International Journal of Engineering Research & Management Technology
ISSN: 2348-4039
Volume 9, Issue-1 Jan-Feb 2022
Email: editor@ijermt.org
www.ijermt.org

OPTIMIZING RESOURCE ALLOCATION IN CLOUD COMPUTING AUTOMATED CLOUD DEFRAGMENTATION DISTRIBUTION TECHNIQUE

Safia Khanam
Research Scholar
Deptt. of Computer Science and Engg.
SJJTU, Jhunjhunu.

Dr. Prasadu Peddi
Supervisor
Deptt. of Computer Science and Engg.
SJJTU, Jhunjhunu.

ABSTRACT:

Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) are common XaaS layers in cloud architecture (IaaS). The previous research looked at how to allocate resources efficiently in cloud computing to meet the goals of cloud users, IaaS providers, and SaaS providers. The composition of distinct cloud layers, such as IaaS and SaaS, and their joint optimization for optimal resource allocation are proposed in this paper. Subproblems are used to solve the efficient resource allocation optimization problem. Our planned work focuses on efficient scheduling and resource allocation to help cloud customers, IaaS providers, and SaaS providers achieve their goals. Early task scheduling algorithms were solely concerned with decreasing make span, with no methods in place to lower the monetary cost of cloud setup.

KEYWORDS: Resource Allocation, Optimization Problem, Efficient Scheduling

1.1 INTRODUCTION

Cloud computing is a computer framework that consists of a collection of networked computing nodes, servers, as well as programming administrations and applications that are powerfully provided to customers. It focuses on providing end-users with reliable, safe, viable, and adaptable administrations, platforms, and infrastructure. The goals of cloud computing frameworks are to allow for unlimited data storage and computation, as well as to hide massive amounts of complex data from consumers. The services are accessed via the internet, a private network, or both. The cloud computing model, according to the National Institute of Standards and Technology (NIST), allows convenient, ubiquitous, and on-demand services to access a shared pool of configurable computer resources such as servers, networks, services, and other resources that can be quickly provisioned and released with minimal service provider interaction. Three service models, five major characteristics, and four deployment types make up the cloud computing model.

1.2 CLOUD ARCHITECTURE

The architecture of cloud computing includes several segments and sub-segments, which can be divided into two categories: front end and back end.

The internet or a virtual network connects the two ends. Cloud computing architecture is divided into two parts: cloud resources and middleware. The user, or customer's computer system, that accesses the cloud system, sees the front end. The various cloud systems have various user interfaces. On the other hand, the service provider establishes cloud computing services on the back end using various data storage, virtual machines, computers, and servers, among other things. Various types of apps make up the back end structure. The back end side's main responsibilities include providing security to mechanisms, using protocols to connect networked computers for communications, and monitoring network traffic and protocols.

1.3 ESSENTIAL CHARACTERISTICS AND SERVICES OF CLOUD COMPUTING

In general, cloud computing refers to the delivery of services over the internet to anyone, anywhere. Many businesses have adopted the cloud strategy, and many vendors are producing applications and services for other businesses. Cloud Computing administrations have proven to be more cost-effective and adaptable than traditional solutions. Cloud administrations are among the associations in this regard. The following are the key aspects of cloud computing:

- **Self-service on demand:** A computer facilitates communications with all service providers, and services are automatically provided whenever there is a demand, such as network or server applications, emails, and so on.
- **Resource pooling:** CSP processing assets are pooled to serve several customers in a multi-tenant paradigm, with various dynamic virtual and physical assets allocated and reallocated according to client demand. Memory, virtual machine (VM), processor, storage, and other resources are all included.
- **High elasticity:** Cloud services are provisioned fast and with great flexibility. Automatic cloud services are sometimes swiftly scaled out and scaled in. The quick elasticity feature ensures that the user application's correct capacity is available at all times.
- **Measurable service:** This service manages and monitors cloud resources, as well as providing transparency to both the provider and the consumer. Metering capability is used in cloud computing services to monitor and increase resource use. Measuring capability in the cloud

architecture determines the right level of abstraction for various sorts of services such as bandwidth, processing, storage, and so on.

- **Multi-tenancy:** The Cloud Security Alliance is a proponent of cloud computing. It specifies the needs of multiple customer populations, including isolation, service levels, policy-driven enforcement, chargeback/billing schemes, governance, and so on.

1.4 RESOURCE ALLOCATION

The centralized resource manager oversees managing cloud computing resources. The tasks are assigned to the required VMs by the unified resource administrator. Clients/applications can access cloud DC assets via virtual machines (VMs). Virtual machines (VMs) are used to meet the asset requirements and provide run-time support for the applications.

Specifically, running an application for a required asset can be accomplished in two stages: creating a case for the VM as required by the application (VM provisioning) and booking the demand to real assets (asset provisioning). The VM is a product that resembles the components of a PC framework (genuine machine). Because of its numerous advantages, such as server combination, live movement, and security seclusion, the VM innovation has become well-known in recent years in server farms and distributed computing circumstances. Distributed computing is based on the concept of virtualization, which encapsulates several administrations that might suit a client's need in a cloud computing environment. Virtual machines (VMs) are designed to run on a server and provide a variety of operating systems with the help of various applications. A minimum of one VM can be set or conveyed on a PM that meets the VM's requirement. The errand can be scheduled using representation innovation and dynamic load adjustment between the hosts in distributed computing situations.

1.5 ENERGY EFFICIENCY IN CLOUD ENVIRONMENT

In large DC and cloud providers, energy efficiency has emerged as a unique issue. The DC has a huge number of web servers, commonly known as PMs. The main challenge in content distribution machines and maximum dispersed systems is power consumption (Cloud systems). These needs are made up of networked computing resources from one or more organisations on distributed computing centers (DCs) that are becoming more prevalent around the world. In today's DC and cloud computing systems, this consumption is a significant design element. This usage is also likely to increase, as the DC is forecast to grow in both size and number.

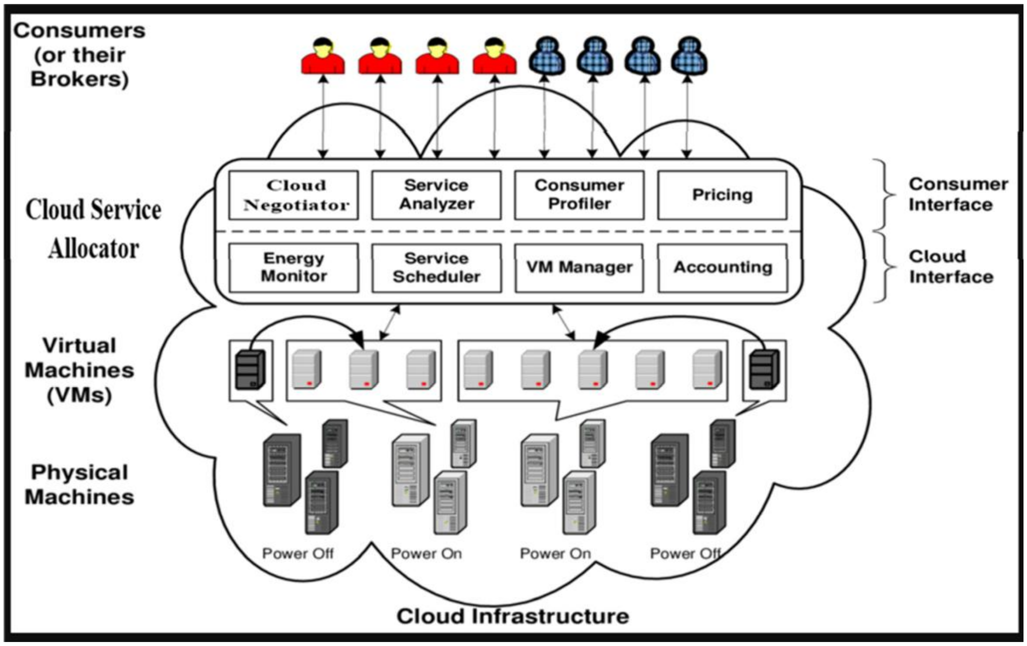

FIGURE.1.1- ENERGY EFFICIENCY IN CLOUD ENVIRONMENT

1.6 MOTIVATION

The cloud computing environment satisfies the criteria by establishing virtual machines (VMs) and assigning assets to them based on the needs for the applications that must be deployed. These applications cover a wide range of architectures and configurations, including service delivery, social networking, and web hosting, among others.

As a result, successful management of the cloud computing life cycle necessitates the use of all assets. Load balancing is an important method for deploying applications effectively and successfully in the cloud life cycle. The cloud service provider delivers services based on the service level agreement (SLA) agreed with the customer, therefore the cloud datacenter deploys very high performance servers and other infrastructures to maintain the SLA and assure high availability of the service. These datacenters are particularly costly to run since they demand a lot of electricity and additional power is used to cool the servers and other computational devices. As a result, resource allocation and scheduling procedures in cloud computing must be prioritized.

1.7 PROBLEM STATEMENT

All data files are saved in the cloud storage system in cloud computing. Each data file is broken down into multiple sections. The data are not stored in a contiguous manner, resulting in fragmentation. As a result, unallocated space is not available for other files, and the storage system is squandered. As a result, both the computational and operational costs rise. Due to an imbalanced workload, the task execution time (make span) is long. Furthermore, physical computing servers in cloud datacenters are never idle; they typically function at 10- 15% of their maximum capacity, and usage rarely exceeds 100%. The difficulty with this utilization strategy is that even when the servers are turned down, they consume 70% of their maximum power. This datacenter's energy inefficiency can be avoided by utilizing virtualization technology, which produces numerous virtual machines on a single physical server, maximizing resource use. It can also be avoided by employing live virtual machine migration, which consolidates VMs to a smaller number of servers based on their current resource usage pattern and switches the remaining idle servers to low-power modes like sleep or hibernation. The main issue with live VM migration is that the resource usage pattern of current service applications cannot be foreseen, therefore frequent live VM migration can result in performance degradation and a breach of the service level agreement between the client and the cloud provider. The aforementioned issues have clearly stated the need for efficient resource scheduling, which can efficiently allocate virtual machines to available servers and execute users' requests, such as tasks, on the appropriate virtual machine while maintaining the energy-performance trade-off in the cloud computing system.

2.1 REVIEW OF LITERATURE

P. Pradhan [2020] presented a modified RR method for resource allocation in cloud computing. Cloud computing is a computer model that allows users to access resources on demand. Users of the cloud can access their data and applications at any time and from any location. An organization can lease assets from the cloud for capacity and other computing functions, lowering the overall cost of their framework. By reducing the waiting time, the modified RR asset allotment mechanism satisfies client requests and application requirements. This technique is made up of several aspects, including load prediction, hot spot mitigation, and green computing. The resource allocation in the cloud platform is established on two levels. Initially, the method's application is uploaded to the cloud, and the load balancer is allocated to the computers' requested instances. Finally, the varied sorts of multiple incoming orders or requests are realized, and the computational burden is balanced by assigning those requests to each individual utility or application at a time.

For the process of scaling Software-as-a-Service utility over cloud infrastructures, **Javier Espadas et al. [2020]** created a resource allocation model based on tenant attributes. The Cloud architecture gives essential access to a set of computing resources as well as some basic system business models; fetches systemic providers indefinitely while monitoring service rate. Its underlying infrastructures enable the construction of a variety of virtual machines and instances based on the application's demand. The flexibility to scale up or down the capability of application resources is the key benefit of SaaS providers. The system's ability to consume or pay, which is the most important role in cloud computing at the time, is limited.

When compared to other forms of traditional regular hustings, the approach should be correctly applied, and the cost is lower. Even large-scale applications can be built with the help of a pay-per-use cloud, but the cost and scalability efficiency is not achieved. The reason for this is because idle steps and CPU resources are not utilized yet are charged as resource expenses. The excess and under provisioning of cloud resources is also unresolved, and it has been identified as a problem. Without the use of an elasticity model, the methodology accurately predicts peak loads, but resources in terms of expenses are squandered during nonpeak hours, or potential profits from customers are missing after attaining poor service or saturation. This research has produced a wide range of formal metrics for the provisioning of some virtualized elements of cloud infrastructure, particularly the SaaS platform and its deployments. This strategy caters for multi-tenancy and the cost-effective establishment of scalable domains by using a resource allocation model for the deployment of SaaS applications on cloud computing data.

Anton Beloglazov et al. [2019] offered a scenario of heuristics-based energy-aware resource allocation approaches for major management of information DCs in cloud computing architectures. Cloud computing has provided application-oriented software services to people all around the world. However, DCs and their hosting have consumed a significant quantity of electrical power, resulting in higher operating expenses for the environment. As a result, Green Cloud computing has reduced not only operational expenses but also environmental effect. An architectural framework is defined as a result of this research, as are the principles of energy-efficient approach. Many open access research challenges and resource provisioning methods have been discovered, and the cloud system has been motivated, thanks to the architecture. The allocation heuristics from the DC function to the client set of applications have improved energy in a cost-effective manner. The system's implementation achieves the negotiated approach of Quality of Service (QoS). The conclusion was obtained by conducting a survey and comparing the findings of previous studies. The end result is (a) architectural principles for the management of energy-efficient Clouds; (b) efficient policies for resource allocation and scheduling mechanisms that consider QoS needs and electricity characteristics of the devices used; and (c) a variety of open researches and challenges, with significant benefits to both providers and consumers of resources. The Cloud Sim toolkit and its operations are used to evaluate and judge the performance of the generated algorithm.

3.1 OBJECTIVES OF THE STUDY

- An efficient cloud data resource allocation methodology is used, which is based on an automatic cloud defragmentation distribution technique.

4.1 PROPOSED METHODOLOGY

For handling multi-cloud storage in cloud computing, an automated cloud storage dynamic defragmentation model is presented. The suggested defragmentation methodology allows for the distribution of data over several cloud servers, ensuring that no unallocated cloud space remains underutilized and is converted into useable resource space. The proposed model is explained in the section below. The defragmentation model, on the other hand, boosts the cloud market by giving clients some more storage space with the allocated resource. This approach for a single provider with multiple consumers likewise extends to a multi-provider environment.

4.2 DEFRAGMENTATION BASED MODEL

Defragmentation employs fragmentation, which is a procedure that reorganizes data on a disc into contiguous squares. Because less effort is necessary to get at content, this results in better disc drive and overall framework performance, resulting in lower Input/Output (I/O) requirements. Defragmenting during read/write operations extends the life of the disc drive, putting less strain on the device. Defragmenting is an important technique for any server, client system, or other disk-based device. As a result, disc access times are significantly slower than memory access times.

Defragmentation-based model in cloud environment consists of three parts: cloud providers, cloud consumers, and the broker. Users tell the broker what kind of resource they want and how much they want to bid on it. The broker performs the most important function, and the benefit of utilizing a broker is that the broker will have all of the auction's trade information. The broker majorly have two responsibilities such as, initially calculate the highest bidder for the respective type of resource and second to allocate the resource to highest bidder. Defragmentation is performed on the allocated storage if necessary.

4.3 ALGORITHM

Start

Input: VM settings (v); C: Capacity to allocate.

Result: Allocation Computed for VM.

Variables: i=1…n, a, b, u, v

For each i=1…n do

a(i)=v(i)

b(i)=u(i)-v(i)

E=C-Sum (v (i).

For each I=1…n do a(i)=a(i)+((b(i)*E))/Sum(b(i))).

End

5.1 EXPERIMENTAL RESULT AND DISCUSSION

The Java platform is used to implement the suggested automatic dynamic defragmentation model and load prediction algorithms. For the cloud defragmentation model simulation, all of the procedures have been designed to the best possible degree. For the entire web development, the proposed model utilizes open-source development tools like Eclipse JEE, while the server backend is Tomcat. We were able to achieve our goal by utilizing sophisticated tools such as Eclipse JEE and Tomcat. We may use the open source Eclipse JEE to create dynamic enterprise-based web applications. To create a web application, Apache Tomcat is used, which is a framework that runs on both a host and a server. To run java scripts, a java HTPP web server environment is necessary, which is configured at local host 8080.

5.2 PERFORMANCE ANALYSIS

In terms of energy usage, live migration, and SLA Violation, the suggested automatic defragmentation cloud storage model is compared to the existing physical cloud storage model.

TABLE-1.1 ENERGY CONSUMPTION PERFORMANCE OF EXISTING AND PROPOSED CLOUD STORAGE MODELS

Number of Work loads	Existing Cloud Storage Model	Proposed Cloud Storage Model
1	24.43	22.54
2	35.79	33.90
3	22.24	21.35
4	20.35	17.66
5	25.36	22.46
6	28.90	27.74

The amount of energy required to finish the operation varies depending on the number of workloads. The energy consumption in the DC varies depending on the amount of tasks. The proposed automated defragmentation cloud storage strategy reduces the DC's energy usage significantly.

The graph depicts the performance of the traditional cloud storage model and the proposed cloud storage model in terms of energy consumption. In various types of workloads, the proposed defragmented model consumes less power. The defragmented cloud storage model outperforms the conventional cloud storage paradigm. The overall cost of the proposed model will be reduced if the power usage is lower.

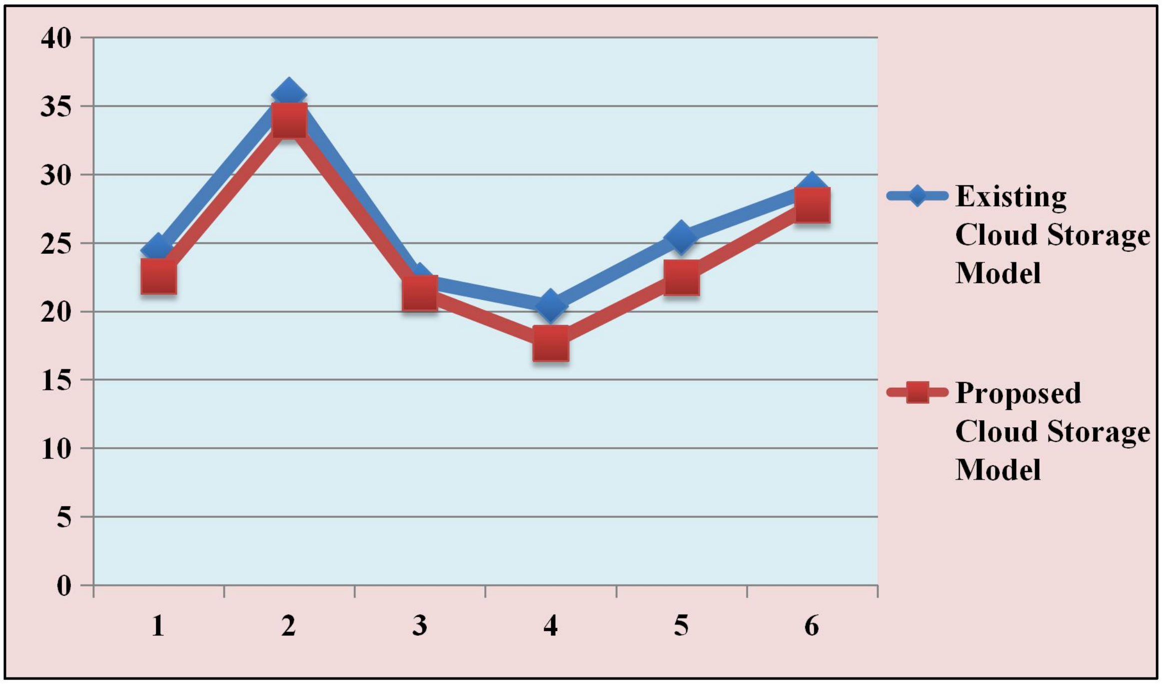

FIGURE.1.2: ENERGY CONSUMPTION PERFORMANCE

TABLE – 1.2 PERFORMANCES OF EXISTING AND PLANNED CLOUD STORAGE MODELS FOR LIVE MIGRATION

Number of Work loads	Existing Cloud Storage Model	Proposed Cloud Storage Model
1	2690	3346
2	1780	2770
3	1885	3696
4	1445	2385
5	1470	3226
6	1645	2669

The performance of the number of live migrations is shown in Table. The live migrations differ depending on the workload. The migration schedule varies during each consolidation period. During the first consolidation period, a higher number of live migrations reduce the number of active physical computers. The number of live migrations falls during the other consolidation period.

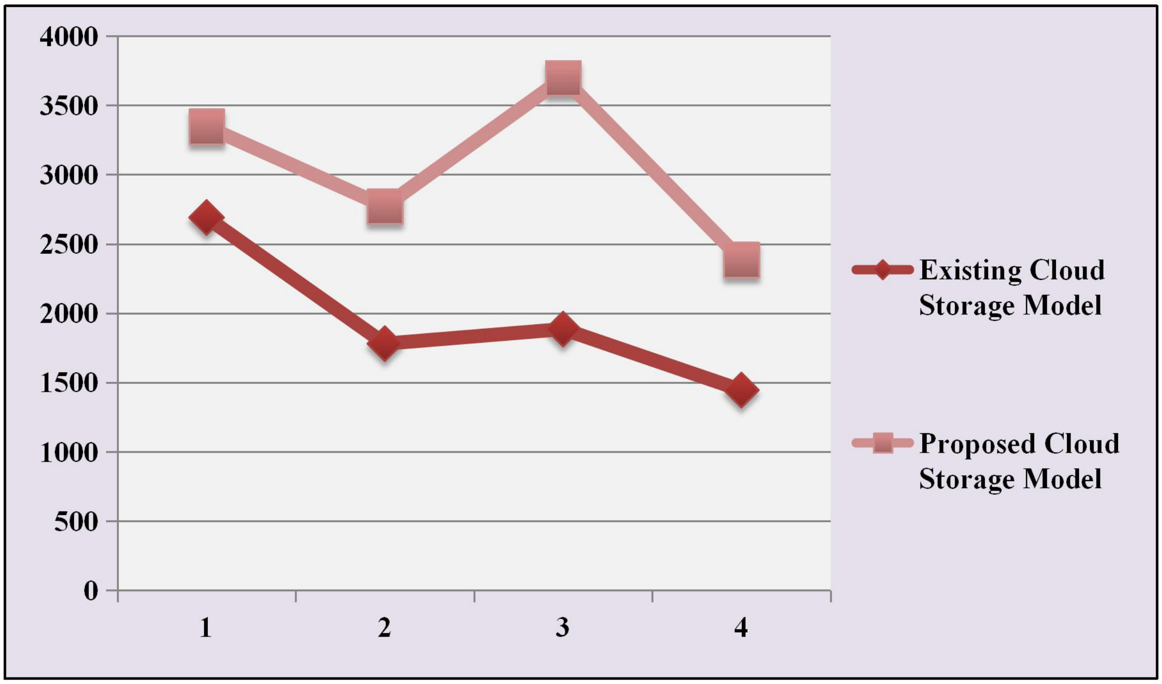

FIGURE-1.3. PERFORMANCE OF LIVE MIGRATIONS Six different workloads are used to test the performance of live migration in various models. If numerous physical servers become available, live migration from less-used to more-used physical servers occurs. In this case, the VM is migrated many times for different iterations. Figure shows that the proposed cloud storage strategy outperforms the existing cloud storage model in terms of live migration.

The cloud provider can manage and distribute user data more efficiently thanks to the automatic dynamic defragmentation approach. This model divides data into manageable chunks and stores them in cloud-based distributed servers. Unallocated cloud space is converted to useful resource space using the defragmentation paradigm. The broker model plays a crucial role in resource allocation with a single provider in this paradigm, which explains the defragmentation model. It's also detailed how the entities communicate during resource allocation. In the future, it would be interesting to work on multi-provider and multi-user scenarios to adaptively alter resource prices and storage space to examine the benefits of the defrag model over a fixed-price and-space model for resource allocation.

The proposed cloud load balancing system for automated dynamic defragmentation. Load balancing functions as a layer between the VMs and the client, allowing the task in demand to be executed as quickly as feasible. The scheduled load balancing between the available VMs clearly saves money and energy. The results of the experiments suggest that load balancing can improve performance and reduce the spread in a cloud computing environment.

6.1 CONCLUSION

Cloud computing is a rapidly evolving processing paradigm that provides clients with whatever they need as a service. Cloud computing has evolved as a viable means of allocating resources such as memory, storage, and processors for the execution of tasks and administrations for a geographically scattered group of suppliers. Every client benefits from the suggested distributed storage defragmentation paradigm, which improves consistency, accessibility, and cloud data storage selections. In a cloud context, an energy-efficient load balancing solution balances workloads and decreases the make span. The outcomes of effective job scheduling techniques play a critical role in reducing energy use. When comparing the task consolidation approach for scheduling tasks to the VM to current ways, the generated result clearly indicates how the suggested approach outperforms the existing technique. When compared to the results of the original LRR-MMT approach, modifications made to the task scheduler of the LRR-MMT technique using the task consolidation approach, the experiment results show that the number of VMs to be migrated is significantly reduced, resulting in lower energy consumption in the DC. The GA algorithm-based VM allocation strategy maps the VM and assigns the task to the appropriate hosts, maximizing resource usage and reducing cloudlet execution time in a cloud environment.

6.2 FUTURE WORK

In a cloud computing environment, a defragmentation-based dynamic storage allocation mechanism improves the cloud market by providing users with additional storage capacity. It can be extended in the future for many providers with multiple users to alter the resource price and storage space for fixed assigned resources.

REFERENCES

- Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment Chunlin Li • Layuan Li Published online: 7 ebruary 2013 © Springer Science+Business Media New York 2013.
- Chunlin Li • Layuan Li Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment Published online: 7 February 2013 © Springer Science+Business Media New York 2013.
- SaaS (Software as a Service) On Cloud by Sanjay Chaudhary International Conference on Modern Trends in Science, Engineering and Technology 2013.
- Chaisiri S, Lee B, Niyato D (2012) Optimization of resource provisioning cost in cloud. IEEE Trans Serv Comput 5(2):164–177.
- Rimal BP, Choi E (2012) A service-oriented taxonomical spectrum, cloudy challenges and opportu-nities of cloud computing. Int J Commun Syst 25(6):796–819.
- Younge AJ, von Laszewski G, Wang L (2012) Efficient resource management for cloud computing environments.
- Garg SK, Yeo CS, Anandasivam A, Buyya R (2011) Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers. J Parallel Distrib Comput 71(6):732–749
- Gulati A, Shanmuganathan, G, Holler A (2011) Cloud scale resource management: challenges and techniques. In: USENIX Hot Cloud, Portland.
- Kouki Y, Ledoux T, Sharrock R (2011) Cross-layer SLA selection for cloud services. In: First international symposium on network cloud computing and applications, pp 143–147
- Wu L, Garg SK, Buyya R (2011) SLA-based resource allocation for software as a service provider in cloud computing environments. In: Proceedings of the 11th IEEE/ACM international symposium on cluster computing and the grid (CCGrid 2011), Los Angeles, USA, 23–26 May 2011
- Addis B, Ardagna D, Panicucci B (2010) Autonomic management of cloud service centers with availability guarantees. In: IEEE 3rd international conference on cloud computing, pp 220–227
- Chang F, Ren J, Viswanathan R (2010) Optimal resource allocation in clouds. In: IEEE 3rd international conference on cloud computing, pp 418–425
- Ferretti S, Ghini V, Panzieri F, Pellegrini M, Turrini E (2010) QoS–aware clouds. In: IEEE 3rd international conference on cloud computing, pp 321–328

- Lee YC, Wang C, Zomaya AY, Zhou BB (2010) Profit-driven service request scheduling in clouds. In: 10th IEEE/ACM international conference on cluster, cloud and grid computing, pp 15–249
- Mohd Yusoh Z, Tang M (2010) A cooperative co evolutionary algorithm for the composite SaaS placement problem in the cloud. In: Proceedings of the neural information processing, theory and algorithms, pp 618–625
- Underutilizing resources for HPC on clouds by J. Napper, P. Bientinesi, R. Iakymchuk Aachen Institute for Advanced Study in Computational Engineering Science Preprint: AICES-2010/06-1 10/June/2010