Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 2, August 2023
Paper : 09

Literature Review: Improving the Quality ofServices in Cloud Computing Environment

Manila Gupta
Deptt. Of Computer Science & Engineering
IFTM UNIVERSITY
Moradabad, INDIA
manilagupta24@gmail.com

Dr. Devendra Singh
H.O.D, Deptt. Of Computer Science & Engineering
IFTM UNIVERSITY
Moradabad, INDIA
dev625g@gmail.com

Dr. Bhumika Gupta
Deptt. Of Computer Science & Engineering
G.B.P.I.E.T
Pauri-Garhwal, INDIA
mail2bhumikagupta@gmail.com.com

Abstract— Cloud computing has become a prominent method for delivering data and IT services, where users can access resources over the internet. It introduces a pay-per-use model for computing services, which has led to an increased demand for cloud computing services. As companies transition to the cloud for their services, cloud providers must ensure that their offerings meet the expected quality requirements. Managing the quality of services (QoS) in cloud computing is a critical task, involving the allocation of resources to applications to ensure performance, availability, and reliability. The purpose of this paper is to provide a comprehensive survey on QoS in cloud computing, covering the techniques used as well as their advantages and disadvantages. The survey incorporates all relevant previous works on QoS in the cloud computing environment. By studying these previous works, the paper aims to offer insights into the various QoS methods employed in cloud computing. The survey likely explores a range of QoS techniques that have been employed in the context of cloud computing. These techniques may include resource allocation algorithms, load balancing strategies, fault tolerance mechanisms, performance monitoring, and optimization approaches, and more. Each technique will have its own advantages and disadvantages, which the paper aims to discuss in detai .By conducting this survey, the authors intend to provide a comprehensive overview of the existing QoS methods in cloud computing. This can serve as a valuable resource for researchers, practitioners, and decision-makers who are interested in understanding and implementing QoS mechanisms in the cloud.

Keywords— Cloud Computing, Quality of Service (QoS), Service level Agreement (SLA), Software Defined Network (SDN).

I. INTRODUCTION

The evolving trend in technology today is cloud computing, which is a service-oriented technology delivered over the internet. It offers computing as a service, making it a unique form of internet-based service. Cloud computing can be considered a new computing model and is widely used by many users who require various services based on their needs. However, due to limited resources, it becomes challenging for cloud providers to fulfill all the desired services for their users. Cloud providers aim to allocate cloud resources in a coherent manner to meet the satisfaction and quality of service (QoS) requirements of cloud users. The internet plays a vital role in the development of various fields such as business and research, and cloud computing provides a wide range of services in the form of software, hardware, and platforms. Cloud computing can be defined as a combination of two words: "cloud" and "computing." The cloud refers to a collection of services comprising interfaces, hardware, networks, and software, which are made available to end users. These services in the cloud can be categorized as software, platform, and infrastructure as a service, delivered to clients via the internet. With cloud computing, users can access resources without the need to purchase them outright. It enables users to perform tasks on computers or other internet-connected devices without the need to buy software, as they can easily access and utilize it through the cloud. Users may pay a fee to gain access to specific services. Before delivering services to clients, a service level agreement (SLA) is typically signed between the client and the vendor. The SLA outlines the terms and conditions of the service, including the agreed-upon level of service quality and other relevant details [1].

Models of cloud computing

Cloud computing is based on the SPI model, which stands for Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). These models provide different levels of services to users. Here is a description of each model:

a) Software-as-a-Service (SaaS):

SaaS is a software distribution model where applications are hosted by service providers and made available to clients over the internet. With SaaS, users can access and use software applications without the need to install or execute them on their local machines. Examples of SaaS providers include Amazon Web Services, Salesforce, Google Apps, and Facebook.

b) Platform-as-a-Service (PaaS):

PaaS is a computing platform that eliminates the need for users to purchase or install hardware and software infrastructure. It provides a platform with various computing

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 2, August 2023
Paper : 09

applications required by consumers. PaaS also offers resource management functions for scheduling processes in real time.

c) Infrastructure-as-a-Service (IaaS):

IaaS provides users with various resources such as storage, network, and processing capacity. Users can control and manage the cloud environment as a service, paying only for the duration they utilize the services. This model enables faster service delivery at a lower cost. With IaaS, users can use the service without being aware of the underlying storage location. It is sometimes also referred to as Hardware-as-a-Service (HaaS) [3].

These three models—SaaS, PaaS, and IaaS—offer different levels of abstraction and services to users, catering to their specific needs and requirements in the cloud computing environment.

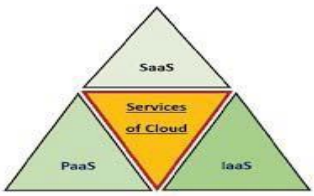

Fig 1: Services Provided by a Cloud

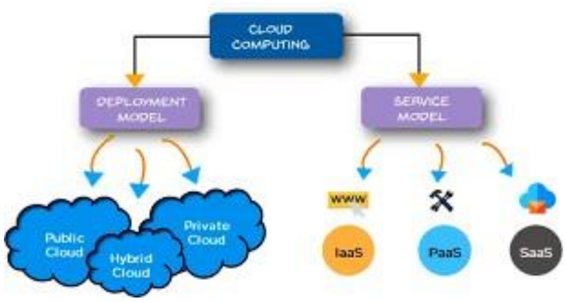

Fig 2: Cloud Development Model

Cloud Deployment Models

Cloud computing offers various deployment models to cater to different organizational and user requirements. The commonly recognized cloud deployment models are as follows:

a) Private cloud:

A private cloud is owned and operated by a single organization. It allows the organization to leverage cloud computing technology for centralizing access to IT resources within their infrastructure. The administration of a private cloud can be carried out by internal IT staff or outsourced to a third-party provider. Private clouds are valued for their enhanced security measures and control over data and applications.

b) Public cloud:

A public cloud is a cloud environment that is publicly accessible and owned by a third-party cloud provider. IT resources in the public cloud are typically provisioned and offered to users at a cost. The cloud provider is responsible for creating and maintaining the public cloud infrastructure and its resources. Public clouds offer scalability, cost-effectiveness, and convenience for users, as they can access resources on-demand without the need for infrastructure management.

c) Hybrid cloud:

A hybrid cloud combines two or more different cloud deployment models, usually a mix of private and public clouds. This deployment model allows businesses to take advantage of the security benefits of hosting sensitive applications and data on a private cloud, while also benefiting from the cost-effectiveness and scalability of the public cloud for shared data and applications. Hybrid clouds provide flexibility and enable organizations to optimize their IT infrastructure based on specific workload requirements.

d) Community cloud:

A community cloud is similar to a public cloud but with restricted access limited to a specific community of cloud users. The community cloud may be jointly owned and operated by the community members or provisioned by a third-party cloud provider. It serves the needs of a particular industry, interest group, or community by providing shared infrastructure and services. Community clouds enable collaboration, resource sharing, and cost savings within a specific user community.

These different deployment models offer organizations and users the flexibility to choose the most suitable cloud environment based on their specific needs, preferences, and security requirements [2].

II. QUALITY OF SERVICE (QoS) TECHNIQUES

Quality of Service (QoS) plays a crucial role in cloud computing, as it determines the level of performance, reliability, and availability that users expect from an application and the underlying platform or infrastructure. Cloud users rely on service providers to deliver the desired QoS characteristics, while providers need to balance QoS levels with operational costs [4].

Determining the optimal trade-off for QoS is a challenging decision-making process for cloud providers. It involves considering Service Level Agreements (SLAs), which define the QoS targets and economic penalties associated with SLA violations. Service providers are bound by SLA contracts, which outline the rewards and penalties based on the achieved performance level [5].

Service Level Agreements (SLAs) serve as agreements between the service provider and the customer, outlining the terms and conditions of the service. SLA violations act as significant constraints, as they can have financial and reputational consequences for both parties involved. Monitoring mechanisms are implemented to track and detect SLA violations, allowing for appropriate actions to be taken [6].

Ensuring QoS in cloud computing requires continuous monitoring, performance management, and adherence to SLAs to maintain customer satisfaction and meet service provider obligations. DEDCA (Dynamic Enhanced Distributed Channel Access): This mechanism focuses on channel access in wireless networks and utilizes SDN technology in 802.11-based wireless networks to enhance QoS by controlling the contention window size of wireless terminals [7].

QoS in SDN (Software-Defined Networking): Several papers explore the relationship between QoS and SDN. They discuss the benefits of SDN in improving QoS, reducing latency, and achieving cost savings in comparison to traditional networks [8], [9], [10].

Controller Selection in SDN: One paper proposes a two-step approach for selecting SDN controllers based on qualitative features and performance analysis to improve QoS [11].

QoS Management in Cloud Computing: Researchers are investigating automated methods for managing QoS in cloud platforms, considering the performance, heterogeneity, and resource isolation challenges [12].

QoS Monitoring and Management in Enterprise Systems: An approach called SOS (Service-Oriented Architecture for Monitoring, Management, and Response) is proposed for monitoring and managing QoS in enterprise systems deployed in cloud environments [13].

Cloud Federation and QoS: A framework is presented that enables multiple independent cloud providers to collaborate and offer scalable QoS-assured services, with the aim of maximizing QoS targets and minimizing SLA violations [14].

QoS-oriented Cloud Services Design and Analysis: A prescribed model is proposed for designing and analyzing QoS factors in ECC (Enterprise Cloud Computing) services, incorporating SaaS, PaaS, and IaaS layers [15].

S. Lee et al. [18] proposed an architecture that employed agent technology to handle requested QoS requirements and service level agreements in cloud environments. The agent technology analyzed resource allocation and deployment, focusing on verification and validation. However, this work lacked a self-learning algorithm for determining the optimal allocation of system resources.

L. Bin et al. [19] proposed a QoS-aware data imitation and deletion approach to optimize disk space and maintenance costs for distributed storage systems. The DRDS (Dynamic Replica Deletion Strategy) algorithm aimed to guarantee the availability and performance quality of service requirements while saving costs. However, data replication introduced increased overhead on data updates and irregularity.

P. Zhang et al. [20] presented a QoS framework for mobile cloud computing and an adaptive QoS management process to ensure QoS assurance in mobile cloud computing environments. They introduced a QoS management model based on fuzzy cognitive maps (FCM). However, no detailed evaluation of the model construction was provided.

Y. Xiao et al. [21] proposed a reputation-based QoS provisioning scheme that aimed to reduce the cost of computing resources while rewarding the desired QoS metrics. They considered the arithmetic probability of the response time as a practical metric instead of the typical mean response time. However, the QoS provisioning algorithm did not incorporate security and privacy metrics.

M. Xu et al. [22] introduced a QoS-inhibited scheduling strategy called MQMW (Multi-QoS Multi-Workflow) to address the issue of multiple workflows with diverse QoS requirements. The strategy enabled the scheduling of multiple workflows occurring simultaneously, but it did not include availability and reliability as QoS restrictions for workflows

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 2, August 2023
Paper : 09

S.No.	TechniqueUsed	Simulator Used	Algorithm Used/Parametercounts	Result Obtained
[1]	Simulation	Cloud analytic simulator	Equally Spread Current Execution(ESCE)	Better result in overall response time and datacenter processing time.
[2]	Simulation	CloudSim	Cluster Based Load Balancing Algorithm	Clustering can also be implemented at the client side. We can divide our tasks/cloudlets into different clusters depending upon their task length, cost and priority.
[3]	Simulation	CloudAnalyst	Round-Robin, Throttled, ESCE	The Throttled algorithm used in load balancer in cloud computing simulations using CloudAnalyst is the average response rate is still within the average range between UB1 and the other.
[4]	Simulation	CloudSim	Works on CSA gateway.	Results prove that the CSA(Cloud Service Agent) based approach fared better in terms of throughput and service time as compared againstthe already existing approaches.
[5]	Simulation	CloudSim	Proposed an algorithm named Reverse filling LB Algorithmand Compare with Round Robin Algorithm.	Proposed an algorithm that reduces the waiting time of the machines along with the overheads inthe system resulting in lesser cost than existing approach. The future work includes modifications in the system leading to greater resource utilization and performance.
[6]	Analysis	Proposed a method forLoad Re-balancing.	Honeybee Foraging Algorithm, Active clustering algorithm and Ant Colony Optimization.	None of the methods has considered the execution time of any undertaking at the runtime. In this manner there is an objective to grow such load balancing method that can enhance the execution of cloud computing alongside most extreme asset usage.
[7]	Analysis	Proposed an improvedACO	Ant colony optimization(ACO)	The advantage of the approach lies in the fact that the task of each ant is specialized rather thanbeing general and the task depends on the type of first node that was encountered whether it wasoverloaded or underloaded.
[8]	Simulation	CloudSim	Novel dynamic load balancing algorithm	Execution analysis of the simulation shows that change of MIPS will effects the response time. Increase in MIPS vs. VM decreases the response time.
[9]	Simulation	MATLAB R 2010 toolkit, Cloud Analyst	Improvised Genetic Algorithmand Round Robin Algorithm.	Experiment has been conducted by varying the number of nodes in a VM available in a cloud configuration of single data center. The results show the use of other soft computing techniques is needed to be studied for further improvement.
[10]	Analysis	Proposed a Load Balancing Algorithm	Round Robin Algorithm	By developing new load balancing algorithms and model, the dynamic situations on servers canbe easily handled and overloading problem can be avoided.
[11]	Simulation	Cloud Analyst	Throttled Algorithm, RoundRobin Algorithm, Active Monitoring Load BalancingAlgorithm	Throttled Algorithm proves to be better compared with other two algorithms and gives better response time.
[12]	Analysis	Adopting CLBVM(Central LoadBalancing Policy For Virtual Machines).	Central Load Balancing Policy	By Adopting the CLBVM policy the throughputof the system will increase.
[13]	Simulation	Cloud Analyst,CloudSim	Round Robin and EquallySpread Current ExecutionLoad.	The response will become very less with MY Load Balancer .The parameters considers are overall response time and Data Centre Processing time.
[14]	Simulation& Analysis	Cloud Analyst Simulation toolkit	Round-Robin and Throttled Algorithm	With the modified Throttled Algorithm, if the Number of VM increases it will reduces the response time and processing time of cloud datacenters.

TABLE 1: Comparison Table for QoS Techniques

III . CONCLUSION

The review mentioned in Table 1 discusses various Quality of Service (QoS) methods in cloud computing environments. The authors analyzed existing research work aimed at improving QoS in cloud computing architecture. They studied the techniques used so far to enhance quality and examined the advantages and disadvantages of these methods.

The review also suggests that there is still scope for further research to overcome the limitations identified and address areas where the system requires improvement in terms of QoS. The authors propose future work that involves considering specific parameters and designing new, efficient algorithms to achieve better QoS in cloud computing environments.

In summary, the review highlights the existing research efforts in enhancing QoS in cloud computing and emphasizes the need for continued work to reduce limitations and develop improved approaches for achieving higher quality standards.

REFERENCES

[1] Rathore. Rahul, Gupta.Bhumika, Sharma. Vaibhav,Gola. Kamal Kumar" A New Approach For Load Balancing in Cloud Computing", IJCT, Vol13. No.12, (2014).

[2] Helen Anderson Akpan, B. Rebecca JeyaVadhanam "A Survey on Quality of Service in Cloud Computing", ISSN: 2231- 2803, IJCTT – volume 27 Number 1 – September 2015

[3] Zhang Q, Cheng L, BoutabaR: Cloud computing: state-of-the-art and research challenges. *J Internet ServAppl* 2010, 1(1):7– 18. 10.1007/s13174-010-0007-6.

[4] D. Ardagna, G. Casale, M. Ciavotta, J. F. Perez, W. Wang, "Quality-of-Service in cloud computing: modelling techniques and their applications", Journal of Internet Services and Applications, Volume: 5, Issue: 11, pp.1-13, 2014.

[5] D. Ardagna, B. Panicucci, M. Trubian, L. Zhang, "Energyaware autonomic resource allocation in multitier virtualized environment", IEEE Transactions on Services Computing, pp 1-18, Volume: 5, Issue: 1, 2012.

[6] V. Rajesekaran, A. A. Ashok, R. Manjula, "Novel sensing approach for predicting SLA violations", International Journal of Computer Trends and Technology (IJCTT), Volume:10, no.1, pp.25-26, 2014.

[7] PilarManzanares-Lopez, JosemariaMalgosa-Sanahuja and Juan Pedro Muñoz-Gea, "Software-Defined Networking Framework to Provide Dynamic QoS Management in IEEE 802.11 Networks", Sensors 2018, 18, 2247.'

[8] D'Souza, Derrick. Sundharan P Krishna, LokanathSavithru, Mittal Vivek:" Improving QoS in a Software-Defined Network", Capstone Research Paper, April 30, 2016.

[9] Hossain .MdAlam, sheikh. Mohammad Nowsin Amin, Halder. Monish Anker, Biswas .Sujan, Arman.MdAriful Islam," Quality of Service in Software Defined Networking", ISSN: 0975-4172 & Print ISSN: 0975-4350, Volume 18 Issue 3 Version 1.0 Year 2018.

[10] Karakus.Murat, Durresi.Arjan," Quality of Service (QoS) in Software Defined Networking (SDN): A Survey", (JNCA)December 8, 2016.

[11] JehadAliI, Byeong-heeRoh, SeungwoonLeeID "QoS improvement with an optimum controller selection for software-defined Networks", PLoS ONE 14(5):e0217631.

[12] D. Petcu, G. Macariu, S. Panica, C.Crcium, "Portable cloud applications-From theory to practice", Future Generation Computer Systems, Volume: 29, Issue: 6, pp. 1-5, 2013.

[13] P. C. Hershey, S. Rao, C. B. Silio and A. Narayan, "System of systems for Quality-of-Service observation and response in cloud computing environment", IEEE Systems Journal, Volume: 9, Issue: 1, pp. 1-5, 2015.

[14] M. Salam and A. Shawish, "A QoS-oriented inter-cloud federation framework", IEEE Systems Journal, pp. 642-643, 2015.

W. C. Chu, C. Yang, C. Lu, C. Chang, N. Hsueh, T. Hsu, S. Hung, "An approach of quality of service assurance for enterprise cloud computing (QoSAECC)", in Proc. Int. Conf. on Trustworthy Systems and their Applications, pp. 7-13, 2014.

[15] M. M. Hassan, B. Song, M. S. Shamin, and A. Alamri, "QoSaware resource provisioning for big data processing in cloud computing environment", in Proc. Int. Conf. on Computational Sc. and Computational Intelligence, pp. 107-112, 2014.

[16] R. Karim, C. Ding, A. Miri, "An end-to-end QoS mapping approach for cloud service selection", in Proc. IEEE Ninth World Congress on Services, pp. 341-348, 2013.

[17] S. Lee, D. Tang, T. Chen, W. C. Chu, "A QoS assurance middleware model for enterprise cloud computing", in Proc. IEEE 36th Int. Conf. on Computer Software and Application Workshops, pp. 322-327, 2012.

[18] L. Bin, Y.Jiong, S. Hua, N. Mei, "A QoS-aware dynamic data replica deletion strategy for distributed storage systems under cloud computing environments", in Proc. Second Int. Conf. on Cloud and Green Computing, pp. 219-225, 2012.

[19] P. Zhang, and Z. Yan, "A QoS-aware system for mobile cloud computing", in Proc. of IEEE, pp. 518-522, 2011.

[20] Y. Xiao, C. Lin, Y. Yiang, X. Chu, X. Shen, "Reputation-based QoS provisioning in cloud computing via Dirichletmultinomial model", IEEE ICC Proceedings, pp. 1-5 2010.

[21] M. Xu, L. Cui, H. Wang, Y. B. Bi, "A Multiple QoSconstrained scheduling strategy of multiple workflows from cloud computing", IEEE International Symposium on Parallel and Distributed Proceeding with Applications, pp. 629-633, 2009.