A Survey on Quality of Service in Cloud Computing

Helen Anderson Akpan[#1], B. RebeccaJeya Vadhanam[#2]

[#1]MSc Student, Department of Information Technology,[#2]Assistant Professor, Department of Computer Science, SRM University, Chennai, Tamil Nadu, India.

Abstract-*Cloud computing systems enables pay-per-usage pricing model for computing services delivered to users across the globe over the internet. Currently the demand for cloud services has increased as enterprises and individuals have now migrated to the cloud and cloud providers need to offer services based on the expected quality requirements. One of the challenges posed by cloud application is quality of service (QoS) management, which is the problem of allocating resources to the applications to guarantee service based on performance, availability and reliability. In this paper, we have presented a survey on the quality of service in cloud computing with respect to techniques used, advantages and disadvantages.*

Keywords-cloud computing, quality of Service (QoS), Service level Agreement (SLA).

I. INTRODUCTION

Cloud Computing is an emerging computing paradigm that may change the way information services are provisioned. Clouds represent a new step in evolutional computing and communication technologies development chain by introducing new type of services and a new abstraction layer for the general services virtualization and mobility [1].

Many cloud service providers (e.g, Amazon, Google, Microsoft, IBM, etc) are now available in the market to provide cloud services such as Governance as a Service (GaaS), Business as a Service (BaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Cloud technology stack has also become mainstream in enterprise data centers, where private and hybrid cloud architectures are increasingly adopted [2].

However, though cloud computing has tremendous advantages; there are challenges in the area of Quality of Service (QoS). QoS denotes the levels of performance, reliability and availability offered by an application and by the platform or infrastructure that hosts it. QoS is fundamental for cloud users, who expect providers to deliver the advertised quality characteristics, and for cloud providers, who need to find the right tradeoffs between QoS levels and operational costs [2]. Any violation of service level agreement (SLA) entails a loss for both cloud providers and cloud users [3]. Over provision is often adopted by providers as an approach to satisfy the SLA, but it fails to optimize the resource utilization, especially for private clouds [4].

Due to complexities which have affected QoS assurance, many researchers have been prompted to explore automated QoS management techniques that can leverage the challenges encountered by the users and the cloud service providers. In this paper, we surveyed related publications from 2009 to 2015 to ascertain if the present techniques and implementations have been able to resolve the issues of quality of service in cloud computing.

II. OVERVIEW OF CLOUD COMPUTING

A. Definition of Cloud Computing

Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to shared resources (e.g. network, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [5]. Cloud computing can also be described as a systematically devised mechanism wherein the users can use the computing applications as and when they need and they are made accessible in 'cloud' through a browser or any other web-based tools [6].

B. Cloud Models

The cloud model is composed of five essential characteristics which include three service models, and four deployment models [5].Currently, cloud model has extended to layers of Business as a Service (BaaS) and Governance as a Service (GaaS) using multidimensional and service-oriented architectural model [7].

i) Cloud Service Models

Software as a Service (SaaS): SaaS refers to a software distribution model in which applications are hosted by a vendor or service provider and made available to internet users through a browser. The cloud providers manage the infrastructure and platforms that run the applications. SaaS represents the potential for a lower-cost way for businesses to use software on demand rather than buying a license for every computer.

Platform as a Service (PaaS): PaaS is a paradigm for delivering operating systems and associated services over the internet without downloads or installation. The defining factor that makes PaaS unique is that it lets developers build and deploy web applications on a hosted infrastructure. PaaS allows customers to leverage the seemingly infinite compute resources of a cloud infrastructure. It provides development environment as a service to the users.

Infrastructure as a Service (IaaS): IaaS involves outsourcing the equipment used to support operations, including storage, hardware, servers and networking components. IaaS enables developers to avoid capital expenditure on hardware and human resources. IaaS delivers infrastructure as a service to the users.

Business as a Service (BaaS): Many companies have migrated to the cloud and run their entire business in the cloud. Cloud computing present new economic models that companies can use to provision IT and services. More enterprises are now employing third parties to reuse their solutions rather than lock their valuable capital in sourcing hardware and software themselves. This model enables enterprises to get more efficient, lower cost and achieve business agility across channels, market and customer segments [8].

Governance as a Service: Cloud services governance is a general term for applying specific policies or principles to the use of cloud computing services. Governance is one of the important challenges for collaborative enterprises [9]. There is a need to remove policy enforcement from applications and platform and ascribe those responsibilities to an intermediate governance layer. Cloud service governance can be viewed as an extension of SOA governance. A single authority provides governance services to multiple heterogeneous administrative domains in which SOA-based applications enable business and collaborative service that supports end users who are producing and consuming data using software and infrastructure services [7]. Ideally, cloud services governance is integrated into existing governance processes and viewed as an ongoing process but not a product. Governance as a service layer reside in the background defensively governing content throughout the enterprise.

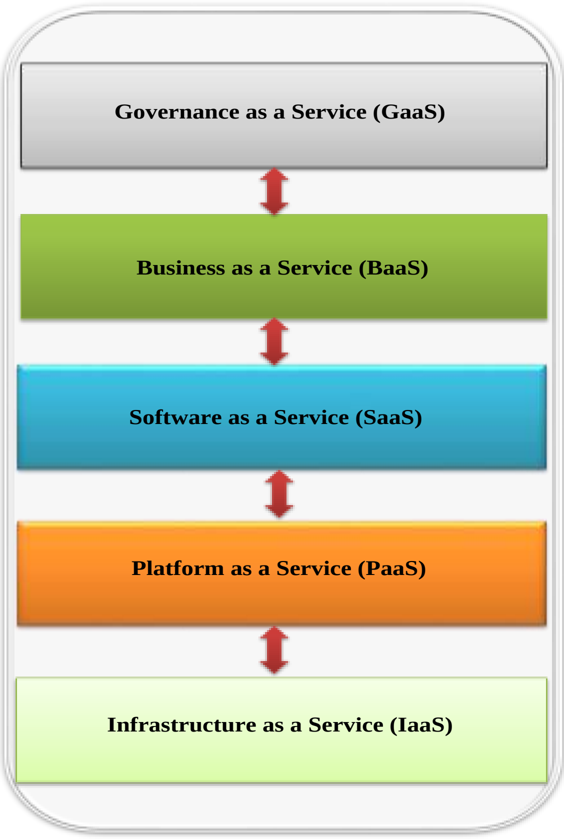

Figure 1: Cloud Service Models

ii) Cloud Deployment Models

Private cloud: A private cloud is owned by a single organization. Private cloud enables an organization to use cloud computing technology as a means of centralizing access to IT resources in the organization. The actual administration of a private cloud environment may be carried out by internal or outsourced staff. This model has tremendous value from a security point of view.

Public cloud: A public cloud is a publicly accessible cloud environment owned by a third party cloud provider. The IT resources on cloud computing are usually provisioned and offered to cloud providers at a cost. The cloud provider is responsible for the creation and ongoing maintenance of the public cloud and its IT resources.

Hybrid cloud: A hybrid cloud is a cloud environment comprised of two or more different cloud deployment models. This deployment model helps business to take advantage of secured applications and data hosting on private cloud, while still enjoying the cost benefit by keeping shared data and applications on the public cloud.

Community cloud: A community cloud is similar to public cloud except that its access is limited to a specific community of cloud users. The community cloud may be jointly owned by the community members or third-party cloud provider that provisions a public cloud with limited access.

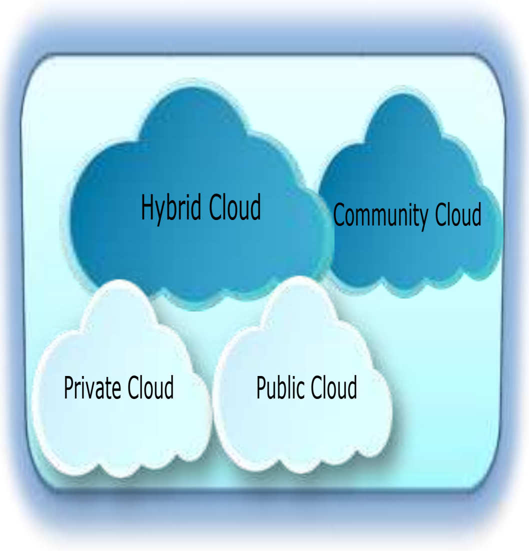

Figure 2: Cloud Deployment Models

III. QUALITY OF SERVICE TECHNIQUES

Quality of Service (QoS) denotes the level of performance, reliability, availability offered by an application and by the platform or infrastructure that hosts it. QoS is fundamental to cloud users, who expect providers to deliver the advertised characteristics, and for cloud providers, who need to find the right tradeoff between QoS levels and operational cost [10]. Finding out the optimal tradeoff is not an easy decision problem because it involves Service Level Agreements (SLAs) which specifies QoS targets and economic penalties associated with SLA violations. Service providers need to comply with SLA contracts which determine the revenue and penalties on the basis of the achieved performance level [11]. Service Level Agreements (SLA's) are signed between the service provider and the Customer where SLA violation acts as major constraints. SLA violation is reduced through mechanisms involving monitoring [12].

Though QoS properties have received constant attention even before the evolution of cloud computing, performance and heterogeneity and resource isolation mechanisms of cloud platforms have significantly complicated QoS analysis, prediction and assurance. Thus, several researchers are investigating automated QoS management methods that can leverage the high programmability of hardware and software resources in the cloud [13].

P. C. Hershey et al. [7] proposed a SoS approach to enable QoS monitoring, management and response for enterprise systems that deliver computing as a service through a cloud environment. Enterprise Monitoring, Management and Response Architecture in Cloud Computing Environments (EMMRA CC) extended previous work to provide structure from which to identify points within the administrative domains where QoS metrics may be monitored and managed. A concrete example was provided for applying the new SoS approach to a real world scenario (viz, distributed denial of service (DDoS)). The approach is very effective but it was not applied to federated clouds in real time.

M. Salam et al. [14] proposed a QoS-oriented federated cloud computing framework where multiple independent cloud providers can cooperate seamlessly to provide scalable QoS-assured services. The key elements for enabling cloud federation used were Cloud Coordinators (CC) and Federation Coordinators (FC). The distinct feature of the proposed federation framework is its QoS-orientation that can trigger the on-demand resource provisioning across multiple providers, hence helping to maximize QoS targets and resources usage, eliminate SLA violations and enhance SLA formalization. However, Complex services were not constructed using a mixture of services from different cloud providers and no provision was made for distributed denial of service (DDoS) attacks.

W. C. Chu et al. [15] proposed a formal model to assist not only the ECC services design and construction through SaaS, PaaS, IaaS but also the simultaneous monitoring and dynamic analysis on the QoS factors for the promises from QoS service providers and the service level agreements (SLA) for multiple ECC consumers. Based on the formal model, analysis model and testing model was generated to support automatic testing as well as runtime monitoring to assure the satisfaction to the requirements/SLA constraints. This work had some limitations such as not adapting the features and solutions of IOT into the framework as well as the field experiment.

M. M. Hassan et al. [16] studied and tested the workload of Big data by running a group of typical Big data jobs on Amazon cloud EC2. They created a large simulation scenario and compared the proposed method with other approaches. Though, the proposed approach was cost effective, performance metrics such as delay, delay variable and throughput were not taken into consideration.

R. Karim et al. [17] proposed a mechanism to map the users' QoS requirements of cloud services to the right QoS specifications of SaaS then map them to the best IaaS service that offers the optimal QoS guarantees. The end-to-end QoS values was calculated as a result of the mapping. They proposed a set of rules to perform the mappings process. The QoS specifications of cloud services was hierarchically modelled using the analytic hierarchy process (AHP) method. The AHP based model helped to facilitate the mapping process across the cloud layers and to rank the candidate cloud services for the end users. A case study was used to illustrate and validate the solution approach. No performance of evaluation was done based on real QoS datasets of cloud services.

S. Lee et al. [18] proposed an architecture that employed the agent technology to handle the monitoring of requested Quality of Service requirements and service level agreements, to support verification and validation. Furthermore, the agent technology dynamically analyzed resources allocation and deployment. This work's weak point was lack of self-learning algorithm to determine the timing of automatic allocation of system resources.

L. Bin et al. [19] et al. proposed a novel QoS-aware dynamic data replicas delete strategy for disk space and maintenance cost saving purpose. Experimental results demonstrated that the DRDS algorithm can save disk space and maintenance costs for distributed storage system while the availability and performance quality of service requirements are ensured. However, increased overhead on update and inconsistency of data is usually associated with data replication.

P. Zhang et al. [20] presented a QoS framework for mobile cloud computing and an adaptive QoS management process to manage QoS assurance in mobile cloud computing environment. Furthermore, they presented a QoS management model based on fuzzy cognitive map (FCM). No good model with a suitable configuration was generated.

Y. Xiao etal. [21] and presented an efficient reputation-based QoS provisioning scheme, which can minimize the cost of computing resources, while satisfying the desired QoS metrics. They considered the statistical probability of the response time as a practical metric rather than the typical mean response time. More so, QoS provisioning algorithm was not used to integrate security and privacy metrics.

M. Xu et al. [22] introduced a multiple QoS constrained scheduling strategy of multi-workflows (MQMW) to address the issue of multiple workflows with different QoS requirements. The proposed strategy could schedule multiple workflows which were started at any time though QoS constraints such as availability and reliability were not added to workflows.

TABLE 1: ANALYSIS OF QOS TECHNIQUES IN CLOUD COMPUTING

Title	Technique Used	Advantages	Disadvantages
System of Systems for Quality-of-Service Observation and Response in Cloud Computing Environment	Enterprise Monitoring, Management and Response Architecture in Cloud Computing (EMMRA CC) and System of System Approach (SoS)	(i) Enhanced Qos Performance (ii)Prevents distributed denial of service (DDoS) attacks.	The solution cannot be applied to federated clouds because cloud providers' servers were not integrated in real time.
A QoS–Oriented Inter-Cloud Federation Framework	QoS Oriented Cloud Computing Framework, Cloud Coordinators (CC) and Federated Coordinators (FC)	(i)Enables providers to dynamically act as backup for each other in peak times. (ii)Protect the Providers from any possible SLA violation	(i) Complex services were not constructed using a mixture of services from different cloud providers. (ii) No provision was made for distributed denial of service (DDoS) attacks.
An Approach of Quality of Service Assurance for Enterprise Cloud Computing (QoSAECC)	Multi agent model	Established an integrated cloud data service to support QoS and SLA manipulation.	The features and solutions of IOT was not adapted into the framework as well as the field experiments.
QoS-Aware Resource Provisioning for Big Data Processing in Cloud Environment	Heuristic algorithms	Cost effective and dynamic Vm allocation model to handle big data tasks.	Performance metrics such as delay, delay variation and throughput were not taken into consideration

An End-To-End QoS Mapping Approach for Cloud Service Selection	AHP based ranking algorithm	Presented a new way of computing end-to-end values in cloud environment	No performance evaluation was done based on real QoS datasets of cloud services
A QoS Assurance Middleware Model for Enterprise Cloud Computing	Agent technology	(i) Automated resource allocation (ii) Enhance transmission of cross layer control information with respect to some services performance such as SLA.	No self-learning algorithm was used to determine the timing of automatic allocation of system resources
A QoS-Aware Dynamic Data Replica Deletion Strategy for Distributed Storage Systems under Cloud Computing	DRDS Algorithm	Save disk space and maintenance cost for distributed storage system	Increased overhead on update and inconsistency of data
A QoS-Aware System for Mobile Cloud Computing	Fuzzy cognitive map and QoS Prediction Algorithm	Facilitates QoS prediction, establishment, assessment and assurance	No good model with suitable configurations was generated.
Reputation-based QoS Provisioning in Cloud Computing via Dirichlet Multinomial Model	Dirichlet Multinomial model	The proposed management framework provides an efficient QoS provisioning scheme for cloud computing	QoS provisioning algorithm was not used to integrate security and privacy metrics
A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing	Scheduling Algorithm	Produced better scheduling results	QoS constraints such as reliability and availability was not added to workflows.

IV. CONCLUSION

In this paper, we surveyed various QoS techniques in cloud computing to ascertain the extent to which QoS challenge has been resolved. Though Many researchers have provided scheduling techniques, admission control, traffic control, dynamic resource provisioning, etc in order to handle the issue of QoS in cloud computing. To the best of our knowledge cloud computing is still having QoS-related challenges. Thus, we propose a Performance monitoring application using System of Systems (SoS) approach which can monitor and enhance the Quality of Service in cloud environment .Our approach will involve integration of n number of cloud servers using system of system (SoS) approach. The QoS metrics will be monitored to enable the cloud users to access resources from n number of cloud servers.

ACKNOWLEDGEMENT

We are grateful to the Almighty God for the grace to complete this work. We express our gratitude to our families for their support and encouragement. We extend a hand of fellowship to our Colleagues for their encouragement.

REFERENCES

[1] S. Dubey and S. Agrawal, "Methods to ensure quality of service in cloud computing environment", International Journal of Advanced Research in Computer Science and Software Engineering, Volume: 3, Issue: 6, pp. 405-411, 2013.

[2] D.Ardagna, G. Casale, M. Ciavotta, J. F. Perez and W. Wang, "Quality-of-Service in cloud computing: modeling techniques and their applications", Journal of Internet Services and Applications, Volume: 5, Issue: 11, pp. 1-12, 2014.

[3] V. Stantchev and C. Schropfer, "Negotiating and enforcing QoS and SLAs in grid and cloud computing", in Proc. Int. Conf. on Advances in Grid and Pervasive Computing, GPC '09, pp. 25-33, 2009.

[4] Z. Yuchao, D. Bo, P. Fuyang, "An Adaptive QoS-aware cloud", in Proc. of Int. Conf. on Cloud computing, Technologies, Applications and Management, pp. 160-163, 2012.

[5] P. Mell and T. Grance, "The NIST definition of cloud computing", Natl. Inst. of Standards of Technology (NIST), U.S Dept. of Commerce, Gaithersburg, MD, USA, NIST Special Publication 800-145, pp. 2-3, 2011.

[6] E. B. K. Manash and T. U. Rani, "Cloud computing- A potential area for research", International Journal of Computer Trends and Technology (IJCTT), Volume: 25, no.1, pp.10-11, 2015.

[7] 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.
[8] K. Ashar, "The Next Generation Enterprise: Business as a Service in cloud", Linkedin.com, pp. 1-2, 2015.
[9] J. Li, F. Bennier, and Y. Amghar, "Governance as a Service for collaborative environment", in Proc. IFIP Advances in Information and Communication Technology, Volume: 380, pp. 687-690, 2012.
[10] 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.
[11] D. Ardagna, B. Panicucci, M. Trubian, L. Zhang, "Energy-aware autonomic resource allocation in multitier virtualized environment", IEEE Transactions on Services Computing, pp 1-18, Volume: 5, Issue: 1, 2012.
[12] 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.
[13] 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.
[14] M. Salam and A. Shawish, "A QoS-oriented inter-cloud federation framework", IEEE Systems Journal, pp. 642-643, 2015.
[15] 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.
[16] M. M. Hassan, B. Song, M. S. Shamin, and A. Alamri, "QoS-aware resource provisioning for big data processing in cloud computing environment", in Proc. Int. Conf. on Computational Sc. and Computational Intelligence, pp. 107-112, 2014.
[17] 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.
[18] 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.
[19] 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.
[20] P. Zhang, and Z. Yan, "A QoS-aware system for mobile cloud computing", in Proc. of IEEE, pp. 518-522, 2011.
[21] 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.
[22] 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.