International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

A Provably Secure Distant Data Integrity Checking With Homomorphic Hash Function in Secure Cloud Storage

Manasa M J, S Usha , Rajesh K S

Abstract: ***The cloud storage with user facilities like great data storage quality, higher computing, scalable and flexible has been one of the major application of cloud computing. Large number of data holders is subcontracting the files to the cloud. The cloud server being a public domain is not very reliable thus the data holders are required to find a reliable and trustworthy way to check the ownership of the data files that they subcontract on the cloud server which is present in a remote location. To tackle this drawback number of Distant Data Integrity Checking (DDIC) protocols is in the literature but these protocols have liabilities with respect to the data dynamics and the efficiency. This work recommends a novel DDIC protocol built across homomorphic-hash-function. This system gives a resistance against number of attacks like replay, replaces and falsifications. This work introduces the Operation-record-table(ORT) which is an optimized table resulting in keeping the constant cost for supporting the data variations and keep tracks of blocking the file operations.***

Keywords : Cloud computing, DDIC: Distant Data Integrity Checking, Operation-record-table, Data Dynamics, Homomorphic-hash-functions

I. INTRODUCTION

Cloud computing appears as a new paradigm of calculation after the grid. The resources across the cloud have to be managed as it possesses very high speed computing and a large memory which is Virtual [1]. The Service provider of the cloud should always give the users with the cost effective, trustworthy and scalable service for data storage. The services are levied to the users at the metered cost and the cost will be calculated based on the demands of the users for the storage and the computation. This model is also called "Pay-Per-Use" Model and the users will be compelled to use the required IT infrastructure with a initial investment which might come down in cost significantly on a long run [2]. The users can also compromise on their rented resources by changing the number of outsourcing they will be doing.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Manasa M J[*], Department of Computer Science Engineering RajaRajeswari College of Engineering, Bangalore, India Email: mj.manasa17@gmail.com
S Usha, Department of Computer Science Engineering RajaRajeswari College of Engineering, , Bangalore, India Email: sakthivelusha@gmail.com
Rajesh K S , Department of Computer Science Engineering RajaRajeswari College of Engineering, , Bangalore, India. Email: rajeshks_hrr@yahoo.com

The cloud service supplier is trying to provide a promising data storage service, saving users the asset and resource expenses. On the other hand, cloud storage also includes different security problems for outsourced data [3]. Even though few security issues have been determined, significant issues related to data falsification and data loss are as yet present in cloud storage. From one perspective, the crash disk error or the cloud storage server (CSS) equipment collapse can cause surprising fraud of outsourced files. However, CSS isn't completely reliable from the perspective of the data owner; it can effectively remove or alter files for enormous economic advantages. Simultaneously, CSS can conceal the data owner's incorrect behavior and data loss incidents in order to keep a good quality status [4], [5]. Hence, it is critic also as to the data owner uses an effective means of verifying the integrity of data outsourcing [6].

The main intention of the CSP is to provide the required computation and storage while being cost effective for the users. The main issue in the cloud is the security of the data dumped in the cloud by the user [7]. There is large number of security protocols implemented for the cloud data but the problem of data loss and the imitation of the data or the wrong data is still persisting. The outsourced files may be completely lost if the storage hardware crashes or the server crashes and the recovery maybe next to impossible. This poses a threat to the user with respect to this data as it can be modified or deleted for a cause of profit gain and he finds the cloud storage unreliable. The Cloud storage Server (CSS) is also having the capacity to hide the identity of the data owner his misbehavior and also any incidents of the data loss for the sake of the service provider's reputation [8]. Thus the data holder or the source should always use an effective integrity check process before outsourcing the data.

The DDIC protocols are very efficient protocols for checking the integrity of the stored files on the CSS. It gives an option to the user to make sure the service provider is faithful and he will maintain the integrity of the data. In these protocols the user will actually challenge the CSS on the file he has outsourced and in turn the CSS has to produce the proof of its sincerity. It shows that it indeed maintains the original data files and no procedure for recovering is used.

The main goal of these DDIC protocols is to check the file integrity without getting hold of the complete original file [9]. One more necessity is that the protocols must support the various data operations. The operations may be like the user should be able to modify, delete, add the new data or the block of data as required the user.

While subscribing for these services the user should be intimated about the complexity of extra calculations and the message sharing between the protocol and this leads to extra cost and it needs to be considered for the real time application.

II. RELATED WORK

In [10], the author proposes the Provable Data Model (PDM) for the verification of the data integrity. The verification of data integrity is largely on the go as the cloud space and computations are being used to process and store the user's data. This leads to a requirement of data integrity from user's side to ensure the CSP is genuine and do not hamper the data. The PDM model sends the untrusted CSP the data for storage and keeps some of the data with itself. The user then challenges the SP to prove itself genuinely and check if the data stored is the original copy and not he tampered or the modified data. These models are used only for the files which has dynamics of only appending.

Here the PDP model is extended to Dynamic PDP model with a better efficiency framework. This is used to update the data stored and prove authenticity of the CSP. The cost of the dynamic updates may vary for different size of the data files but the amount of misbehavior from the CSP may be same. Here the author shows that to prove the trust of the SP for a 1GB data the proof size of 415KB and the computational requirement up to 30ms is used. The work also shows how DPDP can be used to check the CSP for the user's data integrity and also the Version Control System (CVS).

The [11], is a keynote article which presents the vision to the 21st century processing, different processing platforms to distribute the computing needs of the various users. It also describes the CC architecture for generating the CC model which is focused at the cloud market where the resources of the CSP clouds is rented in the form of the virtual machines. The keynote also focuses on how to manage the leveraged resources which includes both user requirement management and also the risk that will be encountered while maintain the SLA oriented allocation of the resources. The author proposes some cloud platforms that can be used in the present industry and also gives the overview of the platform being used by the present industries. It shows how to interface the proposed platform with the platform the current industry is using. The author even highlights on the use of 3rd generation Aneka enterprise grid technology for renting cloud services and how the cloud environment for dynamic environment is created. The author concludes that with the combining effort of all the IT paradigms in the CC the better services could be leveraged to the users in the coming century.

The paper [12], concentrates on dealing with the data replication at the CSP's cloud. The CSP displays it has the multiple copies of the user's data which might be harmful for the user. In reality it will be possessing only one copy of the data. This leads to checking the SP for the trust of keeping the user's data safe and intact. The work uses the multiple-replica provable data possession (MR-PDP) scheme in order to overcome this problem. In this scheme the user is allowed to check the trust by posing the challenge to the server. Here it is done in three ways. When the challenge is thrown at the server it may generate the unique replica. The CSP stores multiple times (as required by the storage system) to save the single replica.The scheme uses multiple replicas hence it is computationally efficient compared to the single replica storages. Here the every replica of the file is encrypted differently before storing. It can also produce more copies if the need be by the user and for those copies different encrypting strategies will be used. In [13], the authors highlight that the cloud model where the services are leveraged to the clients are in greater hype in today's world. The client's store their data on the cloud and replicate the data and store them in the multiple servers to give their clients efficient and durable service. The clients and the CSP are not in the same trusted circle and hence there is always the risk for the client's data. The CSP provides the security for the data of the client via a third party authentication and maintains the data integrity. The CSP provides the data replication at a very high cost and this needs the client to be doubly confirmed about the CSP's service and the too follow the same level of service that is mentioned in the contract. This is done by the SP for the replication and also the updates needed and it should be reflected on all the servers where the data of the client is stored. The previous methods shows that this is achieved by the focusing on the static updating and if the dynamic updating is made it will cost the client very high. This work proposes dynamic multi-replica provable data possession scheme (DMR-PDP) for solving the problem of data updating and knowing if the SP is trustworthy. Here the method maintains the data integrity and keeps the data of the client confidential and also if the SP is trying to cheat it will be prevented by keeping less number of copies than the client is paying for so that data tampering can be avoided. The method also communicates only the updates required rather than displaying the operations performed on the data to fulfill the client's requirement for the data replication. The operations are kept secret hence the data updates will not be available to anyone other than intended receivers of the client. The proposed wok DMR-PDP also provides the dynamic operations like avoiding deletion and insertion or modification options on the replicated file copies without the consent of the authorized client. The working of the protocol shows that this method is very efficient and gives good performance over the other methods in the literature. In [14], again the problem with the cloud storage being on an untrusted domain is shown which poses a problem to the client data which is outsourced on the cloud. To maintain its integrity is the biggest requirement. The author proposed a model wherein it provides public verification and the data dynamics is privatized. The major drawback of this method is there is no protection from the attack of the proactive opponent and the protection can be given over the fixed block size. Thus as improvement to this model the author proposed a model which supports the variable block sizes and also gives the improved data security. Along with these features the method retains that earlier advantage of public verification and dynamic data integrity of the earlier model. In [15], the authors promote a method wherein a set of servers will act together and prove that they maintain data integrity when the client avails the data storage facility.

It uses a method called High-Availability and Integrity Layer (HAIL). When the client challenges the CSP the servers calculate the proof and the length of the proof is very compact and calculated on the very efficient platform. The length of the proof is very compact irrespective of the largest file or the smallest file of data. This algorithm protects the data from the active opponent by allocating different blocks of file to different servers to calculate the proof [16]. The attacker on failing to get the data may try to corrupt the servers which are involved in calculating the proofs. The proposed method severely analyzes the consequences and chooses the right parameter set to improve the quality of the security. This is installed on every server to ensure efficient security system.

A.Existing Methodology

There are large group of data holders outsourcing their data on the cloud storage. The CSS is unreliable and cannot be completely trusted this brings in a necessity of the client to check if their files are safe by themselves. For this purpose number of systems is designed for checking of data integrity remotely, but these are inefficient when it comes to data dynamics [17]. A expanding number of file owners choose to outsource the files to the CS. The services offered by these systems are

- The client has to use some means to keep a check on the data integrity and it is very important requirement of the client.
- There are two methods (S-PDP, E-PDP) in the existing system which is having very good performance but they lack to cope with the data dynamics.
- The data integrity needs to be checked remotely the existing system lacks in this. Does not provide efficiency in remote data integrity checking.
- It is very costly.
- The flexibility provided by these systems are very low and are insufficient.

III. PRELIMINARIES

A.System Structure

The requirement of the client from the CC is the high quality, flexible and scalable storage and computing facilities. Numerous clients are using the option of data outsourcing which can be rented at a lower cost than investing in the high cost storage and computing local facilities. The CSS on the other hand may not be trustworthy and hence the clients have to formulate their own method of verifying the CSS for its truthfulness. The DDIC algorithm gives security to the data against various attacks and hence proves to be a powerful tool for the clients. To support the verification the ORT is built and it can block the operations based on clients requirement. Both the participants in this application are benefitted like the CSS earns good amount by leveraging the access services after storing client data and the client in turn gets the service in low cost

B. Security Requirements

The CSS cannot be trusted completely as it may try to shade the occurrences of misbehavior like data corruption or the data loss for its own status in the market. The CSS can present the following attacks on the proposed protocol scheme.

- Forge Attack: The CSS may generate the tag for a block and might disrupt the data.
- Repeated Attack: The CSS may get hold of the results of the tests conducted by the client without having the valid tag.
- Replace Attack: The CSS may replace the proof of one block and tag with the other in order to prove itself.

C. Homomorphic Hash Function

The two main components of the DDIC scheme is the HHF and the ORT. The HHF calculates a hash value whose added sum of the two blocks considered should be equal to the product of two blocks. The table is used to enter the details of all the operations that is taking place on the data stored by the client on the CSS [18]. To implement this doubly linked list with the tables are used which will give the optimal ORT. The operations on the data can be addition, deletion or modification. The ORT will keep the cost value constant so that the confusion can be avoided. The system which uses the DDIC protocol provides the data security against many attacks. The data is safe from reproduction and replacement with this security model.

IV. PROPOSED METHODOLOGY

The proposed method is an efficient and bendable distributed schema with unambiguous dynamic data maintains, involves updating, deleting and more adding blocks. Adapting the designed protocol through the distributed authentication of erased coded data and the system performs public validate and data dynamics in opposition to third-party valuators that illustrates the finding of data fraud throughout verification of storage space accuracy on the distributed servers. The protocol presents the some security and activities assurance: Public audit ability, Storage correction, Confidentiality protecting, Light and Batch audit [19]. The propose model aspires to secure the cloud data from unreliable service suppliers. This method includes the data owners, cloud service providers and data users. Data owners accumulate data in the cloud and forward each contribute to data admissions to service providers.

We present a novel and well-organized DDIC protocol. The new system is surely secure against falsification attacks, replaces attacks, and replays attacks basis of a distinctive security model. We also provide a novelperfectexecution for the ORT.The ORT which isbuilds the price of access to this process almost constant. We do the complete routine analysis that displays in our system havebenefits in computing and communication expenses. The execution of the prototype in addition to the researches showsto the scheme is reasonable for genuine applications. Some of the main advantages of the proposed systems are:
The various experiments conducted displays that the scheme has good performance and is better in the group and is very feasible to be used in real time applications.

- The scheme uses data operations like blocking or deleting or inserting or appending based on the ORT buildr0 using the advanced DDIC protocol.

- Less Computation Costs.
- The data owners can perform dynamic activities for the files stored on CSS.

A. Schema of DDIC Protocol

The cloud storage method that includes two contestants: the CSS and the owner of the data. CSS has powerful storage and compute resources, accepts requests from data owners to store data files outsourced [20]. A stored data file offers right to use service. The data owner benefits from the CSS service and puts a large amount of CSS files without local backup. Since it is not assumed that CSS is reliable and sometimes does not work for instances by altering or removing in complete data files, the data owner be able to effectively validating the integrity of outsourced data in cloud computing.

The CSS is a computing device with very high computing and storing ability. It takes the data that the client submits and stores it in its storage and levies the access to the client to meet his requirement. The client or the data holder on the other hand has large amount data which he gives to the CSS to store and he does not even possess a copy of that data in his local hard disk. Since the CSS is not in the trustable environment of the client the client has to make arrangements for the timely check on the data to ensure it is not tampered [21]. In Fig 1 it shows the complete working of the DDIC protocol. The black lines indicate the verification process and the dotted lines specify the dynamic data operations that can be performed by the client on the file he has stored and all its replicas. A DDIC schema involves the following procedures. There are File Block Mechanism, Tag Generation, Challenging the Cloud, Proof Generation from Cloud, Integrity Checking, Prepare Update, and Execute Update.

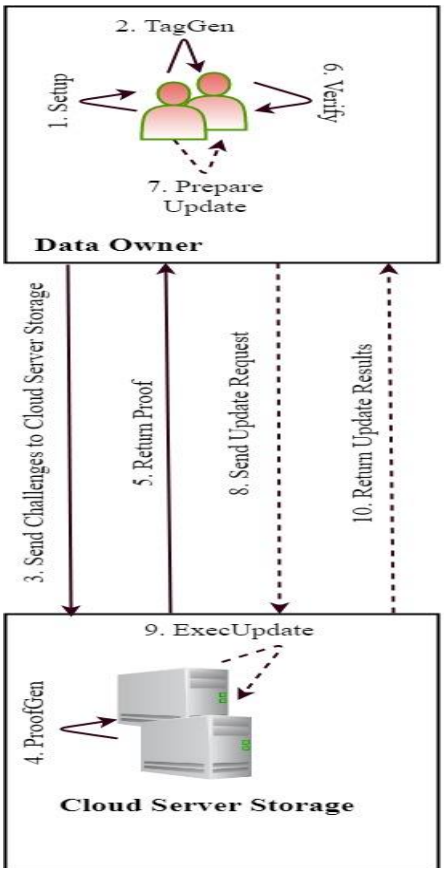

Figure 1: Schema and Designing of new DDIC Protocol

The complete workflow of DDIC protocol is illustrated in the Figure 1, in the figure solid lines and the dashed lines represent the verification of the integrity of the data and the dynamic functions on the file.

B. File Block Mechanism

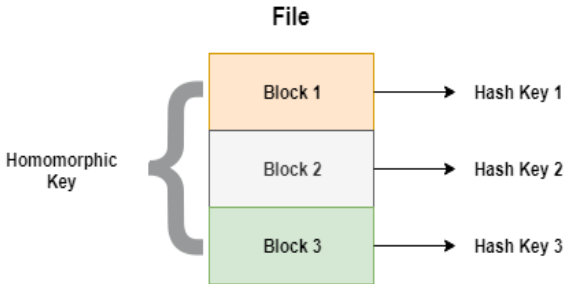

Figure 2: File Block Mechanism

As shown in the figure 2, by this mechanism the file will be divided into three blocks based upon the contents in the file [22]. For each block the unique hash key will be generated. And there will be a homorphickey for the whole file to distinguish between the blocks of a file.

C. Tag Generation

The client runs this algorithm to get a set of tags that can be used with each block of data. The algorithm takes in K, Sk and the data file as the input and generates T number of tags.

D. Challenge Generation

This algorithm is also run by the client to challenge the CSS for its truthfulness. The challenge is computed based on the disputed block and sent to the CSS.

E. Proof Generation

This algorithm is executed by the CSS to prove its truthfulness to the client. This algorithm takes the disputed file F, the labels generated at client T and the challenge posed by the client and provides the evidence P.

F. Integrity Checking

This algorithm is used by the client to verify the truthfulness of the CSS. The algorithm takes in to account K, Sk, chal and the P sent by the CSS and calculates the verification value. If the verification value generated is 1 then the CSS is genuine if the value generated is O he is not genuine [23].

G. Prepare Update

When the client wants to perform some operation on the data like addition, deletion or modification he runs this update algorithm [24]. This algorithm takes the file for updating Fi, the location of the block i and the update type and it will give out the information of the update request.

H. Execute Update

On reception of the Update request from the client the CSS will execute this algorithm [25].

On successful completion of this algorithm the data operation requested by the client will be witnessed, otherwise it will send an unsuccessful message to client.

V. CONCLUSION

In this paper there is a brief description about the issues related to integrity of the data files stored by the client in the CSS. The work proposes a very secure and efficient protocol DDIC which incorporates the data dynamics also. This scheme will use the HHF for verifying the data integrity of the files stored in the remote server. Since the client himself checks for the correctness of the data the cost incurred by the client for this is low. This system is built with the data structure requiring low computation for processing which giving the efficient results. The data structure used here is a hybrid one and when the file operations are being done the operations are done on the reduced block set. The security model of the proposed system is very efficient and is proved to secure the data efficiently. When the performance related issues are considered it has low cost on all the three important factors like computation, storage and community. This system through experiments has proved to be good for the practical implementations.

REFERENCES

1. Hao Yan, Jiguo Li, Jinguang Han, Member, IEEE and Yichen Zhang, "A Novel Efficient Remote Data Possession Checking Protocol in Cloud Storage", IEEE 2017.
2. Ari Juels and Michael Szydlo, "Attribute-Based Encryption: Using Identity-Based Encryption for Access Control", RSA Laboratories Bedford, MA 01730, 17 June 2004.
3. Meenal Jain and Manoj Singh, "Identity Based and Attribute Based Cryptography: A Survey", International Journal of Engineering, Management & Sciences (IJEMS) ISSN-2348 – 3733, Volume-2, Issue-5, May 2015.
4. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg and I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility," Future Generation Computer Systems, Vol. 25, No. 6, 2009, pp. 599-616
5. John Oredo, James Njihia, XN Iraki. The Role of Organizing Vision in Cloud Computing Adoption by Organizations in Kenya. American Journal of Information Systems. 2017; 5(1):38-50. doi: 10.12691/ajis-5-1-6.
6. Encryption with Immediate Attribute Revocation for FineGrained Access Control in Cloud Storage", School of Computer Science and Engineering, University of Electronic Science and Technology of China, No.2006, Xiyuan, IEEE, 2013.
7. Giuseppe Ateniese, Roberto Di Pietro, Luigi V. Mancini and Gene Tsudik, "Scalable and Efficient Provable Data Possession", 2008.
8. inguang Han, Willy Susilo, Yi Mu and Jun Yan, "PrivacyPreserving Decentralized Key-Policy Attribute-Based Encryption", IEEE Transactions on Parallel and Distributed systems, VOL. 23, NO. 11, NOVEMBER 2012
9. Ning Cao, Cong Wang, Ming Li, Member, KuiRen and Wenjing Lou, Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data, IEEE Transactions on Parallel and Distributed Systems, VOL. 25, NO. 1, JANUARY 2014.
10. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic,"Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Gener. Comp. Sy., vol. 25, no. 6, pp. 599 – 616, 2016.
11. H. Qian, J. Li, Y. Zhang and J. Han, "Privacy preserving personal health record using multi-authority attribute-based encryption with revocation," Int. J. Inf. Secur., vol. 14, no. 6, pp. 487-497, 2015.
12. H. Qian, J. Li, Y. Zhang and J. Han, "Privacy preserving personal health record using multi-authority attribute-based encryption with revocation," Int. J. Inf. Secure., vol. 14, no. 6, pp. 487-497, 2015. J. Li, X. Lin, Y. Zhang and J. Han, "KSF-OABE: outsourced attribute-based encryption with keyword search function for cloud storage," IEEE Trans. Service Comput., DOI: 10.1109/TSC.2016. 2542813.
13. J. Li, Y. Shi and Y. Zhang, "Searchable ciphertext-policy attribute-based encryption with revocation in cloud storage," Int. J. Commun. Syst., DOI: 10.1002/dac.2942.
14. J.G. Han, W. Susilo, Y. Mu and J. Yan, "Privacy-Preserving Decentralized Key-Policy Attribute-Based Encryption," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no.11, pp. 2150-2162, 2012
15. R. Mukundan, S. Madria and M. Linderman, ''Efficient integrity verification of replicated data in cloud using homomorphic encryption,'' Distrib. Parallel Dat., vol. 32, no. 4, pp. 507-534, 2014.
16. C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession,'' in Proc. 16th ACM Conf. on Comput. And Commun. Security (CCS), 2009, pp. 213-222.
17. R. Curtmola, O. Khan, R. Burns, and G. Ateniese, ''MR-PDP: Multiple-replica provable data possession,'' in Proc. 28th IEEE Conf. on Distrib. Comput. Syst. (ICDCS), 2008, pp. 411-420.
18. J.G. Han, W. Susilo, Y. Mu and J. Yan, "Privacy-Preserving Decentralized Key-Policy Attribute-Based Encryption," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no.11, pp. 2150-2162, 2012
19. Z. J. Fu, X. M. Sun, Q. Liu, L. Zhou, and J. G. Shu, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," IEICE Transactions on Communications, vol. E98-B, no. 1, pp.190-200, 2015.
20. K. D. Bowers, A. Juels, and A. Oprea, "Hail: A high-availability and integrity layer for cloud storage," in Proc. 16th ACM Conf. on Comput. andCommun. Security (CCS), 2009, pp. 187–198.
21. E. Zhou and Z. Li, ''An improved remote data possession checking protocol in cloud storage,'' in Proc. 14th Int'l Conf. on Algs. andArchs. for Parall Proc. (ICA3PP), 2014, pp. 611-617
22. T.J. Ren, J. Shen,J. G. Wang, "Mutual verifiable provable data auditing in public cloud storage".
23. K. Kiran Kumar, K. Padmaja, P. Radha Krishna, "Automatic Protocol Blocker for Privacy-Preserving Public Auditing in Cloud Computing", International Journal of Computer science and Technology, vol. 3 pp, ISSN. 0976-8491(Online), pp. 936-940, ISSN: 2229- 4333 (Print), March 2012
24. Dhiyanesh "A Novel Third Party Auditability and Dynamic Based Security in Cloud Computing", International Journal of Advanced Research in Technology, vol. 1,no. 1, pp. 29-33, ISSN: 6602 3127, 2011
25. XU Chun-xiang, HE Xiao-hu, Daniel Abraha, "Cryptanalysis of Auditing protocol proposed by Wang et al. for data storage security in cloud computing", http://eprint.iacr.org/2012/115.pdf.