International Journal of Computer Science and Engineering (IJCSE)
ISSN 2278-9960
Vol. 2, Issue 3, July 2013, 13-20
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

SECURED DATA STORAGE AND RETRIEVAL IN MULTICLOUDING USING SHAMIR'S SECRET SHARING ALGORITHM

JAYASHRI V. BHARAMBE & RICHA K. MAKHIJANI

Department of Computer Science and Engineering, S.S.G.B.C.O.E.T., Bhusawal, Maharashtra, India

ABSTRACT

The use of cloud computing has increased rapidly in many organizations. The end of this decade is marked by a paradigm shift of the industrial information technology towards a pay-per-use service business model known as cloud computing. But ensuring security is considered to be one of the most critical aspects in a cloud computing environment due to the sensitive and important information stored in the cloud for users. Cloud providers should address privacy and security issues as a matter of high and urgent priority. In this paper, we propose a multi-cloud secret sharp model in cloud computing which holds an economical distribution and retrieval of data among the available SPs in the market to provide customers with data availability as well as secure data storage.

General Terms: Security

KEYWORDS: Cloud Computing, Cloud Security, Data Integrity, Service Availability, Multi Cloud

INTRODUCTION

The use of cloud computing has increased rapidly in many organizations. One of the prominent services offered in cloud computing is the cloud data storage, in which subscribers do not have to store their data on their own servers, where instead their data will be stored on the cloud service provider's servers. In cloud computing, subscribers have to pay the service providers for this storage service.

This service does not only provides flexibility and scalability for the data storage, it also provide customers with the benefit of paying only for the amount of data they need to store for a particular period of time, without any concerns for efficient storage mechanisms and maintainability issues with large amounts of data storage. In addition to these benefits, customers can easily access their data from any geographical region where the Cloud Service Provider's network or Internet can be accessed.

Security is considered to be one of the most critical aspects in a cloud computing environment due to the sensitive and important information stored in the cloud for users [1]. Users are wondering about attacks on the integrity and the availability of their data in the cloud from malicious insiders and outsiders, and from any collateral damage of cloud services [1]. These issues are extremely significant but there is still much room for security research in cloud computing.

RELATED WORK

HAIL (High Availability and Integrity Layer) [6] is example of a protocol that controls multiple clouds. HAIL is a distributed cryptographic system that permits a set of servers to ensure that the client's stored data is retrievable and integral. HAIL provides a software layer to address availability and integrity of the stored data in an intercloud [6]. HAIL does not guarantee data confidentiality, it needs code execution in their servers, and it does not deal with multiple versions of data.

None of these limitations are found in DepSky [4], whereas the RACS system differs from the DepSky system in that it deals with “economic failures” and vendor lock-in and does not address the issue of cloud storage security problems [5]. In addition, it also does not provide any mechanism to ensure data confidentiality or to provide updates of the stored data.

Bessani et al. [4] use Byzantine fault-tolerant replication to store data on several cloud servers, so if one of the cloud providers is damaged, they are still able to retrieve data correctly. Data encryption is considered the solution by Bessani et al. [4] to address the problem of the loss of privacy.

They argue that to protect the stored data from a malicious insider, users should encrypt data before it is stored in the cloud. As the data will be accessed by distributed applications, the DepSky system stores the cryptographic keys in the cloud by using the secret sharing algorithm to hide the value of the keys from a malicious insider.

But Rocha and Correia [3] present the limitations of this work which occurs due to the fact that DepSky is only a storage service like Amazon S3, and does not offer the IaaS cloud model. On the other hand, this system provides a secure storage cloud, but does not provide security of data in the IaaS cloud model. This is because it uses data encryption and stores the encrypted key in the clouds by using a secret sharing technique, which is inappropriate for the IaaS cloud model [3].

Cachin et al. [2] identify two layers in the multicloud environment: the bottom layer is the inner-cloud, while the second layer is the inter-cloud. In the intercloud, the Byzantine fault tolerance finds its place.

They first summarize the previous Byzantine protocols over the last three decades. Cachin et al. [2] present a design for intercloud storage (ICStore), which is a step closer than RACS and HAIL as a dependable service in multiple clouds. Cachin et al. [2] develop theories and protocols to address the CIRC attributes (confidentiality, integrity, reliability and consistency) of the data stored in clouds.

PROBLEM STATEMENT

Customers’ stored data at cloud service providers is vulnerable to various threats. In our work, we consider four types of threat models. First is the single point of failure [7], which will affect the data availability that could occur if a server at the cloud service provider failed or crashed, which makes it harder for the customer to retrieve his stored data from the server. Availability of data is also an important issue which could be affected, if the cloud service provider (CSP) runs out of service.

Our second threat discussed in this paper is data integrity. Integrity is a degree confidence that the data in the cloud is what is supposed to be there, and is protected against accidental or intentional alteration without authorization. Such worries are no more beneficial issues; therefore, a cloud service customer can not entirely rely upon a single cloud service provider to ensure the storage of his vital data [1].

To illustrate this threat we use an example in Fig. 1. Let us assume that three customers (C1, C2 and C3) stored their data on three different service providers (CSP1, CSP2 and CSP3) respectively.

Each customer can retrieve his own data from the cloud service provider who it has a contract with. If a failure occur at CSP1, due to internal problem with the server or some issues with the cloud service provider, all C1’s data which was stored on CSP1’s servers will be lost and cannot be retrieved. One solution for this threat is that, the user will seek to store his data at multiple service providers to ensure better availability and data integrity of his data.

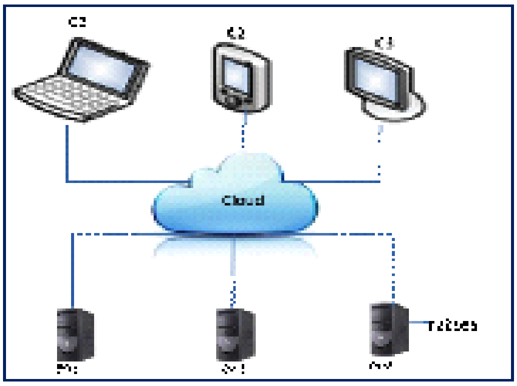


Figure 1: CSP Failure

Our third threat discussed in this paper is data intrusion. Another security risk that may occur with a cloud provider, such as the Amazon cloud service, is a hacked password or data intrusion [1]. If the hacker hacked one cloud provider’s password or even two cloud provider’s passwords, they still need to hack the third cloud provider (in our case) to know the secret which is the worst and the hardest case scenario. Hence, replicating data into multi-clouds by using a multi share technique may reduce the risk of data intrusion.

Our fourth threat discussed in this paper is malicious insiders. Administrators manage the infrastructure and as they have remote access to servers, if the administrator is a malicious insider, then he can gain access to the user’s data [1].

SHAMIRS SECRET SHARING SCHEME

Shamir's secret sharing scheme is based on polynomial evaluations. The central party is the dealer that performs share computation operations on input secrets and distributes the resulting shares to other parties. When the secret has to be reconstructed, the parties give their shares to the dealer, which can then combine the shares and retrieve the secret.

An intruder needs to retrieve at least three values to be able to find out the real value that wants to hide from the intruder. This depends on Shamir’s secret sharing algorithm with a polynomial function technique which claims that even with full knowledge of (k – 1) clouds, the service provider will not have any knowledge of vs (vs is the secret value). In other words, hackers need to retrieve all the information from the cloud providers to know the real value of the data in the cloud. Therefore, if the attacker hacked one cloud provider’s password or even two cloud provider’s passwords, they still need to hack the third cloud provider (in the case where k = 3) to know the secret which is the worst case scenario. Hence, replicating data into multi-clouds by using a multi-share technique may reduce the risk of data intrusion and increase data integrity [1].

SECURED DATA STORAGE AND RETRIEVAL

Proposed Model

We propose a multi-cloud secret sharp model in cloud computing which holds an economical distribution and retrieval of data key among the available CSPs in the market to provide customers with data availability as well as secure data storage.

We use an example in Fig. 2 to illustrate our proposed model. In this example we assume that we have five cloud service providers (CSP1, CSP2… CSP5). Let us assume that a customer (C1) has encrypt his own data then the available encoded key is split and he wish to store on some CSP’s servers into five pieces of encoded data key. A customer required retrieving at least three pieces of encoded data key from different SPs to reconstruct his own data to get the full information, where in our example, three CSPs will participate in the encoded data key retrieval (CSP1, CSP3, and CSP5).

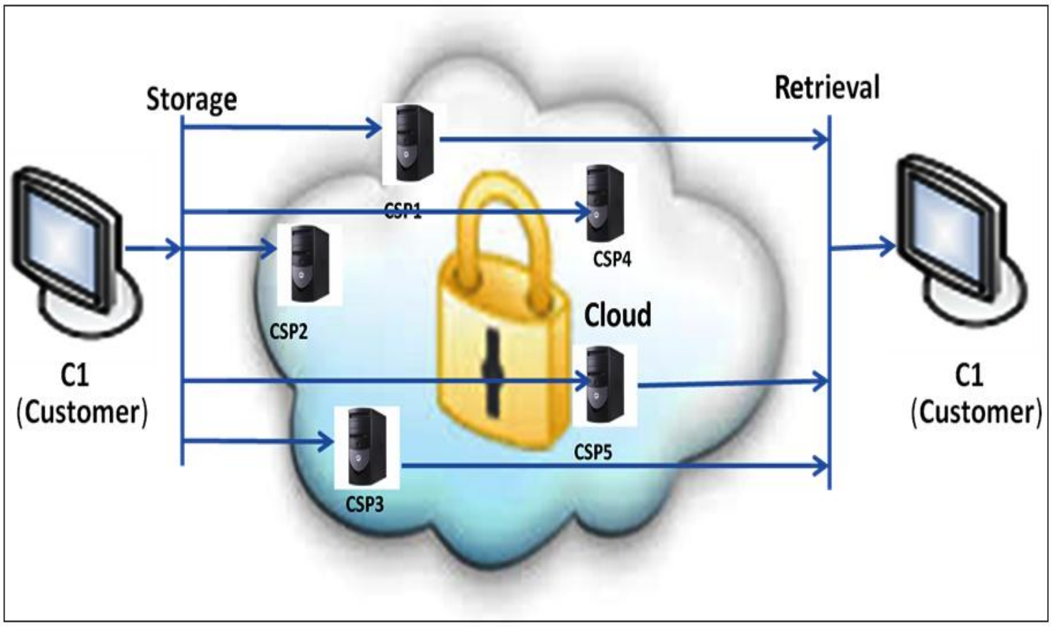


Figure 2: Data Storage and Retrieval

Data Security Risks

Data Integrity

One of the most important issues related to cloud security risks is data integrity. The data stored in the cloud may suffer from damage during transition operations from or to the cloud storage provider [14]. Integrity is a degree confidence that the data in the cloud is what is supposed to be there, and is protected against accidental or intentional alteration without authorization.

Data Intrusion

According to Garfinkel [8], another security risk that may occur with a cloud provider, such as the Amazon cloud service, is a hacked password or data intrusion.

Service Availability

Lastly, another most problematic issue is service availability, as several companies using cloud computing have already experienced downtime (Amazon servers subject to what appeared to be a denial of service attack). Other things to keep in mind are contract policies between clients and vendors, so that data belongs only to the client at all times, preventing third parties to be involved at any point. Also, authentication should be backed by several methods like password plus flash card, or password plus finger print, or some combination of external hardware and password. Cloud technologies can increase availability through widespread internet-enabled access, but the client is dependent on the timely and robust provision of resources. Availability is supported by capacity building and good architecture by the provider, as well as well-defined contracts and terms of agreement.

Implementation and Evaluation

This section explains the experimentation to examine the multi-cloud secret sharp model. The experiment provides evaluation of Shamir’s Secret Sharing algorithm. We implemented the generation of secrets using Shamir’s Secret Sharing algorithm. Our experimentation is conducted using Java to simulate encoded data key storing in multi-cloud providers, encoded data key retrieval from different cloud providers on a system with an Intel Core 2 processor running at 2GHz, 2GB of RAM.

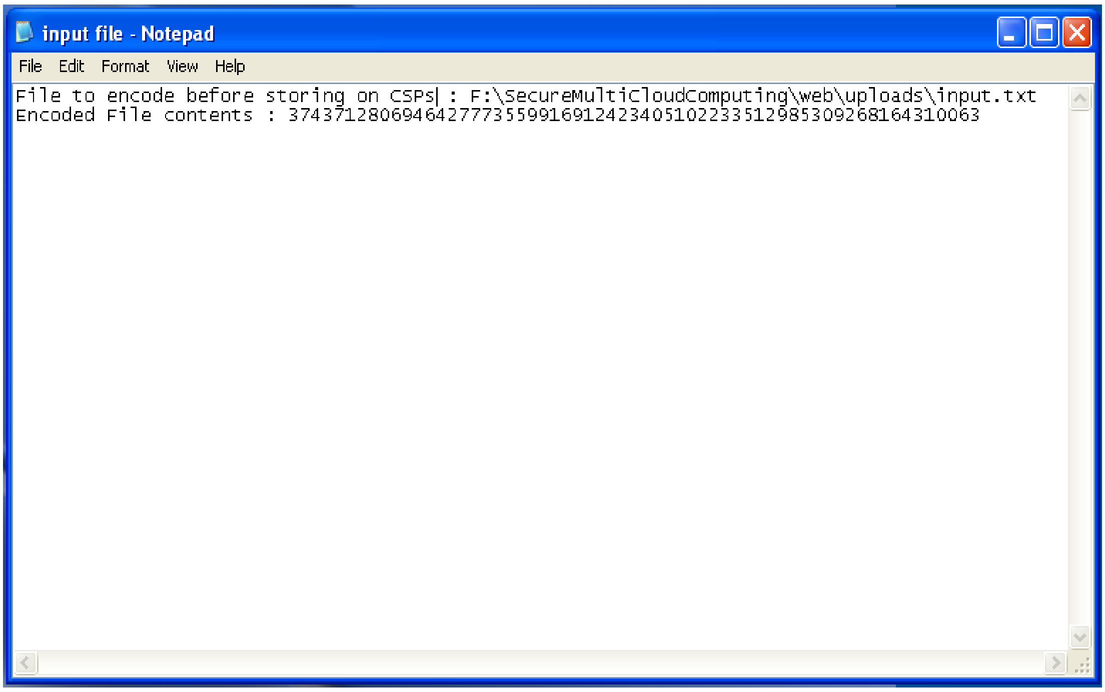
Encoded Data Key Storing Procedure: Encoded data key storing in multi-cloud secret sharp model involves data distribution from the data source to different cloud providers. This is done after executing the polynomial functions on the data.

To analyze the effect of a number of shares in our model, we perform experimentation for data storing in multi-cloud secret sharp using static data size (176 KB).

First we encode data input using RSA algorithm (See Figure 3). Encoded data key storing in our experimentation involves distribution of encoded data key from the data source to six different cloud providers which are available in markets like Google Drive, Drop Box, and Sky Drive.

This is done after executing the polynomial functions on the data. Figure 4 shows that secrets of data input in the form of polynomial function values. This solves the problem of malicious insiders or outsiders because if the attacker hacked one cloud provider’s data or even two cloud provider’s data, they still need to hack the third cloud provider (in the case where k = 3) to know the secret which is the worst case scenario.

Increasing the number of shares will improve the security level of the hidden value of the data from un-trusted cloud provider due to the fact that the CSPs need more numbers of *k* to know the details of the data. If the number of shares decreases to fewer than 3, then it might not be very effective for privacy purposes.

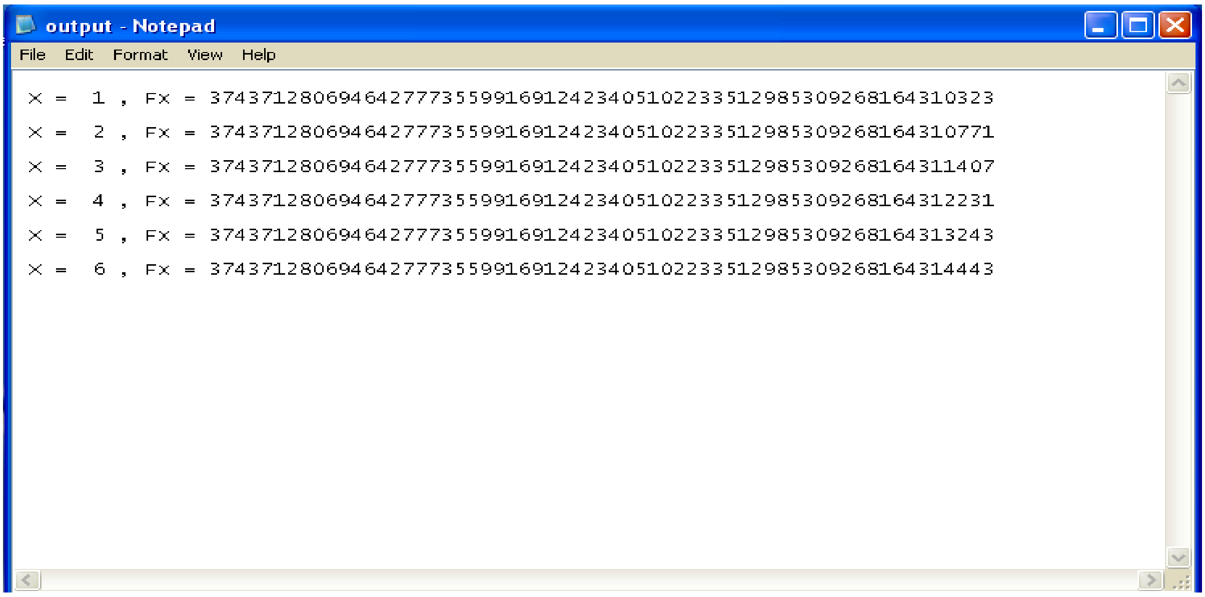


```
File to encode before storing on CSPs : F:\SecureMultiCloudComputing\web\uploads\input.txt
Encoded File contents : 3743712806946427773559916912423405102233512985309268164310063
```

Figure 3: Encoded File Contents

Encoded Data Key Retrieval Procedure: The encoded data key retrieval process in the multi-cloud secret sharp model starts from six different cloud providers by retrieving any three numbers of shares as show in figure 3, one for each CSP, after retrieving any three numbers of shares; reconstruct these shares using Lagrange functions. Decode that file and download original file.

We argue that increasing the number of shares will also increase the security level of data because the malicious insiders in CSPs will need to retrieve more values from more shares in order to be able to determine the hidden information in CSPs.



```
X =  1 , Fx = 3743712806946427773559916912423405102233512985309268164310323
X =  2 , Fx = 3743712806946427773559916912423405102233512985309268164310771
X =  3 , Fx = 3743712806946427773559916912423405102233512985309268164311407
X =  4 , Fx = 3743712806946427773559916912423405102233512985309268164312231
X =  5 , Fx = 3743712806946427773559916912423405102233512985309268164313243
X =  6 , Fx = 3743712806946427773559916912423405102233512985309268164314443
```

Figure 4: Polynomial Values after Secret Creation of Encoded Input Data

Table 1: Comparison between Amezon, Depsky and Multi-Cloud Secret Sharp Models

Security Issues → ↓ Cloud Security Architectures		Data Integrity	Data Intrusion	Service Availability	Malicious Insiders
Amazon		If data hacked?	If password hacked?	If system down?	If malicious insider is present?
Data Status	**Safe**				
	Lost	√	√	√	√
Depsky		If data hacked from one CSP?	If password hacked from one CSP?	If one cloud down?	If malicious insider is present?
Data Status	**Safe**	√	√	√	
	Lost				√
Multi-Cloud Secret Sharp		If data hacked from one CSP?	If password hacked from one CSP?	If one cloud down?	If malicious insider is present?
Data Status	**Safe**	√	√	√	√
	Lost				

As a result of the three above arguments for data integrity, data intrusion, and service availability, our newly proposed Multi-Coud Secret Sharp model is better in addressing the three security factors than in Amazon and Depsky cloud service and more secured in protecting user's data from untrusted cloud service providers and from the malicious insider especially when Amazon cloud service ask the users to encrypt their data before storing it in their instances, whereas, Multi-Cloud Secret Sharp take responsibility of this task. Table 1 summarizes the differences between Amazon, Depsky and our proposed Multi-Cloud Secret Sharp model in terms of the four security factors that may occur in cloud computing environment.

CONCLUSIONS

It is clear that although the use of cloud computing has rapidly increased; cloud computing security is still considered the major issue in the cloud computing environment. Customers do not want to lose their private information as a result of malicious insiders and outsiders, and from any collateral damage of cloud services. In addition, the loss of service availability has caused many problems for a large number of customers recently. The purpose of this work is to propose a new model called Multi-Cloud Secret Sharp which use Shamir's secret sharing algorithm with multiclouds providers instead of single cloud.

At this stage we compared our proposed multiclouds model with Amazon, Depsky cloud service model. As a result of this comparison, it has shown that the multi-clouds model is superior to single cloud model in addressing the security issues in cloud computing.

REFERENCES

1. Mohammed A. AlZain , Eric Pardede , Ben Soh , James A. Thom, "Cloud Computing Security: From Single to Multi-Clouds", 2012 45th Hawaii International Conference on System Sciences, 2012, pp. 5490-5499.

2. C. Cachin, R. Haas and M. Vukolic, "Dependable storage in the Intercloud", Research Report RZ, 3783, 2010.

3. F. Rocha and M. Correia, "Lucy in the Sky without Diamonds: Stealing Confidential Data in the Cloud", Proc. 1stIntl. Workshop of Dependability of Clouds, Data Centers and Virtual Computing Environments, 2011, pp. 1-6.

4. A. Bessani, M. Correia, B. Quaresma, F. André and P. Sousa, "DepSky: dependable and secure storage in a cloud-of-clouds", EuroSys'11:Proc. 6thConf. On Computer systems, 2011, pp. 31-46.

5. H. Abu-Libdeh, L. Princehouse and H. Weatherspoon, "RACS: a case for cloud storage diversity", SoCC'10:Proc. 1st ACM symposium on Cloud computing, 2010, pp. 229-240.

6. K.D. Bowers, A. Juels and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage", CCS'09: Proc. 16th ACM Conf. on Computer and communications security, 2009, pp. 187-198.

7. C. Wang, Q. Wang, K. Ren and W. Lou, "Ensuring data storage security in cloud computing", ARTCOM'10: Proc. Intl. Conf. on Advances in Recent Technologies in Communication and Computing, 2010, pp. 1-9.

8. S. L. Garfinkel, An evaluation of amazon's grid computing services: EC2, S3, and SQS, Citeseer, 2007, pp. 1-15.