Lett., vol. 5, pp. 165–172, 2011.
[22] O. L. Mangasarian, "Privacy-preserving horizontally partitioned linear programs," Optim. Lett., vol. 6, no. 3, pp. 431–436, 2012.
[23] S. Goldwasser, Y. Kalai, and G. Rothblum, "Delegating computation: interactive proofs for muggles," in Proc. 40th Annu. ACM Symp. Theory Comput., 2008, pp. 113–122.
[24] P. Golle and I. Mironov, "Uncheatable distributed computations," in Proc. Conf. Topics Cryptol.: The Cryptographer's Track RSA, 2001, pp. 425–440.
[25] D. Szajda, B. G. Lawson, and J. Owen, "Hardening functions for large scale distributed computations," in Proc. IEEE Symp. Secur. Privacy, 2003, pp. 216–224.
[26] D. Szajda, B. G. Lawson, and J. Owen, "Hardening functions for large scale distributed computations," in Proc. IEEE Symp. Secur. Privacy, 2003, pp. 216–224.
[27] Shephali Singh, Puneet Sharma, Dr. Deepak Arora, "Secure Outsourcing of Linear Programming in Cloud Computing:A Review,IJERA,www.ijera.com/papers/Vol7_issue4/Part6/L0704066468.pdf.
[28] Parisha, Pooja Khanna, Puneet Sharma, Sheenu Rizvi, "Data Partitioning Technique In Cloud: A Survey On Limitation and Benefits",IJERA,http://www.ijera.com/papers/Vol7_issue4/Part-6/M0704066975.pdf.
[29]Mehul A. Shah, Ram Swaminathan, Mary Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," IACR, 2008, http://eprint.iacr.org/2008/186.pdf
[30] CATTEDDU, D. & HOGBEN, G. (2009): Cloud Computing: Benefits, risks and recommendations for information security; European Network and Information Security Agency (ENISA)

 