Privacy Preserving Approach using Encrypted Cloud Databases

Shital H. Dinde[1] , Arati M. Dixit[2]
1. Department of Computer Engineering, PVPIT/JSPM, Bavdhan, Pune, Maharashtra, India
2. Department of Technology, Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract- **Rapid advances in storage, communications, and processing allow us to move all data into cyberspace. Data management systems began by automating traditional tasks like keeping record of business transactions. This data consisted primarily of numbers and character strings. Today the automated systems offer the infrastructure for our society, allowing quick, reliable, secure, and automatic access to information distributed throughout the world. One of that service provider system is cloud. Currently the enterprises are moving towards less cost, more availability, agility, managed risk, all of that is accelerated towards Cloud Computing. Cloud is not a specific product, but a simplest way of delivering IT services that are serviced on demand, elastic to rescale as needed, and follow a pay-for-usage model. Out of the three common kinds of cloud computing service models, Infrastructure as a Service (IaaS) is a service model that provides servers, computing power, network bandwidth and Storage capacity, as a service to their subscribers. Cloud will relate to many things but without the basic storage items, that is provided as a service particularly Cloud Storage, none of the other applications is feasible. But there are many security issues in cloud storage. The solution for that is to work on encrypted data i.e. store the data into the cloud database in encrypted format and operate on that encrypted data and directly connect distributed client to encrypted cloud database i.e. without intermediate proxies.**

Keywords – Cloud database, Privacy, cloud storage, metadata, PLA, PBA, PLAC.

I. INTRODUCTION

Computers will store all varieties of data like documents, images, sound recordings, videos, scientific information, and lots of new information formats. It creates strides to capture, store, manage, analyze, and visualize this information. These tasks are called as data management[1]. Data management systems generally store large quantities of information representing the historical records of a organization. it's vital that the recent data and applications still work as new data and applications are value added. The systems are in constant amendment. Indeed, most of the larger information systems operational nowadays were designed many decades past and have evolved with technology. A historical perspective helps to know current systems. There are six distinct phases in data management[1]. Initially, data was manually processed. Next evolution was punched card instrumentation and electro-mechanical machines which is used to type and tabulate lots of records. The next phase was used to hold on data on magnetic tape and to perform execution on serial files. The fourth generation introduced the conception of a database schema and on-line navigational relational databases and value-added distributed and client server process. The early stages of sixth generation systems that have more data types, notably documents, images, voice, and videos. These sixth generation systems are the storage engines for the rising internet and Intranets.

Now a day's information is increasingly important in anyone's daily live. All have become information dependent, so living in an on command, on demand world, which means, individual need information when and where it is required. Everyone access the internet every day to perform searches, participate in social networking, send and receive emails, share pictures and videos. This will create a huge information, but it has no value until it is shared with others. To be shared, this information need to be uploaded to central data repositories (data centers) via network. Businesses are also depends on fast and reliable access to information. The increasing dependence of businesses and an individual on information has amplified the challenges in storing, protecting and managing data. Organizations usually maintain one or more data centers to store and manage information. The data center is a facility that contains information storage and other physical information technology resources for computing, networking and storing information. So, the cloud computing is one of that data center which brings in a fully automated request fulfillment process that enables users to rapidly obtain storage and other IT resources on demand. One of the services provided by cloud computing is the cloud database which stores the data online, on demand. A cloud database is a database that generally runs on a cloud computing platform, like Amazon EC2, GoGrid, Salesforce, Rackspace, and Microsoft Azure. There are two deployment models, the first one is users can run databases on the cloud independently, using a virtual machine image, and the second is they can purchase access to a database service which is maintained by a cloud database provider.

Cloud computing is a tool that offers enormous benefits to its subscribers. As it is a tool, it comes with its set of problems and inefficiencies. The main and major concern is security and privacy in cloud. By leveraging a remote cloud based infrastructure, an organization essentially gives away private data and information and other things that might be sensitive and confidential. It is then depends on the cloud service provider to manage, protect and retain them. An organization's existence might be in danger, so therefore all doable alternatives should be explored before a decision. Similarly, privacy in the cloud is another major issue.

ISSN: 2278 – 1323

Organizations and users have to trust their cloud service vendors that they are going to protect their data from unauthorized users. The various stories of data loss and password leakage in the media does not help to reassure some of the most concerned users.The public nature of cloud computing poses significant implications to data privacy and confidentiality. Cloud data is often stored in plain text, and few companies have an absolute understanding of the sensitivity levels their data stores hold. A recent report by the Cloud Security Alliance lists data loss and leakage as one of top security concerns in the cloud. Recent laws, regulations and compliance frameworks compound the risks; offending companies can be held responsible for the loss of sensitive data and may face heavy fines over data breaches. To lose data security practices also harm on a personal level. Lost or stolen medical records, credit card numbers or bank information may cause emotional and financial ruin. Sensitive data stored within cloud environments must be safeguarded to protect its owners.

II. LITERATURE REVIEW

A. Evolution of Cloud Storage

Rapid data growth and the need to keep it safe will require organizations to integrate how they manage and use their data, from creation to end of life. Now we can store all our data in the internet storage space i.e. cyberspace. These storages are provided and maintained by the third parties through the Internet which is represented in Fig. 1[2]. Cloud storage offers a large pool of storage was available for use, with three significant attributes: access via Web services APIs on a non persistent network connection, immediate availability of very large quantities of storage space, and pay for what you have used. It supports rapid scalability [2].

Cloud storage is an offering of cloud computing. Fig. 2 [2]shows the evolution of Cloud Storage based on traditional network storage and hosted storage. One of the advantages of cloud storage is the access of your data from anywhere.

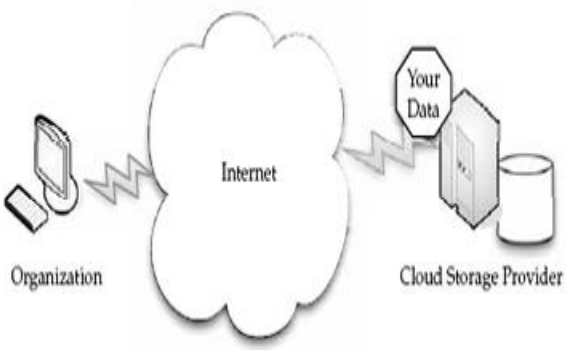

Figure 1. Simple cloud storage model

Cloud storage providers provide storage varying from small amount of data to even the entire warehouse of an organization. User can pay to the cloud storage provider for what they are using and how much they are transferring to the cloud storage. In the cloud storage environment the user copies the data into one of the data server of the cloud storage provider. This data will be made available to the remaining data servers on the cloud which will be result in high availability of the data.

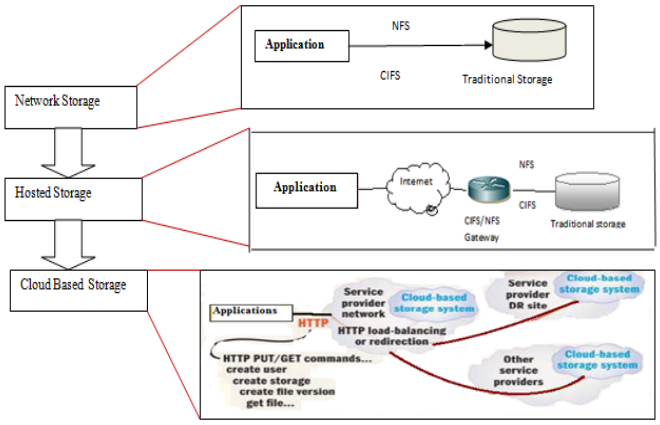

Figure 2. Evolution of Cloud Storage

B. Related Work

Cryptographic file systems and secure storage will guarantee confidentiality of the data and the integrity of data which will be stored on untrusted cloud. DBMS engines offers encryption of data using Transparent Data Encryption (TDE) [3].It possible to build a trusted DBMS over untrusted storage by using this technique. But, in the DBaaS context the DBMS engine is not trusted because it is controlled by the cloud provider, hence the TDE approach is not suitable for the cloud database services. An approach to preserve data confidentiality in scenarios where the DBMS is not trusted. However it requires a modified DBMS engine that is not compatible with commercial and open source DBMS software adopted by cloud providers. On the other hand, the we proposed architecture is compatible with standard DBMS engines, and allows customers to make a secure cloud database by leveraging cloud DBaaS readily available.

The proposal in [4] uses encryption to control accesses to encrypted data stored in a cloud database. This solution is not applicable to usage contexts in which the structure of the database changes, and does not support concurrent accesses from multiple clients possibly distributed on a geographical scale.

The following are three types of architectures are defined to preserve the privacy.

1. Proxy-based architectures (PBA)

The proxy-based architectures[5] shown in Fig. 3[5] do not satisfy our design requirements because the proxy is a bottleneck and a single-point-of-failure that limits availability, scalability and elasticity of the cloud DBaaS. Since the proxy must be trusted, it cannot be outsourced to the cloud and has to be deployed and maintained locally. Moreover, proxy-based architectures cannot scale trivially by increasing the number of proxies. Such a naive solution would imply the replication of metadata among all the proxies, but this would require synchronization algorithms and protocols to guarantee consistency among all the proxies.

ISSN: 2278 – 1323

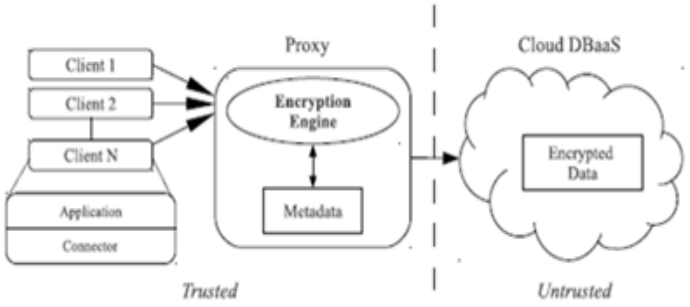

Figure 3. PBA Architecture

2. Proxy-less architectures that store metadata in the clients (PLA)

The Proxy-less architectures that store metadata in the clients[4] does not use an intermediate proxy and metadata are stored at the client side. So the clients can connect directly to the cloud database, this architecture provides availability, scalability and elasticity. So, each client has its own encryption engine and manages a local copy of metadata. So, this solution can represent a sub-case of the proxy-based architecture, in which a different proxy is deployed within each client. A similar architecture for cloud accesses would suffer from the same consistency issues of proxy-based architectures. This architecture is given in Fig. 4[4].

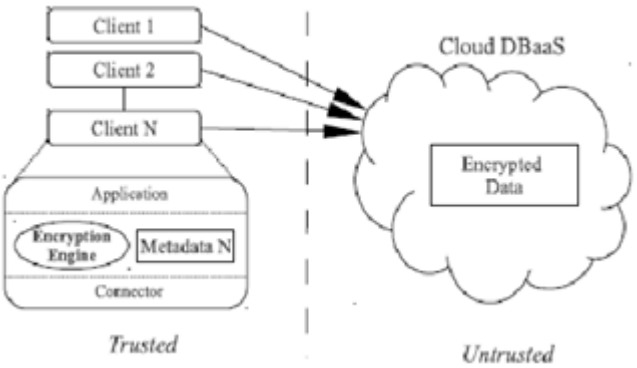

Figure 4. PLA Architecture

3. Proxy-less architectures that store metadata in the cloud database (PLAC)

The third architecture is proxy-less architectures[6] shown in Fig. 5[6] that store metadata in the cloud database. In this the metadata is stored to the cloud database, but the encryption engine is executed by each client. As metadata are not shared among all the clients there is no need of synchronization. Client machines execute a client software component that allows a user to connect and issue queries directly to the cloud DBaaS. This software component retrieves the necessary metadata from the untrusted database through SQL statements and makes them available to the encryption engine at the client. Multiple clients can access the untrusted cloud database independently, with high availability, scalability and elasticity. The drawback of this architecture is bottleneck and the single point of failure.

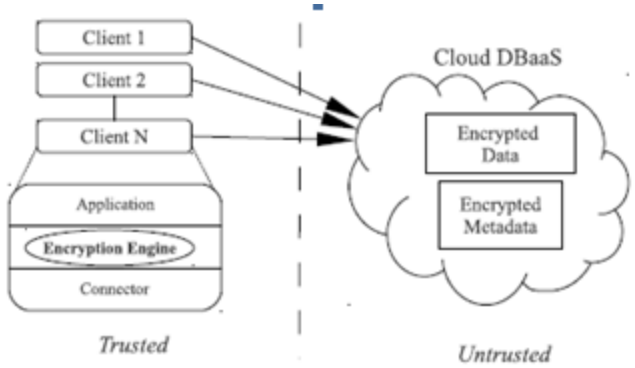

Figure 5. PLAC Architecture

III PROPOSED DD-PLAC ARCHITECTURE

Protecting privacy is very much difficult in a computerized world where individuals, devices, and sensors are associated and information is created, accessed and shared widely with one another. To ensure the clients' security, governments additionally came up with legitimate measures, for example, the US federal law called The Secret Data Assurance and Measurable Productivity Act (CIPSEA). In the same endeavors, organizations have utilized different information de-ID routines, for example, pseudonymization, encryption and so on to remove/hide any data that recognizes people. However these de-ID strategies have not been totally ready to secure the client's protection.

If anyone wants to store the personal or confidential data in the cloud, these are securely encrypted before storing them to the cloud. Encrypting the data will safeguard the privacy of your data; especially important when you are storing sensitive corporate data or personal information that should never fall into the wrong hands.

ISSN: 2278 – 1323

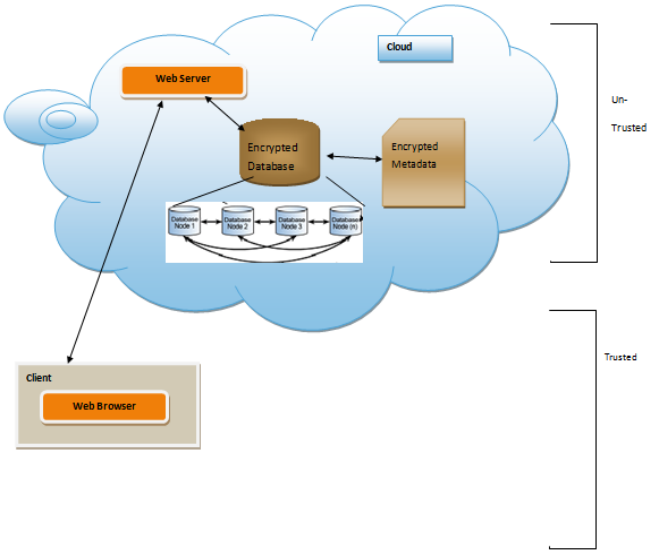

Figure 6. DD-PLAC Architecture

The limitations in PLAC architecture are bottleneck and the single point of failure. These are avoided in the proposed DD-PLAC architecture. The DD-PLAC architecture shown in Fig. 6 is same as proxy-less architectures that store metadata in the cloud database but to avoid single point of failure the distributed cloud database is used where the data is distributed over the cloud, which will allow the databases to truly support the elastic requirements of cloud computing applications. Databases have been distributed in terms of instances running on servers that have access to a high-speed network for a while. It also increases the availability of the data.

In the encryption algorithm, the encryption key is to be designed based on the data to be stored in to the database. For example, if we store the birth information of child in the database then the encryption key may contain the some part of information which is required at the time of birth registration like birth date, name of child, birth place, pincode, etc. Every time new combination is being taken. Due to this for every record the different encryption key is being generated so the data is being safe and remain private in cloud. These encryption keys, database name, table name and unique record id is to be stored in again encrypted format in metadata. The structure of metadata is given in Fig. 4. Using this metadata it is easier to find the encryption key of record being stored in database for the application. Concurrent read and write operations that do not modify the structure of the encrypted database. This encryption algorithm ensures the data privacy and the distributed cloud database will ensures the continuous service from the cloud service provider as well as confirms the data is safe in hands of cloud database.

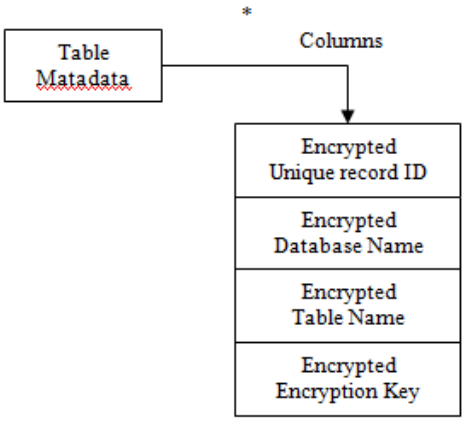

Figure 7. Metadata Structure

IV CONCLUSION

With speed-of-light technological innovation, information privacy is becoming more complex by the minute as more data is being collected and exchanged. As the technology gets more sophisticated (indeed, invasive), so do the uses of data. And that leaves organizations facing an incredibly complex risk matrix for ensuring that personal information is protected. As a result, privacy has fast-emerged as perhaps the most significant consumer protection issue, if not citizen protection issue, in the global information economy. Encrypting the data will safeguard the privacy of your data; especially important when you are storing sensitive corporate data or personal information that should never fall into the wrong hands.

The DD-PLAC architecture guarantees the data privacy which is saved into cloud databases. All the data which is stored on the cloud provider are encrypted through cartographic algorithms which allow the execution of standard SQL queries on encrypted data. This architecture is also provide direct, independent and concurrent access to the cloud database. It does not rely on a trusted proxy that represents and also avoids the single point of failure and a system bottleneck, which in turn increases the availability and scalability of cloud database services.

V REFERENCE

[1] Daniel J. Abadi, Data Management in the Cloud: Limitations and Opportunities, IEEE Data Engineering Bulletin, Volume 32, March 2009, 3-12.

[2] James Broberg, Rajkumar Buyya, Zahir Tari, MetaCDN: Harnessing Storage Clouds or high performance content delivery, Journal of Network and Computer Applications, 1012–1022, 2009.

ISSN: 2278 – 1323

[3] Oracle corporation: Oracle advanced security (October 2012), http://www.oracle.com/technetwork/database/options/advanced-security

[4] Damiani, E., De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Metadata Management in Outsourced Encrypted Databases. In: Jonker, W., Petkovi´c, M. (eds.) SDM 2005. LNCS, vol. 3674, pp. 16–32. Springer, Heidelberg (2005)

[5] Popa, R.A., Redfield, C.M.S., Zeldovich, N., Balakrishnan, H.: CryptDB: protecting confidentiality with encrypted query processing. In: Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, SOSP 2011, pp. 85–100. ACM, New York (2011)

[6] Luca Ferretti, Michele Colajanni, and Mirco Marchetti: Distributed, Concurrent, and Independent Access to Encrypted Cloud Databases. IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014.

[7] "Xeround: The Cloud Database," Xeround, http://xeround.com, Apr. 2013.

[8] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M.Newbold, M. Hibler, C. Barb, and A. Joglekar, "An Integrated Experimental Environment for Distributed Systems and Networks," Proc. Fifth USENIX Conf. Operating Systems Design and Implementation, Dec. 2002.

ISSN: 2278 – 1323