A Survey: Different Data Hosting Schemes

Dr. N. L. Udaykumar[#1], Rima Thokchom[#2]
[#1]*Assistant Professor*, [#2]*Post Graduate Student, Department of CSE*
Sri Siddhartha Institute of Technology
Tumkur, Karnataka, India

Abstract:-*Cloud computing is a resource provisioning mechanism, which represents a convenient way for users to access different computing resources. It is quickly becoming the platform of choice for many web services and virtualization is the key underlying technology enabling cloud providers to host services for a large number of customers. A cloud must provide services to many users at the same time and different users have different QoS requirements. Here in this paper we look at different cost-efficient data hosting schemes in cloud. We also present our own framework.*

Keywords: *Cloud Computing, Renting Resources, Dynamic Allocation, Cost and Time Efficiency.*

I. INTRODUCTION

Cloud computing is a possible solution for providing a flexible, on demand computing infrastructure for a number of applications. Many companies and research institutes show great interests in cloud computing. A cloud computing environment has several key features[1]:

- It is massively scalable,
- Can be encapsulated as an abstract entity that delivers different levels of services to customers outside the Cloud,
- It is driven by economies of scale, and
- The services can be dynamically configured (via virtualization or other approaches) and delivered on demand.

Nowadays more enterprises and organizations are hosting their data into the cloud as it helps to reduce the IT maintenance cost and enhance the data reliability. There are three entities existing in a cloud architecture namely a cloud service provider, a data owner/user and a third party auditor. A cloud service provider is the one who owns the data centers, cloud computing system and provides services to user who requests for services. User is the one who stores data in data center. A third party auditor is the one trusted entity which acts like an intermediary entity between user and cloud service provider. It acts only upon request by user. User can directly communicate with the cloud service provider or may get services through the third part auditor.

Three different types of cloud exists such as public, private and hybrid cloud[10]. In public cloud, access security is low i.e. anyone can access the information and the cost is very much less compared to other clouds. Amazon EC2, Microsoft Azure and Google Cloud are some popular examples of public cloud. While in private cloud, accessibility is limited to the organization i.e. the information can only be accessed by the organization that owns the cloud/information thus enhancing the security. The cost of private cloud is high. Hybrid cloud, like the name indicates, is a combination of public and private cloud where the confidential information are kept in private cloud and non-confidential information are kept in public cloud, thus allowing the user to use both public and private cloud cost-effectively.

Cloud computing provides access to different sharable computing resources (e.g., networks, servers, storage) which are owned by Cloud Service Providers (CSP). In cloud users do not need to own resources and frees them from expensive purchases and running and maintenance costs. Users need to pay whenever they rent and make use of the cloud computing services for their requirements.

In general, two resource renting options are provided by CSPs: on-demand and reserved[12]. The on-demand option allows users to pay for computing capacity by the hour with short term commitment, and spares the users from the costs and complexities of hardware planning, purchase and maintenance. It is suitable for real time applications as large fixed costs can be replaced by much smaller variable costs. The reserved option is a long-term strategy which enables users to make a low, onetime payment for each resource they need and keep the rented resources for a long time. . However, CSPs exploit the fact that the rented resources cannot be kept busy all the time and utilization rate is low. Because of this reason, users receive a significant discount on the hourly charge for each resource. A different planning horizon results in different renting alternatives, i.e., long-term (year/months) planning with the reserved alternative and/or short-term (hour/minutes) with the on-demand alternative. Renting the appropriate resources in different intervals of a time horizon is helps to minimize the total resource renting cost.

However, customers may be bemused with choosing the cloud for storing the data and the data hosting strategy to use as there are numerous cloud vendors providing different cloud server with different pricing policies. Based on comprehensive analysis of various state-of-the-art cloud vendors, this paper proposes a novel data hosting scheme which integrates two key functions desired.

The first is selecting several suitable clouds and an appropriate redundancy strategy to store data with minimized monetary cost and guaranteed availability. The second is triggering a transition process to re-distribute data according to the variations of data access pattern and pricing of clouds.

II. RELATED WORK

In this section, we are going to discuss related works about the different adapted algorithms that influence hosting data and retrieval of data from the cloud server with keeping track of total resource renting costs. Human beings tend to keep things simple by moving the complex aspects to computing. As a consequence, we prefer to go to one or a limited number of sources for all our information needs. In contemporary scenario where information is replicated, modified (value added), and scattered geographically; retrieving information in a suitable form requires lot more effort from the user and is difficult. We noticed that researchers use different approaches regarding this problem.

In [3], J. Li, B. Li, Z. Du, and L. Meng make use of learning systems (Neural Network based) that can intelligently decide and retrieve the information that we need by going directly to the source of information. This approach reduces single point of failure, eliminates bottlenecks in the path of information flow and reduces the time delay. It makes efficient information retrieval approach for collaborative cloud computing.

As cloud computing becomes widely deployed, one of the challenges faced involves the ability to orchestrate a highly complex set of subsystems (compute, storage, network re- sources) that span large geographic areas serving diverse clients. To ease this process, in [4], C. Liu, B.T. Loo, and Y. Mao have presented COPE (Cloud Orchestration Policy Engine), a distributed platform that allows cloud providers to perform declarative automated cloud resource orchestration. In COPE, cloud providers specify system-wide constraints and goals using COPElog, which is a declarative policy language geared towards specifying distributed constraint optimizations. COPE takes policy specifications and cloud system states as input and then optimizes compute, storage and network resource allocations within the cloud such that provider's operational objectives and customer SLAs can be better met. However it has yet to incorporate into the optimization framework, recent models on resource provisioning and deployment analysis in the cloud. For example, [5] explores the feasibility of migrating enterprise applications to the cloud, based on a cost-benefit analysis. It still needs to explore the applicability of COPE to enable such resource provisioning and cost-benefit analysis

In [2], L. Chen, X. Li and R. Ruiz concerns with the challenging problem of determining the right amount of resources for multiple periodical workflow applications. They have considered a scheduling periodical workflow applications on cloud resources with the objective of minimizing the total renting cost by developing a Precedence Tree based Heuristic (PTH) which consists of three components: Workflow Combination and Parameter Initialization(WCPI), initial schedule Construction Methods(CM) and Schedule Improvement Procedure(SIP) where WCPI considers the features and constraints of different workflows and combines them into a big single workflow and relative parameters are initialized and used in CM and SIP. Different types of rules are proposed in CM to construct the initial schedule for the considered problem. SIP contains two main improvement procedures, which decrease the resource renting cost by mode and resource adjustments. However it doesn't consider the problems with resources that are not sharable between workflows.

As Internet clouds work as service factories built around Web-scale data centers, the elastic cloud resources and huge datasets being processed are subject to security breaches, privacy abuses, and copyright violations. The provisioned cloud resources on-demand is especially vulnerable to cyber attacks. The cloud platforms built by Google, IBM, and Amazon all reveal this weaknesses. In light of this, K. Hwang, S. Kulkarni, and Y. Hu in [6] has proposed a new approach to integrating virtual clusters, security-reinforced data centers, and trusted data accesses guided by reputation systems. Here they have integrated and extended the fuzzy-theoretic trust models by Song, et al [7] and by He, et al [8] in a cloud application environment. A hierarchy of P2P reputation systems is suggested to protect clouds and data centers at the site level and to safeguard the data objects at the file-access level. Different security countermeasures are suggested to protect cloud service models: IaaS, PaaS, and SaaS, currently implemented by Amazon, IBM, and Google, respectively. However, the interoperability and common cloud standards are still wide open problems.

As discussed in the previous section, there are two resource renting options provided by CSPs: on-demand and reserved [12]. Resources provisioned by reservation plan are cheaper than that provisioned by on-demand plan, since cloud consumer has to pay to provider in advance. With the reservation plan, the consumer can reduce the total resource provisioning cost. However, the best advance reservation of resources is difficult to be achieved due to uncertainty of consumer's future demand and providers' resource prices. To address this problem, in [9], S. Chaisiri, B.S. Lee and D. Niyato have proposed an optimal cloud resource provisioning (OCRP) algorithm by formulating a stochastic programming model. The OCRP algorithm can provision computing resources for being used in multiple provisioning stages as well as a long-term plan. The demand and price uncertainty is considered in OCRP and shows that cloud consumer can successfully minimize total cost of resource provisioning in cloud computing environments. For future work the authors have suggested scenario reduction techniques [11] to reduce the number of scenarios and also the optimal pricing scheme for cloud providers with the consideration of competition in the market.

However still most I&O organizations struggle to compare the capital, operational, and staffing cost differences between internal and cloud storage. In [13], A. Reichman breaks down the cost differences by modeling a common workload, file storage, built and deployed traditionally versus consumed through the public cloud. The models reveal a significant cost difference, with the cloud-based model coming in 74% less expensive than I&O running it in-house. Similar to this research, K. Liu and L.J. Dong illustrates how to improve the file storage method in [14], based on eyeOS web Operating Systems which realizes file distributed storage and fault-tolerant control through HDFS technology of Hadoop

For a secure and efficient data hosting onto the cloud, in addition to selecting vendors that provide the proper underlying facilities, infrastructure, and storage management applications, we will need some storage access security measures.In light of this, researches on a secure storage system that ensures data encryption and data availability when data integrity and data are not complete has been done. In [15], R. Wang proposed a data secure storage scheme based on Tornado codes (DSBT) by combining the technique of symmetric encryption and erasure codes and the POR system based on trusted log and combining the DSBT scheme and POR system, the computational efficiency of the POR algorithm is optimized, and the system can provide strong data loss recovery ability and resist the Byzantine fault. In [16], A. Kumar, B.G. Lee, H.J. Lee has used the technique of elliptic curve cryptography encryption to protect data files and proposed a model that has two part in the cloud storage server, Private data section and Shared data section. These two part of the cloud storage server makes the sharing of data easy and secure. This model solves the problem of group sharing of data in the shared data section as only member of group can access the data stored over shared data section. But however, one to many, many to one and many to many communication is not possible.

III. METHODOLOGY

In this paper, a "BROKER" scheme framework is proposed, which will be a cost-efficient data hosting scheme with high availability in heterogeneous multi-cloud. Two important modules in BROKER MODULE are Data Hosting and SMS. Our scheme will attempt to solve the problem of how to combine the two mechanisms, namely heterogeneous pricing policies and the different redundancy mechanisms elegantly so as to reduce cost and also guarantee required availability. Specifically, two popularly used redundancy mechanisms, i.e., replication and erasure coding, will be combined into a uniform model to meet the required availability in the presence of different data access patterns. To help preserving data privacy for outsourced data in cloud server, Identity-Based Encryption (IBE) and Identity-Based Signature (IBS), which is a type of public-key encryption has been addressed in which the public key of a user is some unique information about the identity of the user. This helps to keep the integrity of the data of the user in check. This means that a sender who has access to the public parameters of the system can encrypt a message using e.g. the text-value of receiver's email address as a key.

The Architecture of the proposed system is shown below in the figure 1.

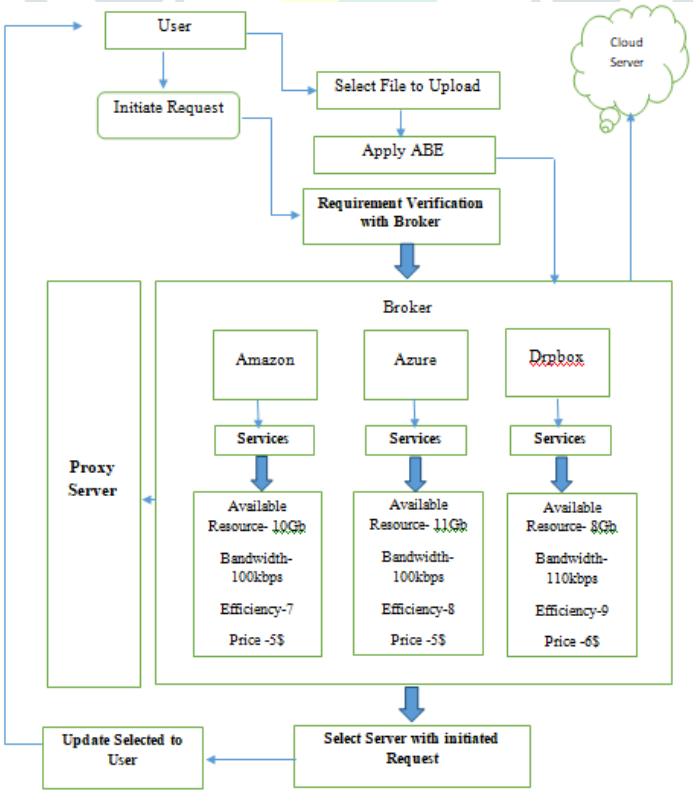

Figure 1: Architecture for uploading a file on cloud.

IV. OUR CONTRIBUTIONS

As a holistic storage system, there are several other factors to be considered, such as cache strategies, geographical data consistency, etc. However, we will only be focusing on the data hosting strategy to minimize monetary cost while meeting flexible availability requirements. The main contributions of this paper are the following:

- Literature survey on different related works on data hosting schemes.
- Proposal of an efficient data hosting strategy based on the survey results.
- A strategy where data stored in cloud can be accessed by only authorized users with valid attributes.
- Authentication of users who store and modify their data on the cloud.
- Integrity of the data stored in the cloud be preserved.
- Integrity of the file stored in the cloud will be checked by each cloud users.

The admin should update the service parameter in every time interval and the updated details should be displayed to cloud user to get access of the service through Broker module. It should be able to provide a transparent relation between the clients and the CSPs. Figure 2 below shows service details displayed. It shows which cloud service provider provides what kind of services, through the broker.

Provider Name	Amount Of Storage	Storage Weight	Operation_Price_Get	Operation_Price_Put	Bandwidth	Bandwidth Weight	Availability	Availability Weight
Amazon	200.0	10.0	0.3	0.4	100.0	100.0	9.0	9.0
azure	20.0	10.0	0.5	0.4	100.0	100.0	8.0	8.0
dropbox	200.0	10.0	1.0	1.0	110.0	110.0	9.0	9.0

Figure 2: Different CSPs with their different storage, cost, bandwidth and availability range.

According to the availability and costs of the resources provided by the different CSPs and the resource requirement of the user, the broker should suggest to the user the most suitable cloud server for uploading his/her files.

V. CONCLUSION

Here we look at the different data hosting schemes, researched and developed over the years. Over time, the user requirements and the services provided by the CSPs differs. In this paper we also present a model framework that will check the customer requirements and select the most preferable and convenient cloud service platform in an efficient manner and be also able to resolve any future migration issues. It also addresses the security and privacy issues and use of and ID-based encryption method to encrypt the messages while uploading or downloading the documents for the cloud servers.

REFERENCES

[1]. Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu, "Cloud Computing and Grid Computing 360-Degree Compared", Grid Computing Environments Workshop 2008(GCE '08).

[2]. Long Chen, Xiaoping Li, Senior Member, IEEE, Ruben Ruiz, "Resource renting for periodical cloud workflow applications", IEEE TSC, Citation:DOI 10.1109/TSC.2017.2677450.

[3]. J. Li, B. Li, Z. Du, and L. Meng, "CloudVO: Building a Secure Virtual Organization for Multiple Clouds Collaboration," Proc. 11th ACIS Int'l Conf. Software Eng. Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD), 2010.

[4]. C. Liu, B.T. Loo, and Y. Mao, "Declarative Automated Cloud Resource Orchestration," Proc. Second ACM Symp. Cloud Computing (SOCC '11), 2011.

[5]. Hajjat, M., Sun, X., Sung, Y.-W. E., Maltz, D., Rao, S., Sripanidkulchai, K., and Tawarmalani, M. Cloudward bound: planning for beneficial migration of enterprise applications to the cloud, in SIGCOMM (2010).

[6]. K. Hwang, S. Kulkarni, and Y. Hu, "Cloud Security with Virtualized Defense and Reputation-Based Trust Management," Proc. IEEE Int'l Conf. Dependable, Autonomic and Secure Computing (DASC), 2009.

[7]. S. Song, K. Hwang, R Zhou, and Y.K. Kwok, "Trusted P2P Transactions with Fuzzy Reputation Aggregation", IEEE Internet Computing, Special Issue on Security for P2P and Ad Hoc Networks, Nov/Dec. 2005.

[8]. R. He, J. Niu, M. Yuan, and J. Hu, "A Noval Cloud-Based Trust Model for Pervasive Computing", The Fourth International Conference on Computer and Information Technology, Sept. 14-16 2004.

[9]. Sivadon Chaisiri, Student Member, IEEE, Bu-Sung Lee, Member, IEEE, and Dusit Niyato, Member, IEEE, "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE TSC, VOL. 5, NO. 2, APRIL-JUNE 2012.

[10]. R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, and R. Buyya, "Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, 2011.

[11]. H. Heitsch and W. Romisch, "Scenario Reduction Algorithms in Stochastic Programming," J. Computational Optimization and Applications, vol. 24, 2003.

[12]. AmazonEC2, "Amazon elastic compute cloud (Amazon EC2)," http://aws.amazon.com/ec2/pricing, 2014.

[13]. Andrew Reichman, "File Storage Costs Less In The Cloud Than In-House", for Infrastructure & Operations Professionals, August 25, 2011.
[14]. Kun Liu and Long-Jiang Dong, "Research on cloud data storage technology and its architecture implementation." , International Workshop on Information and Electronics Engineering (IWIEE), 2012.
[15]. Rongzhi Wang, "Research on data security technology based on cloud storage", 13th Global Congress on Manufacturing and Management, GCMM 2016, Procedia Engineering 174 (2017) 1340 – 1355.
[16]. Arjun Kumar, Byung Gook Lee, Hoon Jae Lee, "Secure Storage and Access of Data in Cloud Computing" , IEEE ICTC, 2012.