Comprehensive Study of Quality of Service in Cloud: An Initiative Towards Brain mapping

Rakesh S. Deore1, Madhusudan V Amrutsagar2
*[1]Assistant Professor, Department of Computer Science,
S.S.V.P.S's L.K.Dr.P.R.Ghogrey Science College, Dhule
*[2]Assistant Professor, Department of Zoology,
S.S.V.P.S's L.K.Dr.P.R.Ghogrey Science College, Dhule

ABSTRACT

Clouds are a large puddle of computing and storage resources. They located with the help of standard protocols with a general interface. This technology blends several existing technologies, such as the internet of things, service-oriented architectures, utility computing, outsourcing, grid computing, etc. Cloud is a new computing paradigm that promises the reliable quality of service satisfying the customer requirements. Quality of Service (QoS) is a primary requirement in essential applications and distributed computing for multimedia and. QoS is a pathway to render cloud services that are acceptable by customers in cloud computing. In this study, we like to write a detailed review of QoS modeling approaches.

Keywords: Internet of Things , Cloud Computing, Grid Computing, ,Quality of Service, Utility Computing,

I. INTRODUCTION

The delivery of services and information evolved due to cloud computing. According to National Institute of Standards and Technology (NIST), "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1,2]. According to Foster, "cloud computing is a large-scale distributed computing paradigm driven by economies of scale, in which a pool of abstracted virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on-demand to external customers over the Internet" [3]. Gartner thought it "style of computing where scalable and elastic IT inclinations rendered as a service to multiple external customers using Internet technologies" [4].

Qualitative Cloud characteristics describe the qualities of cloud computing with precise technological necessities. The qualitative feature perceived in reoccurring ways. The quality of each feature depends on various service providers [5].

Elasticity: It means that the acquisition of services is adaptable and elastic. It authorizes the customers to demand the service without design for more loads. In order to match a perfect customer usage, these services estimated in fine-grain. The Performance of the system is continually observed and is consistent [6].

• Availability. It is a suitable ability that meets the particular demands of the outsourced services. The QoS metrics throughput and response time must be confirmed to ensure availability[7].

Reliability. It depicts the expertise to assure constant system operation without agitation. Employing tautological sites, the probability of missing data decreases. It ensures that cloud computing is well suited for disaster recovery and business continuity. Reliability is a primary QoS element, focusing upon the repression of impairment [8].

• Agility. It is a fundamental necessity for cloud computing. Cloud service providers must be skilled in online regressions to changes in environmental conditions and resource demands. In the same context, the client must be aware of in house infrastructure capabilities. Agility expects both sides to accommodate self-management capabilities [9].

QoS depends on the type of reliability, availability, and even it also estimates the performance offered by the platform and infrastructure [10]. The cloud customers always demand cloud providers to deliver quality features; meanwhile, the cloud providers find the correct trade-off between qos levels and operational costs [11]. However, finding an optimal exchange between customer and provider is still a debatable decision problem.

II. Literature Review

QoS is an essential role in delivering acceptable cloud services .These systems have capability to gather large number of internationally demobilized users at any time instance. Each users may ask for freakish kinds of services having peculiar demands. This depends on the type of services involved , type of users, and resources [12,13]. The next few paragraphs we focus on findings and methodologies of few authors which addresses the innovative and new solutions for handling efficient QoS.

In [14], Alhamazani et al. give the QoS parameter for virtualization. He principally concentrates on SNMP based QoS monitoring. He showed that service monitoring benefited cloud users and cloud providers.

Due to this advantage, they host their applications with more investment and efficiency. It is also possible to detect the changes in application performance, dynamic configuration changes, and SLA violation.

Buyya et al. [15]. proposed the resource allocation algorithms in the cloud for SaaS cloud providers. This algorithm focused on some service attributes like response time and initiation time to satisfy customers but minimizing the use of hardware resources. This approach reuses the existing virtual machines to optimize cost. But it may introduce the problem of security as the leftover information in the VMs can be used against them.

Chen and Zhang [16] have studied the workflow scheduling algorithm with a base of Particle Swarm Optimization. He invented a method to optimize seven features. This innovative algorithm gives an efficient result. However, this approach has a lack of monitoring on QoS parameters

Den Bossche et al. [17], developed cost-effective heuristics scheduling constraint for a hybrid cloud system. This approach focused on more utilizing local resources and minimizing the use of external resources without sacrificing the QoS requirement of the application. The set of heuristic considers the cost of computation and data transfer. It analyses the different cost factors and workload characteristics. The main advantage of this innovative approach is the optimization of resources in public and private clouds. The main drawback of this system is, it does not consider the failure after scheduling has done.

Emeakaroha et al. [18] have invented schedules based on SLA attributes like storage capacity, network capacity, and CPU time for application deployments in the Cloud environment. These attributes have limited applications in the real-world as they required during deployment. The drawback of this scheme is that users are more interested in performance attributes.

Fatma Omara [19], The service composition problem NP-hard problem. The service composition is economically driven and long term based. They designed an enhanced QoS-based Service Composition approach in the multi-Cloud environment. Selecting accurate services have elevated attention to interest and importance for the Cloud consumer in the multi-Cloud environment. This suggested method has executed using a real QoS dataset. It is found that the proposed plan has achieved a high degree of optimality with low time complexity compared to the existing models.

Haotang Cao et al. [20] observed that virtual network algorithms are mostly static. Some of the existing dynamic algorithms just focus on embedding virtual networks but there is lack of foucs on QoS performance of each accepted virtual networks.They investigate a dynamic virtual network algorithm with focus on QoS performance. They improve performance of system by 13%.

Halal Hasan et al. [21] explains the importance of trust management to improve the quality of service. They introduce a novel QoS based model for evaluating all criteria of trust. In this approach, the model calculates the value of the accumulative trust which is updated dynamically. Experimental result shows that these dynamic calculation help to improve the QoS parameters.

Ismail Zhraddeen[22] designed the Priority Based Delay Time Scheduling technique. The aim of this method is to maximize resource utilization, and it also ensures the QoS performance. It gives a good result as compared to the Min-Min and Max-Min method.

Mathew Turner et al. [23] focus on task scheduling and the problem of service placement. They design an algorithm to work on a three-tiered edge cloud platform. They deal to solve the problem of time-sensitive applications. They developed methods to solve the NP-HARD problem of meeting users' deadlines. In this method, two novel heuristics deciding on a local and global scale are invented. These proposed heuristics works near to optimal and also reduces the complexity.

Li Yang et al. [24] studied the bandwidth allocation among multiple concurrent live virtual machine migration. The prime aim of this study is to reduce the migration tile for a single live virtual machine migration. Here they maximize the quality of service. They formulate this problem as Non Linear Programming and also show that it is NP-HARD.

Ian Tao Zohu [25] observed that different cloud services have different QoS. Also, the QoS requirement of different users is different. The problem of finding the best QoS strategy to satisfy user is NP-HARD. Here they used the Swarm Intelligence method to solve the multiobjective optimization problem. In this paper, they compare multiple algorithms based on swarm intelligence.

Madnesh K Gupta et al. [26] proposed a new QoS-aware Virtual Machine placement method called vScale. The prime aim of this paper is to minimize the service level agreement violation penalty for user application. They show that this method is quite better than the existing methods.

Shivangi Dhariwal et al. [27] worked to improve the utilization of maximum resources. They focused on the waiting time, the response time of jobs. They implemented three algorithms to find maximal possible profit based on current server speed and size. The performance of the system identified in terms of net profit, optimal profit, and maximal profit.

Zuhang Zaho et al. [28] designed a new method based on adjusted matrix factorization to predict the missing QoS values. In this method, the training performed on the prefilled QoS matrix. Their experimental results outperform both the MF-based methods and PCC-based CF methods. This method also has a reasonable stability rate.

III. Conclusion

In this review paper, we reviewed a few current proposed frameworks for improving QoS management. We also review the methods used to improve QoS in terms of different parameters of services. We observed that most of the problems in providing services according to the customer perspective are NP-HARD.

IV. References

[1] Demarest, G., Wang, R. Oracle cloud computing. An Oracle white paper. http://www.oracle.com/us/technologies/cloud/oracle- cloud-computing-wp-076373.pdf (2010)

[2] Mell, P., Grance, T.: The NIST definition of cloud computing.NIST (National Institute of Standards and Technlogy) Special Publication 800-145.

[3] Foster, I., Zhao, Y., Raicu, I., and Lu, S. (2008). Cloud computing and grid computing 360-degree compared. In GCE'08: Proceedings of the Grid Computing Environments Workshop, 2008, pages 1–10.

[4] Plummer, D., Cearley, D., and Smith, D. (2008). Cloud computing confusion leads to opportunity. Technical Report G00159034, Gartner Research.

[5] Peng, D.-T. and Shin, K.-G. (1993). A new performance measure for scheduling independent real-time tasks. Journal of Parallel Distributed Computing, 19:11–26.

\[6] Masrur, A. and Chakraborty, S. (2011). Nearoptimal constant-time admission control for DM tasks via non-uniform approximations In RTAS: Proceedings of the 17th IEEE Real-Time and Embedded Technology and Applications Symposium, Chicago, IL, USA.

[7] Manabe, Y. and Aoyagi, S. (1995). A feasibility decision algorithm for rate monotonic scheduling of periodic real-time tasks. In Proceedings of the IEEE Real-Time Technology and Applications Symposium, pages 212–218.

[8] Hyman, J., Lazar, A., and Paci□ci, G. (1991). Real-time scheduling with quality of service constraints. IEEE Journal on Selected Areas in Communications, 9:1052–1063.

[9] Harter, J. and Paul, K. (1987). Response times in level structured systems. *ACM Transaction on Computer Systems*, 5:232–248.

[10] Hyman, J., Lazar, A., and Paci□ci, G. (1991). Real-time scheduling with quality of service constraints. IEEE Journal on Selected Areas in Communications, 9:1052–1063.

[11] Dimitrios, Z., Dimitrios (2012), L.: Addressing cloud computing security issues. Future Gener.Comput. Syst. 28, 583–592(Elsevier)

[12] Zaigham, M.(2011): Data location and security issues in cloud computing. In: International Conference on Emerging Intelligent Data and Web Technologies, IEEE (2011).

[13] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson,D., Rabkin, A., Stoica, I., Zaharia, M (2020).: A view of cloud computing. Mag. Commun. ACM 53 50– 58

[14] Alhamazani, K., Ranjan, R., Rabhi, F., Wang, L., Mitra, K.: Cloud monitoring for optimizing the QoS of hosted applications. In: Proceedings of 4th International Conference on Cloud Computing Technology and Science, pp. 765–770 (2012)

[15] Buyya, R.K., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I (2019).: Cloud computing and emerging IT platforms: reality, vision, and hype for delivering computing as the 5th utility.J. Future Gener. Comput. Syst. 25(6), 599–616 (2019).

[16] Chen, W-N., Zhang, J.: A set based discrete PSO for cloud workflow scheduling with user-defined QoS constraints. In: IEEE International Conference on Systems, Men and Cybermetics, pp. 773–778 (2012)

[17] Van den Bossche, R., Vanmechelen, K., Broeckhove, J.: A set of Heuristics scheduling deadline constrained workloads on hybrid cloud System. In: Proceedings of the 3rd IEEE International Conference on Cloud Computing Technical and Science, pp. 320–327 (2011).

[18] Emeakaroha, V.C., Brandic, I., Maurer, M., Breskovic, I.: A Scheduling heuristic that takes multiple SLA parameters when deploying applications in cloud. In: Proceedings of 35th IEEE Annual Computer Software and Applications Conference Workshops, pp. 298–303 (2011)

[19] Haytamy and F. Omara, "Enhanced QoS-Based Service Composition Approach in Multi-Cloud Environment," 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 2020, pp. 33-38,

[20] H. Cao, S. Wu, G. S. Aujla, Q. Wang, L. Yang and H. Zhu, "Dynamic Embedding and Quality of Service-Driven Adjustment for Cloud Networks," in IEEE Transactions on Industrial Informatics, vol. 16, no. 2, pp. 1406-1416, Feb. 2020, doi: 10.1109/TII.2019.2936074.

[21] H. Hassan, A. I. El-Desouky, A. Ibrahim, E. M. El-Kenawy and R. Arnous, "Enhanced QoS-Based Model for Trust Assessment in Cloud Computing Environment," in IEEE Access, vol. 8, pp. 43752-43763, 2020, doi: 10.1109/ACCESS.2020.2978452.

[22] I. Z. Yakubu and C. Malathy, "Priority Based Delay Time Scheduling for Quality of Service in Cloud Computing Networks," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-5, doi: 10.1109/ic- ETITE47903.2020.379.

[23] M. Turner and H. Khamfroush, "Meeting Users' QoS in a Edge-to-Cloud Platform via Optimally Placing Services and Scheduling Tasks," 2020 International Conference on Computing, Networking and Communications (ICNC), Big Island, HI, USA, 2020, pp. 368-372, doi: 10.1109/ICNC47757.2020.9049749.

[24] L. Yang, D. Yang, J. Cao, Y. Sahni and X. Xu, "QoS Guaranteed Resource Allocation for Live Virtual Machine Migration in Edge Clouds," in IEEE Access, vol. 8, pp. 78441-78451, 2020, doi: 10.1109/ACCESS.2020.2989154.

[25] Y. Wang, J. Zhou, Y. Jiao and X. Song, "Comparative Analysis of Evolutionary Algorithms Based on Swarm Intelligence for QoS Optimization of Cloud Services," 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Porto, Portugal, 2019, pp. 434-439, doi: 10.1109/CSCWD.2019.8791926.

[26] M. K. Gupta, A. Jain and T. Amgoth, "QoS-aware Virtual Machine Placement for Infrastructure Cloud," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp. 353-357, doi: 10.1109/GUCON.2018.8674979.

[27] S. Dhariwal and H. Makwana, "Maximize the Cloud Profit to Improved QoS in Cloud Computing: Design and Analysis," 2019 Third International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2019, pp. 279-285,

[28] C. Mao and Z. Zhao, "Predicting QoS for Cloud Services through Prefilling-Based Matrix Factorization," 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), Suzhou, China, 2019, pp. 25-30, doi: 10.1109/CBD.2019.00015.