Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

CLOUD BROKERS SYSTEM FOR CLOUD SERVICE PROVIDERS IN CLOUD COMPUTING

NAGASATISH AISETHY[1], Dr. B. V. S. VARMA[2], Dr. G. SATYANARAYANA[3]

#1 M.Tech Scholar and Department of Computer Science Engineering,
#2 Professor, Department of Computer Science and Engineering, DNR College Of Engineering and Technology, Bhimavaram, AP, India.
#3 Professor, HOD Department of Computer Science and Engineering, DNR College Of Engineering and Technology, Bhimavaram, AP, India.

Abstract:

Cloud is one of the emerging technologies in computer engineering. Several companies move around on the way to this technology due to lessening maintenance cost. Numerous organizations offer cloud service such as SaaS, IaaS, PaaS. Different organization provides same service with different service charges and waiting time. So customers can select services from these cloud providers according to their criteria like cost and waiting time. By using 'demand pricing' strategy, providers can provide services with minimum cost without loosing any income or valuable resource time. But the existing system does not provide any automated job scheduling considering consumer cost, provider benefit, consumer waiting and provider idle time. This paper proposes a multi objective genetic algorithm for solving this multivariable optimization problem. This system provides a new cloud brokering mechanism with cloud service discovery using this optimization technique. This paper considers IaaS. In this system user submit a job to cloud. Cloud provides infrastructure to run this job and gave output to user. Here aim of user is to obtain output with minimum time and minimum cost. At the same time aim of provider is to increase the income. For that provide run more job within unit time. So We have to minimize consumer cost, consumer waiting time and provider idle time, and maximize provider benefit.

Keywords— Cloud computing, multiserver system, pricing model, profit, response time, ,service charge, SLA, waiting time, server configuration, QoS

I. Introduction

The cloud is a next generation platform that provides dynamic pools of resource, virtualization, and high accessibility. Today, we utilize scalable, distributed computing environments within the margins of the Internet put into practice known as cloud computing. Cloud computing helps to decipher the daily computing problems, of hardware resource and software availability unhurried by computer users. The cloud computing provides a straightforward and non-ineffectual solution for daily computing. Prevailing cloud systems mainly focus on finding an effective solution for the resource management.

Cloud Computing is Internet based computing where virtual shared servers provide software, infrastructure, devices, platform and other resources .They also provide hosting to customers on a pay-as-you-use basis. The cloud makes it possible for user to access your information from anywhere at any time. Cloud Computing enables a user what you need and pay for what you use model. This will enable businesses to invest on innovative solutions that will help them address key customer challenges instead of worrying about operational details.

"Cloud computing is a model that enable suitable, on-demand network access to a pool of shared configurable computing resources (e.g., servers, storage, networks, applications, and services) that able to be rapidly provisioned and released with minimal management effort or service provider interaction."

More specifically, cloud describes the use of a set of information, services, applications, and infrastructure.

Infrastucture in cloud mainly consist of a great pool of storage, computer, network, information, and resources. These can be rapidly scaled up or down providing an on-demand utility-like model of allocation and consumption. Cloud enhances collaboration, scaling, availability, and provides the potential for cost reduction through optimized and efficient computing.

Dynamic resource management is one of the crucial problems in the existing cloud environment due to changing resource demands of large computational tasks. They require knowledge of the resource needs for service requests of various kinds, and these needs may change over time. The blend of heterogeneous computing and cloud computing is emerging as a influential new standard to meet the requirements for high-performance computing (HPC).Cloudbased, heterogeneous[8] computing represents a significant step toward solving large computational tasks. Efficient load balancing[21] in a cloud is challenging since running machines have the problem of load imbalance due to resource variation in heterogeneous environment. For heterogeneous systems nodes have different processing capabilities, dynamic load balancing methods are preferred. This approach makes load balancing decision based on the current load status which varies on each machine . distribute load on the nodes at run time Modern parallel computing hardware demands increasingly specialized attention to the details of scheduling and load balancing across heterogeneous execution resources in cloud.

This research work mainly concentrating on developing an efficient dynamic load balancing method in heterogeneous cloud environment and hence achieve effective utilization of resources and thereby increase the performance of the system. , we propose a new framework for calculation provider benefit ,consumer benefit, consumer cost , quality of service and power consumption.

II. Related work

Zahra Ali,Raihan ur Rasool,Peter Bloodsworth. “Social Networking for Sharing Cloud Resources”. In our approach we have linked a social network with the computational cloud to create a social cloud (SC) so that users can share their part of the cloud with their social community. A prototype system has been deployed on a social network by using the bartering resource trading mechanism. It is anticipated that this may help users to share their dedicated resources without the need for money changing hands and different communities. D. Shivalingaiah, Sheshadri K N.“Applications of Cloud computing for resource sharing in academic libraries”. In this paper authors have made an attempt to discuss the types, applications, advantages and disadvantages of implementing cloud computing mainly for resource sharing in an academic library environment. The purpose of this article is to look specifically at how cloud computing can be employed by libraries and what are the potential areas need to be considered before moving into a cloud computing solution implementation. Ashwin R. Bharambe Cormac Herley Venkata N. Padmanabhan. “Analyzing and Improving a BitTorrent Network’s Performance Mechanisms”. Our results confirm that BitTorrent performs near-optimally in terms of uplink bandwidth utilization, and download time except under certain extreme conditions. We also show that low bandwidth peers can download more than they upload to the network when high bandwidth peers are present. We find that the rate-based tit-for-tat policy is not effective in preventing unfairness. We show how simple changes to the tracker and a stricter, block-based titfor-tat policy, greatly improves fairness. A Profit Maximization Scheme with Guaranteed Quality of Service in Cloud Computing. 2008.In a cloud computing environment, there are always three tiers, i.e., infrastructure providers, services providers, and customers. An infrastructure provider maintains the basic hardware and software facilities. A service provider rents resources from the infrastructure providers and provides services to customers. A customer submits its request to a service provider and pays for it based on the amount and the quality of the provided service. In this paper, we aim at researching the multiserver configuration of a service provider such that its profit is maximized. Like all business, the profit of a service provider in cloud computing is related to two parts, which are the cost and the revenue. For a service provider, the cost is the renting cost paid to the infrastructure providers plus the electricity cost caused by energy consumption, and the revenue is the service charge to customers. In general, a service provider rents a certain number of servers from the infrastructure providers and builds different multiserver systems for different application domains. Each multiserver system is to execute a special type of service requests and applications. Hence, the renting cost is proportional to the number of servers in a multiserver system. Profit Maximization for Geographical Dispersed Green Data Centers, 2016".Complying with such a growing demand in an environmentally friendly manner calls for innovations across different disciplines. Therefore, recently, studies on data centers have focused on reducing the energy consumption and accordingly the cost of electricity. These studies can be largely categorized into two main approaches: power management techniques and green data centers. The rst approach, which investigates CPU and memory power consumption, aims at reducing the carbon footprints and the cost of electricity. In particular, a Dynamic Voltage/Frequency Scale (DVFS) scheme is deployed in to reduce CPU power, and Felter et al. in proposed re-budgeting the available power between processor and memory to maintain a server budget within constrained power budgets.

III. Proposed Methodology

The pricing model of a service provider in cloud computing is based on two components, namely, the income and the cost. For a service provider, the income is the service charge to users, and the cost is the renting cost plus the utility cost paid to infrastructure vendors. A pricing model in cloud computing includes many considerations, such as the quantity of a service (the amount of a service), the application environment's workload, of a multiserver system's configuration(the size and the speed), the penalty cost of a low-quality service, the consumer's satisfaction (the expected service time),the SLA, the quality of a service (the task waiting time and the task response time), , the cost of renting and energy consumption, the service provider's margin and profit. The profit (i.e., the net business gain) is the income minus the cost. In order to maximize the profit, it is essential that a service provider should be aware of both service charges and business costs, and also, how they are resolute by the of the application's characteristics and the multiserver system's configuration..

A.The Model

This system provide a pricing model, that takes following factors such as the workload of each and every application environment, the multiserver system's configuration, , the SLA , the approval of a consumer, the penalty cost of a low-quality service , the amount of a service, the QoS parameters of a service, the cost of renting, the cost of energy consumption.It also considers the service provider's margin and profit too.The Multiserver system is treated as an M/M/m queuing model, so as to solve the profit maximization problem analytically.

Main objective of this project is to allocate user submitted job to one of these N servers based on following criteria

1. Power consumption
2. Quality of service
3. Consumer satisfaction and mainly maximum profit

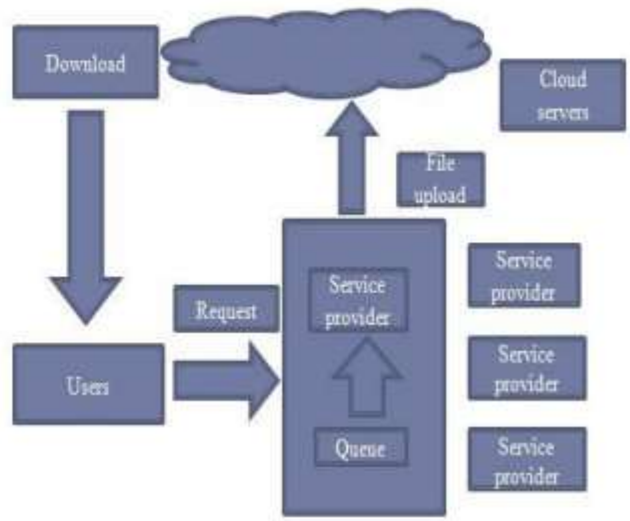

Fig. Proposed Architecture diagram

To achieve the objective of adaptive resource allocation for satisfying the service requests of customers, we use the architecture namely cloud booster archietecture.

B. M/M/m queuing model

The multiserver system is treated as an M/M/m queuing model, so as to formulate and solve the optimization problem can analytically . The system considers two server speed and power consumption models. They are the idle-speed model and the constant-speed model. The the waiting time of a newly arrived service request is obtained by calculating the probability density function . from that function the expected service charge to a service request is calculated. From the expected service charge to a service request calculate the expected net business gain in single unit of time, and thus determine the optimal server size and the server speed numerically.

c. Ant Colony Optimization

This section describes the ACO algorithm, which can be used for proper scheduling . Ant colony optimization is based on the technique known as Swarm Intelligence, which is a part of Artificial Intelligence.The ACO system contains two rules:

1. Local pheromone update rule, which applied whileconstructing solutions. 2. Global pheromone updating rule, which applied after all ants construct a solution.Furthermore, an ACO algorithm includes two more mechanisms: trail evaporation and, optionally, daemon actions. Trail evaporation decreases all trail values over time, in order to avoid unlimited accumulation of trails over some component. Daemon actions can be used to implement centralized actions which cannot be performed by single ants, such as the invocation of a local optimization procedure, or the update of global information to be used to decide whether to bias the search process from a non-local perspective At each step, each ant computes a set of feasible expansions to its current state, and moves to one of these in probability. The probability distribution is specified as follows. For ant k, the probability of moving from state t to state n depends on combined checking of two values :• the ant movement attractiveness , it is computed by some heuristic indicating the priori desirability of that move;• the movement , this indicates how capable it has been in the past to make that particular move: therefore it represents a posteriori indication of the movement desirability.

D. Cloud Booster Architecture

We use the following architecture To achieve the objective of adaptive resource allocation for satisfying the service requests of customers .It consist of mainly the following

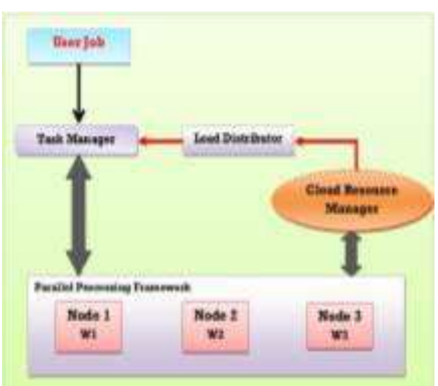

Figure 2: Cloud Booster Architecture

Users/Brokers: Users or brokers acting on their behalf submit service requests to the cloud via cloud controller for processing.

Cloud Controller: It acts as the interface between the cloud service provider and external users/brokers. It acts similar to the Queen in the ant colony. Virtual Machines(VMs): This is where the applications of customers will be deployed. We can dynamically create, start, stop and migrate these VMs depending on our requirement, from one physical machine to another. Physical Machines: These are the physical computing servers that will provide hardware infrastructure for creating virtual machines.

Cloud controller maintains a queue(Q) for storing the service requests for hosting the applications. It enqueues each of the service request received, in this queue. It generates the tester, scout, cleaner and worker ants periodically. The movement of these ant agents is modelled in the following way.

Each ant except Queen & Worker maintains a Visited Node list which is initially empty. Each node in the cloud maintains a list of neighbouring node's information. Whenever an ant reaches a node, it updates the controller about the current utilisation and randomly chooses an unvisited neighbouring node. When all the nodes are covered, it makes the Visited Node list empty and continues again in the same way.

We can change the number of ants that will be produced so that it will yield better results depending on our requirement. The next subsection describes the method used by worker ants for accepting or rejecting the service requests.

Worker ant:

Whenever a service request received in the queue, one of the worker ants creates a VM with a specific CPU processing power and memory etc, if accepted. So, worker ants are always looking in the queue to check if there are some pending requests to be processed. If such a request is found, it dequeues the request and calls Algorithm 1.

Since most of the CPUs are work conserving, we are creating a VM with specific CPU processing power and memory. Depending on the load, more intensive applications can use the resources of the other VMs having less load. The worker ant is only responsible for deploying the request on a VM . Load balancing decisions are taken by tester ant. After deploying, it creates a Service Level Agreement(SLA) monitor agent that monitors the hosted application. In the next subsection, we provide the details about the SLA monitor agent.

SLA Monitor Agent:

It calculates the Avg. response time and throughput of the hosted application by continuously monitoring it. It passes this information to the hypervisor on that host in the form of a variable(SLAM) which is calculated depending on the performance of the application.

IV. Conclusion and future scope:

In this project a brand-new approach referred to as resources allocation deals with methodology for service suppliers is projected to confirm the standard of service demands yet on enhance the profit and to reduce the wastage of resources, this methodology includes each the short term dealing methodology yet as long run dealing methodology. To vary the system sizes, a queuing model is enforced for multi-server system. The information hosting formula is enforced before the DQG methodology, to form the appropriate selection of cloud yet on enhance the profit of service supplier. After that, a difficulty of best configuration of profit increasing is outlined in these many aspects area unit determined. This projected double-dealing strategy is enforced to heterogeneous cloud setting. we have a tendency to use double dealing system so as to extend the profit of supplier. This project conjointly focuses on the scope of improvement in each strategy followed in regard with resource allocation methodologies in cloud setting. to reinforce the standard services to the consumer, a tool is to be developed that may track the information usage and storage areas of all the users and supply flexibility within the dealing to the consumer.

References

[1]P.deLangenandB.Juurlink,“Leakage-aware multiprocessor scheduling,” J. Signal Process. Sys. vol. 57, no. 1, pp. 73–88, 2009.

[2]G. P. Cachon and P. Feldman, “Dynamic versus static pricing in the presence of strategic consumers,” Tech. Rep., 2010.

[3]Y. C. Lee, C. Wang, A. Y. Zomaya, and B. B. Zhou, “Profitdriven scheduling for cloud services with data accessawareness,” J. Parallel Distr. Com., vol. 72, no. 4, pp. 591–602, 2012.

[4]M. Ghamkhari and H. Mohsenian-Rad, “Energy and performance management of green data centers: a profit maximization approach,” IEEE Trans. Smart Grid, vol. 4, no. 2, pp. 1017–1025, 2013.

[5]Odlyzko, “Should flat-rate internet pricing continue, ”IT Professional, vol. 2, no. 5, pp. 48–51, 2000.

[6]G. Kesidis, A. Das, and G. de Veciana, “On flat-rate and usage-based pricing for tiered commodity internet services,” in 42nd Annual Conf. Information Sciences and Systems. IEEE, 2008, pp. 304–308.

[7]S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, “The price of simplicity,” IEEE J. Selected Areas in Communications, vol. 26, no. 7, pp. 1269–1276, 2008.

[8]H. Xu and B. Li, “Dynamic cloud pricing for revenue maximization,” IEEE Trans. Cloud-Computing, vol. 1, no. 2,pp. 158–171, July 2013.

[9]M. Mac´ıas and J. Guitart, “A genetic model for pricingin cloud computing markets,” in Proc. 2011 ACM Symp. Applied Computing, 2011, pp. 113–118.

[10]“Amazon EC2,” http://aws.amazon.com, 2014.

Authors

NagaSatish Aisethy pursuing M.Tech in Department of Computer Science and Engineering from D.N.R College of Engineering & Technology, Bhimavaram, Andhra Pradesh, West Godavari District, 534201, India. His area of interest in Cloud Computing, Services.

Dr. B.V.S Varma is working as Professor & Vice-Principal in the Department of Computer Science and Engineering in D.N.R College of Engineering & Technology, Bhimavaram, Andhra Pradesh, West Godavari District, 534201, India.

Dr. G. Satyanarayana is working as professor & HoD in the Department of Computer Science and Engineering in D.N.R College of Engineering & Technology, Bhimavaram, Andhra Pradesh, West Godavari District, 534201, India.