Data Duplication in Cloud for Optimal Performance and Security

Manjunath M
2nd year M.tech Student
Branch: computer science and engineering
A.P.S College of Engineering

Ms. Shwetha S. M
B.E M.tech
Assistant professor, Department of CSE
A.P.S College of Engineering

***Abstract*- The Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that collectively approaches the security and performance issues. Data to a third-party administrative control, as is done in cloud computing, gives rise to security concerns. The data compromise may occur due to attacks by other users and nodes within the cloud. Therefore, high security measures are required to protect data within the cloud. To secure the data in cloud by using fragmentation and replication. In DROPS methodology, divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a one fragment of data file that ensures even in case of a successful attack, no meaningful information is revealed to the attacker.**

Index Terms- Cloud security, cryptography, fragmentation, replication, performance

I. INTRODUCTION

Cloud computing is a trending technology in the field of information technology as it allows sharing of resources over a network. Cloud computing is nothing but a specific style of computing where everything from computing power to infrastructure, business apps etc., Cloud computing is a model for enabling convenient, on-demand network access to a share pool of configurable computing service (for ex: networks, servers, storage, applications and services) that can be provisioned rapidly and released with minimal management effort or services provider.

Cloud storage provides online storage where data stored in the form of virtualized pool that is usually hosted by third parties. The off-site data storage cloud utility requires users to move data in cloud's virtualized and shared environment that may result in various security concerns. However, the benefits of low cost, negligible management and greater flexibility come with increased security.

The data compromise may occur due to attacks by other users and nodes within the cloud. Therefore, high security measures are required to protect data within the cloud. However, the employed security strategy must also take into account the optimization of the data retrieval time.

The Division and Replication of Data in cloud provide the Optimal Performance and Security (DROPS).

II. LITERATURE SURVEY

Various approaches have been put for thin the literature survey to point out the problem of security and access control in cloud and various storage devices. Personal health record (PHR) **[1]** that enables patients to manage their own medical records in centralized way, which greatly facilitates the access, storage and sharing of personal health information. Under encryption, it is most challenging to achieve fine-grained access control to data Personal health record in a scalable and efficient way. For all patients, the PHR data should be encrypted so that it is increase with the number of users having access. Due to the more number of users and owners in the PHR system, potentially huge computational and management risk on the entities in the system can be incurred, which will limit the PHR system usability and data accessibility.

The Proposed Framework for Patient-Centric Data Access Control: There are multiple categories of security domains SDs: **personal domains (PSDs)** and **public domains (PUDs).** A public domain usually contains a more number of professional users, and multiple *public attribute authorities* (PAA) that distributive governs a disjoint subset of attributes to erase key escrow. An owner encrypts her Personal Health Record data so that authorized users from both her PSD and PUDs may read it. Users belonging to a public domain only need to obtain credentials from the corresponding public authorities, without the need to interact with any PHR owner, which greatly decrease the key management overhead of users and owners.

Cloud storage that enables users to remotely store their data and enjoy the on-demand more quality cloud applications [2] without the burden of local software and hardware management. Though the benefits are clear, such a service is also relinquishing user's physical possession of their shared environment, which inevitably poses security risks towards the correctness of the data in cloud. In order to address this problem and further achieve a safe and dependable cloud storage service. The proposed design allows users to audit the cloud storage with very lightweight computation and communication cost. The auditing result not only ensures strong cloud storage correctness guarantee, but also simultaneously achieves faster way to data error localization, i.e., the identification of misbehaving server. Secure outsourcing of computation to an untrusted (cloud) service **[3]** provider is becoming more important. Pure cryptographic solutions based on fully verifiable encryption and homomorphism, recently proposed, are promising but suffer from very high latency. Trusted computing is another approach that uses trusted hardware and software components on computing platform to provide useful mechanisms such as attestation allowing the data owner to verify the integrity of the cloud and its computation.

Fully Homomorphic Encryption technique [5] to ensure that the cloud is not able to read the data while performing computations on them. Holomorphic encryption scheme which means performing the operations on the encrypted data. Homomorphism encryption can be applied in any system by using different public key algorithms. When the data is moved to the public area, there are some other encryption algorithms to secure the storage and the operations of the data. But to process data avilable on remote server and to preserve privacy, homomorphic encryption is more useful that allows the operations on the cipher text, which can provide the same accurate results after calculations as the working directly on the raw data. This decryption refreshes the data without exposing it, allowing an infinite number of computations on the same.

Cipher text-Attribute based encryption CP-ABE [6] in which the receiver has the access policy in the form of a tree structure, with attributes as leaves and monotonic access structure with AND, OR operations and other threshold gates. Here central authority generates the global key and issues the secret key (SK) for the user. They use decryption key is in form of secret key. The decryption key is shared by multiple users who possess the corresponding attributes, so that any malicious owner of a decryption key would have the intention to leak partial or even his entire decryption privilege for financial interest. SHA-1 is one of other cryptographic hash functions, most often used to verify that a file has been same. SHA is short for Secure Hash Algorithm. File verification using SHA-1 is accomplished by comparing the checksums created after running the algorithm on the other two files you want to compare.

In a cloud environment, total file is stored in a single node leads to a point of failure [17]. A successful attack on a node might put the data confidentiality, integrity, or both at risk. The aforesaid scenario can occur both the case of intrusion or accidental errors. In such systems, performance in terms of retrieval time can be enhanced by employing replication strategies. However, replication increases the number of file copies within the cloud storage. Thereby, increasing the probability of the node holding the file to be a victim of attack as discussed in Section 1&2. Security and replication are more essential for a large-scale system, such as cloud, as both are utilized to provide services to the end users. Replication and security must be balanced such that one service must not lower the service level of the other.

III. EXISTING SYSTEM

The off-site data storage cloud utility requires users to move data in cloud's virtualized and shared environment that may result in various security concerns. Pooling and elasticity of a cloud, allows the physical resources to be shared among many users. The data outsourced to a public cloud must be secured. Unauthorized data access by other users and processes (whether accidental or deliberate) must be prevented as discussed above, any weak entity can put the whole cloud at risk. In such a scenario, the security mechanism must substantially increase an attacker's effort to retrieve a reasonable amount of data even after a successful intrusion in the cloud.

IV. PROPOSED SYSTEM

The data is fragmented and replicated to achieve both security and ideal performance. The Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that judicially fragments user files into pieces and replicates them at strategic locations within the cloud. The division of a file into fragments is performed based on a given user criteria such that the individual fragments do not contain any meaningful information. Each of the cloud nodes (we use the term node to represent computing, storage, physical, and virtual machines) contains a distinct fragment to increase the data security. The non-cryptographic nature of the proposed scheme makes it faster to perform the required operations such as placement and retrieval of data on cloud, it significantly improves the security. Firstly, in this methodology user sends the data file to cloud. The cloud manager system upon receiving the file performs: (a) fragmentation, (b) first cycle of nodes selection and stores one fragment over each of the selected node, and (c) second cycle of nodes selection for fragments replication. The cloud manager maintains record of the fragment placement and is assumed to be a secure entity.

A. Architecture Design

The DROPS methodology utilizes the concept of data fragmentation for securing the user data within the cloud. To further enhance the security, the fragments are not stored on the adjacent nodes. To separate the storage of fragments by given distance, the concept of T-colouring is used. To improve the retrieval time of fragments, the fragments are stored on the most central nodes. The selection of central nodes is carried out by evaluating the centrality measures for the nodes. The working of the DROPS methodology is shown as a high-level work flow.

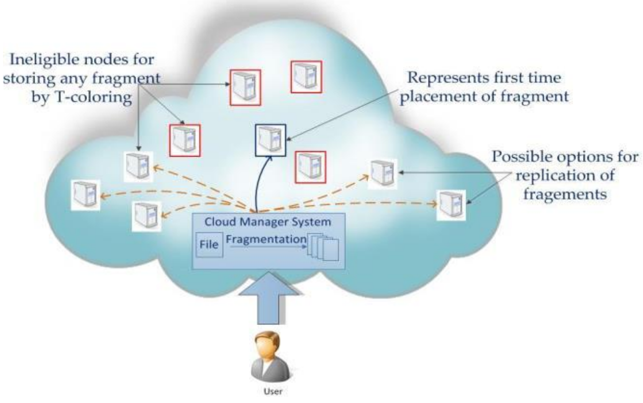

Fig.1. *The DROPS Methodology.*

B. *Methodology*

1) *Data Fragmentation:*

The amount of compromised data can be reduced by making fragments of a data file and storing them on separate nodes. Compromising a single file will require the effort to penetrate only a single node. A successful intrusion on a single or few nodes will only provide access to a portion of data that might not be of any significance. If an attacker is uncertain about the locations of the fragments, the probability of finding fragments on all of the nodes is very low.

2) *Centrality*:

The objective of improved retrieval time in replication makes the centrality measures more important. The centrality of a node in a graph provides the measure of the relative importance of a node in the network. The node is important in the network if it: (a) interconnects more nodes than others, (b) can be reached easily by other nodes, or (c) can reach other nodes easily. In this method, we used only three centrality.

They are: 1) Closeness.
2) Betweenness.
3) Eccentricity.

3) *T-coloring:*

The nodes are selected in such a manner that they are not adjacent and are at given distance from each other. The node separation is ensured by the means of the T-coloring. a graph G = (V,E) and a set T containing non-negative integers including 0. The T-coloring is a mapping function f from the vertices of V to the set of non-negative integers, such that $|f(x) - f(y)| \neq T$, where $(x; y) > E$. The mapping function f assigns a color to a vertex. The T-coloring Problem for channel assignment assigns channels to the nodes, such that the channels are separated by a distance to avoid interference.

V. CONCLUSIONS

We proposed the DROPS methodology, outsourcing data to cloud storage is mainly concentrate on security and performance in terms of recapture time. The data file was divided and the divided files are distributed over multiple nodes. The nodes are separated in terms of T-coloring concept. The fragmentation and distributed ensured that no significant information was obtainable by an adversary in case of a successful attack. In cloud no one node, stored more than a one fragment of the same file. The performance of the DROPS methodology was compared with full-scale replication techniques. The results of the simulations revealed that the simultaneous focus on the security and performance resulted in increased security level of data accompanied by a slight performance drop.

REFERENCES

[1] M. Li, S. Yu, K. Ren, and W. Lou, “Securing Personal Health Records in Cloud Computing: Patient-Centric and Fine-Grained Data Access Control in Multi-Owner Settings,” Proc. Sixth Int’l ICST Conf. Security and Privacy in Comm. Networks pp. 89-106, 2010.

[2] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, “Toward Secure and Dependable Storage Services in Cloud Computing,” IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.- June 2012.

[3] A.-R. Sadeghi, T. Schneider, and M. Winandy, “Token-Based Cloud Computing,” Proc. Third Int’l Conf. Trust and Trustworthy Computing (TRUST), pp. 417-429, 2010.

[4] D.F. Ferraiolo and D.R. Kuhn, “Role-Based Access Controls,” Proc. 15th Nat’l Computer Security Conf., 1992

[5] C. Gentry, “A Fully Homomorphic Encryption Scheme,” PhD dissertation, Stanford Univ., http://www.crypto.stanford.edu/ craig, 2009.

[7] J. Bethencourt, A. Sahai, and B. Waters, “Cipher-text-policy attribute-based encryption: An expressive, efficient, and provably secure realization” in Proceedings: Public Key Crypto Proc. IEEE Symp. Security and Privacy, pp. 321-334, 2007. B. Raja Sekhar, B. Sunil Kumar, and V. Poorna Chandra, proposes a “CP-ABE Based Encryption for Secured Cloud Storage Access” International Journal of Scientific & Engineering Research, Volume 3, Issue 9, September-2012

[8] X. Boyen, “Mesh Signatures,” Proc. 26th Ann. Int’l Conf. Advances in Cryptology (EUROCRYPT), pp. 210-227, 2007.

[9] Ms. Shubhra Sagar, Dr. R.K. Datta proposes a “An improved RSA Encryption Algorithm for Cloud Computing Environments: Two key Generation Encryption” pp. 427-439, 2006

[10] Hemalatha, Dr.R.Manickachezian proposes a “Security Strength of RSA and Attribute BasedEncryption for Data Security in Cloud Computing” IEEE ,pp.252-352- Trans. June-2009

[11] S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” Proc. 14th Int’l Conf. Financial Cryptography and Data Security, pp. 136-149, 2010.

[12] J. Hur and D. Kun Noh, “Attribute-Based Access Control with Efficient Revocation in Data Outsourcing Systems,” IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 7, pp. 1214-1221, July 2011.

[13] S. Ruj, A. Nayak, and I. Stojmenovic, “DACC: Distributed Access Control in Clouds,” Proc. IEEE 10th Int’l Conf. Trust, Security and Privacy in Computing and Communications (TrustCom), 2011.

[14] M. Chase and S.S.M. Chow, “Improving Privacy and Security in Multi-Authority Attribute-Based Encryption,” Proc. ACM Conf. Computer and Comm. Security, pp. 121-130,

[15] A. Beimel, “Secure Schemes for Secret Sharing and Key Distribution,” PhD thesis, Technion, Haifa, 1996.

[16] K. Bilal, S. U. Khan, L. Zhang, H. Li, K. Hayat, S. A. Madani, N. Min-Allah, L. Wang, D. Chen, M. Iqbal, C. Z. Xu, and A. Y. Zomaya, “Quantitative comparisons of the state of the art data center architectures,”Concurrency and Computation: Practice and Experience, Vol. 25, No. 12, 2013, pp. 1771-1783.

[17] K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Zomaya, “On the characterization of the structural robustness of data center networks,” IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013, pp. 64-77.

[18] D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, and A. Y. Zomaya, “Energy-efficient data replication in cloud computing datacenters,” In IEEE Globecom Workshops, 2013, pp. 446-451. . [19] Y. Deswarte, L. Blain, and J-C. Fabre, “Intrusion tolerance in distributed computing systems,” In Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy, Oakland CA, pp. 110-121, 1991.

[20] B. Grobauer, T.Walloschek, and E. Stocker, “Understanding cloud computing vulnerabilities,” IEEE Security and Privacy, Vol. 9, No. 2, 2011, pp. 50-57.