

A Survey Paper on Quality of Service in Cloud Computing

Poonam Khot¹, S. D. Satav²

¹M.E (Computer) Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007

²Prof (Computer) Department of Information Technology, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007

Abstract: As Cloud computing turns out to be more prominent, comprehension the financial aspects of Cloud computing turns out to be basically critical. To amplify the benefit, an administration supplier ought to comprehend both administration charges and business costs, and how they are dictated by the qualities of the applications and the arrangement of a multi-server framework. The issue of ideal multi-server setup revenue driven augmentation in a Cloud computing environment is considered. Our valuing model takes such figures contemplations as the measure of an administration, the workload of an application situation, the setup of a multi-server framework, the service-level agreement, the fulfillment of a buyer, the nature of an administration, the punishment of a low-quality administration, the expense of leasing, the expense of vitality utilization, and an administration supplier's edge and benefit. Our methodology is to regard a multi-server framework as a M/M/m lining model, such that our streamlining issue can be figured and illuminated logically. Two server speed and power utilization models are considered, to be specific, the unmoving pace model and the consistent pace model. The likelihood thickness capacity of the waiting time of a newly arrived service request is derived. The expected service charge to a service request is calculated.

Keywords: Cloud computing, multi-server framework, queuing model

1. Introduction

As a viable and proficient approach to unite figuring assets and registering administrations, obfuscating figuring has turned out to be more mainstream. Distributed computing brings together administration of assets and administrations, what's more, conveys facilitated administrations over the Internet. The equipment, programming, databases, data, and all assets are focused and gave to purchasers on-interest. Distributed computing transforms data innovation into normal products and utilities by the pay-per-use valuing model. In a distributed computing environment, there are constantly three levels, i.e., framework suppliers, administrations suppliers, and clients. A base supplier keeps up the fundamental equipment and programming offices. An administration supplier rents assets from the base suppliers and gives administrations to clients. A client presents its solicitation to a administration supplier and pays for it in light of the sum and the nature of the gave administration. In this paper, they go for investigating the multi-server setup of an administration supplier such that its benefit is expanded. Like all business, the benefit of an administration supplier in cloud registering is identified with two sections, which are the expense and the income. For an administration supplier, the expense is the leasing 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 multi-server systems for different application domains. Each multi-server 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 multi-server system. The power consumption of a multi-server system is linearly proportional to the number of servers and the server utilization, and to the square of

execution speed. The revenue of a service provider is related to the amount of service and the quality of service. To summarize, the profit of a service provider is mainly determined by the configuration of its service platform. To configure a cloud service platform, a service provider usually adopts a single renting scheme. That's to say, the servers in the service system are all long-term rented. Because of the limited number of servers, some of the incoming service requests cannot be processed immediately. So they are first inserted into a queue until they can handled by any available server. However, the waiting time of the service requests cannot be too long. In order to satisfy quality-of-service requirements, the waiting time of each incoming service request should be limited within a certain range, which is determined by a service-level agreement (SLA). If the quality of service is guaranteed, the service is fully charged, otherwise, the service provider serves the request for free as a penalty of low quality. To obtain higher revenue, a service provider should rent more servers from the infrastructure providers or scale up the server execution speed to ensure that more service requests are processed with high service quality. However, doing this would lead to sharp increase of the renting cost or the electricity cost. Such increased cost may counterweight the gain from penalty reduction. In conclusion, the single renting scheme is not a good scheme for service providers. In this paper, we propose a novel renting scheme for service providers, which not only can satisfy quality-of-service requirements, but also can obtain more profit.

2. Literature Survey

In paper [1] author explained why a Firm may prefer static pricing over dynamic pricing when consumers are strategic and decide whether to consider to purchase based on the Firm is chosen pricing strategy. By charging a static price a

Firm imposes a rationing risk on consumers whereas a Firm that changes prices dynamically imposes a price risk on consumers. Imposing a rationing risk on consumer can dominate, especially when consumers valuations for the product are highly variable. The problem with dynamic pricing is that the Firm may charge a high price that leaves consumers with zero surplus, so the Firm can improve its revenues by implementing a pricing strategy that leaves consumers with a positive surplus in all states of demand. Overall, we conclude that even though dynamic pricing responds better to demand conditions, charging a static price can be the preferable pricing strategy when consumers are strategic. However, constrained dynamic pricing is an even better strategy - charge either a reasonable list price or mark down from that list price, but never mark up.

In paper [2] author explained the problem of optimal multi-server configuration for profit maximization in a cloud computing environment. the approach is to treat a multi-server system as an M/M/m queuing model, such that our optimization problem can be formulated and solved analytically. they consider two server speed and power consumption models, namely, the idle-speed model and the constant-speed model. the main contributions are as follows. they derive the probability density function (pdf) of the waiting time of a newly arrived service request.

In paper [3] author explained Distributed computing, the long-held long for processing as an utility, can possibly change an expansive piece of the IT industry, making programming much more alluring as an administration and forming the way IT equipment is composed and obtained. Engineers with inventive thoughts for new Internet benefits no more require the expansive capital expenses in equipment to send their administration or the human cost to work it. They need not be worried about over provisioning for an administration whose notoriety does not meet their expectations, in this way squandering immoderate assets, or under provisioning for one that turns out to be fiercely famous, in this way missing potential clients and income. Also, organizations with huge group arranged undertakings can get results as fast as their projects can scale, since utilizing 1000 servers for one hour costs close to utilizing one server for 1000 hours. This flexibility of assets, without paying a premium for expansive scale, is remarkable ever.

In paper [4] author explained As Clouds are designed to provide services to external users, providers need to be compensated for sharing their resources and capabilities. In this paper, we have proposed architecture for market-oriented allocation of resources within Clouds. We have discussed some representative platforms for Cloud computing covering the state-of-the-art. We have also presented a vision for the creation of global Cloud exchange for trading services. The state-of-the-art Cloud technologies have limited support for market-oriented resource management and they need to be extended to support: negotiation of QoS between users and providers to establish SLAs; mechanisms and algorithms for allocation of VM resources to meet SLAs; and manage risks associated with the violation of SLAs. Furthermore, interaction protocols needs to be extended to support interoperability between different Cloud service providers. As Cloud platforms

become ubiquitous, we expect the need for internetworking them to create a market oriented global Cloud exchange for trading services. Several challenges need to be addressed to realize this vision. They include: market-maker for bringing service providers and consumers; market registry for publishing and discovering Cloud service providers and their services; clearing house and brokers for mapping service requests to providers who can meet QoS expectations; and payment management and accounting infrastructure for trading services. Finally, we need to address regulatory and legal issues, which go beyond technical issues.

In paper [5] author explained that Cloud computing is an evolving paradigm. The NIST definition characterizes important aspects of cloud computing and is intended to serve as a means for broad comparisons of cloud services and deployment strategies, and to provide a baseline for discussion from what is cloud computing to how to best use cloud computing. The service and deployment models defined form a simple taxonomy that is not intended to prescribe or constrain any particular method of deployment, service delivery, or business operation.

In paper [6] author explained the use utility theory leveraged from economics and develop a new utility model for measuring customer satisfaction in the cloud. Based on the utility model, they design a mechanism to support utility-based SLAs in order to balance the performance of applications and the cost of running them. We consider an infrastructure-as-a-service type cloud platform (e.g., Amazon EC2), where a business service provider leases virtual machine (VM) instances with spot prices from the cloud and gains revenue by serving its customers. Particularly, we investigate the interaction of service profit and customer satisfaction. In addition, we present two scheduling algorithms that can effectively bid for different types of VM instances to make tradeoff between profit and customer satisfaction. We conduct extensive simulations based on the performance data of different types of Amazon EC2 instances and their price history. Our experimental results demonstrate that the algorithms perform well across the metrics of profit, customer satisfaction and instance utilization.

In paper [7] author explained Power-aware scheduling reduces CPU energy consumption in hard real-time systems through dynamic voltage scaling(DVS). The basic idea of power-aware scheduling is to find slacks available to tasks and reduce CPU's frequency or lower its voltage using the found slacks. In this paper, we introduce temporal workload of a system which specifies how much busy its CPU is to complete the tasks at current time. Analyzing temporal workload provides a sufficient condition of schedulability of preemptive early-deadline first scheduling and an effective method to identify and distribute slacks generated by early completed tasks. The simulation results show that proposed algorithm reduces the energy consumption by 10-70% over the existing algorithm and its algorithm complexity is $O(n)$. So, practical on-line scheduler could be devised using the proposed algorithm.

In paper [8] author explained At the point when top execution is superfluous, Dynamic Voltage Scaling (DVS)

can be utilized to decrease the dynamic force utilization of inserted multiprocessors. In future advances, be that as it may, static power utilization because of spillage current is normal to increment fundamentally. At that point it will be more viable to confine the quantity of processors utilized (i.e., turn some of them off), or to utilize a mix of DVS what's more, processor shutdown. In this paper, spillage mindful planning heuristics are exhibited that decide the best exchange off between these three strategies: DVS, processor shutdown, and discovering the ideal number of processors. Exploratory results acquired utilizing a open benchmark set of assignment charts and genuine parallel applications demonstrate that our methodology decreases the aggregate vitality utilization by up to 46% for tight due dates ($1.5\times$ the basic way length) and by up to 73% for free due dates ($8\times$

the basic way length) looked at to a methodology that just utilizes DVS. We additionally think about the vitality devoured by our booking calculations to two outright lower limits, one for the situation where all processors consistently keep running at the same frequency, and one for the situation where the processors can keep running at distinctive frequencies and these frequencies may change after some time. The outcomes demonstrate that the vitality diminishment accomplished by our best approach is near these hypothetical limit

3. Survey Table

Table 1: Survey Table

Sr no.	Paper	Technique	advantages	disadvantages	Result
1	Dynamic versus Static Pricing in the Presence of Strategic Consumers	a firm may prefer static pricing over dynamic pricing when consumers are strategic and decide whether to consider to purchase based on the firm's chosen pricing strategy	Dynamic Pricing responds better to demand conditions		dynamic pricing responds better to demand conditions, charging a static price can be the preferable pricing strategy when consumers are strategies
2	Optimal Multi-server configuration for profit maximization in cloud computing	optimal multi-server configuration for profit maximization in a cloud computing environment	Optimization problem are solved effectively. Can obtain optimal server size	-	optimization problem are formulated and solved analytically
3	Above the Clouds: A Berkeley View of Cloud Computing	provide simple formulas to quantify comparisons between of cloud and conventional Computing	Provides great efficiency		Data is protected and saved without any loss of content
4	Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities	Internetworking them to create a market oriented global Cloud exchange for trading services.	Data is safe and has great efficiency	need to address regulatory and legal issues,	market-oriented allocation of resources within Clouds.
5	The NIST Definition of Cloud Computing	Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources	Provides great efficiency		Data is protected and saved without any loss of content
6	Tradeoffs between Profit and Customer Satisfaction for Service Provisioning in the Cloud	developed a utility model for measuring customer satisfaction	enable service providers to dynamically optimize their profit.	-	investigated the service provisioning problem at business service level in the cloud.
7	temporal workload analysis and its application to power-aware scheduling	They introduce temporal workload of a system which specifies how much busy its CPU is to complete the tasks at current time.	reduces the energy consumption	-	Power-aware scheduling reduces CPU energy consumption in hard real-time systems
8	Leakage-Aware Multiprocessor Scheduling	leakage-aware scheduling heuristics are presented that determine the best trade-off between these three techniques	reduces the total energy consumption		Dynamic Voltage Scaling (DVS) can be utilized to decrease the dynamic force utilization of inserted multiprocessors

4. Conclusion

In order to guarantee the quality of service requests and increase the profit of service providers Survey is done on a novel Double-Quality-Guaranteed (DQG) renting scheme for service providers. This scheme combines short-term renting with long-term renting, which can reduce the resource waste greatly and adapt to the dynamical demand

of computing capacity. An M/M/m+D queuing model will be created for our multi-server system with varying system size. And then, an optimal configuration problem of profit maximization is formulated in which many factors are taken into considerations.

References

- [1] G. P. Cachon and P. Feldman, "Dynamic versus static pricing in the presence of strategic consumers," Tech. Rep., 2010.
- [2] J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, "Optimal multiserver configuration for profit maximization in cloud computing," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1087–1096, 2013.
- [3] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "Above the clouds: A berkeley view of cloud computing," Dept. Electrical Eng. and Comput. Sciences, vol. 28, 2009.
- [4] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Gener. Comp. Sy., vol. 25, no. 6, pp. 599–616, 2009.
- [5] P. Mell and T. Grance, "The NIST definition of cloud computing. national institute of standards and technology," Information Technology Laboratory, vol. 15, p. 2009, 2009.
- [6] J. Chen, C. Wang, B. B. Zhou, L. Sun, Y. C. Lee, and A. Y. Zomaya, "Tradeoffs between profit and customer satisfaction for service provisioning in the cloud," in Proc. 20th Int'l Symp. High Performance Distributed Computing. ACM, 2011, pp. 229–238.
- [7] J. Mei, K. Li, J. Hu, S. Yin, and E. H.-M. Sha, "Energy aware preemptive scheduling algorithm for sporadic tasks on dvs platform," MICROPROCESS MICROSY., vol. 37, no. 1, pp. 99–112, 2013.
- [8] P. de Langen and B. Juurlink, "Leakage-aware multiprocessor scheduling," J. Signal Process. Sys., vol. 57, no. 1, pp. 73–88, 2009.

Author Profile



Miss. Poonam. P. Khot, is currently pursuing M.E (Computer) from Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007. she received his B.E (Information Technology) Degree from Tatyasaheb Kore Institute of Engineering Technology, Warananagar, India. Shivaji University, Kolhapur, Maharashtra, India -416113. Her area of interest is cloud computing, web mining.



Sandip Satav received the M.E (CSE/IT) degree from the Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, MAH, India in 2004. He is currently working as Asst. Professor with Department of Information Technology, Jayawantrao Sawant College of Engineering, Pune, MAH, India. His research interests include Image Processing, Networking,