International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 01-05

Encrypted Big Data with Redundant Information in Cloud

P. Aiswariya,
Assistant Professor [1]Department of Computer Science / SaradaMahavidyalayam Arts and Science College for Women, Ulundurpet

Abstract: *In the Internet, Cloud computing is a common term for the delivery of hosted services. Cloud computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet. The most important and popular cloud service is data storage. The deduplication process reduces the amount of data in a storage system, but dedupe in the cloud may be more valuable to the cloud provider than the customer. In the cloud and other storage platforms deduplication process remove the data to reduce the physical data amount by which frequent data is stored in a system or machine. In primary storage, deduplication helps to reduce the amount of physical space consumed by removing identical blocks of data and using metadata to associate the logical copies of data to the physical ones. The deduplication capabilities of the storage platform are not showing to the user in the public cloud. If the provider chooses to implement deduplication in cloud computing, then that benefit is retained for the cloud provider. This is because storage space is allocated based on logical capacity used rather than the physical capacity and any reduction in savings is used by the service provider to offer a cheaper service or to reduce its costs. We propose a plan to deduplicate knotted information put away in the cloud using RSA algorithm and Kerberos authentication protocol. The results of the system produce more efficient authentication to servers and delegated authentication.*

Keywords: *big data, cloud computing, data deduplication, Kerberos authentication, encryption*.

I. Introduction

Cloud computing is an innovative service model. Throughout the network to the required resources (hardware, platform, and software), virtual integration into a consistent and high performance of computing platform. In cloud computing, all user data are stored in the cloud resources like cloud nodes. The Cloud computing is the use of computing resources that are delivered as a service over a network. The resources are made available on the internet as managed third-party services. One of the most fundamental services offered by cloud providers is data storage. The data is available at any time because the cloud computing is an internet-based computing. The cloud service providers are providing various services to the users. The conclusion distributes to the user through the network when the user needed the security. Although cloud computing has the aim to mature service model, and have large commercial, cloud computing is still facing many problems. Cloud computing still facing the number of major challenges: Safety, Stability, and Performance issue. Including the security and memory management problem concerns the most. Real cloud computing securities incidents have profound acknowledge the urgency of cloud security and authentication issues, example for users cannot access their email and other personal data. More important, due to the technical personnel not to be make backups of their data, the conclusion of Microsoft cannot recover data. Although the cloud storage service can realize multiple copies of files fault tolerance and backup automatically, it is also guaranteed 100 percent security and authentication. Deduplication has proved to achieve high-cost savings storage needs for backup applications. In this paper, we propose a scheme based on data authorization and using Kerberos authentication to manage encrypted data storage with deduplication.

II. Existing System

A technique which has been proposed to meet conflicting requirements is convergent encryption whereby the encryption key is usually the result of the hash of the data segment. Although convergent encryption seems to be a good candidate to achieve confidentiality and deduplication at the same time, it, unfortunately, suffers from various well-known weaknesses including dictionary attacks: an attacker who is able to guess or predict a file can easily derive the potential encryption key and verify whether the file is already stored at the cloud storage provider or not. In our project, we cope with the inherent security exposures of convergent encryption and propose Cloud Deduplication, which preserves the combined advantages of deduplication and convergent encryption. The security of Cloud Deduplication relies on its new architecture whereby in addition to the basic storage provider, a metadata manager and an additional server are defined: the server adds an additional encryption layer to prevent well-known attacks against convergent encryption and thus protect the confidentiality of the data; on the other hand, the metadata manager is responsible for the key management task since block-level deduplication requires the memorization of a huge number of keys. Therefore, the underlying deduplication is performed at block-level and we define an efficient key management mechanism to avoid users to store one key per block.

III. Proposed System

The user wants to upload the file to the server then the system shows to upload their file is successfully otherwise the data is same them shows the data duplication on the server. The above Fig 1 shows the block diagram of the proposed system architecture. The content of data is not same then store the data or file in the cloud. The new user wants to upload their files to this cloud they want to register on the server and get the user ID and password. Login to this ID and password and upload the data to storage. The authorized user is after login view to access the files. The only authorized user can access the files and this user can download the files. The user can upload files, download files, share files and revoke files. The user share files with an encrypted key to the other user. The authorized user can download this share files with the encrypted key.

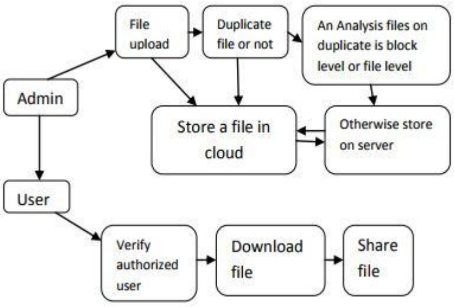

Fig 1: Block diagram of proposed system

RSA algorithm is asymmetric cryptography algorithm. Asymmetric actually what is that it works on two different keys i.e. Public Key and Private Key. As the name indicates that the Public Key is given to everyone and the Private key is kept in private.

An example of asymmetric cryptography:

1. A client (for example browser) sends its public key to the server and requests for some data.
2. The server encrypts the data using client’s public key and sends the encrypted data.
3. The client receives this data and decrypts it.

Since this is an asymmetric algorithm, nobody can decrypt else except browser can decrypt the given data even if a third party has the public key of the browser.

The central theme of RSA is based on the fact that it is very difficult to factorize a large integer. The public key consisting of two numbers where one number is the product of two large prime numbers. And the private key is also derived from the same two prime numbers. So if someone can factorize the large number, the private key is compromised. Therefore the strength of an encryption totally depends on the key size and if we double the key size or triple the key size, the strength of encryption increases exponentially. In below figure 2 RSA keys can be naturally 1024 or 2048 bits long, but experts believe that 1024 bit keys could be broken in the near future. But till now it appears to be an infeasible task.

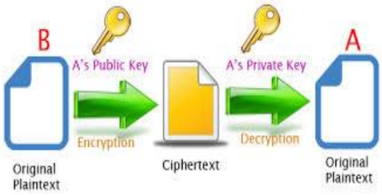

Fig 2: RSA Block Diagram

Let us learn the mechanism behind RSA algorithm:

Generating Public Key :

- Select two prime numbers. Suppose **P = 53 and Q = 59**.
- Now First part of the Public key:**n = P*Q = 3127**.
- We also need a small exponent say **e**
- But e must be

One Day National Conference On “Internet Of Things - The Current Trend In Connected World” NCIOT-2018 2 | Page

- An integer.
- Not be a factor of n.
- **1 < e** <Φ(n)
- Let us now consider it to be equal to 3.
- Our Public Key is made of n and e

Kerberos Authentication

The authentication protocol Kerberos contributes a lot to the mechanism for authentication and shared authentication between a client host machine and a server host machine or between one server and another server. Windows Server 2003 trappings the Kerberos V5 protocol as a security support provider (SSP), which can be accessed through the Security Support Provider Interface (SSPI). The Kerberos Key Distribution Center (KDC) best uses the Active Directory Service Database domains as its security version for the database. Active Directory is required for evasion NTLM and Kerberos implementations.

The Kerberos V5 protocol assumes that initial transactions between clients and servers take place on an open network in which packets transmitted along the network can be monitored and modified at will. The suspected environment, in other words, is very much like today's Internet, where an attacker can easily pose as either a client or a server, and can readily eavesdrop on or interfere with communications between genuine clients and servers.

IV. System Architecture

The architecture preserves the combined advantages of deduplication and convergent encryption, we manage with the innate security showing of proposing Cloud Deduplication and convergent encryption. The security of Cloud Deduplication relies on its new architecture whereby in addition to the basic storage provider, a metadata manager and an additional server are defined: the server adds an additional encryption layer to prevent well-known attacks against convergent encryption and thus protect the privacy of the data; on the other hand, the metadata manager is responsible for the key management task since block-level deduplication requires the memorization of a huge number of keys. Therefore, the underlying deduplication is performed at block-level and we define an efficient key management mechanism to avoid users to store one key per block. In this above architecture, there are following working sides. One is the client side and another one is the server side. Using following methods we will acquire the cloud deduplication in a programmatic manner. As shown in the Fig.3 Data holder is one who uploads and save its data in CSP (Cloud Service Provider). It is feasible to have a number of suitable data holders that could save same encrypted raw information at CSP. The data holder that produces or creates the file is regarded as data owner. It has higher priority than other normal data holders.CSP offers storage services. It permits information owner to keep any kind of information and cannot be fully trusted by users. User fully trusts AP. AP verifies data ownership and handles data deduplication.

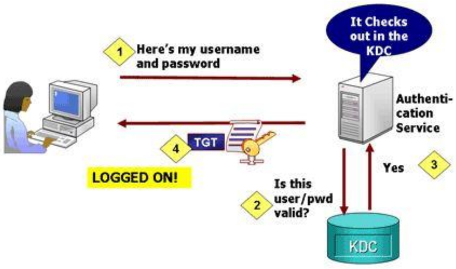

Fig: 3 Kerberos Authentication Process

Our scheme contains the following main aspects as shown in below fig 4.

Encrypted Data Upload:

The data holder wants a need to store the data then it encrypts its data using a randomly selected symmetric key DEK to ensure the security and privacy of data. To stores the encrypted data at CSP as used for data duplication check. The data holder encrypts DEK with pkAP and passes the encrypted key to CSP.

Data Deduplication:

Data holder tries to store the same data that has been stored already at CSP when Data duplication occurs at the time. The comparison is checked by CSP through VCS algorithm. If the assessment of data is equal or same as before, CSP contacts AP for deduplication and the data holder's PRE public key. The AP face the challenge is to checks the eligibility of the data holder data ownership and then issues a re-encryption key. It can adapt the encrypted DEK to a form that can only be decrypted by the eligible data holder.

Data Deletion:

When the data holder deletes data from CSP. It can firstly manage the records of duplicated data holders it removing the duplication record of this user. The rest of the remaining records in CSP will not delete the stored encrypted data. The block data access from the holder that requests data deletion.

Data Owner Management:

The real data owner uploads the data later than the data holder. The CSP can manage to save the data encrypted by the data owner. In the cloud mechanism, when the owner generated DEK and after, AP supports re-encryption at CSP for eligible data holders.

Encrypted Data Update:

The data owner requirements are to update data new encrypted raw data is provided and replace old storage for the reason of security. CSP is issuing the new re-encrypted to all the data holders with the support of AP.

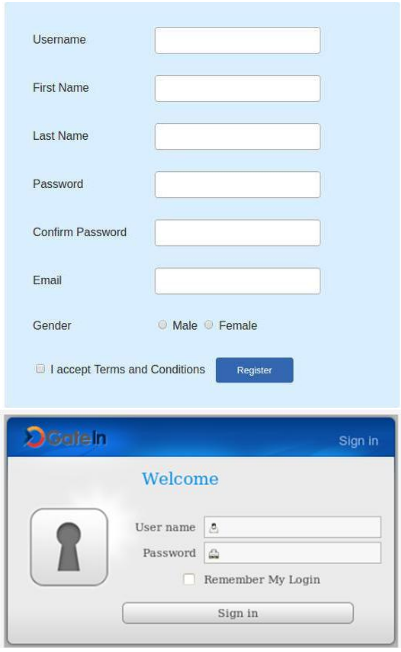

Fig.4 Registration and Login System

V. Conclusion

The final results are to encrypt the data with deduplication and are achieving a successful cloud storage service, for big data storage. In this paper, proposed a scheme to manage the encrypted big data in the cloud with deduplication based on RSA and Kerberos Authentication Protocol. This scheme can flexibly support data update and share with deduplication even when the data holders are offline. Only authorized data holders can able to use the symmetric keys used for data decryption by encrypted data can securely get accessed. The performance analysis and graphs showed that their scheme is secure under the described security model for big data deduplication. In fig 5 the results of the computer simulations further showed in the variation of the graph. Future work is to include optimizing there design and implementation for big data storage in the cloud. To ensure that CSP behaves as expected in deduplication management Studying verifiable computation. The part of cloud computing has brought many researchers from different fields; yet, much effort remains to reach use and the broad acceptance of cloud computing technology. In the future scope propose a big data deduplication. For the future environment, the system can focus on personalizing search on user feedback sessions as well as recommendation base on user point of interest with database security is the interesting part of the system.

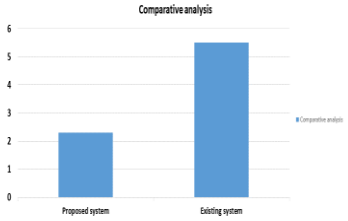

Fig: 5 Comparison graph for proposed system and existing system

References

[1] Maneesha Sharma, Himani Bansal and Amit Kumar Sharma, " Cloud Computing: Different Approach & Security Challenge", IJSCE, Volume-2, Issue-1, March 2012.

[2] Kangchan Lee, "Security Threats in Cloud Computing Environments", International Journal of Security and Its Applications, Vol. 6, No. 4, October 2012.

[3] Sashank Dara, "Cryptography Challenges for Computational Privacy in Public Clouds", International Journal of Security and Its Applications,

[4] David Pointcheval, "Asymmetric Cryptography and Practical Security", International Journal of Security and Its Applications, Volume 4,2002.

[5] Yogesh Kumar, Rajiv Munjal, and Harsh Sharma, "Comparison of Symmetric and Asymmetric Cryptography with Existing Vulnerabilities and Counter-measures", International Journal of Computer Science and Management Studies, Vol. 11, Issue 03, Oct 2011.

[6] Jan Stanek, Alessandro Sorniottiy, Elli Androulakiy, and Lukas Kencl, "A Secure Data De-duplication Scheme for Cloud Storage", IBM Research, Zurich, May 1994.

[7] Fatema Rashid, Ali Miri, Isaac Woungang., “Secure Enterprise Data Deduplication in the Cloud” IEEE Sixth International Conference on Cloud Computing, 367-374, 2013.

[8] Pasquale Puzio, Efimova, Melek O nen, Sergio Loureiro., “ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage” IEEE CloudCom, 2013.

[9] T. T. Wu, W. C. Dou, C. H. Hu, and J. J. Chen, “Service mining for trusted service composition in the cross-cloud environment,” IEEE Systems Syst. J., vol. PP, no. 99, pp. 1–12, 2014, doi:10.1109/ JSYST.2014.2361841.

[10] T. Y.Wu, J. S. Pan, and C. F. Lin Improving accessing efficiency of cloud storage using deduplication and feedback schemes, IEEESyst.J.,vol.8.

[11] C. Fan, S. Y. Huang, and. C. Hsu, Hybrid data deduplication in a cloud environment, in Proc. Int. Conf. Inf. Secure. Intell. Control,2012, pp. 174177.

[12] Ms.P.Aiswariya, Pondicherry, has completed in B.Sc Computer Science from Rajeswari Arts & Science College for Women, Puducherry (2007-2010). M.Sc Computer Science from KanchiMamunivar Center for Post Graduate Studies (Autonomous), Puducherry. (2010-2012). Now she has completed an M.phil degree from TheivanaiAmmal College for Women (Autonomous), Villupuram and working as Lecturer in SaradaMahavidyalayam Arts and Science College for Women, Ulundurpet. Her research interest is Image Processing and program analysis.