Qos and Load Balancing in Cloud Computing- an Access for Performance Enhancement using Agent Based Software

Geeta, Santosh Gupta, Shiva Prakash

Abstract:- Cloud computing, one of the advanced and emerged technologies in the field of computer science has been embraced by different organizations of various sizes. The purpose of organizations moving into cloud is manifold, out of which , performance enhancement and cost optimizer are the primary ones. Generally, when an organization moves their operations into cloud, Cloud Service Providers (CSPs) provision various machine images to different users based on their requirements within the organization. Also, CSPs, potentially offer Anything as a Service (XaaS) to organizations with the help of distributed and connected server farms available at geographically separate locations. QoS parameters in terms of service time, as specified in the Service Level Agreements(SLAs) between CSPs and organizations must be adhered strictly. As an effort towards maintaining QoS, within the cloud, the operational approach of load balancing across multiple distributed servers play a vital role. This paper presents a novel load balancing algorithmic framework with the help of software agent that runs in the gateway system between cloud consumers and cloud service providers. This software agent in the gateway system is vested with the responsibility of diverting the incoming work process to the appropriate servers, based on their current workload and resource utilization. The efficiency of this approach is tested using CLOUDSIM by creating different number of cloudlets and hosts.

KEYWORDS: Cloud , SLAs, Load balancing, QoS, CloudSim, Resource utilization, Agent

I. INTRODUCTION

In today's information technology driven world, cloud computing and big data adoption by businesses decide their competitive ranking. Especially, every business is expected to provide better QoS [13] than their competitors through maximum throughput with minimum response time. In a cloud environment, when a client gives a service request to the service providers, the distributed server farm on the CSP side, must ensure provisioning of appropriate services.The cloud consumers and cloud service providers are generally connected through communication backbone network. An important component in the basic cloud architecture as shown in Fig.1 is the Distributed Server Farm (DSF). DSF can be characterized as a geographically distributed server farm, where in a single farm, many servers are connected with each other using appropriate networking cables and protocols. In the context of distributed server farm, one of the important tasks to be carried out by the CSPs is load balancing across multiple servers. Load balancing is very much essential in cloud's operation as large amount of data are exchanged between client and server , server and server, server and storage. The benefits of the cloud can be leveraged with the virtualization technology. Load balancing [13] in cloud computing indicates the process of distributing workloads and computing resources across distributed servers. Load balancing in any system is generally maintains system firmness, improves system performance and protects against system failures. Azure's Traffic Manager [1] and AWS's Elastic Load Balancing (ELB) [2] technology are some of the representative examples of cloud load balancers. In this paper, a novel effort has been made to study the effect of using an Agent based load balancer running in a separate system considered to be a gateway.

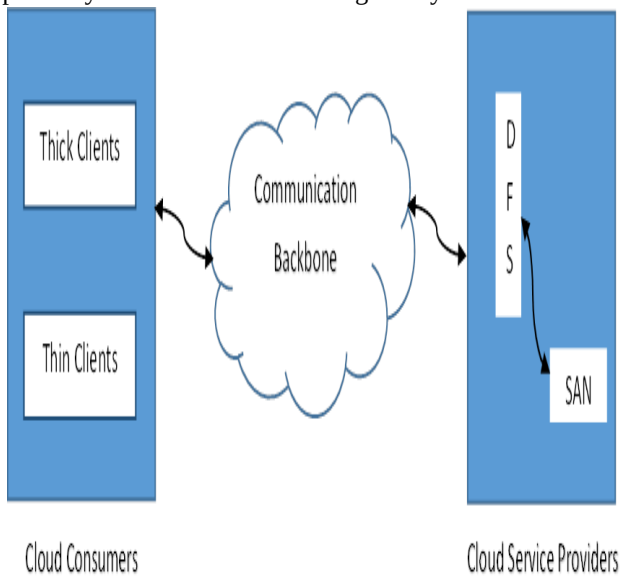

Fig.1 Basic Cloud Architecture

The rest of the paper is organized as follows:

Section II deals with related literature pertaining to load balancing and corresponding QoS metrics in Cloud computing.

Section III deals with the design principles of gateway system which has agent based load balancer installed.

Section IV deals with the experimental simulation and corresponding results.

Section V provides conclusion and future work.

II. RELATED STUDY

This section of the paper discusses the various work that were carried out related to improvement of QoS and various approaches adopted to implement novel load balancing algorithms.

Geeta et al.[13] had discussed various QoS and Load balancing Techniques and also had addressed key performance challenges and different modeling with their applications for QoS management and simulation toolkits in cloud computing.

Devi.D.C. et al.[3], had implemented a load balancing method which utilized improved weighted round robin algorithm for non pre-emptive task. In this work, the authors have taken the capabilities of virtual machine into account and have followed a 3 stage process, that included static scheduler, dynamic scheduler and load balancer.

Domanal et al.[4], optimized the load balancing in cloud using virtual machine by improving the utilization of VM. The authors have taken an approach of efficiently allocating the incoming service requests to various virtual machines.

Maguluri et al.[5], utilized the approach of stochastic processes in load balancing and dynamic scheduling in cloud computing. The arrived jobs based on stochastic process rerquest for virtual resources and the resource allocation is purely performed based on load balancing and scheduling. The approach used , improves the throughput.

Xu et al.[6], developed an intelligent load balancer for cloud computing using classic MapReduce model using agent aid layer and abstracting work load requests through tokens. MapReduce is utilized by the authors for complex job decomposition. In this work, classic mapReduce is utilized along with agent software code for appropriate allotment of resources based on service requests.

Gasior et al.[7], utilized spatially generalized prisoner's dilemma game for effective coalition formation. It is a game theoretic approach in two dimensional cellular automata space. It works by formations of temporal coalitions of participants.

ViolettaN.Volkova et al.[8], had analysed various load balancing algorithms in cloud using cloud analyst and a comparison chart is provided.

Younis et al.[9], followed a hybrid approach for load balancing algorithm in Heterogeneous cloud environment. The authors have considered the current resource information and CPU capacity factor and takes advantage of both random and greedy algorithms. The authors have claimed improved response time and processing time, in comparison with other algorithms.

Geeta et al.[12] reviewed various virtualization techniques in Cloud Computing Environment and also explain about how to maintain the virtualization with optimized resources such as storage, network, application, server, and client in cloud computing.

With these approaches put together, a novel effort is made in this paper to introduce and design an agent code in a separate gateway system that is communication backbone in one way i.e., incoming service requests in the cloud. The approach is simulated using cloud sim [10] and the QoS metrics of response time , throughput and delay were analysed with varying setup as discussed in section IV.

III. DESIGN PRINCIPLES OF GATEWAY IN CLOUD

The entire novelty in the paper is present in the addition of the new component called as Gateway system before the requests can move in and the responses can come out of the CSPs. This component is called gateway, since this system has a piece of software code installed, called as Cloud Service Agent (CSA). This CSA s written and developed using Python libraries and is primarily responsible for load balancing as per the current load informationof various servers in a single cluster. As a client raises a service request, the following steps will be taken before the service can be transferred to the appropriate server:

Client issues a service request to the gateway system.

As the service request reaches the gateway, the request is transferred to CSA through an Application Programming Interface (API)

When CSA receives the service request, it examines the resources required

After the resource manager in CSA, decides on the resources, CSA verifies its resource and load database of various servers in the clusters.

After verification, CSA moves the service request to the appropriate server with optimal load.

CSA's placement in the cloud architecture is shown in Fig.2.

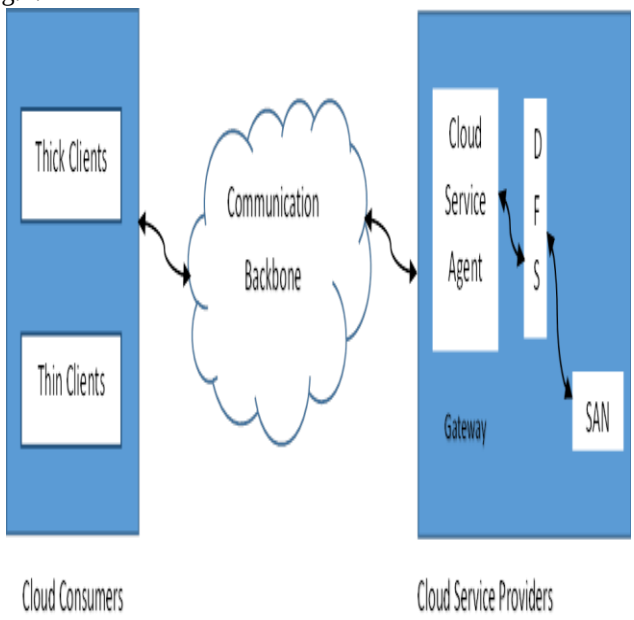

Fig.2 CSA in the Cloud Architecture

CSA's load balancing algorithm functions based on the threshold alue of CPU utilization and RAM usage of various servers. The efficiency of this approach is measured using various QoS metrics such as throughput, latency and service time with many service requests sent to different servers. The experimental simulation of the gateway system within the cloud is carried out and the results are discussed in next section.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experimental simulation of the cloud computing is very much important from the results perspective. We simulated a cloud using CLOUDSIM.

The simulation consists of 100 cloudlets with 100 hosts (which is incrementally varied from 20 cloudlets to 100 cloudlets in steps of 20) along with a single system called as gateway running CSA software, written using Python libraries [11].

Also, the load balancing approach is compared with Domanal etal., and Xu etal.

The obtained throughput per second is measured using Tflops. Latency experienced per request based on different set of requests in multiples of 100 requests and service time per request is analysed using the obtained values as shown in the table. The plotted graph shows that the throughput and service time values of our approach yielded a better value compared to the latency. This is due to the fact that an additional component is introduced in the architecture.

Table 1- Obtained Throughput (in Tera Flops)

No.of Cloud lets	Domanal et al	Xu et al	CSA approach
20	10.23	13.47	17.83
40	9.47	14.54	16.23
60	9.03	12.36	15.45
80	8.87	10.23	13.23
100	8.63	9.87	12.96

Table 2- Obtained Service Time (in milliseconds/request)

No.of Service requests	Domanal et al	Xu et al	CSA approach
100	5.46	6.35	5.03
200	6.45	6.87	5.68
300	6.95	7.13	6.13
400	7.07	7.74	6.84
500	8.83	8.23	7.39

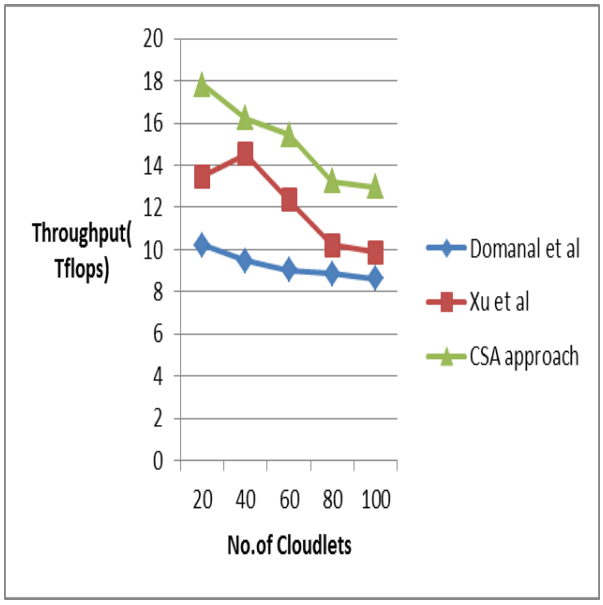

Fig.3 Obtained Throughput

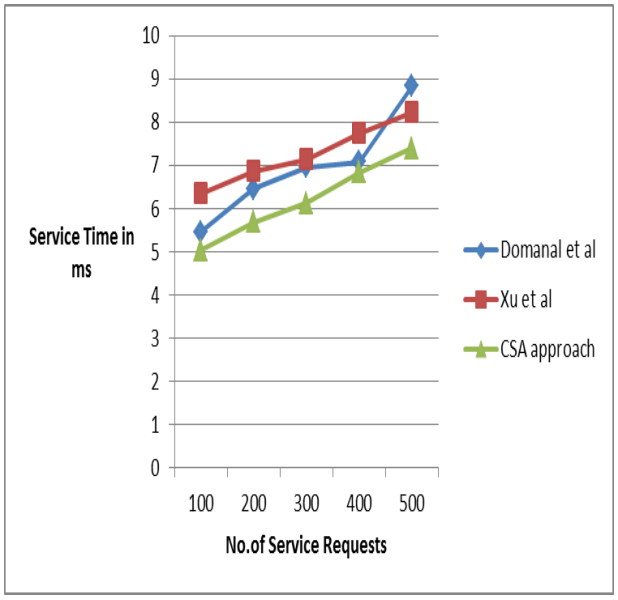

Fig.4 Obtained Service Time

Fig.3 and Fig.4 provide the graphical representation of the obtained values of throughput and service time with respect to two already existing approaches and our CSA approach.

V. CONCLUSION

In this paper, one of the key aspects of cloud computing namely load balancing is taken into account. We have analysed various approaches and a new method in the form of Gateway system with Cloud service agent is utilized to perform resource allocation and load balancing in cloud . The results obtained through simulation and the parameters of QoS were analysed. These results prove that the CSA based approach fared better in terms of throughput and service time as compared against the already existing approaches.

VI. REFERENCES

1) https://azure.microsoft.com/en-in/services/traffic-manager/
2) https://aws.amazon.com/elasticloadbalancing/
3) **D.C.Devi, V.R.Uthariaraj, "**Load Balancing in Cloud Computing Environment Using Improved Weighted Round Robin Algorithm for Nonpreemptive Dependent Tasks", The Scientific World Journal, Volume 2016
4) S.G. Domanal and G.R.M. Reddy, "Load Balancing in Cloud Computingusing Modified Throttled Algorithm," Cloud Computing in Emerging Markets (CCEM), 2013 IEEE International Conference on, pp. 1-5, 2013.
5) S.T.Maguluri, R.Srikant, L.Ying, "Stochastic Models of Load Balancing and Scheduling in Cloud Computing Clusters", 2012 IEEE Infocomm
6) Xu, Y., Wu, L., Guo, L., Yang, Z.C. and Shi, Z. (2008). An Intelligent Load Balancing Algorithm Towards Efficient Cloud Computing, in Proc. of AI for Data Center Management and Cloud Computing: Papers, from the 2011 AAAI Workshop (WS-11-08), pp.27–32.
7) J Gasior, F.Serendynski**, "**A Decentralized Multi-agent Approach to Job Scheduling in Cloud Environment", Advances in Intelligent systems and Computing, 2014,

pp.403-414
8) V.N.Volkova, A.Loginova, E.N.Deistrynikova, "Simulation modeling of a technological breakthrough in the economy", 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)
9) H.J.Younis, A.A.Halees, M.Radi, "Hybrid Load Balancing Algorithm in Heterogeneous Cloud Environment", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-3, July 2015
10) http://www.cloudbus.org/cloudsim/
11) https://docs.python.org/3/library/
12) **Geeta**, Prakash S. (2019) Role of Virtualization Techniques in Cloud Computing Environment. In: Bhatia S., Tiwari S., Mishra K., Trivedi M. (eds) Advances in Computer Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol 760. Springer, Singapore
13) Geeta, Prakash S. (2018) "A Literature Review of QoS with Load Balancing in Cloud Computing Environment" , In: Aggarwal V., Bhatnagar V., Mishra D. (eds) Big Data Analytics. Advances in Intelligent Systems and Computing, vol 654. Springer, Singapore.