Efficient Framework for Resource Allocation and Monitoring approach for Data Centers and Virtual Environment

Prof. S. Padmavathi[1] C. Vaishnavi[2]
[1]Head of Dept. [2]M. Phil. Student
[1,2]Department of Information Technology
[1,2]Marudupandiyar College, Thanjavur, India

Abstract— In our thesis Virtualization technology to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. We introduce the concept of “skewness” to measure the unevenness in the multidimensional resource utilization of a server. By minimizing skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources. Resource allocation process must be scalable both in the number of machines in the cloud and the number of sites that the cloud hosts. Green scheduling can determine which server to be in running state. Control a different set of technique is used to predict non-stationary workloads of the system. We develop a set of heuristics that prevent overload in the system effectively while saving energy used. Trace driven simulation and experiment results demonstrate that our algorithm achieves good performance.

Key words: Skewness, Green Scheduling, Virtualization, Green Computing

I. Introduction

A cloud storage system consists of robustness, confidentiality and functionality. The proxy re-encryption scheme supports encoding operations over encrypted messages as well as forwarding operations over encoded and encrypted messages. To provide data robustness is to replicate a message such that each Storage server stores a copy of the message. It is very robust because the message can be retrieved as long as one storage server survives. The number of failure servers is under the tolerance threshold of the erasure code, the message can be recovered from the codeword symbols stored in the available storage servers by the decoding process. This provides a tradeoff between the storage size and the tolerance threshold of failure servers.

A decentralized erasure code is an erasure code that independently computes each codeword symbol for a message. A decentralized erasure code is suitable for use in a distributed storage system. A storage server failure is modeled as an erasure error of the stored codeword symbol. We construct a secure cloud storage system that supports the function of secure data forwarding by using a threshold proxy re-encryption scheme. The encryption scheme supports decentralized erasure codes over encrypted messages and forwarding operations over encrypted and encoded messages. Our system is highly distributed where storage servers independently encode and forward messages and key servers independently perform partial decryption.

For years the Internet has been represented on network diagrams by a cloud symbol until 2008 when a variety of new services started to emerge that permitted computing resources to be accessed over the Internet termed Cloud Computing.

Fig. 1: Cloud computing technologies

Cloud computing encompasses activities such as the use of social networking sites and other forms of interpersonal computing; however, most of the time cloud computing is concerned with accessing online software applications, data storage and processing power.

Cloud computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructure, training new personnel, or licensing new software. It extends Information Technology’s (IT) existing capabilities. In the last few years, cloud computing has grown from being a promising business concept to one of the fast growing segments of the IT industry.

Cloud computing is an on-demand Internet-based virtual service model where data is stored and accessed using distributed on-demand elastic services in a transparent manner. Internet has grown by leaps and bounds and is growing at an unimaginable pace. Every business on the planet has its persona on the Internet. With the advent of cloud computing, business houses have found it easy to host their applications on a virtual environment hosted and managed by a cloud service provider.

The notion of on-premise application deployment is fast fading out. This is known as Software-As-A-Service in cloud computing world. This gives organizations the flexibility to choose the best cloud hosting provider based on parameters such as cost, reliability, performance, storage, security, availability etc. Enterprises believe that it is possible to reduce the Total Cost of Ownership (TCO) by adopting the SaaS model. The cloud computing model has three service delivery models as shown in (See Figure 2).

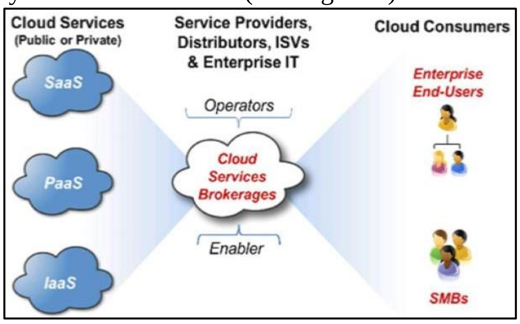

Fig. 2: Cloud computing services

They expect the cloud hosting provider to get their systems and practices certified by authorized certifying bodies before signing Service Level Agreements (SLAs). Although hosting providers try to combat security risks by strengthening their infrastructure by means of regular security updates, it becomes insufficient and obsolete when new threats and attacks emerge.

Moreover, application vendors and enterprises might have stringent security requirements which the hosting providers may not be able to deliver. The security standards prevailing today are for regular non-cloud deployments which cannot be directly applied to a cloud environment. This is because of the nature, operational artifacts and the architecture of the cloud environment.

Security officers should be aware of concerns experienced in the cloud environment. Jay Heiser et al has conducted an investigation regarding information security issues that should be considered when dealing with cloud computing. In their investigation report, they have highlighted seven security issues with respect to access, compliance, data loss and recovery etc.

Security as a Service (Sc-aaS) is a mechanism to deliver security solutions by cloud vendors to hosted applications. It is an area under research by CSA which is evolving. Policy-based Security as a Service (PSc-aaS) is an approach towards provisioning security among the cloud hosting providers. It refers to security services entrusted by a security policy using the SaaS model.

The architecture consists of four different entities: data owner, user, cloud server (CS), and TPA. Here the TPA is the trusted entity that has expertise and capabilities to assess cloud storage security on behalf of a data owner upon request. Under the cloud paradigm, the data owner may represent either the individual or the enterprise customer, who relies on the cloud server for remote data storage and maintenance, and thus is relieved of the burden of building and maintaining local storage infrastructure. In most cases cloud data storage services also provide benefits like availability (being able to access data from anywhere), relative low cost (paying as a function of need), and on demand sharing among a group of trusted users, such as partners in a collaboration team or employees in the enterprise organization.

For simplicity, assume a single writer/many readers scenario here. Only the data owner can dynamically interact with the CS to update her stored data, while users just have the privilege of file reading. Within this scope, it focuses on how to ensure publicly auditable secure cloud data storage services. (See Figure 3)

As the data owner no longer possesses physical control of the data, it is of critical importance to allow the data owner to verify that his data is being correctly stored and maintained in the cloud. Considering the possibly large cost in terms of resources and expertise, the data owner may resort to a TPA for the data auditing task to ensure the storage security of her data. Assume the TPA, who is in business of auditing, is reliable and independent, and thus has no incentive to collude with either the CS or the owners during the auditing process.

Fig. 3: Cloud Storage

The TPA should be able to efficiently audit the cloud data storage without local copy of data and without any additional online burden for data owners. Besides, any possible leakage of an owner's outsourced data toward a TPA through the auditing protocol should be prohibited.

A cloud storage system is considered as a large-scale distributed storage system that consists of many independent storage servers. Data robustness is a major requirement for storage systems. There have been many proposals of storing data over storage servers. One way to provide data robustness is to replicate a message such that each storage server stores a copy of the message. It is very robust because the message can be retrieved as long as one storage server survives. Another way is to encode a message of k symbols into a codeword of n symbols by erasure coding. To store a message, each of its codeword symbols is stored in a different storage server. A storage server failure corresponds to an erasure error of the codeword symbol. As long as the number of failure servers is under the tolerance threshold of the erasure code, the message can be recovered from the codeword symbols stored in the available storage servers by the decoding process. This provides a tradeoff between the storage size and the tolerance threshold of failure servers. The main advantage of using cloud storage is that it is proves to be a flexible option for those enterprises which cannot immediately estimate the amount of storage capacity required.

II. PROBLEM DEFINITION

Resource Management is critical in cloud computing. With improper resource management, applications might experience network congestion, long time wait, CPU waist, overused CPU and memory, and security problems. To maximized cloud computing infrastructure utilization and minimize total cost of both the cloud computing infrastructure and running applications, resources need to be managed properly. To overcome this there are kinds of resources in the large-scale computing infrastructure need to be managed, CPU load, network bandwidth, disk quota, and even type of operating systems.

To provide better quality of service, resources are provisioned to the users or applications, via load balancing mechanism, high availability mechanism and security and authority mechanism. To maximize cloud utilization, the capacity of application requirements shall be calculated so that minimal cloud computing infrastructure devices shall be procured and maintained. Given access to the cloud computing infrastructure, applications shall allocate proper resources to perform the computation with time cost and infrastructure cost minimized.

For all the potential of the cloud storage model, there are still a few disadvantages. By utilizing cloud storage, users relinquish direct control over their data. Although a cloud storage provider is far less likely to have its data lost or compromised than most individuals or organizations, people are still more comfortable knowing precisely where their critical data is located and who is personally responsible for it. Access to data stored on a distributed network is also constrained by the access to the Internet. Lower Internet bandwidth will result in decreased performance and an Internet service outage will completely sever an organization's access to its information.

III. Implementation

In our proposed system considers the process of resource management for a large-scale cloud environment. Such an environment includes the physical infrastructure and associated control functionality that enables the provisioning and management of cloud services.

The perspective we take is that of a cloud service provider, which hosts sites in a cloud environment. The cloud service provider owns and administrates the physical infrastructure, on which cloud services are provided. It offers hosting services to site owners through a middleware that executes on its infrastructure. Site owners provide services to their respective users via sites that are hosted by the cloud service provider. Therefore, the user demands are transformed to this virtual cloud server. Through this efficient method, the user's demands will be satisfied successfully by serving the customer without waiting. Therefore, the resources will be allocated dynamically. This work contributes towards engineering a middleware layer that performs resource allocation in such a cloud environment, with the following design goals:

A. Performance Objective

We consider computational and memory resources and the objective is to achieve max-min fairness for computational resources under memory constraints.

B. Adaptability

The resource allocation process must dynamically and efficiently adapt to changes in the demand for cloud services.

C. Scalability

The Resource allocation process must be scalable both in the number of machines in the cloud and the number of sites that the cloud hosts. This means that the resources consumed per machine in order to achieve a given performance objective must increase sublinearly with both the number of machines and the number of sites.

We present the design and implementation of an automated resource management system that achieves a good balance between the two goals. We make the following contributions.

We develop a resource allocation system that can avoid overload in the system effectively while minimizing the number of servers used. We introduce the concept of "skewness" to measure the uneven utilization of a server. By minimizing skewness, we can improve the overall utilization of servers in the face of multidimensional resource constraints. We design a load prediction algorithm that can capture the future resource usages of applications accurately without looking inside the VMs. The algorithm can capture the rising trend of resource usage patterns and help reduce the placement churn significantly.

Cloud computing is a usage of very large scalable and virtualized resources in a dynamic way over the internet. Due to the rapid growth of cloud environment usage many tasks require to be executed by the available resources .At the same time it should be possible to achieve better performance, optimizing the servers, reduce migration, support green computing, better resource utilization etc. so resource allocation using virtual machine plays a most important role in cloud environment because it should allocate proper resource to proper resources to various machines to get maximum benefit. In this paper we are going to study different set of resource allocation process and their concerns using virtual machines in cloud computing.

IV. Approaches and Algorithm

A. Skewness Algorithm

Skewness is used to quantify the unevenness in utilization of multiple resources on the server. By minimizing the skewness leads to combine of different combine different workloads and improve utilization of server. Skewness consists of three steps: load prediction, hot spot migration, and green computing.

B. Green Scheduling Algorithm

Green scheduling can determine which server to be in running state. It will turn on and turnoff servers based on load and virtual machine is allocated. Server must be in four states: OFF, ON, SHUTTING, RUNNING. Based on platform any of the state is triggered.

C. Control Algorithm

In the control algorithm a different set of technique is used to predict non-stationary workloads of the system. In this two set of process are used Markov Host Overload Detection (MHOD) and Optimal Markov Host Overload Detection (MHOD-OPT).

D. Dynamic Resource Allocation

Resource Allocation (RA) is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation starves services if the allocation is not managed accurately. Resource provisioning solves that problem by allowing the service providers to manage the resources for each individual module.

E. Cloud Service Provider

The cloud service provider is responsible for maintaining an agreed-on level of service and provisions resources accordingly. A CSP, who has significant resources and expertise in building and managing distributed cloud storage servers, owns and operates live Cloud Computing systems, It is the central entity of cloud.

Cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment so as to meet the needs of the cloud application It requires the type and amount of resources needed by each application in order to complete a user job. The order and time of allocation of resources are also an input for an optimal resource allocation.

F. Cloud Consumer

Cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from, a cloud provider. Users, who stores data in the cloud and rely on the cloud for data computation,

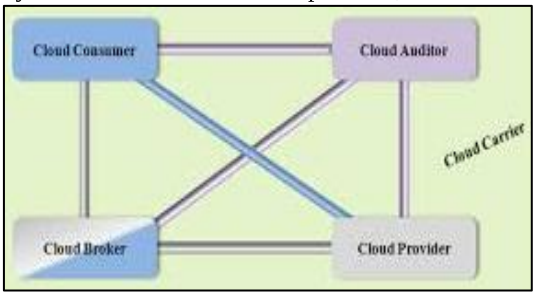

Fig. 4: Cloud provider interact to cloud consumer

Cloud consists of both individual consumers and organizations. Cloud consumers use Service-Level Agreements (SLAs) for specifying the technical performance requirements to be fulfilled by a cloud provider. A cloud provider may also list in the SLAs a set of restrictions or limitations, and obligations that cloud consumers must accept.

G. Virtual Machine Environment

Virtualization provides an efficient solution to the objectives of the cloud computing paradigm by facilitating creation of Virtual Machines (VMs) over the underlying physical servers, leading to improved resource utilization. Virtualization refers to creating a virtual version of a device or a resource such as a server, a storage device, network or even operating system where the mechanism divides the resource into one or more execution environments.

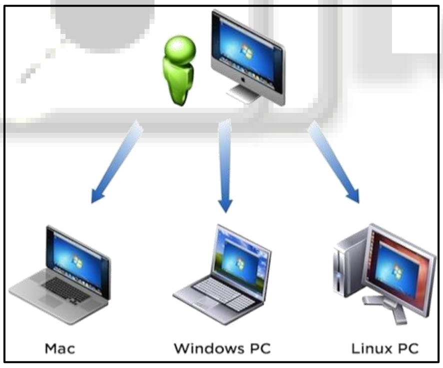

Fig. 5: Virtual Machine Environment

When a physical server is considered to be overloaded requiring live migration of one or more VMs from the physical server under consideration. Selection of VMs that should be migrated from an overloaded physical server. VM selection policy algorithm has to be applied to carry out the selection process. Finding a new placement of the VMs selected for migration from the overload and physical servers and finding the best physical.

H. Resource Manager

Service management in this context covers all the data center operations activities. This broad discipline considers the necessary techniques and tools for managing services by both cloud providers and the internal data center managers across these physical, IT and virtual environments. The availability of Service computing clouds gives researchers access to a large set of new resources for running complex scientific applications. However, exploiting cloud resources for large numbers of jobs requires significant effort and expertise.

I. Performance Evaluation

In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. Some of the strategies discussed above mainly focus on CPU, memory resources .secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

In our proposed system considers the process of resource management for a large-scale cloud environment. Such an environment includes the physical infrastructure and associated control functionality that enables the provisioning and management of cloud services. The perspective we take is that of a cloud service provider, which hosts sites in a cloud environment. The cloud service provider owns and administrates the physical infrastructure, on which cloud services are provided. It offers hosting services to site owners through a middleware that executes on its infrastructure.

V. IMPLEMENTATION RESULTS

I have completed the design, implementation, and evaluation of a resource management system for cloud computing services. Our system multiplexes virtual to physical resources adaptively based on the changing demand. We use the skewness metric to combine VMs with different resource characteristics appropriately so that the capacities of servers are well utilized. Our algorithm achieves both overload avoidance and green computing for systems with multi resource constraints.

VI. CONCLUSION

An effective and flexible distributed scheme with explicit dynamic resource allocation to distributed environment. In this paper, we present the design and implementation of an automated resource management system that achieves a good balance between the two goals. We make the following contributions: . We develop a resource allocation system that can avoid overload in the system effectively while minimizing the number of servers used. . We introduce the concept of "skewness" to measure the uneven utilization of a server. By minimizing skewness, we can improve the overall utilization of servers in the face of multidimensional resource constraints. . We design a load prediction algorithm that can capture the future resource usages of applications accurately without looking inside the VMs. The algorithm can capture the rising trend of resource usage patterns and help reduce the placement churn significantly.

VII. FUTURE ENHANCEMENT

In achieves more efficient and realistic user friendly approach for resource allocation scheme in distributed environment and also resolves load balancing problem and minimizing the number server used in distributed environment. Currently, our project runs on homogeneous environment, In future I am going to perform heterogeneous environment and also i am going to try web services and big data and hadoop environment in same works. In addition I am going to create better resource allocation setup using several new technologies (Microsoft framework, oracle framework, IBM framework) and etc. The future work will be concerned with the development of the better allocation algorithm which is in heterogeneous and works in dynamic environment using virtual machines.

REFERENCES

[1] M. Armbrust et al., "Above the Clouds: A Berkeley View of Cloud Computing," technical report, Univ. of California, Berkeley, Feb.2009.
[2] L. Siegele, "Let It Rise: A Special Report on Corporate IT," The Economist, vol. 389, pp. 3-16, Oct. 2008.
[3] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the Art of Virtualization," Proc. ACM Symp. Operating Systems Principles (SOSP '03), Oct. 2003.
[4] "Amazon elastic compute cloud (Amazon EC2)," http://aws.amazon.com/ec2/, 2012.
[5] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I.Pratt, and A. Warfield, "Live Migration of Virtual Machines," Proc. Symp. Networked Systems Design and Implementation (NSDI '05), May 2005.
[6] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast Transparent Migration for Virtual Machines," Proc. USENIX Ann. Technical Conf., 2005.
[7] M. McNett, D. Gupta, A. Vahdat, and G.M. Voelker, "Usher: An Extensible Framework for Managing Clusters of Virtual Machines," Proc. Large Installation System Administration Conf. (LISA '07), Nov. 2007.
[8] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-Box and Gray-Box Strategies for Virtual Machine Migration," Proc. Symp. Networked Systems Design and Implementation (NSDI '07), Apr. 2007.
[9] C.A. Waldspurger, "Memory Resource Management in VMware ESX Server," Proc. Symp. Operating Systems Design and Implementation (OSDI '02), Aug. 2002.
[10] G. Chen, H. Wenbo, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services," Proc. USENIX Symp. Networked Systems Design and Implementation (NSDI '08), Apr. 2008.
[11] P. Padala, K.-Y. Hou, K.G. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal, and A. Merchant, "Automated Control of Multiple Virtualized Resources," Proc. ACM European conf. Computer Systems (EuroSys '09), 2009.
[12] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Virtual Machines for Managing SLA Violations," Proc. IFIP/IEEE Int'l Symp. Integrated Network Management (IM '07), 2007.
[13] "TPC-W: Transaction Processing Performance Council," http://www.tpc.org/tpcw/, 2012.
[14] J.S. Chase, D.C. Anderson, P.N. Thakar, A.M. Vahdat, and R.P.Doyle, "Managing Energy and Server Resources in Hosting Centers," Proc. ACM Symp. Operating System Principles (SOSP '01),Oct. 2001.
[15] C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A Scalable Application Placement Controller for Enterprise Data Centers," Proc. Int'l World Wide Web Conf. (WWW '07),