IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. V. (Jul.-Aug. 2017), PP 08-12
www.iosrjournals.org

Secure Distributed Big Data Storage Using Cloud Computing

*Sana Khan, Ekta Ukey
[1](Computer, Pillai HOC Rasayani/ Mumbai University, India)
[2](Computer, Pillai HOC Rasayani/ Mumbai University, India)
Corresponding Author: *Sana Khan*

Abstract: *The cloud is increasingly being used to store and process the big data. Many researchers have been trying to protect big data in* ***cloud computing*** *environment. Traditional security mechanisms using encryption are neither efficient nor suited to the task of protecting big data in the Cloud. first discuss about challenges and potential solutions for protecting big data in cloud computing. Second,* ***'Secure distributed big data storage in cloud computing'*** *Architecture for protecting* ***Big Data*** *in Cloud Computing Environment. Model ensures efficient processing of big data in cloud computing environment and gains more business insights. The data security and privacy has become a critical issue that restricts many cloud applications. One of the major concerns in security and privacy is caused by the fact that cloud operators have chances to reach the sensitive data. This concern dramatically increases user's anxiety and reduces the adoptability of cloud computing in many fields, such as the financial industry and governmental agencies. It focuses on issues and approach, by which the cloud service operators cannot directly reach partial data. The approach divides the file and separately stores the data in the distributed cloud servers. An alternative approach is designed to determine whether the data packets need a split in order to shorten the operation time.*

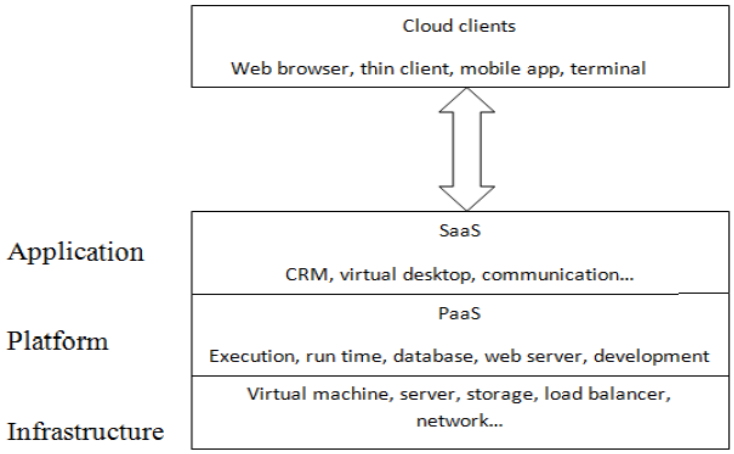
Date of Submission: 24-07-2017 Date of acceptance: 05-08-2017

I. Introduction

Cloud computing:

Cloud computing can be defined as Multi –tenancy , Massive Scalability (Shared Resources), Elasticity, Pay as You go and Self-Provisioning of resources. Cloud computing enables user to use the remote servers hosted on the internet to process and store the data. Service models of cloud is classified into three types such as SaaS, PaaS, IaaS and different deployment models are classified into Private, Public, and Hybrid. Due to the high availability of cloud to all end users, cloud computing faces more security challenges. These challenges are classified into two broad categories as security issues faced by cloud providers and security issues faced by Customers.

Figure 1: End User Accessing Applications and Data in Distributed Cloud

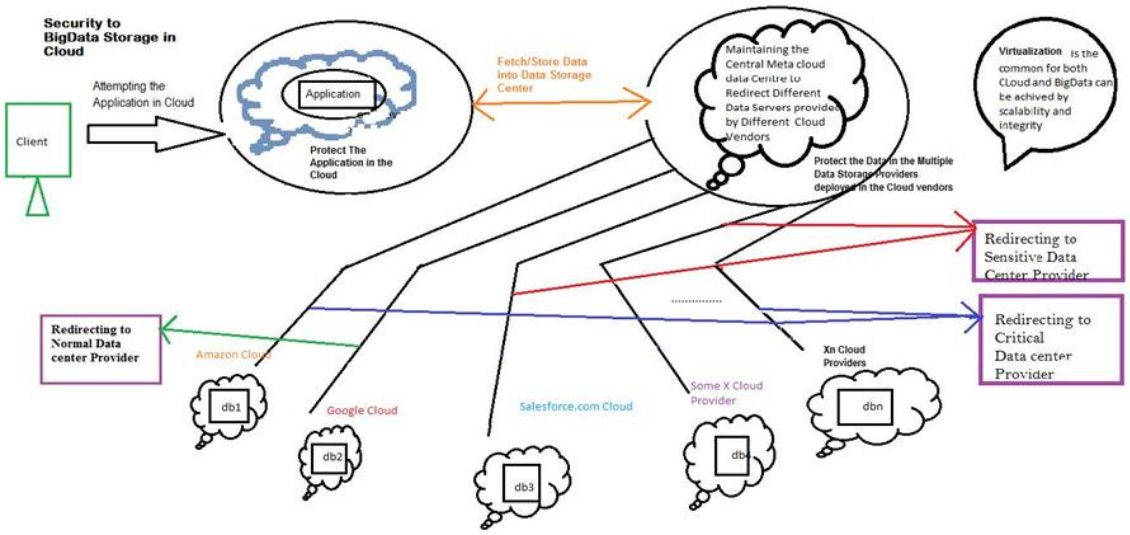


Big Data: Big data can be defined as collection of huge size of the data sets with different types so that data becomes difficult to process by using traditional data processing platforms and algorithms. Recently the number of data provisions has increased, such as social networks, sensor networks, high throughput instruments, satellite and streaming machines and these environments produce huge size of data. Big data used in many application health care , education, natural resources, social networking and so on.

The three main terms that generally signify Big Data are:
Volume: This has to do with the amount of data generated on a daily basis which is so large and keeps increasing with time. [2]
Variety: Today data is created in different type, form and formats such as emails, video, audio, transactions etc. [2]
Velocity: This has to do with the speed it takes to produce data and how fast this data produced needs to be processed on time to meet individual demand.[2]

II. Literature survey

Figure 2: End User Accessing Applications and Data in Distributed Cloud [1]



As one of the significant technologies used in cloud computing, the distributed storage has enabled the mass remote data storage via Storage-as-a-Service (STaaS) service model. This cloud service model has broadly become an acceptable approach in big data along with the development of Web services and networks. Many cloud vendors have given attractive storage service offerings that provide giant and scalable cloud-based storage spaces for users, such as Amazon, Drop box, Google Drive, and Microsoft's One Drive. However, the security issue caused by the operations on cloud side is still an obstacle of using STaaS for enterprises. Many cloud users concern about their sensitive data to which the cloud operators have the access. This matter embarrasses contemporary implementations of STaaS, even though many prior researches have addressed this field. Moreover, Mass Distributed Storage (MDS) has been explored to scale up the data storage size in recent years. The high level performances of the scalable computation are considered benefits of implementing MDS. One aspect that needs improvements is to secure distributed data storage, in which the threats come from a variety of sides. The distributed storage manner can result in more chances of malicious attacks or abuse activities , such as attack during data transmissions. Currently, the unexpected operations can also occur at the cloud server side, which are mainly constrained by laws and regulations. Meanwhile, it is difficult to balance functionality and security performances due to cost concerns. Therefore, it is a challenging issue to efficiently secure distributed data in cloud systems, since the risks deriving from different network layers are hardly fully addressed. Secure distributed big data storage using cloud computing architecture concentrates on the problem of cloud operators abuse issues and attempts to avoid cloud users' data release from cloud servers. Architecture proposes an intelligent cryptography approach, named Security-Aware Efficient Distributed Storage (SAEDS) model that is designed to obtain an efficient MDS service, as well as high level security protections. Proposed mechanism aims to encrypt all data and distributive store the data to the different cloud servers without causing big overheads and latency. user's data are assessed by an alternative process in which searchable named-data-packets techniques are applied. The solid arrow lines represent the data splits and storage operations. The broken arrow lines represent the operational directions of the data retrievals. Normal data will be assigned to a single cloud server. Meanwhile, the data with sensitive information are split into two parts that are assigned to two cloud servers, Cloud A and Cloud B. This process is mainly supported by our proposed algorithm, Alternative Data Distribution (AD2) algorithm. Moreover, splitting data process is accomplished by the main algorithm, Secure Efficient Data Distributions (SED2) Algorithm, which is designed to spilt data in order to prevent sensitive information from leaking on the cloud side using minimum costs. The sensitive data retrieval needs a decryption process that is supported by proposed algorithm, Efficient Data Conflation (EDCon)

algorithm. The significance of the proposed mechanism is that provides an adaptable approach for those enterprises that intend to use STaaS but require a high level data storage security, such as the financial service industry. The main problem solved by proposed scheme is preventing cloud providers from directly reaching users' original data. The main contributions of model are twofold:

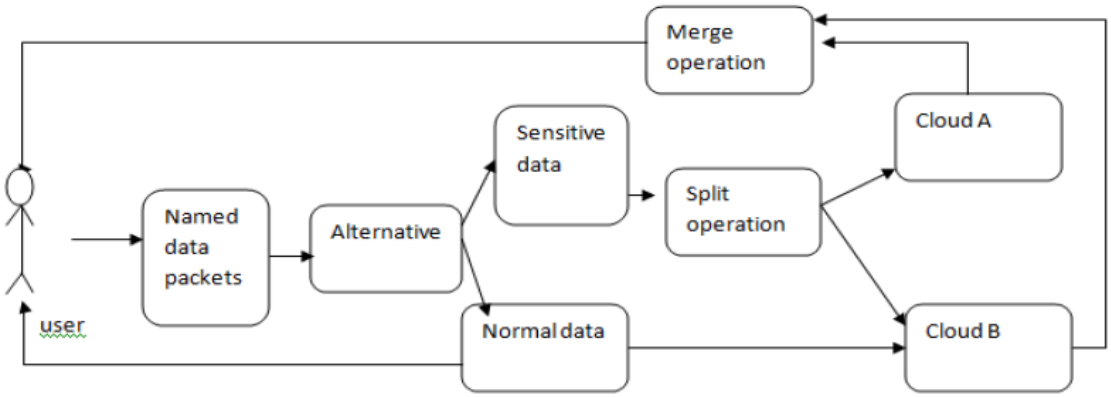
- A novel cryptography approach for delivering mass distributed storage by which users' original data cannot be directly reached by cloud operators. The proposed method is an effectual cryptography means for defending malicious activities occurred on the cloud server.

- An efficient data split mechanism that does not produce big overheads, as well as ensures data retrievability.

III. Proposed Framework:

The proposed framework will distribute all data parts in different storage service providers, and each provider holds some of the data parts. In order to provide high availability and robustness, the proposed framework will store multiple copies of same data on different cloud storage providers. Though big data is split and stored in different data center, the administrator of the entire system will keep the storage index information for each data parts. When there is a problem in some data parts on the cloud storage, propose framework can find another copy of the data parts according to their storage index information. The proposed security algorithm which is shown below that protects the unauthorized access when trying to login into application which has been deployed in cloud. Although this algorithm updates the following tables such as 1) threat updated table, 2) meta data storage cloud table, 3) Amazon cloud data storage table, 4) Google cloud data storage table, 5) xcloud data storage table, 6) xncloud data storage table. Threat updated table will store the entry related to malicious attempt, whereas Meta data storage cloud table stores information regarding the data storage entry of different vendors. Critical, sensitive and non-sensitive data are stored in other tables.

Figure 3: Architecture of secure distributed big data storage in cloud computing



This paper concentrates on the problem of cloud operators abuse issues and attempts to avoid cloud users ' data release from cloud servers. user's data are assessed by an alternative process in which searchable named-data-packets techniques are applied. The solid arrow lines represent the data splits and storage operations. The broken arrow lines represent the operational directions of the data retrievals. Normal data will be assigned to a single cloud server. Meanwhile, the data with sensitive information are split into two parts that are assigned to two cloud servers, Cloud A and Cloud B. This process is mainly supported by algorithm, Alternative Data Distribution (AD2) algorithm. Moreover, splitting data process is accomplished by the main algorithm, Secure Efficient Data Distributions (SED2) Algorithm, which is designed to spilt data in order to prevent sensitive information from leaking on the cloud side using minimum costs. The sensitive data retrieval needs a decryption process that is supported by our proposed algorithm, Efficient Data Conflation (EDCon) algorithm [3].

Figure 4: The security architecture for the secure distributed bid data storage in Cloud computing

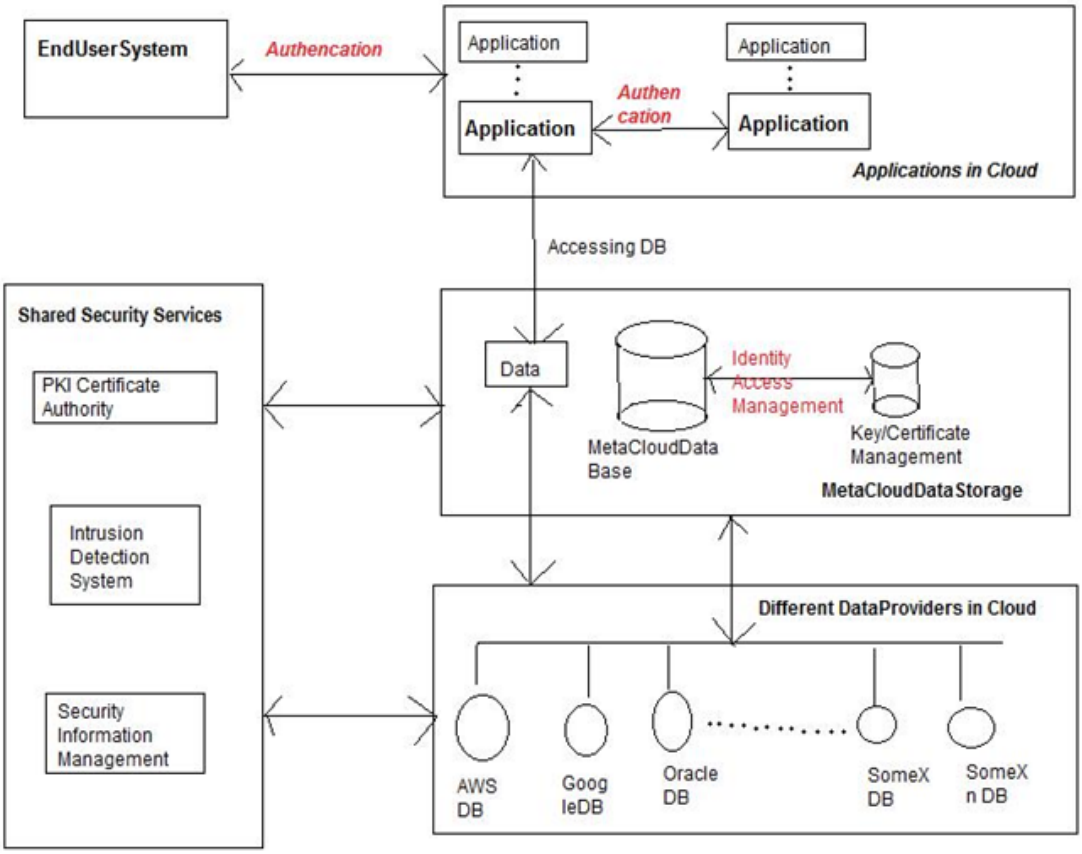
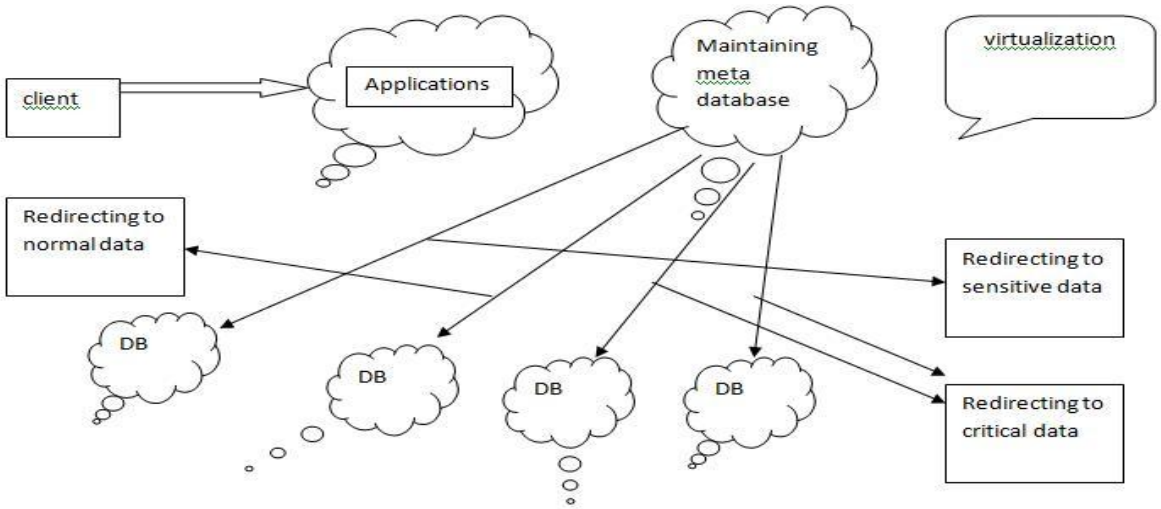


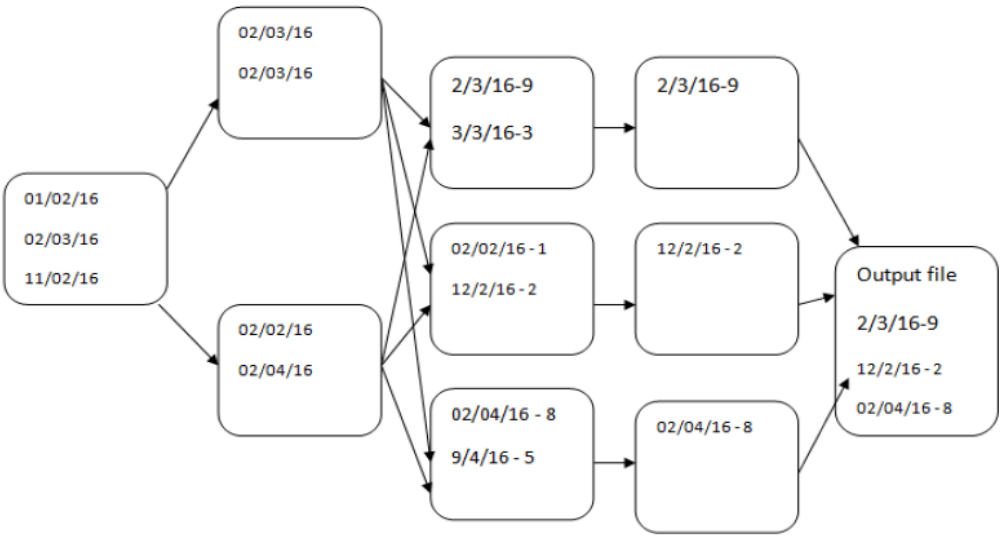
Figure 5: End User Accessing Applications and Data in Distributed Cloud



Although, the proposed framework will distribute all data parts in different storage service providers, and each provider holds some of the data parts. In order to provide high availability and robustness, the proposed framework will store multiple copies of same data on different cloud storage providers. Though big data is split and stored in different data center, the administrator of the entire system will keep the storage index information for each data parts. When there is a problem in some data parts on the cloud storage, propose framework can find another copy of the data parts according to their storage index information. Figure 1 shows how the end user will access the applications and data in distributed cloud. The proposed security algorithm which is shown below that protects the unauthorized access when trying to login into application which has been deployed in cloud. Although this algorithm updates the following tables such as 1) Threat updated Table, 2)

Meta Data Storage Cloud Table, 3) Amazon Cloud Data Storage Table, 4) Google Cloud Data Storage Table, 5) Xcloud Data Storage Table, 6) Xncloud Data Storage Table. Threat updated table will store the entry related to malicious attempt, whereas Meta Data Storage Cloud table stores information regarding the data storage entry of different vendors. Critical, Sensitive and non-sensitive data are stored in other tables. [1]

Figure. 6: Map Reduce framework for processing log files



Map Reduce is a programming model or framework that process tasks in parallel across a huge size of systems. It contains two functions such as Map and Reduce. Map function splits the huge size of input data into <key, value> pairs. Intermediate <key, value> pairs will be created bases on aggregating several input key value pairs from the Map phase. Finally, Reduce takes the intermediate key value pairs and produces the output <key, value> pairs that can be easily understood by the end user. In this proposed architecture, Map Reduce framework is used to find the number of users who were logged in to the cloud data center. Proposed Map Reduce pseudo code can efficiently process the huge size of log file in which it contains users who were logged in with date and the log in time duration. As shown in the Figure 6, the first process is map phase in which each date that represents the key is assigned a value of one initially. While reduce phase, the key values are summed up to find out the number of users logged in. For example, three users were logged in 01-02-2016, whereas two users were logged in 02-02-2016 [1].

IV. Conclusions

In this architecture we proposed secure distributed big data storage in cloud computing Architecture for protecting Big Data in Cloud Computing Environment. Map Reduce framework has used to find the number of users who used to logged into the cloud data center. Proposed framework protects the mapping of various data elements to each provider using implemented architecture. Though this proposed approach requires high implementation effort, it provides valuable information for cloud computing environment that can have high impact on the next generation systems. Our future work is to extend the proposed secure distributed big data storage in cloud computing Architecture for real time processing of streaming data.

References

[1]. Gunasekaran Manogaran, Chandu Thota. MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing. Procedia Computer Science 87 (2016).
[2]. Awodele, Izang A.A. Big Data and Cloud Computing Issues. International Journal of Computer Applications (0975 – 8887) Volume 133 – No.12, January 2016.
[3]. Yibin Li , Keke Gai. Intelligent cryptography approach for secure distributed big data storage in cloud computing. Software School, Henan University, Kaifeng, Henan, 475000, China.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Sana Khan. "Secure Distributed Big Data Storage Using Cloud Computing." IOSR Journal of Computer Engineering (IOSR-JCE) 19.4 (2017): 08-12.