Data Leakage Optimization in Multi-cloud Storage Services

Krishna Vitthal Nil
PG Student
SSVPS's B.S Deore College of Engineering
Dhule, 424005, India

ABSTRACT
Now a days the use of cloud computing has increased speedily in many organizations. Cloud computing provides many advantages in terms of easy accessibly of data and also low cost. Data security is a major factor in cloud computing, as cloud users often store sensitive information in provided cloud storage providers. Recently the use of a single cloud is reduced because, the risks of service availability failure and the possibility of malicious insiders in the single cloud. So, many organizations preferred storing data on multiple clouds. Multi-Cloud Storage suggests to use the different distributed storage services. It is distributing information over different cloud storage providers and also it automatically provides a certain degree of data leakage control. But due to a large amount of data and unstructured distribution of information chunks can lead to high information disclosure even though using multiple clouds. It studies about the important data leakage problem caused by unstructured and less secure data distribution in mutlicloud storage services. To optimize the data leakage, it utilizes StoreSim, which data leakage aware storage system in the mutlicloud. StoreSim achieves this goal by using novel algorithms, SHA-1 algorithm efficiently generated signatures for data chunks and this signature computes the data leakages. Next, AES Algorithm is an effective data encrypted algorithm which reduce the attackability of sensitive data and also minimal data leakages across multiple clouds. The final analysis of proposed system are, reduces the risk of wholesale data leakage and also control data leakages.

General Terms
Algorithm, cloud security, multi-cloud security

Keywords
Data leakage, Multi-cloud Storage, Cloud Storage Servers, Encryption, Decryption

1. INTRODUCTION
Now a days increasingly rapid intake of devices such as a tablet, mobile, laptop users requires ubiquitous and huge network storage to hold their ever-growing digital lives. To meet these demands, many cloud-based storage and file sharing services such as Amazon S3,Google drive ,Dropbox, have gained popularity due to the easy-to-use interface and low storage cost. However, these centralized cloud storage services are criticized for lay hold of the control of users' data, which allows storage providers to run systematic advertising and also in marketing. Also, the users information can be leaked by malicious insiders, backdoors, bribe and coercion.

Multicloud storage service is a multiple public cloud and storage services in single network architecture. Also multi cloud is a collection of servers that cloud user access over the internet. typically, each cloud manages by a cloud provider. It provide multiple cloud storage provider such as Amezon s3,Dropbox,Goggle drive etc. All CSP is being used for storing a large amount of user data. Multi-Cloud computing can increase storage space and improve information sharing and this viewpoint will be an incredible help to clients. Clients share their data in the cloud. In any cloud computing model, security is the most important factor due to the risk of their sensitive and privet user's information or data which is stored in a cloud.

In recent every Organization is a force to increase their data sharing systems. Most cloud services are not free and more expensive and also have different sizes [5]. For instance, the use of Single Cloud Storage has storage limitation which makes it disadvantageous in comparison to multi-cloud storage. Single cloud environment is used a single cloud service. Multicloud is more popular than a single cloud in cloud customer. The important advantage of multi-cloud storage is it will increase performance and higher security for the transmission of data. In the single cloud storage data remains on the centralized storage which can be easily accessed by the attacker. Many organization transfers their privet data to client, supplier as per priority. It has needed more security no single point of attack can leak their data.

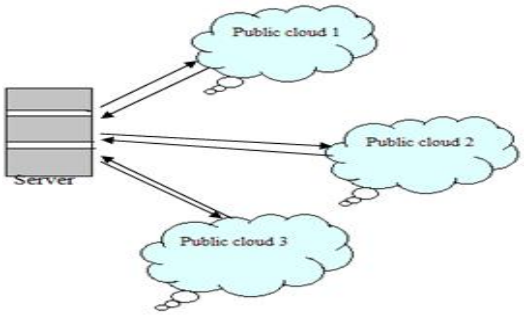

Fig1: Multicloud

As shown in fig1 organization of sharing of data to Multicloud storage. But, this situation is not simple many CSPs is synchronize local files to remote files in their cloud. Each local file is divided into small chunks and these chunks are uniquely identify by hash base signature generate algorithm such as SHA-1,MD5.for every updated local file and their generated signature will be upload to the cloud. This synchronization is based on the signature which comparing the two same files line by line if both files are the same then it uploaded to the cloud. However, this signature is to check duplicate data. Therefore,

we required more powerful techniques to find out to reduce the data leakage in the Multicloud storage system.

1.1 Contributions

In a Multicloud system, we can store a large amount of privet data and sensitive information without any proper optimization of the user information. We will be focused on reducing data leakages to every could storage provider and provide a better technique for distributed user privet information over different CSPs in a leakage aware manner. We will provide two main things first we provide a signature generation algorithm that checks the similar data chunks in the Multicloud system and efficiently synchronizes similar chunks together in a Multicloud. Next based on the security of data we will provide a cryptographic function which makes the data more secure.

Specifically, the contributions of the propose system in this paper is as follow:

- It presents our proposed system, in this proposed system we will formulate data leakage optimization.
- It propose an approximate algorithm, SHA-1 to generate signatures for data chunks and find similar data chunks. And we also used the AES algorithm for encrypting the data for more security purpose
- Finally, It sharing two text files to existing and our proposed system and we show the effectiveness and efficiency of our proposed scheme for reducing information leakage across multiple clouds. Furthermore, our analysis of the system attackability demonstrates that our proposed system makes attacks on data much more complex.

2. RELATED WORK

Yashaswi Singh, Farah Kandah, Weiyi Zhang[2011] In this paper, they proposed a secured cost-effective multicloud storage (SCMCS) in cloud computing, which seeks to provide each customer with a better cloud data storage decision, taking into consideration the user budget as well as providing with the best quality of service. The model has shown its ability of providing a customer with a secured storage under his affordable budget. It gives a superior choice to clients as indicated by their accessible spending plans. In this model, the client divides his data to several SPs which is based on their budget and availability of the market.

Nicolas Bonvin and Karl Aberer[2012] They presented Scalia, a system that continuously optimizes the placement of data stored at multiple cloud providers, based on their access statistics. Details of various layer approaches and our scalable technique for adaptive data placement.

Shilpashree Srinivasamurthy and David Q. Liu[2012] Every calculation is gone for unraveling a specific hazard. Anyway distributed computing is as yet battling in its earliest stages, with positive and negative remarks made on its conceivable usage for a vast estimated venture. Its security insufficiencies and advantages should be painstakingly weighed before settling on a choice to actualize it. However along with these advantages, storing a large amount of data including critical information on the cloud motivates highly skilled hackers thus creating a requirement for the security to be considered as one of the best issues while considering Cloud Computing. The cloud is just usable through the Internet so Internet dependability and accessibility is basic.

Zapater, Jose L. Ayala[2013] Reducing the energy consumption of enterprise servers in data centers continues to be a major challenge. This paper has displayed a test technique to investigate the impact of spillage temperature tradeoffs on the vitality effectiveness of big business servers. Method can be stretched out to genuine outstanding tasks at hand by utilizing a bigger arrangement of execution counters to portray runtime elements and applying measurable investigation to determine vitality execution. Their controller reduces energy consumption by up to 9% for a set of test workloads. It minimize the energy consumption.

Emil Stefanov and Elaine Shi[2013] They described a practical two-cloud Oblivious RAM protocol that reduces the client server Bandwidth cost to about 2:6 times that of simply reading or writing the block from non-oblivious cloud storage. They proposed a checksum encryption it allows our multi-cloud ORAM protocol to efficiently protect the privacy of the access pattern against one malicious cloud.and also it provide a full-edged implementation of 2-cloud ORAM system, and report results from a real-world deployment over Amazon EC2 and Microsoft Azure. In practice, each cloud can spread the data across multiple servers. For simplicity, they will first regard each cloud as a single logical entity; then in the full online version

Rohit Bhadauria, Rituparna Chaki, Nabendu Chaki and Sugata Sanyal[2016]In order to keep the Cloud secures, these security threats need to be controlled In addition information dwelling in the cloud is additionally inclined to various dangers and different issues like classification and honesty of information ought to be considered while purchasing stockpiling administrations from a cloud specialist organization. In this paper different security worries for Cloud processing condition from numerous point of view and the answers for anticipate them have been exhibited analyzed and ordered. In this it discus uncertain issues in cloud..Utilization dynamic groups conspire, where by predicates are analyzed over encoded information and multiparty registering

3. MULTICLOUD STORAGE SERVICES

In Multicloud storage services, we will first introduce a representation of the distribution of data and optimization of multicloud storage service. Then we will focus data synchronization technique among three distributed entities in multicloud storage services.

3.1 Distribution and Optimization

Cloud storage services such as Dropbox and Google Drive, Amazon S3 it providing the storage space to the user it may be free or low cost. While the users used these storage services, they also lose their control over the data. And lose their important data. Some other cloud storage services such as Wuala, Spider Oak employ client-side encryption to encrypt all the data before uploading the data i.e. it provides security to the user. But some time encryption key is exposed So, the user's entire data can be easily divulged [12]. The same situation can be somewhat suffering by using multiple cloud services so that no single CSP has access to the user's entire data. So, these works show that data distribution over multiple CSPs can avoid a single point of failure, thereby improving the service availability and fault-tolerance and reduce the attack ability. Besides, it also optimization on different metrics such as attack ability, uploading time of file[12,14] .

3.2 Data synchronization of cloud storage services

In a multi-cloud storage system, there is 3 entity which synchronizes users' data from the remote client to the cloud:

Client-In this pre-processing the users' data for optimization is charge in the client, such as chunking i.e., it is dividing files into single chunks of a maximum size data unit, deduplication it avoiding storing and re-transmitting the same content already available on the remote servers, bundling i.e., the transmission of multiple small files as a single object and encryption and decryption of data.

Metadata Server-It is a server that is used to store the metadata database about the information of files, which is managing the metadata operation of an entire file system, CSPs, and users, which usually are plane data representing the whole cloud file system.

Storage servers-This server store the raw data blocks which can be both structured data or unstructured data. named Times.[7,12,6]

4. ARCHITECTURE

In this, describe the architecture of proposed system. Then we explain the system terms of CSP models and metadata. Finally, we formulate the information leakage optimization problem in the multicloud system.

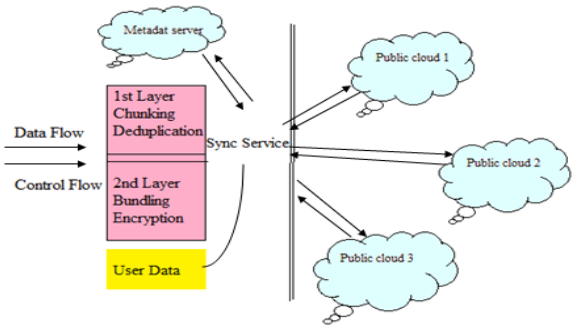

Fig2:Architecture of System[12]

Fig 2 shows the architecture of the proposed system. In the proposed system you can see that there is a boundary between the storage server and metadata server this boundary is a trust boundary We will assume that clients and metadata servers, which are inside the trust boundary, which are trustable by users. and the remote server is outside the trust boundary which is untrustworthy. E.g. the metadata server can be stored in servers which is a private database server while storage servers can be located in public cloud storage providers such as Amazon S3, Dropbox, and Google Drive. Accessed storage servers can be through standard APIs (Application Programming Interfaces).

As shown in Figure 2 Inside the trust boundary all control flows are available while data flows can cross the trust boundary. Optimize the data leakage, Design two components in the proposed system. The first part is the Leakage Measure layer which is used to checking the information leakage and further to generate the storage plan which maps data chunks to different clouds. The other part is the Cloud Manager layer that semantically provides cloud internal operation.

5. ALGORITHMS

In this two algorithms are used like AES algorithm and SHA-1algorithm which is increasing performance and also providing more security.

5.1 AES Algorithm

Advanced Encryption Standard(AES) algorithm is a symmetric key algorithm, The main purpose of this algorithm is to provide more security to user privet data which is store in multiple public cloud storage device and also it is one the most efficient symmetric algorithm.AES algorithm not only for security purposes but also for great speed. It provides different keys and the block size. In this, it encrypts the user plan text by using the AES algorithm and then ciphertext which we have got will again encrypt likewise there will be various rounds in the AES algorithm includes 10, 12 and 14 rounds with 128, 192, and 256 key bits. As there are various rounds in this algorithm encryption performs many time on plain text and this helps the data to have the security. The benefit provides strong security from attackers. In this paper It will use AES 128 for making encryption of user data. AES algorithm has four steps[16].

1. Substitute Bytes-Substitute byte means it replaces each byte of input data to another byte from the substitution table

2. Shift Rows-In this, the byte in each row of the state is shifted to the left. The number of places each byte is shifted dissimilar for each row

3.Mixing Columns-In the MixColumns step, each column of the state is multiplied by a fixed polynomial (x)

4.AddRoundKey-using ⊕ operation each byte of the state is combined with a byte of the round subkey.The algorithm describes as follows-

AES Algorithm(Encryption Algorithm)

```
1. Void Cipher(byte[]in ,byte[]out, word[]w){
2. byte[][]state=newbyte[4][Nb];
3. state=in;//actual component-wise copy
4. AddRoundKey(state,w,0.Nb-1);
5. For(int round=1;round<Nr;round++)
6.  SubByte(state);
7.  ShiftRows(state);
8.  MixColumn(state);
9.  AddRoundKey(state,w,Round*Nb,(round+1)*Nb-1);
10.  }
11.  SubBytes(state)
12.  ShiftRows(state)
13.  AddRoundKey(state,wNr*Nb,(Nr+1)*Nb-1);
14.  out=state
15.  End
```

AES Algorithm(Decryption Algorithm)

```
1.  Void InvCipher(byte[]in ,byte[]out,word[]w){
2.   byte[][]state=newbyte[4][Nb];
3.   state=in;//actual component-wise copy
4.   AddRoundKey(state,w,0.Nb-1);
5.   For(int round=1;round<Nr;round++)
6.   InvSubByte(state);
7.   InvShiftRows(state);
8.   AddRoundKey(state,w,Round*Nb,(round+1)*Nb-1);
9.   InvMixColumn(state);
10.  }
11.  InvShiftBytes(state)
12.  InvSubRows(state)
13.  AddRoundKey(state,w,0,Nb-1);
14.  out=state
15.  End
```

5.2 HAMC SHA-1 Algorithm

HAMC SHA-1 algorithm is a cryptographic algorithm that is commonly used in cryptographic applications and environments where the need for data integrity is high. It is to

identify checksum errors and data corruption. It is based on the signature, with the help of signature it synchronizes the chunk and finds out deduplication. It also used to index hash functions. It is a one-way cryptographic function it not encryption it cannot be back to the original text in the SHA-1 algorithm. The algorithm describes as follows-

SHA-1 Algorithm

1. function hmac (key, message)
2. if (length(key) > blocksize) then
3. key = hash(key) // **keys longer than blocksize are shortened**
4. end if
5. if (length(key) < blocksize) then
6. key = key || [0x00 * (blocksize - length(key))] // **keys shorter than blocksize are zero-padded (where || is concatenation)**
7. end if
8. o_key_pad = [0x5c * blocksize] || key // **Where blocksize is that of the underlying hash function**
9. i_key_pad = [0x36 * blocksize] || key // **Where || is exclusive or (XOR)**
10. return hash(o_key_pad || hash(i_key_pad || message)) // **Where || is concatenation**

end function.

6. IMPLEMENTATION

It have implemented the System prototype using Java and used Netbeans IDE 8.0.2 and it includes two layers. These both layers performing chunking, data deduplication, bundling, and encryption/decryption first Layer is performing Chunking and deduplication and the second Layer enables the system to communicate with multiple CSPs, it also encrypts and bundling the data. In the proposed system has common fixed-size chunking, the chunk is identified by SHA-1 signature, which is also used for data deduplication, for better network transmission it can be measured leakage optimization, encrypted before the chunk is synchronized. The system synchronizes new chunks which are identified by SHA-1 signatures between two copies. All data are store in a MySQL database. The Second Layer of the system provides uniform APIs such as initialize, connect, upload, download and delete, for different CSPs. With this abstraction, the user can move their data across different CSPs in a transparent way. It have implemented public storage clouds Amazon S3. Amazon S3 creating bucket as per user choice. All the communications between the system and public CSPs are done by using APIs supplied by those CSPs[17].

7. RESULTS

You can see the aim of result is how well the system reaches the goals and fills the requirements. It will be checking the performance and working of the system and find out the information leakage. Fig 3 shows the checking of leakages of information. If any leakages are available in the data then it will not unbundle and decrypt the information..

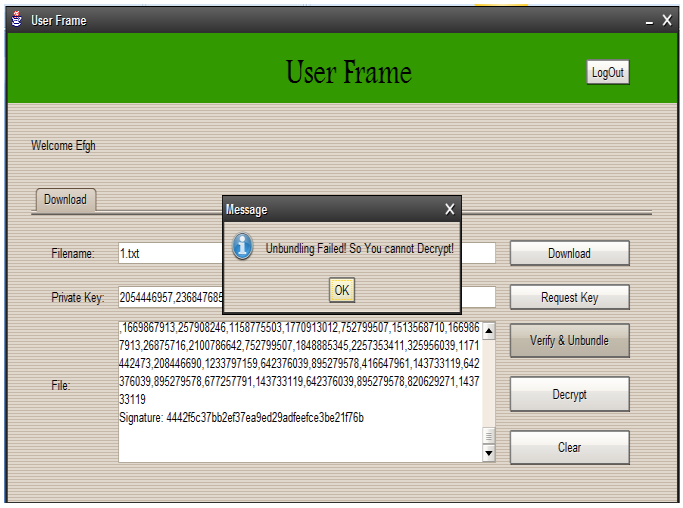

Fig3: Attackability

Performance analysis of time required for uploading file on Multicloud as shown in fig 4 required time of the file is less in our propose system.

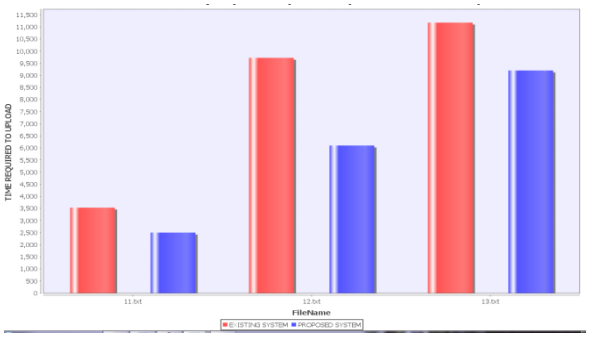

Fig4: Performance analysis

8. CONCLUSION

In the existing system distributing data chunks can leak user's data. To optimize the data leakage, we propose our system, an data leakage aware storage system in the multicloud. Our proposed system achieves this goal by using novel algorithms, SHA-1 and AES algorithm which place the data with minimal data leakage on the same cloud which is base on similarity. Overall evaluation based on our data, our proposed system is more effective and efficient and minimizing data leakage in multicloud storage system. Our proposed system can achieve near-optimal performance and reduce information leakage. Finally, we show the analysis of the algorithm, we further demonstrate that our system not only reduces the risk of data leakage but also makes attacks on data much more complex.

9. REFERENCES

[1] Benjamin Fabian, Tatiana Ermakova, Philipp Junghanns "Collaborative and securesharing of healthcare data in multi-clouds "Information Systems, Volume 48 IssuesC, 2015, pp 132-150

[2] Thanasis G. Papaioannou, Nicolas Bonvin and Karl Aberer"Scalia: An AdaptiveScheme for Efficient Multi-Cloud Storage", IEEE November 10-16, 2012.

[3] Emil Stefanov and Elaine Shi "Multi-Cloud Oblivious Storage", IEEE ACM978-1-4503-2477, November 48, 2013.

[4] Rohit Bhadauria, Rituparna Chaki, Nabendu Chaki and

Sugata Sanyal,"A Surveyon Security Issues in Cloud Computing", Journal of Network and Computer ApplicationsVolume 71, August 2016.

[5] Shilpashree Srinivasamurthy and David Q. Liu, "Survey on Cloud Computing Security",IEEE International Conference on Computing Sciences on 24 December 2012.

[6] Marina Zapater, Jose L. Ayala, Jose M. Moya, Kalyan Vaidyanathan"Leakageand Temperature Aware Server Control for Improving Energy Efficiency in Data Centers",IEEE Conference 06 May 2013.

[7] Jens-Matthias Bohli, Nils Gruschka, Meiko Jensen, Luigi Lo Iacono, And NinjaMarnau "Security And Privacy-Enhancing Multicloud Architectures", IEEE TransactionsOn Dependable And Secure Computing, Vol. 10, No. 4, July/August 2013.

[8] Shubhashis Sengupta, Vikrant Kaulgud, Vibhu Saujanya Sharma "Cloud ComputingSecurity - Trends and Research Directions", IEEE World Congress on Services4-9 July 2011.

[9] Ibrahim Abdullah Althamary, Talal Mousa Alkharobi , "Secure File Sharingin Multi-Cloud using Shamirs Secret Sharing Scheme", Transactions on Network andcommunications Vol 4 issue 6, 2016,pp53-67.

[10] Safaa Salam Hatem, Maged H.Wafy,Mahmoud M.EI-Khouly, "Malware Detectionin cloud Computing", International Journal of Advanced Science and ComputerScience Applications, Vol 5 No 2014.

[11] Danan Thilakanathan, Shiping Chen, Surya Nepal and Rafael A. Calvo,"SecureData Sharing in the Cloud"In Security, Privacy and Trust in Cloud Systems, SpringerBerlin Heidelberg, 2015, (pp. 45-72).

[12] Hao Zhuang, Member, IEEE, Rameez Rahman,Pan Hui, Member, IEEE, and Karl Aberer, Member, IEEE optimizing information leakage in multicloud storage services

[13] Balasaraswathi, V. R., & Manikandan, S. (2014) "TEnhanced security for multicloudstorage using cryptographic data splitting with dynamic approach , In AdvancedCommunication, International Conference on Control and Computing Technologies(ICACCCT), 2014 on (pp. 1190-1194) IEEE.

[14] Mazhar Ali, Revathi Dhamotharan, ErajKhan, Samee U. Khan, Athanasios V.Vasilakos, KeqinLi, Albert. Y. Zomaya "SeDaSC: Secure Data Sharing in Clouds,Systems Journal, IEEE, volume: PP, Issue: 99, 2015, pp 1-10.

[15] Yashaswi Singh, Farah Kandah, Weiyi Zhang "A Secured Cost-effective Multi-Cloud Storage in Cloud Computing",IEEE INFOCOM on Cloud Computing in 2011.

[16] Bih-Hwang Lee, Ervin Kusuma Dewi, Muhammad Farid Wajdi Data security in cloud computing using AES under HEROKU cloud.

[17] Mohammed A. AlZain,Eric Pardede , Ben Soh , James A. Thom Cloud Computing Security: From Single to Multi-Clouds.