SURVEY ON DATA SECURITY BY USING SECURE REPLICATION IN CLOUD

[1]Prof. P.N. Mundhare, [2]Miss. Payal U. Gujare, [3]Prof.S.G.Malas
[1]Assistant Professor, [2]Lecturer, [3]Assistant Professor
[1]Computer Science and Engineering,
[1]Sanmati Engineering College, Washim, Maharashtra, India

Abstract: The world relies on cloud computing to store their public and private information which is needed by the user. Cloud services offered to its users by cloud service providers. Clouds are generally implemented on cluster computers to provide the necessary scale and performance required by such services. Redistributing information to an outsider authoritative control, as is done in distributed computing, offers ascend to security concerns. The information bargain may happen because of assaults by different clients and hubs inside the cloud. Along these lines, high safety efforts are required to ensure information inside the cloud. The single database storage system is a less secure because data remain under a single database servers. This can lead to data loss due to different causes like hacking, server failure issues. If an attacker chooses to attack a specific user, then he can concentrate on a fixed cloud provider, try to have access to the client's information. This makes an easy job of the attackers, and gets the benefit of using data mining to a great extent. Thus single data server storage architecture is the biggest security threat concerning data mining on cloud, so in this paper present the secure approach that encrypt and replicate the data in distributed data server storage system. This approach involves the replication and storage of data.

Index Terms **- Cloud, Distributed, Data Security, Replication, Data Mining.**

I. INTRODUCTION

Cloud services are provided by different famous organizations like Google, Amazon and Microsoft etc. By using these services the client avoid the cost of buying extra resources. Cloud services provide the high computation capacity at low cost. The various data analysis techniques which are used for extracting valuable information from a large volume of data. These different techniques are used by Cloud service provider like Google uses the technique for identifying the user behavior on the basis of search behavior. In previous trend data to store on a single cloud the attacker applies an attack on it and accesses the information which is stored by the client on cloud storage. If the client is a field related to healthcare, shopping, insurance, banking, etc then there is big loss of information access by attackers, so distributed environment handles such kind of problem. The distributed data mart storage is service which is provided by cloud service provider. In Distributed Cloud storage, the information is stored from different kind of devices they only pay for storage as per usage.

The term cloud computing is used to capture vision of computing as a utility. A cloud is defined as a set of Internet-based application, storage and computing services sufficient to support most users' needs, thus enabling them to largely or totally dispense with local data storage and application software. Cloud computing is technology that provides the different services at very low cost. The different client stores data on Cloud storage. Cloud computing provides storage for storing the information and provides the security of that information. Cloud service models are infrastructure as a service, platform as a service, and software as a service and new for cloud is database as a service.

The cloud computing paradigm has reformed the usage and management of the information technology infrastructure. Cloud computing is characterized by on demand self-services, ubiquitous network accesses, resource pooling, elasticity, and measured services. The aforementioned characteristics of cloud computing make it a striking candidate for businesses, organizations, and individual users for adoption. However, the benefits of low-cost, negligible management (from a users perspective), and greater flexibility come with increased security concerns.

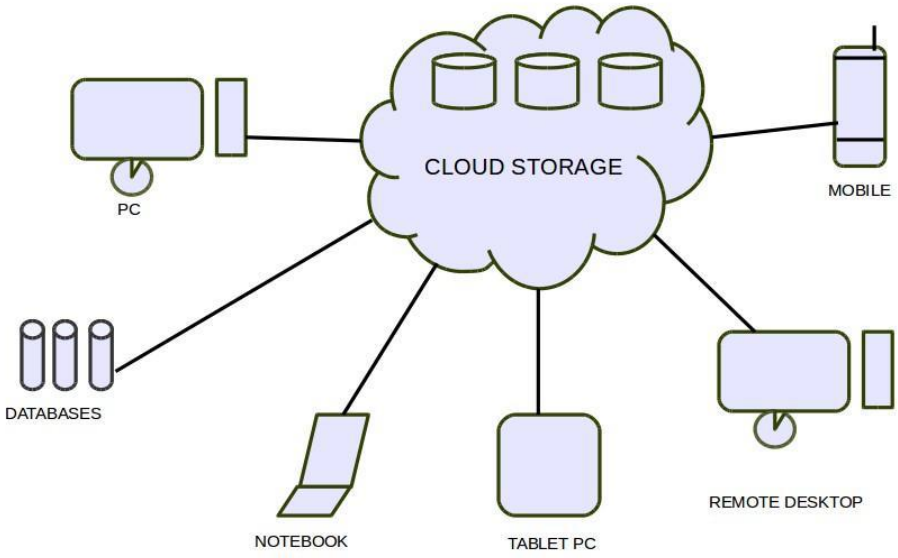
Cloud Storage is a type of service that allow a user to save data on offsite storage system managed by third-party and is made accessible by a web services API. On cloud the data is stored at virtualized pools of storage. Hosting companies operate large space data centers and lease their storage. And companies, organizations and institutes who require their data to be hosted buy or lease storage capacity from hosting companies. The cloud storage system stores multiple copies of data on multiple servers and at different locations. If one system fails, then it only requires tracking the replicated data location [1].

Here purposes a secure approach that replicates the client's data and store on different data servers. Before replication, full copy of encrypted information stores on data warehouse for increasing the availability of information. It will increase the reliability and privacy of data. Storing the data in cloud is not that simple task. Apart from its flexibility and convenience, it also has several challenges faced by the consumers.

The consumers require ability to:

- Provision additional storage on demand.
- Know and restrict the physical location of the stored data.
- Verify- how data was erased

- Have access to a documented process for surely disposing of data storage hardware.



1. Cloud storage and devices interaction

II. RELATED WORK

With distributed computing, every one of your information is put away on the cloud. That is just fine, yet how secure is the cloud? Could other, unapproved clients access your classified information? Hypothetically, information put away in the cloud is curiously protected, reproduced over numerous machines. In any case, in case your information goes missing, you have no physical or nearby reinforcement. Except if you deliberately download all your cloud reports to your very own work area. Cloud-based processing is a rising practice that offers altogether more foundation and money related adaptability than customary figuring models. Cloud suppliers today offer everything from access to crude process or capacity limit assets to out and out application benefits in regions, for example, finance and client relationship the executives.

A number of authentication techniques have been proposed in the recent times that are based upon graphical methods. Text based passwords are most commonly used for authentication; however they are highly vulnerable to several kinds of attacks. Graphical techniques are coming up as an attractive alternative to the conventional methods of authentication. In this paper proposed a graphical method of authentication that employs graphical coordinates along with a novel introduction of time interval between successive clicks. The user needs to recall the coordinates and the time interval of the successive clicks. This leads to the incorporation of the advantages of the recent graphical methods along with the added security achieved through the use of time interval. The proposed scheme has a much higher password space than the other contemporary graphical authentication schemes. The scheme is robust, secure and very convenient to use [2]. Both the administrator and the users should undergo the graphical password text. In view of the shortcomings of the traditional approach to authentication, i.e. alphanumeric passwords, Graphical techniques are gaining importance.

The data owner searches the data from encrypted data bases the search is based on rank keyword. The ranked keyword base search reduces the overhead of the data owner because there is no need to go through from each file. In this technique, the server site is only responsible for the search operation all other responsibilities are taken by the data owner [3]. Cloud Data Protection for Masses proposes a new cloud computing paradigm, data protection as a service. DPaaS is a suite of security primitives offered by a cloud platform, which enforces data security and privacy and offers evidence of privacy to data owners, even in the presence of potentially compromised or malicious applications. Data protection is provided by using three primitives they are access control, key management and logging. Also there is an auditor who audits all the transactions occurred in the system. Auditor finally provides an audit report based on all conversations done [4].

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. Moreover, the probable amount of loss (as a result of data leakage) must also be minimized [5].

Juels and Opera [6] presented a technique to ensure the integrity, freshness, and availability of data in a cloud. The data migration to the cloud is performed by the Iris file system. A gateway application is designed and employed in the organization that ensures the integrity and freshness of the data using a Merkle tree. The file blocks, MAC codes, and version numbers are stored at various levels of the tree. The proposed technique in [6] heavily depends on the User's employed scheme for data confidentiality. Moreover, the probable amount of loss in case of data tempering as a result of intrusion or access by other VMs cannot be decreased. Our proposed strategy does not depend on the traditional cryptographic techniques for data security. Moreover, the DROPS methodology does not store the whole file on a single node to avoid compromise of all of the data in case of successful attack on the node.

III. MATERIAL AND METHODOLOGIES

Mists are commonly executed on group PCs to give the fundamental scale and execution required by such administrations. A bunch PC is a lot of interconnected PCs that participate near give a solitary, coordinated elite registering ability. One distributed computing arrangement is to send the stage as a method for catastrophe recuperation, business coherence, and broadening the server farm. With adaptable "pay-as-you-develop" models, distributed computing can advance with the necessities of your business. In utilizing the cloud, numerous associations are as yet asking – When would it be a good idea for me to utilize the cloud for replication? Replication is totally imperative for numerous individuals taking a gander at similar information in the meantime, contingent upon where they are, and it is vital for calamity recuperation. Before we go into the details of the data replication methodology, we introduce the related concepts in the following sections.

1. Data Fragmentation

The security of a large-scale system, such as cloud depends on the security of the system as a whole and the security of individual nodes. A successful intrusion into a single node may have severe consequences, not only for data and applications on the victim node, but also for the other nodes. The data on the victim node may be revealed fully because of the presence of the whole file [7]. A successful intrusion may be a result of some software or administrative vulnerability [7]. In case of homogenous systems, the same flaw can be utilized to target other nodes within the system. The success of an attack on the subsequent nodes will require less effort as compared to the effort on the first node. Comparatively, more effort is required for heterogeneous systems. However, compromising a single file will require the effort to penetrate only a single node. The amount of compromised data can be reduced by making fragments of a data file and storing them on separate nodes [7], [8]. 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. Moreover, 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 centrality of a node in a graph provides the measure of the relative importance of a node in the network. The objective of improved retrieval time in replication makes the centrality measures more important. There are various centrality measures; for instance, closeness centrality is when a node is said to be closer with respect to all of the other nodes within a network, if the sum of the distances from all of the other nodes is lower than the sum of the distances of other candidate nodes from all of the other nodes. The lower the sum of distances from the other nodes, the more central is the node., degree centrality, betweenness centrality is centrality of a node n is the number of the shortest paths, between other nodes, passing through *n*, eccentricity centrality and Eigen vector centrality [9].

3. DROPS METHODOLOGY

A cloud storage security conspire that all things considered arrangements with the security and execution regarding recovery time. The information document was divided and the sections are scattered over various hubs. The hubs were isolated by methods for T-shading. The fracture and dispersal guaranteed that no critical data was possible by an enemy if there should arise an occurrence of an effective assault. No hub in the cloud, put away in excess of a solitary part of a similar record. The execution of the DROPS strategy was contrasted and full-scale replication strategies. The consequences of the recreations uncovered that the synchronous spotlight on the security and execution brought about expanded security dimension of information joined by a slight act drop. At present with the DROPS approach, a client needs to download the document, update the substance, and transfer it once more. It is key to build up a programmed update system that can recognize and refresh the required sections as it were. The previously mentioned future work will spare the time and assets used in downloading, refreshing, and transferring the record once more. Besides, the ramifications of TCP incast over the DROPS technique should be examined that is applicable to conveyed information stockpiling and access [10].

IV. DISCUSSION AND WORK

In a cloud environment, a file in its totality, stored at a node leads to a single point of failure. A successful attack on a node might put the data confidentiality or integrity, or both at risk. 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. Thereby, increasing the probability of the node holding the file to be a victim of attack. Security and replication are essential for a large-scale system, such as cloud, as both are utilized to provide services to the end user. Security and replication must be balanced such that one service must not lower the service level of the other.

1. Working of Secure Replication:

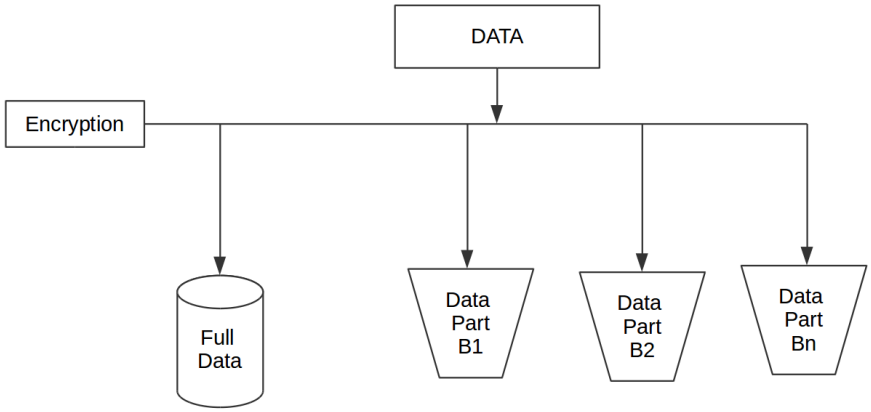


Figure 2: Secure replication system architecture

The process is carried out as follows:-

- Step1: Client sends data to cloud service provider for storing.
- Step2: Cloud provides receive data and perform encryption.
- Step3: Full copy of encrypted data stores on data warehouse.
- Step4: After backup, performing replication and divide the data in parts according to the availability of data bucket (in our system use three data buckets B1, B2, B3)
- Step5: Storing the different part of information on different data bucket.
- Step6: Repeat Steps as per storing request.

In this system client send data to cloud service provider for storing it. The clouds receive data from client and perform encryption on it. After performing encryption full copy of data stores on data warehouse for backup. After full backup, performing replication divide the data in parts according to the availability of data bucket. In current system use three data buckets (B1, B2, and B3) for increasing privacy and availability of client's data. The client's data store on backup warehouse and then divide the data in three parts and store on respective data buckets B1, B2, and B3. If any data bucket lost the part of client's data then it can reload from backup warehouse. In this way replication of client's data on different data buckets increase the availability of information as well as enhance the security of information.

This makes troublesome occupation of the aggressors. The insider aggressor alludes as worker that works under association which is dependable verifying and putting away the customer's data. On the off chance that any information can hack by an assailant, at that point it can get to the main piece of data, for full data there is have to apply assaults on other information cans. The information basin is crashes or down likewise sways on the accessibility of data. The purposed framework additionally expels that disadvantage. On the off chance that any information pail is crashes or down, at that point customer's solicitation additionally ready to extricate the information from reinforcement stockroom. In this situation information container B1 is come up short and not reacting the client demand. For this situation the piece of data is lost. This framework enable client to remove the data from reinforcement product house. The accessibility of information can likewise influence on security of data. If there should arise an occurrence of vast no of information pails the information isolate in more parts and store distinctive parts in various information cans. Every datum pails have little piece of data. In the event that any information can is hacked by aggressor, at that point it can take just little piece of data.

V. CONCLUSION

A considerable lot of cloud clients believe that cloud is secure and simpler. In any case, the greater part of the IT specialists feel that the cloud has loads of issues in the field of information security and protection issues towards the development of distributed computing. No client will exchange their information to the cloud until the trust is worked between the cloud specialist organizations and shoppers. In this paper, we have laid out the general standards of new way to deal with perform secure replication on put away information. This is a most compelling strategy which will give better outcomes to security and accessibility of data. This safe replication system can be helpful so as to construct a safe and dependable appropriated stockpiling. The upgrade done in this procedure will build the quality by various information server have with cloud supplier and store data as per its affectability. This new strategy can be relevant in various cloud suppliers organizations and associations and so forth.

REFERENCES

[1] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," Info. Sci., DOI: 10.1016/j.ins.2015.01.025, 2015.

[2] Sunumol Cherian and Kavitha Murukezhan. "Providing data protection as a service in cloud computing", International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 ,ISSN 2250-3153.

[3] Gupta Sarika, Sangita Rani Satapathy, Mehta Piyush and Tripathy Anupam, "A Secure and Searchable Data Storage in Cloud Computing", 3rd IEEE International Advance Computing Conference (IACC), 2013, page 106-109.

[4] Dawn Song, Elaine Shi, Ian Fischer, and Umesh Shankar. "Cloud data protection for the masses", IEEE Transactions in Computer Society, 45(1):39–45, 2012.

[5] N. Khan, M. L. M. Kiah, S. U. Khan, and S. A. Madani, "Towards secure mobile cloud computing: A survey," Future Gener. Comput. Syst., vol. 29, no. 5, pp. 1278–1299, 2013.

[6] Juels and A. Opera, "New approaches to security and availability for cloud data," Commun. ACM, vol. 56, no. 2, pp. 64–73, 2013.

[7] A. Mei, L. V. Mancini, and S. Jajodia, "Secure dynamic fragment and replica allocation in large-scale distributed file systems," IEEE Trans. Parallel Distrib. Syst., vol. 14, no. 9, pp. 885–896, Sep. 2003.

[8] M. Tu, P. Li, Q. Ma, I.-L. Yen, and F. B. Bastani, "On the optimal placement of secure data objects over Internet," in Proc. 19th IEEE Int. Parallel Distrib. Process.Symp,2005,p.14

[9] M. Newman, Networks: An Introduction. London, U.K.: Oxford Univ. Press, 2009.

[10] Mazhar Ali, Samee U. Khan, Kashif Bilal, Bharadwaj Veeravalli, Keqin Li, and Albert Y. Zomaya," DROPS: Division and Replication of Data in Cloud for Optimal Performance and Security", *in* IEEE Transactions on Cloud Computing, Vol. 3, January 2015.