GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
AN OPTIMUM LOAD BALANCING FOR DATA ON CLOUD USING MACHINE LEARNING

Shailesh B Galande*[1] and Sachin Patil[2]
*[1]GHRCEM, Wagholi, Pune, Maharashtra, India
[2]Assistant Professor, GHRCEM, Wagholi, Pune, Maharashtra, India

ABSTRACT
Cloud storage system enables storing of data in the cloud server competently and makes the user to work with the data without any hitch of the resources. In the existing system, the data are stored in the cloud using dynamic data operation with computation which makes the user need to make a copy for further updating and authentication of the data loss. An efficient distributed storage auditing mechanism is planned which over comes the limitations in handling the data loss. In this paper the partitioning method is implemented for the storage of data which avoids the local copy at the user side by using partitioning method. This technique ensures high cloud storage veracity, improved error localization and easy detection of disobedient server. To achieve this, secluded data integrity checking concept is used to enhance the performance of cloud storage. The data is dynamic in nature in cloud; hence this work aims to store the data in reduced space with less time and computational cost.

Keywords- *Machine Learning, Cloud computing, Load balancing.*

I.   INTRODUCTION
Cloud storage technology is perhaps the most famous and revolutionary technology in this era of technological advancements. It basically consists of logical pools where digital data is stored, multiple servers often located at different locations make up for the physical environment of the system. These servers are often owned and managed by a hosting company. These service (cloud storage) providers are responsible for keeping the data available and accessible at all times and apart from that they are also responsible for keeping the physical environment up and running. The technology is so common that we hardly realize its importance; all of us access our mails, accounts on social sites as well as other things on the move but without any extra effort. All this is possible because of the cloud storage technology. Imagine a scenario where we have to carry our laptops or PCs everywhere we go. That would surely become almost impossible, all that is solved using this technology. As we can see the clouds everywhere, we can carry our data where ever we go without actually having to worry about its details. Cloud storage systems enable storing of data on the cloud server efficiently and easily leaving user with no trouble of managing any resources. Further cloud storage is private, public or hybrid (Fig 1). Private cloud storage is restricted to a particular organization and data security risks are less as compared to public cloud storage.

In today's cloud computing system data is stored on the cloud using dynamic data operation, but the problem lies in the making of an extra copy on user's side for further updating and verification of the data loss. In the proposed system these limitations have been removed. Partitioning method is proposed for the data storage which avoids the creation of local copy at user's side.
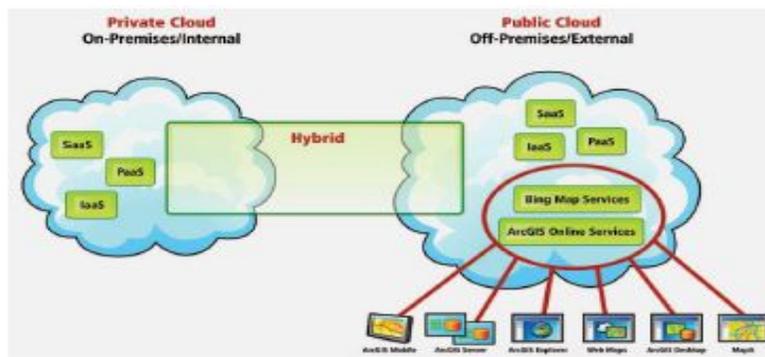

Figure 1: Cloud Services Architecture

Different techniques have been incorporated to ensure high cloud storage integrity, enhanced error localization and easy identification of misbehaving server. Cloud storage performance is enhanced using integrity checking concept. The focus is centered on storage of data in less space, less time and less computation cost.

II.   RELATED WORK

[1] Cloud storage is the broader theme but the system [1] mainly focuses on private cloud storage. Private cloud storage is built using commodity machines. As the storage space is limited, judicious use is a compulsion. Techniques like duplication have been used to create unique instances of data in the storage. Imbalance is caused due to expansion of cluster which is overcome by using centralized load balancing technique.

[2]To remove the overhead of query retrieval time the system comes with the concept of look up tables as a way to provide fine grained partitioning for transactional database applications. Using look up tables query result time is greatly reduced. Twitter, TPC-E and Wikipedia were used for partitioning using look up tables. On the above mentioned applications from 40percent to 300percent performance hike was recorded.

[3] Symmetric key algorithms for example Advanced Encryption Standards (AES) are used in PDP scheme proposed in Data availability i.e. making sure that data is available any time it is accessed and data error recovery are not given much importance in the system.

[4] The paper emphasizes a need on data security and data privacy taking the storage of pictures by users in cloud into picture. As digital images may include sensitive information they need to secure and privacy be maintained. Face recognition and stripping algorithms have been proposed in which sensitive data remains with the Cloud user. There is a problem with the system implementation with true negatives, i.e. the data is classified as non-sensitive but it has to be classified as sensitive. So this needs a further look into it.
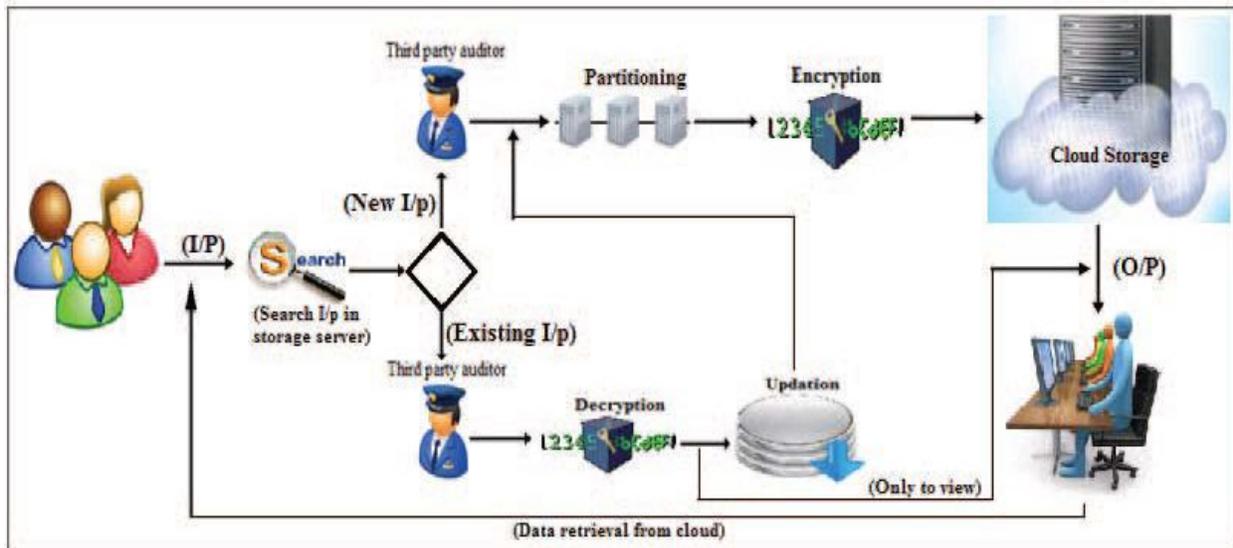
III. EXISTING SYSTEM

Figure 3: Partitioning and Domain Integrity Checking For Data Storage in Cloud Computing

The existing system represents how data is stored and distributed on the cloud using partitioning to provide security and availability across two servers.

3.1 Work Flow
3.1.1. Client Side
a) File Selection: The user/users select a File to upload on the server.
b) Sending File: The client/clients sends respective file to Third Party Auditor (TPA) across the network using a File Transfer Protocol (FTP).

3.1.2. TPA
TPA performs six types of steps on the received file/files from the respective clients for storage across cloud with security and availability.
a) Receiving File
Third Party Auditor Receives File from the client/clients.
b) Partitioning File
TPA partitions file received from the client/clients.
c) Digital Signature Extraction
TPA extracts Digital Signature of each file partition.
d) Secret Key Generation
After partitioning, Third Party Auditor generates Secret keys for each partition respectively.
e) Encryption
TPA encrypts each partition using respective secret keys.
f) Storing Partition Sequence
TPA stores Partition Sequence, Signature, Keys and File attributes on its own server.

3.1.3 Server Side
a) Sending Partitions
Third Party Auditor sends the respective partition to the respective storage. The Respective Storage Server receives the respective File from the Third Party Auditor.
b) Storing
The storage server stores the partition received from the TPA.

IV. PROPOSED SYSTEM
The end user fires a query when he/she needs a certain piece of information simply using their browser. The file is send over the network where it is received by Third Party Auditor (TPA), it receives the file and partitions it into blocks and generates keys for each partition and encrypts each partition using the respective keys. TPA stores partition sequence, signature, keys and file attributes on its own server. Naive Bayes Classifiers are used and the file is forwarded to the database servers where it is stored in the partitioned form for quick access. When a user fires a query for retrieval of data the partitioned file stored on the database are merged together again to form the original file.
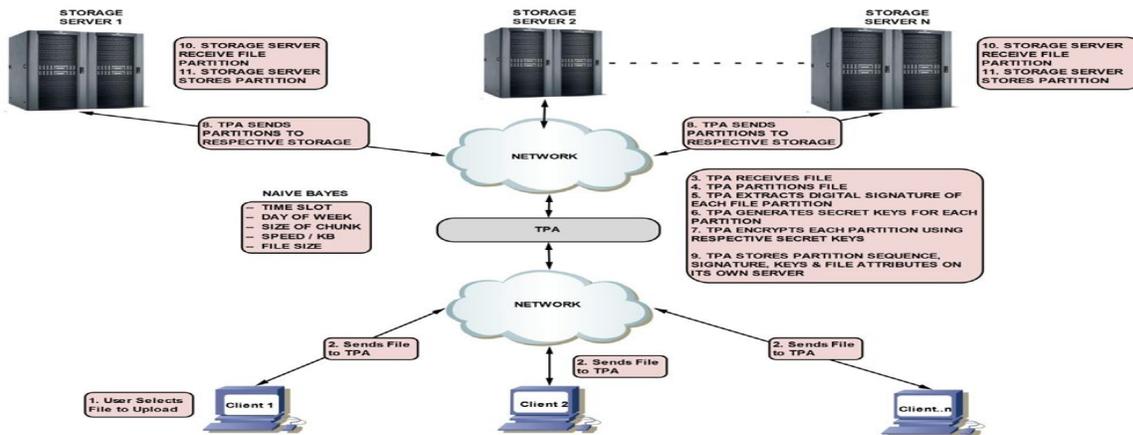

Figure 4: Partitioning and Domain Integrity Checking for Data Storage in Cloud Computing

System consists of end users, a Third party auditor, a network and the multiple servers for storing the data.
• USER: The end user is the actual user which queries the database in the servers using browser installed in his system.

• TPA: Basically is meant for auditing in the system. It uses various algorithms for encryption, decryption and authentication purposes.
• Network: The network used is Internet.
• NBC: Naive Bayes Classifier
• Servers: The servers store data in the partitioned form. After a query is fired the data is organized and sent to user

V.   RESULTS

	Precision	Recall
Relevant Server	0.944444444	0.85
Non Relevant Server	0.882352941	0.75
Total	0.913398693	0.8
Accuracy Percentage	0.8	

VI.   CONCLUSION

The Proposed system is efficient system considering the security and other aspects of the cloud storage technology. Partitioning data makes data access easy and quick. Data stored is highly secured using the cryptography algorithms and digital signatures. It integrates some new concepts like data security, storage optimality, file integrity and authentication access which are not present in the current system. NBC make the system much more robust as compared to existing ones.

REFERENCES

1. *Prabavathy.B, Priya K, Chitra Babu, "A Load Balancing Algorithm for Private Cloud Storage"*
2. *Aubrey L. Tatarowicz, Carlo Curino, Evan P. C. Jones, Sam Madden, "Lookup Tables: Fine-Grained Partitioning for Distributed Databases", Massachusetts Institute of Technology, USA.*
3. *C. Wang, K. Ren, W. Lou, and J. Li "Towards Publicly Auditable Secure Cloud Data Storage Services", IEEE Network Magazine, vol. 24, no. 4, pp. 19-24, July/Aug.2010.*
4. *Rene Leistikow and Djamshid Tavangarian "Secure Picture Data Partitioning for Cloud Computing Services", 27th International Conference on Advanced Information Networking and Applications Workshops 2013.*
5. *Yanhong Zhai and Bing Liu "Web Data Extraction Based on Partial Tree Alignment".*
6. *Yunhua Deng and Rynson W.H. Lau, senior member IEEE,"On Delay adjustment for Dynamic load balancing in distributed virtual environment", April 2012.*
7. *Bin Dong, Xiuqiao Li, Qimeng Wu, Limin Xiao and Li Ruan, "A dynamic and adaptive load balancing strategy for parallel file system with large-scale I/O servers," Journal of Parallel Distribution. Computer.72, pp. 1254–1268, 2012.*
8. *Pragati Priyadarshinee and PragyaJain, "Load Balancing and Parallelism in Cloud Computing," International Journal of Engineering and Advanced Technology, Vol. 1, Issue. 5, 2012, pp. 486-489.*
9. *Bhushan Lal Sahu, Rajesh Tiwari, "A comprehensive study on Cloud computing", International Journal of Advanced Research in Computer Science and Software Engineering, volume 2, issue 9, September 2012, ISSN: 2277 128X.*
10. *Ratan Mishra, Anantjaiswal, Ant colony optimization: A Solution of load balancing in cloud, International J1ournal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012.*
11. *[11]D. Escalnte, Andrew J. Korty, "Cloud Services: Policy and Assessment", Educause review July/August 2011.*

(C) *Global Journal Of Engineering Science And Researches*