View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UUM Repository

A Comprehensive Survey on Quality of Service Implementations in Cloud Computing

Mohamed Firdhous, Suhaidi Hassan, Osman Ghazali

Abstract— Quality of service plays is an important factor in distributed computing. Cloud computing has been the paradigm in distributed computing. Under cloud computing, computing resources are hosted in the internet and delivered to customers as services. Prior to the commencement of services, the customers and cloud providers negotiate and enter into an agreement named service level agreement. The services level agreements clarify the roles, set charges and expectations and provide mechanisms for resolving service problems within a specified and agreed upon time period. Service level agreements also cover performance, reliability conditions in terms of quality of service guarantees. In this paper, the authors present a comprehensive survey on quality of service implementations in cloud computing with respect to their implementation details, strengths and weaknesses.

Index Terms— Cloud Computing, Utility Computing, QoS, SLA.

1 Introduction

Cloud computing has the changed the entire computing landscape by making the resources available over the internet as services. Similar to electricity, water, gas and telephony, computing also becomes a utility under cloud computing [1]. Under the utility computing paradigm, computing resources including hardware, development environment and user applications can be accessed remotely over the Internet and paid for only the usage. In the recent times, due to the popularity of cloud systems the market has been flooded with a large number of cloud service providers [2]. These cloud providers host their services on the Internet and make them available to any customer who would like to purchase them. In [3], Garg, Gopalaiyengar and Buyya state that at any given time, large virtualized systems may host and serve thousands of customers. Though cloud computing systems are advantageous to both customers and service providers in terms of economy and utilization of resources, if the resource provisioning is not carried out optimally it would also become a disaster [4]. Similar to any other subscription based services, prior to the commencement of the services, the service providers and customers enter into an agreement called Service Level Agreement (SLA) [5]. The SLA would contain the roles and responsibilities of the parties involved, scope of services, quality and performance requirements, charges and rates etc. Thus Quality of Services (QoS) plays an important role in making the cloud services acceptable to customers.

Literature has reported many implementations for measuring and ensuring QoS in cloud computing systems. In this paper, the authors carry out a comprehensive survey on the mechanisms and methods proposed by various researchers with respect to their implementation principles, strengths and weaknesses.

2 Cloud Computing

Cloud computing has been identified as the 5th utility after electricity, gas waster and telephony due to the way it transforms, how computing resources can be accessed, used and paid for [1]. Traditional distributed computing where computing resources were leased from business data centers required users to purchase fixed capacities and pay for them irrespective of actual usage. On the other hand, cloud computing only charges for the usage and the resources committed to the users is elastic and closely follow the demand patterns [6]. Hence cloud computing makes the investment on computing an asset that can immediately return the investment made on them.

Cloud computing has been divided into three main layers. They are namely, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [7]. Fig. 1 shows the Cloud computing layers along with the underlying physical computing infrastructure and virtualized computing infrastructure as two distinct layers. The physical hardware is the real workhorse that carries out the processing. The physical hardware is generally provided in the form of computing clusters, grids or individual servers [8]. The virtualized computing infrastructure is created by installing a Virtual Machine Manager (VMM) on the physical hardware [9]. The VMM provides the necessary isolation and security between the multiple virtual machines running in parallel on a single physical computer.

- *Mohamed Firdhous is a Senior Lecturer attached to the Faculty of Information Technology, University of Moratuwa, Sri Lanka. He is currently on study leave and pursues PhD in computer networks attached to the InterNetWorks Research Lab, School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Malaysia, E-mail: mfirdhous@internetworks.my, Mohamed.Firdhous@uom.lk*
- *Prof. Dr. Suhaidi Hassan is a professor in the School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Malaysia. He presently heads the InterNetWorks Research Group and the chairman InterNetWorks Research Lab. Prior to this, he was the Assistant Vice Chancellor of the College of Arts and Sciences, Universiti Utara Malaysia, Malaysia, E-mail: suhaidi@uum.edu.my*
- *Associate Prof. Dr. Osman Ghazali is an associate professor attached to the School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Malaysia and a member of the InterNetWorks Research Group. Previously he was the head of the Department of Computer Science, within the School of Computing and Technical Director of the Universiti Teaching Learning Center, Universiti Utara Malaysia, E-mail: osman@uum.edu.my*

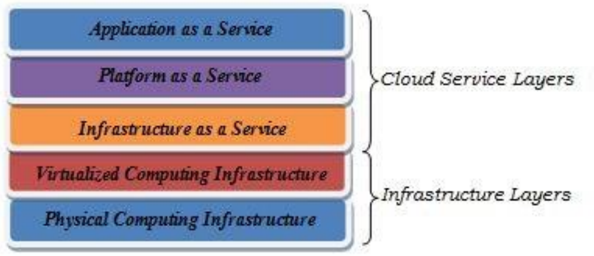

Fig. 1: Cloud Computing Layers

IaaS is the provision of virtual hardware as a service over the Internet. These virtual machines can be brought up and removed on the fly based on customer demand. Once a virtual machine has been purchased, it can be treated as if it is real hardware and any operating system and applications can be installed on it. PaaS is the complete software development environment along with operating system, development and testing tools and application programming interface installed on virtual hardware. PaaS helps web based application developers to reduce the cost and time of bringing their applications to market from the design boards. SaaS is the new paradigm of software marketing and ownership. SaaS enables customers to access web based applications hosted in remote data centers and pay only for the usage. These applications have the capability of managing their own data and configuration information to suit individual user requirements.

3 Quality of Service in Cloud Computing

Cloud computing systems may host thousands of globally dispersed clients at any given time. These clients may access different types of services that have varying requirements depending on the type of clients, services and resources involved. In order to meet the requirements of clients and services, it is necessary to provide a certain level of QoS by the service providers. Nevertheless, providing a guaranteed QoS in such a challenging environment in a widely distributed diverse networks supporting complex hosting of services is not an easy task [10, 11]. Though it is a challenging task, several researchers have undertaken to develop mechanisms, frameworks and systems which could guarantee the QoS requirements of different services. This section takes an in depth look at these mechanisms, frameworks and systems.

In [5], Feng et al have proposed an optimal resource allocation model for revenue maximization. The model has been mathematically derived and tested using both synthetic and traced datasets. The proposed model performs better than heuristic optimization of resources in maximizing profits. But the application of this method is limited as it considers only the mean response time as the QoS attribute to be satisfied. For customers who require guaranteed performance or at least a commitment in terms of a confidence level cannot be served through this model. Hence from the customers' point of view, the model has limited application and may serve only casual users.

Buyya, Garg, and Calheiros have proposed a framework for SLA management with special reference to managing QoS requirements in [11]. The proposed architecture successfully integrates the market based resource provisioning with virtualization technologies for flexible resource allocations to user applications. But the proposed architecture does not support different cloud service offerings such as IaaS, PaaS and SaaS together in a combined manner.

In [12], Liu et al have proposed a generic QoS framework for cloud workflow system. The proposed framework covers all the four stages of cloud workflow namely, QoS requirement specification, QoS-aware service selection, QoS consistency monitoring and QoS violation handling. The shortcoming of this framework is that it does not specifically identify any QoS parameters and also does not discuss how to differentiate clients requiring different QoS levels.

Chen and Zhang have proposed a workflow scheduling algorithm based on Particle Swarm Optimization (PSO) in [13]. The proposed mechanism can optimize up to seven parameters specified the users compared to traditional optimization techniques that consider only the workflow execution time. The downside of the proposed mechanism is that it lacks a monitoring scheme for catching QoS violations or punish the violators.

In [14], den Bossche, Vanmechelen and Broeckhove have proposed a set of heuristics for scheduling deadline-constrained applications in a hybrid cloud system in a cost effective manner. The proposed system attempts to maximize the use of local resources along with minimizing the use of external resources without compromising the QoS requirements of the applications. The optimization heuristics takes the cost of both computation and data transfer along with the estimated data transfer times. The main criteria in optimization is the maximization of cost saving. The effect of different cost factors and workload characteristics on the cost savings have been analyzed along with the sensitivity of the results to the different runtime estimates. The advantages of the proposed methodology is that it can select an optimized set of resources from both in-house (private) and public cloud systems for meeting the QoS requirements. But at the same time it suffers from certain weaknesses. Though it is concerned only about the deadline concerned applications, it does not consider the failures that may occur after the scheduling has been done. The failure will increase the cost of execution and affect the application in terms of quality.

In [15], Emeakaroha et al., have presented a scheduling heuristic that takes multiple SLA parameters for application deployments in the Cloud. The attributes considered include CPU time, network bandwidth and storage capacity for deploying applications. These parameters have limited application in real world systems as they need to be considered only during deployment. Once the applications have been ready for client access, the customers would be more interested in performance parameters such as response time, processing time etc. Hence this heuristic may not have much practical significance in real world business environments.

Li et al in [16] have proposed a novel customizable cloud workflow scheduling model. The authors have incorporated trust into the model in addition to the QoS targets. In order to analyze the users' requirements and design a customized schedule, the authors propose a two stage workflow model where the macro multi-workflow stage is based on trust and micro single workflow stage classifies workflows into time-sensitive and cost-sensitive based on QoS demands. The classification of workflows has been carried out using fuzzy clustering technique. The proposed model restricts the QoS parameters considered to response time, bandwidth, storage, reliability and cost. Also the delivery of QoS is confined only to average values and no guarantee of service delivery is provided at least in terms of a predetermined confidence level. This is a strong limitation of the proposed technique as the users do not have the freedom to select their own QoS parameters and no guarantee of the QoS delivery at the least a statistical validation.

Mushtaq, Augustin, and Mellouk have investigated the effect of different factors on the Quality of Experience (QoE) of multimedia users in a cloud computing network [17]. The authors of this paper have grouped the factors that affect the QoE into four groups. They are namely network parameters, characteristics of videos, terminal characteristics and types of users' profiles. The data collected through different methods have been classified using machine learning techniques such as Naive Bayes, Support Vector Machines, K-Nearest Neighbors, Decision Tree, Random Forest and Neural Networks. Out of these methods they have determined the best method for QoS/QoE correlation after evaluating them. Hence it can be concluded that this paper discusses more about the capabilities of machine learning techniques than about QoS or QoE. The QoS/QoE correlation is a case for evaluating the machine learning techniques.

Alhamazani et al., in [18] have outlined the importance of dynamically monitoring the QoS of virtualized services. they further claim that the monitoring of the services would help both the cloud provider and application developer to maximize the return of their investments in terms of keeping the cloud services and hosted applications operating at peak efficiency, detecting changes in service and application performance, SLA violations, failures of cloud services and other dynamic configuration changes. The paper mainly concentrates on describing the PhD work being carried out in terms of research questions, objectives and methodology. The researchers mainly concentrate on SNMP based QoS monitoring. Since this is a concept paper describing work in progress, no concrete proposal is put forward or evaluated.

In [19], Li et al., have adopted a profit model based on response time to represent the QoS requirements and shown that this model yields different analytical results to that obtained from traditional metrics. It was also shown that both under allocation and over allocation of resources affect profits. The right allocation of resources depends on many factors such as available resources, workload distribution, system configuration, and profit model. This is an innovative method of analyzing the effect of managing QoS on resource utilization. The results only discuss the effect of managing QoS, how to provide an optimal allocation is not discussed.

In [20], Adami et al., have proposed a distributed resource allocation algorithm for cloud and grid systems. The algorithm is capable of handling multiple resource requirements and the criteria for optimization is a tradeoff between the execution time and economic cost with system and network performance parameters as additional factors. The proposed algorithm successfully incorporates many system and network performance parameters but fails to consider the failures that may arise after allocating resources. The failures arising after the allocation increase the cost of computation as they would require more time for execution. Hence the cost based optimization used in the proposed algorithm may not be accurate due this shortcoming.

Gohad, Ponnalagu, and Narendra have proposed an extensible dynamic provisioning framework for multi tenant cloud system [21]. The proposed framework starts with defining a tenancy requirements model for helping map provisioned resources. The other index called the health grading model handles the QoS characteristics of tenants. Together both these indices permit dynamic re-provisioning for existing tenants based on changing tenancy requirements or health grading predictions. The proposed framework is innovative in the dynamic resource provisioning sense, but may not be suitable for applications that have bursty requirements. Also the proposed framework is based on starts small and grows large criterion. But when new tenants arrive, the allocated resources are not deallocated from the existing tenants, this would starve the new tenants of resources.

In [22], Ma, Sun and Abraham and proposed a lightweight framework for monitoring public clouds. The proposed framework is very less resource intensive but does not monitor the QoS parameters such as response time, processing time etc., which the customers may be more interested in.

Zhu and Agrawal have presented a framework for handling adaptive applications in cloud systems in [23]. The proposed framework is based on multi-input-multi-output feedback control model for resource provisioning. But the model is limited to memory and CPU performance only, hence the application may be affected by the underperformance of other resources such as network, disk drives etc.

In [24], Sharma et al., have proposed a cloud resource pricing model balancing QoS requirements and higher profits. This model uses the realistic valuation for underlying resources using the age of resources. The proposed model does not include utilization in computing the cost. Hence it may lead to inaccurate projections.

Stoicuta et al., have developed a client application for monitoring cloud QoS on iOS5 [25]. This application can be used by clients to monitor the performance of their cloud provider. But the application has been designed very narrowly focusing only on available transfer rate and one-way delay. Hence the application has limited applications.

In [26], Goyal et al., have proposed a QoS based trust management model. The proposed model claims to use multiple QoS attribute to compute the trust value, but there is no clear explanation how these parameters are combined. Also there is no prioritization between parameters is possible.

Iyer and Veeravalli have formulated a resource allocation strategy for cloud infrastructure based on bargaining [27]. They have combined the Nash Bargaining Solution and Raiffa Bargaining Solution to arrive at an optimal allocation strategy. The proposed strategy handles the dynamic nature of cloud very well during run time but the system does not permit to manage resources from multiple sources. Hence if a single service provider cannot meet all the requirements of the customer, he will be required to settle for a sub optimal allocation of resources.

In [28], Sanchez et al., have investigated the capability of Markov Arrival Processes based queuing models to predict future workload of cloud systems. The model has tested only with numerical experiments, hence the true capability of the models need to be evaluated with real data traces.

An optimization framework for cross layer cloud services has been proposed in [29] by Kouki, Ledoux, and Sharrock. The optimization across multiple layers has been carried out enforcing the SLA dependencies between them. The framework is very suitable for vendors marketing multitude of services and also takes the dynamic nature of cloud systems. The propose system currently lacks the run time management of QoS performance.

Wu, Garg and Buyya in [30] have proposed some algorithms for resource allocation for SaaS providers to balance the cost of hardware and SLA violations. This proposed algorithm takes certain QoS parameters such as response time and service initiation time for satisfying the customers while minimizing the use of hardware resources. Theses algorithms propose to reuse the already created VMs in order to minimize cost, but it may create security problems for customers as the residual information in the VMs can be used against them.

In [31], Phillips, Engen and Papay have reported the results obtained on the performance of virtualized hardware of two IaaS providers. They have used the Dwarf benchmarks for measuring the performance of these cloud providers and show that the actual performance and show that the labeling such as small, medium or large does not actually reflect the true nature of a system and also they show that certain applications may run better on certain hardware than the other ones.

Chauhan et al., have proposed a process for identifying a cloud service provider for a given set of requirements by matching SLA parameters in [32]. The proposed process finds a match by crating two models called the capability model and requirements model and then translating these models to graphs for matching them for compatibility. Based on the compatibility, each node pair is given a mark between 0 and 1 and the final score is computed by summing them all. This is a good effort for automating the process of finding a match between a customer's requirements and service provider's capability. But, it does not take the dynamic nature of the cloud service into account. It only matches the published capability of cloud providers with customers' requirements. This is a major shortcoming of this process.

Table 1 summaries all the work discussed so far with reference to their strengths and weaknesses along with the proposed model or framework. From Table 1, it can be seen that there is still a lot of scope for future work in this exciting and challenging area.

Table 1: Summary of Strengths and Weaknesses of Proposed Models and Frameworks

Work	Proposed Model/Framework	Strengths	Weaknesses
[5]	Optimal resource allocation model for revenue maximization	Mathematically derived and performs better than heuristics	Only mean performance time is considered, hence not suitable for QoS sensitive applications requiring guaranteed performance.
[11]	A framework for SLA management with special reference to managing QoS requirements	Successfully integrates the market based resource provisioning with virtualization technologies for flexible resource allocations.	Does not integrate IaaS, PaaS and SaaS in a combined manner.
[12]	A generic QoS framework for cloud workflow	Covers all the four stages of cloud workflow.	QoS metrics are not identified and no mechanism for differentiating customers based on requirements.
[13]	A set-based PSO approach scheduling problem in cloud computing	Multiple parameter optimizations are possible.	But no monitoring mechanism is implemented for catching violations.
[14]	A set of heuristics for scheduling deadline-constrained applications in a hybrid cloud system.	The optimization heuristics takes the cost of both computation and data transfer along with the estimated data transfer times and different cost factors and workload characteristics.	It does not consider the failures that may occur after the scheduling has been done. The failure will increase the cost of execution and affect the application in terms of quality
[15]	A scheduling heuristic that takes multiple SLA parameters when deploying applications in the Cloud	Considers deployment attributes such as CPU time, network bandwidth, storage capacity etc., before installation of applications in the cloud system.	Does not consider performance parameters such as response time, performance time etc.,
[16]	A flexible multistage workflow scheduling model.	The proposed model is flexible due to breaking up of the workflow scheduling mechanism into multiple stages and grouping the requests based on the user requirements.	Application is strongly limited due to strict restriction on the type of QoS attributed taken into account and the absence of QoS delivery guarantees.
[17]	The correlation between QoS/QoE has been studied	QoS/QoE correlation has been studied using a selected set of machine learning techniques.	This paper discusses more about the capabilities of machine learning techniques than about QoS or QoE. The QoS/QoE correlation is a case for evaluating the machine learning

			techniques.
[18]	Proposal for monitoring the cloud system for QoS performance	Only the concept and idea based work in progress have been described.	No concrete proposal or evaluation is presented.
[19]	Profit-Based Analysis of Resource Allocation on QoS	An innovative method for analyzing the impact of resource provisioning.	No discussion on how to optimally allocate resources.
[20]	A distributed resource allocation algorithm for cloud and grid systems	Capable of handling multiple resource requirements	Too simple, as it assumes perfect conditions for execution. Failures after allocation of resources are not taken into account.
[21]	Extensible dynamic provisioning framework for multi tenant cloud system	The proposed framework is dynamic and allocates resources depending on the tenant requirements.	May not be capable of handling bursty requirements with short duration and large resource requirements. The new tenants arriving late may suffer from resource starvation.
[22]	Lightweight framework for monitoring public clouds	Less resource intensive.	Does not monitor the real QoS parameters such as response time, processing time etc.
[23]	A framework for handling adaptive applications in cloud systems.	Based on multi-input-multi-output feedback control model for resource provisioning.	Limited only to CPU and memory provisioning. Hence application performance may be affected by other resource constraints such as network, storage etc.
[24]	A resource pricing model for QoS and profit balancing.	Uses realistic values using age as a parameter.	Utilization is not considered in computing cost. Hence may produce inaccurate costs.
[25]	A monitoring application for QoS parameters in iOS5.	Can be used by clients to monitor the performance of service providers.	Very narrow application due to focusing only on available transfer rate and one-way delay as QoS parameters.
[26]	A QoS based trust management model	Multiple QoS parameters can be used.	No clear explanation on how to use the parameters is given nor is there any possibility to prioritize the parameters.
[27]	Resource allocation in a Compute Cloud through bargaining approach.	The proposed strategy handles the dynamic nature of cloud very well during run time.	May lead to sub optimal solutions from a customer's perspective, if a single provider cannot meet all the requirements.
[28]	Investigation of the capability of MAP based queuing models for predicting workload of cloud systems.	Markov arrival processes have the capability fir heavy tail distributions that are common in web applications.	Only numerical experiments have been used to validate the model, hence needs further validation with real data traces.
[29]	An optimization framework for cross layer cloud services.	Suitable for vendors selling products across multiple layers. Dynamic nature of cloud has been considered.	Lacks the run time management of QoS performance.
[30]	Algorithms for resource allocation for SaaS providers for balancing cost and QoS.	It helps reduce the cost of SaaS providers without compromising the QoS of customers.	Due to reuse of already open VMs, it can create security problems for customers.
[31]	Results of an initial investigation of using Dwarf benchmarks to measure the performance of virtualized hardware.	It is shown that general labeling of cloud service providers for size or the number of units used is not sufficient to predict the real capabilities through real experiments.	These are a set of experiments that can be in a laboratory by experts and may not help the general set of customers who are not that tech savvy.
[32]	A process for matching providers' capability with customers' requirements based on SLA parameters.	A good effort for automating the matching process that was hitherto done manually by customers.	Only matches the service providers published capabilities with customer requirements. Thus it cannot track the changes in cloud performance due to dynamic nature of clouds.

4 Conclusions

Cloud computing has been the paradigm shift in distributed computing due to the way the resource provisioning and charging. Managing QoS is a critical task in making such an innovative technology to a larger audience. Several researchers have put forward their ideas for new and innovative solutions for handling this vital area is resource management. In this paper, the authors have carried out a critical review of the most recent work carried out in this area. The findings of the authors in terms of the strengths and weaknesses of the proposed work has been presented in a table for easy reference.

References

[1] 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," *Journal of Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, June 2009.

[2] B.P. Rimal, E. Choi, and I. Lumb, I. "A Taxonomy and Survey of Cloud Computing Systems," *Proc. 5th Int. Joint Conf. INC, IMS and IDC (NCM'09)*, pp. 44-51, 2009.

[3] S.K. Garg, S.K. Gopalaiyengar, and R. Buyya, "SLA-based Resource Provisioning for Heterogeneous Workloads in a Virtualized Cloud Datacenter," *Proc. 11th Int. Conf. Algorithms and Architectures for Parallel Processing (ICA3PP2011)*, pp. 371–384, 2011.

[4] D. Bourcier, and P.D. Filippi, "Cloud Computing: New Research Perspectives for Computers and Law," *Proc. 13th Int. Conf. Artificial Intelligence & Law*, pp. 73–92, 2011.

[5] G. Feng, S. Garg, R. Buyya, and W. Li, "Revenue Maximization using Adaptive Resource Provisioning in Cloud Computing Environments," *Proc. 13th ACM/IEEE Int. Conf. Grid Computing,* pp. 192–200, 2012.

[6] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," *Communications of the ACM,* vol. 53, no. 4. Pp. 50-58, April 2010.

[7] C. Vecchiola, S. Pandey, and R. Buyya, "High-performance cloud computing: A view of scientific applications," *Proc. 10th Int. Symp. Pervasive Systems, Algorithms, and Networks (ISPAN),* pp. 4–16, 2009.

[8] R. Buyya, A. Beloglazov, and J.H. Abawajy, "Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges", *Proc. Int. Conf. Parallel and Distributed Processing Techniques and Applications,* pp.6-20, 2010.

[9] Y. Li, W. Li, and C. Jiang, "A Survey of Virtual Machine System: Current Technology and Future Trends," *Proc. Third Int. Symp. Electronic Commerce and Security,* pp. 332-336, 2010.

[10] J.M. Pedersen, M.T. Riaz, J.C. Júnior, B. Dubalski, D. Ledzinski, and A. Patel, "Assessing Measurements of QoS for global Cloud Computing Services," *Proc. Ninth IEEE Int. Conf. Dependable, Autonomic and Secure Computing,* pp. 682-689, 2011.

[11] R. Buyya, S.K. Garg, and R.N. Calheiros. "SLA-Oriented Resource Provisioning for Cloud Computing: Challenges, Architecture, and Solutions," *Proc. Int. Conf. Cloud and Service Computing,* pp. 1-10, 2011.

[12] X. Liu, Y. Yang, D. Yuan, G. Zhang, W. Li, and D. Cao, "A Generic QoS Framework for Cloud Workflow Systems," *Proc. Ninth IEEE Int. Conf. Dependable, Autonomic and Secure Computing,* pp. 713-720, 2011.

[13] W.N. Chen, and J. Zhang, "A Set-Based Discrete PSO for Cloud Workflow Scheduling with User-Defined QoS Constraints," *Proc. IEEE Int. Conf. Systems, Man and Cybernetics,* pp. 773-778, 2012.

[14] R.V. den Bossche, K. Vanmechelen, and J. Broeckhove, "Cost-Efficient Scheduling Heuristics for Deadline Constrained Workloads on Hybrid Clouds," *Proc. 3rd IEEE Int. Conf. Cloud Comp. Tech. and Sc., (CloudCom),* pp. 320-327, 2011.

[15] V.C. Emeakaroha, I. Brandic, M. Maurer, and I. Breskovic, "SLA-Aware Application Deployment and Resource Allocation in Clouds," *Proc. 35th IEEE Annual Computer Software and Applications Conference Workshops (COMPSACW),* pp. 298-303, 2011.

[16] W. Li, Q. Zhang, J. Wu, J. Li, and H. Zhao, "Trust-based and QoS Demand Clustering Analysis Customizable Cloud Workflow Scheduling Strategies," *Proc. IEEE Int. Conf. Cluster Comp. Workshops,* pp. 111—119, 2012.

[17] M.S. Mushtaq, B. Augustin, and A. Mellouk, "Empirical Study based on Machine Learning Approach to Assess the QoS/QoE Correlation", *Proc. 17th European Conf. Networks and Optical Comm.,* pp. 1-7, 2012.

[18] K. Alhamazani, R. Ranjan, F. Rabhi, L. Wang, and K. Mitra, "Cloud Monitoring for Optimizing the QoS of Hosted Applications," *Proc. 4th IEEE Int. Conf. Cloud Comp. Tech. and Sc.,* pp. 765-770, 2012.

[19] J. Li, Q. Wang, D. Jayasinghe, S. Malkowski, P. Xiong, C. Pu, Y. Kanemasa, and M. Kawaba, "Profit-Based Experimental Analysis of IaaS Cloud Performance: Impact of Software Resource Allocation," *Proc. Ninth IEEE Int. Conf. Services Computing,* pp. 344-351, 2012.

[20] D. Adami, C. Callegari, S. Giordano, and M. Pagano, "A Hybrid Multidimensional Algorithm for Network-aware Resource Scheduling in Clouds and Grids," *Proc. IEEE Int. Conf. Comm.*, pp. 1297-1301, 2012.

[21] A. Gohad, K. Ponnalagu, and N.C. Narendra, "Model Driven Provisioning in Multi-tenant Clouds." *Proc. Annual SRRI Global Conference*, pp. 11-20, 2012.

[22] J. Ma, R. Sun, and A. Abraham, "Toward a Lightweight Framework for Monitoring Public Clouds," *Proc. Fourth Int. Conf. Computational Aspects of Social Networks,* pp. 361-365, 2012.

[23] Q. Zhu and G. Agrawal, "Resource Provisioning with Budget Constraints for Adaptive Applications in Cloud Environments," *IEEE Trans. Services Computing,* vol. 5, no. 4, pp. 497-511.

[24] B. Sharma, R.K. Thulasiram, P. Thulasiraman, S.K. Garg, and R. Buyya, "Pricing Cloud Compute Commodities: A Novel Financial Economic Model," *Proc. 12th IEEE/ACM Int. Symp. Cluster, Cloud and Grid Computing,* pp. 451-457, 2012.

[25] F. Stoicuta, I. Ivanciu, E. Minzat, A.B. Rus, and V. Dobrota, "An OpenNetInf-Based Cloud Computing Solution for Cross-Layer QoS: Monitoring Part Using iOS Terminals," *Proc. 10th Int. Symp. Electronics and Telecomm.*, pp. 167—170, 2012.

[26] M.K. Goyal, A. Aggarwal, P. Gupta, and P. Kumar, "QoS based Trust Management Model for Cloud IaaS," *Proc. 2nd IEEE Int. Conf. Parallel, Distributed and Grid Computing,* pp. 843-847, 2012.

[27] G.N. Iyer and B, Veeravalli, "On the Resource Allocation and Pricing Strategies in Compute Clouds using Bargaining Approaches," *17th IEEE Int. Conf. Networks,* pp. 147-152, 2011.

[28] S.P. Sanchez, G. Casale, B. Scotney, S. McClean, G. Parr, and S. Dawson, "Markovian Workload Characterization for {QoS} Prediction in the Cloud," *Proc. IEEE Int. Conf. Cloud Computing,* pp. 147-154, 2011.

[29] Y. Kouki, T. Ledoux, and R. Sharrock, "Cross-Layer SLA Selection for Cloud Services," *Proc. First Int. Symp. Network Cloud Computing and Applications,* pp. 143-147, 2011.

[30] L. Wu, S.K. Garg, and R. Buyya, "SLA-based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments," *Proc. 11th IEEE/ACM Int. Symp. Cluster, Cloud and Grid Computing,* pp. 195-204, 2011.

[31] S.C. Phillips, V. Engen, and J. Papay, "Snow White Clouds and the Seven Dwarfs," *Proc. Third IEEE Int. Conf. Cloud Comp. Tech. and Sc.,* pp. 738-745, 2011.

[32] T. Chauhan, S. Chaudhary, V. Kumar, and M. Bhise, "Service Level Agreement Parameter Matching in Cloud Computing," *World Cong. ICT,* pp. 564-570, 2011.

IJSER