A Secured Cost Effective Multi-Cloud Storage in Cloud Computing

Prof.V.N.Dhawas,Pranali Juikar,Neha Patekar,Neha Lendghar,Sushant Vartak

(Department Of Computer Engineering ,Sinhgad Institute of Technology,Lonavla)

Abstract— The term "Cloud" is analogical to "Internet".Cloud computing is Internet based computing where virtual shared servers provide software,infrastructure,platform,devices and other resources and hosting to customers on a pay-as-you-use basis.Cloud data storage redefines the security issues targeted on customer's outsourced data.From a customer's point of view relying upon a solo Service Provider for his outsourced data is not very promising.In addition,providing better privacy as well as ensuring data availability,can be achieve by dividing the user's data block into data pieces and distributing them among the available Service Providers in such a way that no less than a threshold number of Service Providers can take part in successful retrieval of the whole data block.In this paper,we propose a secured cost-effectivemulti-cloud storage(SCMCS)model in cloud computing which holds an economical distribution of data among the available Service Providers in the market,to provide customers with data availability as well as secure storage.

Index Terms— Cloud computing, security, storage, cost-effective, cloud service provider, customer.

1 INTRODUCTION

Scientific and commercial applications are leveraging the power of distributed computing and storage resources . These resources are available either as part of general purpose computing infrastructure such as Clusters and Grids, or through commercially hosted services such as Clouds [1]. Clouds have been defined to be a type of parallel and distributed system consisting of inter-connected and virtualized computers. These computers can be dynamically provisioned as per user's requirements [2]. Thus, to achieve better performance and scalability, applications could be managed using commercial services provided by Clouds, such as Amazon AWS, Google AppEngine, and Microsoft Azure. Some of these cloud service providers also have data distribution services, such as Amazon Cloud- Front. However, the cost of computing, storage and communication over these resources could be very high for compute-intensive and data-intensive applications.

The end of this decade is marked by a paradigm shift of the industrial information technology towards a subscription based or pay-per-use service business model known as cloud computing. This paradigm provides users with a long list of advantages, such as provision computing capabilities; broad, heterogeneous network access; resource pooling and rapid elasticity with measured services [3]. A huge amount of data being retrieved from geographically distributed data sources, and non-localized data-handling requirements, creates such a change in technological as well as business model. 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. An example of the cloud computing is shown in Fig. 1. Along with these unprecedented advantages, cloud data storage also redefines the security issues targeted on customer's outsourced data (data that is not stored/retrieved from the costumers own servers).

Since cloud service providers (SP) are separate market entities, data integrity and privacy are the most critical issues that need to be addressed in cloud computing. Even though the cloud service providers have standard regulations and powerful infrastructure to ensure customer's data privacy and provide a better availability, the reports of privacy breach and service outage have been apparent in last few years [4] [5]. Also the political influence might become an issue with the availability of services [6]. In this work we observed that, from a customer's point of view, relying upon a solo SP for his outsourced data is not very promising. In addition, providing better privacy as well as ensure data

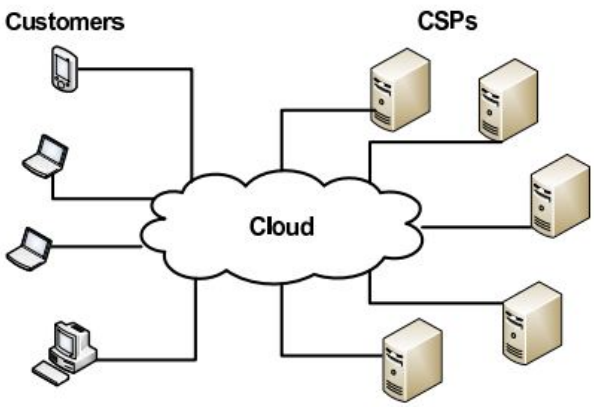

Fig1. Cloud computing Architecture example

availability, can be achieved by dividing the user's data block into data pieces and distributing them among the available SPs in such a way that no less than a threshold number of SPs can take part in successful retrieval of the whole data block To address these issues in this paper, we proposed an economical distribution of data among the available SPs in the market, to provide customers with data availability as well as secure storage. In our model, the customer divides his data among several SPs available in the market, based on his available budget. Also we provide a decision for the customer, to which SPs he must chose to access data, with respect to data access quality of service offered by the SPs at the location of data retrieval. This not only rules out the possibility of a SP misusing the customers' data, breaching the privacy of data, but can easily ensure the data availability with a better quality of service.

Our proposed approach will provide the cloud computing users a decision model, that provides a better security by distributing the data over multiple cloud service providers in such a way that, none of the SP can successfully retrieve meaningful information from the data pieces allocated at their servers. Also, in addition, we provide the user with better assurance of availability of data, by maintaining redundancy in data distribution. In this case, if a service provider suffers service outage [4] or goes bankrupt, the user still can access his data by retrieving it from other service providers.

2 EXISTING SYSTEM

In the cloud computing, the data is stored on an autonomous business party that provides data storage as a subscription service. The users have to trust the cloud service provider (SP) with security of their data. Obtaining information from a third party is much easier than from the creator himself.

Following the pattern of paradigm shift, the security policies also evolved from the conventional cryptographic schemes applied in centralized and distributed data storage, for enabling the data privacy. Many of the cryptographic approaches have been proposed for hiding the data from the storage provider and hence preserving data privacy.

The user's identity is also detached from the data, and claim to provide public auditing of data. These approaches concentrate on one single cloud service provider that can easily become a bottleneck for such services. The sole cryptographic measures are insufficient for ensuring data privacy in cloud computing. In cloud storage needs a hybrid model of privacy enforcement, distributed computing and complex trust ecosystems.

To provide users with better and fair chances to avail efficient security services for their cloud storage at affordable costs, our model distributes the data pieces among more than one service providers, in such a way that no one of the SPs can retrieve any meaningful information from the pieces of data stored on its servers, without getting some more pieces of data from other service providers. Therefore, the conventional single service provider based cryptographic techniques does not seem too much promising.

3 SYSTEM AND THREAT MODEL

First in this section, we will describe our system model and the threat model. Then, formally we will describe our problem statement we are going to study in this paper. Note that, in this work the terms cloud service provider and service providers are interchangeable, the terms cloud storage and cloud data storage are interchangeable, also the terms user and customer are interchangeable.

3.1 System Overview

We consider the storage services for cloud data storage between two entities, cloud users (U) and cloud service - providers (SP). The cloud storage service is generally priced on two factors, how much data is to be stored on the cloud servers and for how long the data is to be stored. In our model, we assume that all the data is to be stored for same period of time. We consider p number of cloud service providers (SP), each available cloud service provider is associated with a QoS factor, along with its cost of providing storage service per unit of stored data (C). Every SP has a different level of quality of

service (QoS) offered as well as a different cost associated with it. Hence, the cloud user can store his data on more than one SPs according to the required level of security and their affordable budgets.

3.2 Threat Model

Customers' stored data at cloud service providers is vulnerable to various threats. In our work, we consider two 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 costumer to retrieve his stored data from the server.

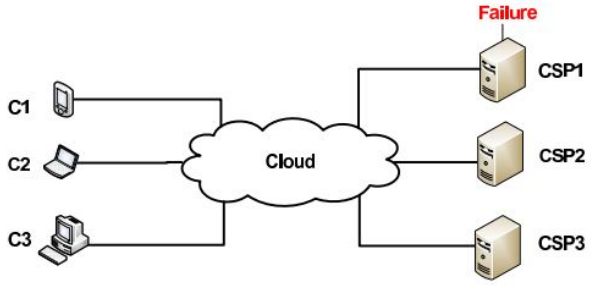

Fig2. CSP failure

Availability of data is also an important issue which could be affected, if the cloud service provider (SP) runs out of business. Such worries are no more hypothetical issues; therefore, a cloud service customer can not entirely rely upon a solo cloud Service Provider to ensure the storage of his vital data.

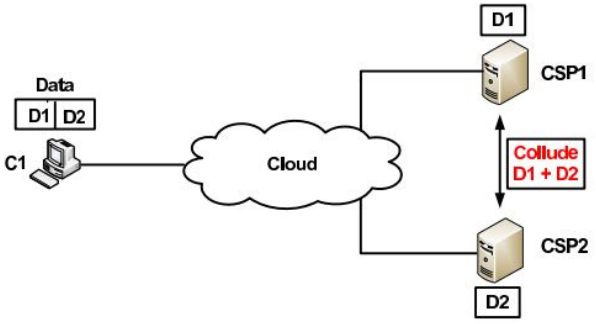

Fig3. Colluding Service provider

To illustrate this threat we use an example in Fig. 2. 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 of his data. Our second threat discussed in this paper is the colluding service providers [8], in which the cloud service providers might collude together to reconstruct and access the user stored data. We illustrate the colluding service providers' threat in Fig. 3. (SCMCS) seeks a distribution of customer's data pieces among the available SPs in such a way that, at least q number of SPs must take part in data retrieval, while minimizing the total cost of storing the data on SPs as well as maximizing the quality of service provided by the SPs.

4 PROPOSED SYSTEM

In this work we observed that, from a customer's point of view, relying upon a solo SP for his outsourced data is not very promising. In addition, providing better privacy as well as ensure data availability, can be achieved by dividing the user's data block into data pieces and distributing them among the available SPs in such a way that no less than a threshold number of SPs can take part in successful retrieval of the whole data block.

We proposed an economical distribution of data among the available Service Provider to provide customers with data availability as well as secure storage. In our model, the customer divides his data among several SPs, based on his available budget. Also we provide a decision for the customer, to which SPs he must chose to access data, with respect to data access quality of service offered by the SPs at the location of data retrieval. This not only rules out the possibility of a SP misusing the customers' data, breaching the privacy of data, but can easily ensure the data availability with a better quality of service.

This approach will provide the cloud computing users a decision model, that provides a better security by distributing the data over multiple cloud service providers in such a way that, none of the SP can successfully retrieve meaningful information from the data pieces allocated at their servers. Also, in addition, we provide the user with better assurance of availability of data, by maintaining redundancy in data distribution. In this case, if a service provider suffers service

outage or goes bankrupt, the user still can access his data by retrieving it from other service providers.

5 LINEAR PROGRAMMING

In this section we describe the setup for the linear programming assignment problem (LP-Assignment) that describes our proposed model. Each cloud customer is provided with p cloud service providers, where each of them offers a QoS level for storage services and required a cost C be paid by the customer per storage unit of data.

TABLE I
NOTATION AND DESCRIPTION

Notations	Descriptions
N	Total number of data units
k	minimum number of data units required for data retrieval
a	Total number of available cloud service providers
b	Minimum number of service providers required for data retrieval
I	i = 1,2,.., a
SP_i	Cloud service provider
Q_i	Quality of Service factor for each service provider
Q_{net}	The QoS achieved at the time of retrieval
c_i	Cost of storing per unit data for i^{th} service provider
C	Total cost of storing the distributed customer data on p service providers
d_i	Number of data units assigned to i^{th} service provider
J	j = 1, 2,.., d_i
$y_{i,j}$	j^{th} data unit on i^{th} service provider

Previous studies in [9] [10] proposed a dividing scheme for user's data in such a way that, the user will divide his data into N data pieces where at-least k data pieces out of N data pieces are required to recover any meaningful information of the data. In addition to this (k, N) threshold, we propose another threshold of (b, a); which states that, at least q number of cloud service providers out of p number of cloud service providers must take part in retrieving users data to provide a successful information retrieval.

5.1 LP-Assignment Problem:

One of the objectives is to minimize the cost of storage of the data pieces over p service providers. If d_i is the number of data pieces stored on i^{th} provider which has a per unit cost of storing the data as c_i.The total cost the customer has to pay is given below:

$$C = \sum_i^a d_i c_i \quad (1.1)$$

In our model, we consider $y_{i,j}$ as a binary variable, which is set to 1 if the j^{th} data piece on i^{th} service provider becomes a candidate in the current data retrieval. Since the Quality of Service factor depends on the physical location of information retrieval, the Quality of Service achieved in retrieving the data can be computed as given in following equation (1.2).

$$Qnet = \sum_i^a \sum_j^{d_i} y_{i,j} * Q_i \quad (1.2)$$

Therefore, the total cost of storing the dis-tributed customer data on a number of service provides must be minimized, and the Quality of Service achieved at the time of retrieval must be maximized. The objective is:

Minimize [C] and Maximize[Qnet] (1.3)

Maximize [Qnet – C] (1.4)

Costraints: Since d_i is the data pieces allocated to stored at i^{th} Service provider, this implies:

$$\sum_{i=1}^a d_i = N \quad (1.5)$$

Referring to the (k, N) threshold and the (b,a) threshold discussed before, the minimum number of pieces that must be chosen for data retrieval is k, for which at least b service providers are required. Thus, we have:

$$\sum_{j=1}^a y_{i,j} \geq b \quad (1.6)$$

and

$$\sum_{j=1}^a \sum_{i=1}^{d_j} y_{i,j} \geq k \quad (1.7)$$

where, N ≥ k and a ≥ b. Now, to make sure that a single Service Provider can not retrieve any meaningful information, the number of data pieces allotted to each Service Provider must be less than k:

$$0 < d_i < k \quad (1.8)$$

Solution: Since we have multiple optimization objectives as well as a set of variables d_i with non-definitive bounds, it seems to be very complex Linear programming problem. The model can be simplified with the help of lemma 1.

Lemma 1. *Given N data pieces to be distributed among a service providers such that, at least b service providers must take part in retrieval of data using at least k pieces from the distribution. This implies* $b_{max} = k$ *and each* $d_i = 1$.

Proof: Let us assume that $b_{max} \neq k$ and there exists at least one service provider, SP_m, that has two data pieces. Since any (at least) k number of data pieces can retrieve the data, we have two situa-

tions. First is that, the m^{th} SP does not take part in data retrieval. In this case, the maximum number of service providers that can be used for successful data retrieval is k, where each service provider has exactly one data piece. Second is, the m^{th} Service provider takes part in the data retrieval, here it only needs k−2 data pieces to retrieve the data successfully. Hence, the maximum number of service providers needed to retrieve these data pieces is k −2. So along with m^{th} SP, we needed only k −1 service providers. Here if we state q = k −1, then it would not be true for any k−1 service providers each having exactly one piece of data. Hence, we can say that, for maximum b, all $d_i = 1$ and this implies that $b_{max} = k$. Using lemma 1, we simplify our Linear programming model to include two binary variables, (p_i) as storage variable and (q_i) as retrieval variable, such that:

$$p_i = \begin{cases} 1 & \text{if } SP_i \text{ is alloted data piece,} \\ 0 & \text{otherwise.} \end{cases} \quad (1.9)$$

$$q_i = \begin{cases} 1 & \text{if } p_i = 1 \text{ and takes part in data retrieval,} \\ 0 & \text{otherwise.} \end{cases} \quad (1.10)$$

Now since, our objective function comprises of multiple objectives, we use goal programming phenomenon to statistically provide weights to each of individual objectives and unite them into a single objective. Hence, our simplified LP problem can be described as bellow:

Maximize $[w_1 * \sum_{i=1}^{N}(q_i * Q_i) - w_2 * \sum_{i=1}^{a}(p_i * C_i)]$

Where $w_1 + w_2 = 1$.

Subject to $\sum_{i=1}^{N} q_i = b$,

$$\sum_{i=1}^{a} p_i = N,$$

and $b = k \leq N \leq a$.

Owing to the non-negativity principle oflinear programming, we have:

$q_i,\ p_i,\ w_1,\ w_2,\ C_i$, and $Q_i \geq 0$.

While, $k > 0$.

Since the basic constraints are mostly equalities, these can easily be disintegrated into two inequalities and can be easily solved in ampl [2].

6 Conclusion

Cloud computing is an emerging technology that allows users to utilize on-demand computation, storage, data and services from around the world. In this paper, we 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 him with the best quality of service (Security and availability of data) offered by available cloud service providers. By dividing and distributing customers data, our model has shown its ability of providing a customer with a secured storage under his affordable budget.

Acknowledgment

This research was supported by Sinhagad institute of Technology, Lonavala Computer department of Pune University.

References

[1] M. Armbrust, A. Fox, R. Grifth, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the clouds :A berkeley view of cloud computing. Technical report,University of California at Berkeley, February 2009.

[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. Cloud computing and emerging it platforms: Vision, Future Generation Computer Systems, 25(6):599– 616, 2009.

[3] P. Mell, T. Grance, "Draft NIST working definition of cloud computing", Referenced on June. 3rd, 2009, Online thttp://csrc.nist.gov/groups/SNS/cloud computing/index.html, 2009.

[4] Amazon.com,"Amazon s3 availability event: July 20, 2008", Online at http://status.aws.amazon.com/s3-20080720.html, 2008.

[5] M Arrington,Gmail Disaster:Reports of mass email deletions"Online at http://www.techcrunch.com/2006/12/28/gmail- isasterreport-ofmass-email deletion/,December,2006

[6] The Official Google Blog, "A new approach to China: an update", online at http://googleblog.blogspot.com/2010/03/new-approach-to- chinaupdate. html, March 2010.

[7] N. Gruschka, M. Jensen, "Attack surfaces: A taxonomy for attacks on cloud services", Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, 5-10 July 2010.

[8] J. Du, W. Wei, X. Gu, T. Yu, "RunTest: assuring integrity of dataflow processing in cloud computing infrastructures", In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS '10), ACM, New York, NY, USA, 293 304.

[9] P. F. Oliveira, L. Lima, T. T. V. Vinhoza, J. Barros, M. M´edard, "Trusted storage over untrusted networks", IEEE GLOBECOM 2010, Miami, FL. USA

[10] A. Shamir, "How to share a secret", Commun. ACM 22, 11(November 1979).

[11] "A Modern Language for Mathematical Programming", Online shttp://www.ampl.com.