QoS Aware Resource Provisioning in Federated Cloud and Analyzing Maximum Resource Utilization in Agent Based Model

Pradeep Kumar V, Kolla Bhanu Prakash

Abstract*: *Federated cloud computing is one form of cloud computing model which supports huge applications needs of Big Data services by managing with collaboration from a different set of cloud service providers. It breaks the application boundary of consumers' access to the large no of resources within specified QoS (Quality of Service) parameters leading to less SLA (Service Level Agreement) Violations. This study is specific to QoS aware resource provisioning in federated clouds to ensure resource provisioning done by cloud broker to provide uninterrupted cloud services and minimize loss of availability of resources, decrease failure rates and meet the application requirements of multiple cloud consumers in managing their resource utilization. Finally, we highlight the open challenges for future research towards the QoS aware resource provisioning in federated clouds and analyze the maximum resource utilization in agent based model.

Index Terms*: *Quality of Service, Resource Provisioning, Cloud Computing, Federated Clouds, Service Level Agreement

I. INTRODUCTION

Cloud computing is a computing paradigm which provides a huge set of resources for satisfying user request at any instance. Large data centers are made available for storing the user transactions and processing their request is managed by flexible virtual machine technology which provides computation resources like memory, CPU's and storage to meet the specific end-user requirements. Apart from providing flexible resources, the cloud has changed the way of providing different services like SaaS, PaaS, IaaS and XaaS (anything as a service) by pay as you use model. The huge demand of enterprises depending on cloud services has become a considerable challenge for cloud providers to meet the requirements of their cloud consumers. Federated cloud is one of the flavors of cloud computing model where more no of cloud providers get collaborated to provide their resources to meet the demand of enterprises that are facing challenges in meeting their end-user requirements [29].

In the federated cloud environment, the cloud provider plays two roles one act as an infrastructure provider and other as the consumer of another provider. All cloud providers target to maximize their revenue by satisfying as many consumers as possible with their available resources or by collaboration with other providers. Resource provisioning is a critical task to manage in a federated cloud as at an instance the request for resources will come from other providers and also consumers and need to provide service without any SLA (Service Level Agreement) violation at cloud broker [27]. The chance of over-provisioning and under-provisioning in the federated cloud is a major challenge which needs to be effectively managed to provide continuous services or resources with awareness of different QoS parameters which are part of SLA specification. The study focuses on QoS aware resource provisioning [28] of federated cloud in which collaborated cloud providers along with consumers are getting resources allocated and serviced without SLA violation. Analyze the implementation models of a federated cloud for resource provisioning and explore the possible open challenges like effective collation formation, QoS aware differentiated resource provisioning and to predict the resource utilization after provisioning resources. The remaining part of this article is divided into three sections. Section 2, presents a detailed survey on the federated cloud and their resource provisioning with QoS awareness and possible implementation models for creating federation among cloud providers with a QoS perspective. Section 3, Lists the open challenges for QoS aware resource provisioning and Section 4, concludes the paper.

Revised Manuscript Received on June 15, 2019.

Pradeep Kumar V, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.

Kolla Bhanu Prakash, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.

II. LITERATURE REVIEW

Federated Cloud

Cloud computing model is designed for solving issues related to cloud providers and generate maximum revenue in providing resources to all cloud users. The cloud consumers who are sticking to single cloud provider need to depend on their services facing challenges like disruption of services, lack of availability of resources, over provisioning and under provisioning of resources. The need for cloud federation has come into existence as mentioned in [30] for reasons like sharing of resources, managing failover, measured QoS values, Cost effectiveness , reduced SLA violation and provider's and consumer's independence while ending contracts. Several flavors of federated cloud, e.g. like hybrid cloud, multi-cloud, and aggregated clouds have been proposed in [29] for managing the resource management without violation of QoS.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The nature of federation implementation is a difficult scenario for managing resources with different API's designed for service composition or configuration with multiple clouds.

QoS Awared Resource Provisioning

In cloud computing, the provision of large set of resources to cloud users without violation of QoS parameter in SLA is a major challenge for cloud providers. An adaptive QoS aware virtual machine provisioning mechanism was proposed in [1] to provide effective resources utilization and VM of similar types of requests are recycled to reduce the VM creation time to serve as many requests made by users. QoS parameters are linked to low-level infrastructure resources for serving all tasks within specified values of SLA. An adaptive recycling and provisioning algorithm in [1] is considered for different tasks assigned for different queues according to priority factor and then checked for provisioning of resources to complete the task without any SLA violation.VM Multiplexing may lead to a violation of the VM allocation policy made by the above algorithm and to manage to get priority factors an efficient memory access mechanism is needed.

The QoS aware resource provisioning for resource reservation based framework was proposed in [11] for measuring user QoS satisfactory through a flexible virtual machine reservation mechanism. This mechanism is integrated with middleware and tested for different user requirements satisfaction but it needs to be tested for higher throughput and less energy consumption. Managing of adjustment of QoS parameters on the fly according to resource utilization made by users while working on parallel and distributed application for scalable storage systems was implemented with hierarchically distributed scheduling mechanism with Coarse-grained CPU mapping and fine-grained CPU scheduling for adjustable computing power was proposed in [2]. This distributed mechanism in [2] was not tested for a large number of cloud users with wide range of resource requirements.

The dynamic management of virtual machine like VM placement and VM migration by aiming to reduce the energy consumption without violating SLA at all data centers was given in [3] where many algorithms were proposed in perspective of energy consumption while performing VM Relocation, VM consolidation and simultaneously checking for violation of SLA by ESDR (Energy reduction and SLAV-reduction dynamic run-time replacement) [3] mechanism. This ESDR [3] mechanism was worked only with CPU loads but not considered for memory and bandwidth loads at a run time and switching between multiple goals approaches to improve the efficiency of a data centre for different workloads need to be considered.

The uncertainty and varied characteristics of QoS values for continuously monitored data of different cloud models need a multi-collaborative QoS prediction to predict the missing QoS values through time series analysis for potential cloud users. A fuzzy analytical hierarchy process (FAHP) was proposed in [4] to determine the similarity of QoS evaluations for multiple periods by assigning an objective weight for every period and compare with neighboring potential users for analyzing and evaluating the missed QoS values. This FAHP mechanism needs to be analyzed for different user's application scenario and requirements for optimal period length.

The SLA aware autonomic management (STAR) was proposed in [5] to focus on automatic reduction SLA violation for effective delivery of cloud services. It is tested for heterogeneous workloads for different QoS parameters like execution time, cost, latency, reliability and availability to optimize them for meeting the user satisfaction of their application requirements by self-management. STAR was only tested for only single cloud but its application of self-management for federated cloud can be beneficial as automation of QoS parameters checking SLA violation rates would reduce some job of monitoring and work on other aspects of Self-healing and Self –configuration.

Cloud consumers are categorized into different classes depending on their application requirements with specified QoS parameter in SLA agreement with cloud providers. In [6] a framework was proposed for SLA management in cloud computing with a specified price control parameter which is used to meet the demands for all classes of consumers. A reinforcement learning (RL) is used to frame a new VM hiring policy in [6] which adapt to changes in the system to guarantee the QoS for all categories of clients.

The need of self adaptive prediction suite for automated resource provisioning to improve the accuracy in measurement of QoS parameters for avoiding less SLA violations by cloud providers is made in [7] and proposed a time series prediction algorithm based on incoming workload pattern by auto-scaling systems in cloud to analyze the risk minimization in terms of allocation of resources with specified QoS parameters and adapt to change of cloud environments automatically for incoming workload pattern. This self-adaptive prediction suite is not tested for federated cloud workload patterns as multiple cloud providers need to share their workload automatically for meeting the multiple user requirements simultaneously.

SLA management has become a prominent task in cloud computing where cloud provider is made responsible for providing functional and non-functional QoS parameters according to the specified SLA made between consumer and provider.SLA based decision strategies during VM scheduling was proposed in [21] for different constraints of resource usage and availability by VM migration among cloud providers. Different scheduling strategies in terms of SLA violations by minimizing the number of migrations was analyzed on a federated clouds framework with real-world measures of business-critical VM's.

Finally, Figure 1 gives the taxonomy of QoS aware Resource provisioning divided in to three categories like QoS aware resource selection, monitoring and allocation. Each category is listed with issues and solutions for their effective management without SLA violation which was briefly discussed and highlighted in this section.

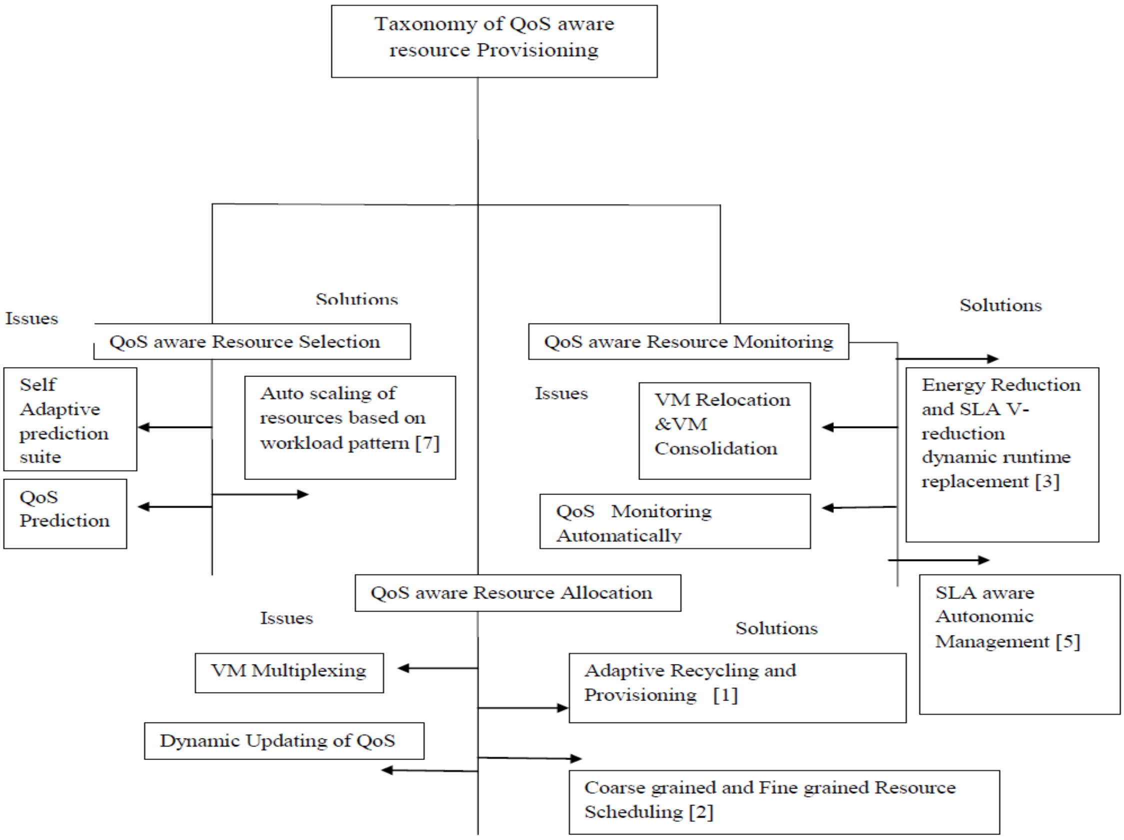


Fig. 1 Taxonomy of QoS Aware Resource Provisioning with issues and Solutions

III. MODELS OF RESOURCE PROVISIONING WITH QOS

A cooperative cloud market model in [8] was used for creating a federated cloud market where interoperability of clouds can be realized, novel two phases coordinated resource reservation and provisioning approach of [8] was used to allocate resources to cloud consumers to minimize user costs with less QoS demands rejection and provide resources without any SLA violations. It is a heuristic coordinated approach of forming a federation and analyzed mathematically. It needs to be tested on the real-time federated cloud test bed for getting the same performance results. Cloud broker plays a prominent role in QoS aware resource management in the federated cloud. In [9] a genetic algorithm called QBROKAGE was proposed for QoS aware resource management for cloud consumers. The cloud brokers deliver services like resource discovering, services comparison with different resource availabilities which are offered by different cloud providers. QBROKAGE of [9] aim is to find resources satisfying different QoS requirements for different application service request made by cloud users. It is also analyzed for a different set of cloud providers and optimizing deployment solution for data transferring cost across multiple clouds. QBROKAGE of [9] was tested for only static applications, it needs to be analyzed for runtime custom-sized applications and support for elasticity.

Failure rate minimization is major criteria for federated clouds wherein availability of resources at any instances is a big challenge and it is dealt in [10] which proposed a hybrid cloud architectural framework for efficient coupling of public and private clouds to create a failure –aware resource provisioning algorithm that is capable of monitoring cloud users QoS requirements and automatically assign to the cloud providers to handle the request. This algorithm need further investigated for a loosely-coupled many task computing applications.

Prediction for small scale and large scale applications resource requests traffic was analyzed in [12] for inter-cloud computing using reward based Markov chain analysis. Federated clouds need to support diverse services for users to access big data of real time applications and real-time working traffic applications. The wide range of resource requirements with different QoS values and SLA specification were managed by reward-based adaptive global cloud resource management by not only predictions of traffic but also maximizing the net profit of providers. This model was only analytical study but need to be implemented real-time federated cloud test bed.A conceptual model of federated cloud was proposed in, [13] which is organized into four layers: business, logical, repository and communication. Similar functionality components are grouped and distributed as segments among these four layers. The segments related to business aspect like namely contracts, business model are grouped to the business layer; segments related to technical aspects like security, orchestration and monitoring are grouped to logical layer and all physical infrastructure components into the repository.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In this conceptual model, there is no specification of what technologies to be used for implementation of layers to analyze the resource provisioning tasks.

A framework of SLA management for a federated cloud is proposed in [14] for which complete SLA life cycle was created to test the SLA violation, SLA Monitor throughout the resource allocation made by cloud broker according to user requirements. The cloud consumer satisfaction was not tested for the above better quality of SLA Management in a federated cloud. Semantic-based federated cloud framework architecture was proposed in [15] for using ontology framework for resource discovery and web semantic for managing interoperability issues for allocation of resources. This ontology framework was not tested for improving SLA quality and managing resource provisioning according to user satisfaction. In [16] monitoring system for federated clouds was proposed in which multiple domain cloud provider's resources are aggregated and provided both network and infrastructure level to cloud users. This monitoring system was validated at the end to end application level using BonFire which is an online source for creating federated cloud infrastructure and detail results have experimented.

Dynamic nature of cloud computing is effectively managed with federated clouds by creating an agent-based model which was proposed in [17] to manage resource provisioning dynamically at any instance. This model integrates between cloud broker and agent organization providers by permitting multi-criteria for migration of a submitted customer request. It is analyzed for different parameters dynamically like network delay; throughput and load balance and many factors are added to check for flexibility of adaptable cloud environment for satisfying the user request. This Model is only simulated with agent-based framework not tested for real-time federated cloud test bed environment. A genetic algorithm based heuristic approach federated cloud model was proposed in [18] for providing big data services by multiple cloud providers sharing the task as assigned by broker based on multi-objectives of tasks. In this model tasks management based on different QoS parameters was analyzed and the reliability of the system was tested for different system failure conditions. This model was also not tested for the real-time federated cloud test bed.

Fairness among cloud providers is very difficult to be managed while scheduling of resources in federated clouds. Fair non-monetary scheduling was proposed in [19] a CloudShare module which was implemented to manage the distributed way of load balancing in federated clouds. The cloud providers fairly schedule resources they also optimize the cost of it by using shapely value function and share profit among themselves. This model was simulated for comparing with fair share algorithms and found it is more efficient in scheduling resources but it was also not tested for real-time federated cloud test bed.

In [20] a web based service model built on SaaS federated cloud to provide services for HPC applications. This model was used to perform economic analysis from the perspective of end users, cloud providers and independent software vendors. It was not tested for different pricing strategies like reputation based cloud selection strategy for achieving more efficiency in provisioning resources for HPC applications.

Interconnected cloud environments of different cloud providers in cloud federation need to share resources to deliver improved service performance. In order to satisfy multiple security requirements of cloud users, a service assignment with optimized multi-objective security was proposed in [22] for which they considered different tradeoffs on three security factors like cost, performance and risk while assigning services by different cloud providers. A simulation model was used to measure the security and performance violation rates to get minimized during service assignment for cloud users. Automated monitoring of resources provisioning among multi-agent was analyzed in [23] using Matlab and show simulation results for multi-cloud environments for round-robin approach in resource provisioning for the better understanding of the federation model of cloud computing. SmartFed is a simulation tool proposed in [24] was built on top of cloudsim to simulate the federated cloud and test federation nature of cloud providers while provisioning resources without SLA violation. A model of SLA based VM scheduling in a federated cloud was analyzed in [25] for different VM Scheduling strategies in distributed cloud federations and a template of scenarios was designed to test complex cloud federation while provisioning resources. A novel workflow management framework called DoFCF(Deploy on Federated Cloud Framework) was proposed in [26] to manage the large scale computations on sensitive data like healthcare using a federated cloud. In these workflows are

Table I. Performance Metrics of QoS Aware Resource Provisioning in Federated Cloud

Author	**QoS Analyzed Parameters (Performance Metrics)**	**Methodology/ Implementation Models**
Hemanth Kumar Reddy et al	Cost, Load Balancing	Two-Phase Coordinated Resource Reservation and Provisioning Approach[9]
Gaeton et al	Cost, Scalability	Genetic algorithm (QBrokerage) [8]
Javadi et al	Availability, Fault Tolerance	Hybrid Cloud Architectural Framework[10]
Chang et al	Cost, Load Aware	Reward Based Markov Chain Analysis[12]
Marcico R et al	Cost	Conceptual Model[13]
Vipual et al	Feedback driven	Framework of SLA Management [14]
Manno et al	Interoperability	Semantic based federated cloud framework [15]
Yahya et al	Load Balancing	BonFire Federated Cloud model [16]
Sofiane et al	Load Balancing, Through Put,	Agent Based Model [17]
Jian Shu et al	Reliability	Genetic Algorithm based Heuristic approach[18]
Milosz et al	Load Balancing, Scalability ,cost	Fair Non monetary Scheduling [19]
Hou et al	Cost ,Elasticity	Web Service Model [20]
Halabi et al	Cost, Load aware	Multi-Objective Security [22]
Nazi et al	Scalability	Multi-Agent Model [23]
Anastasi et al	Cost, Load Balancing	SmartFed [24]
Kohne et al	Scalability, Elasticity	SLA based VM Scheduling [25]

Published By: Blue Eyes Intelligence Engineering & Sciences Publication

dynamically partitioned and distributed to both public/private data centers for minimizing cost, meeting security requirements and handling failure rates within an optimal time. The workflows were tested on a cloudsim framework and analyzed results for different dynamic real-time workflows.

The Federated Cloud Management Framework was developed to manage automated service provisioning for users in [31] was used to monitor services and selection of cloud provider in heterogeneous cloud environments. This framework does list all cloud providers information of federated cloud and provide these details to multiple providers through unified access to make them transparent for heterogeneous cloud environments. The cloud broker uses the repository to hold the details of virtual appliances with service operations and descriptions. Depending on the request made by users for this generic repository a meta-broker service will create queues dynamically and provide reliable services from multiple cloud providers by allocation of virtual appliances.

The QoS aware SLA management was proposed in [32] where a new mechanism was designed for measuring the QoS for preventing SLA violations in federated clouds. The SLA based provisioning algorithm in [32] creates a service tree holding the cloud providers with their QoS threshold values and for every application request the QoS upper and lower bound is specified to check at each node meeting this threshold value and take decision for identifying criteria for SLA violations like single-point failure, redundant and planned down time. It was tested on prototype for single data center environment and cloud federation for checking SLA violations but not implemented for QoS aware resource management.

A virtual network reservation framework was proposed in [33] for managing effective bandwidth provisioning among multiple cloud providers sharing similar SLA for finite time. In cloud computing application components will interact with different network devices grouped in to layers over reliable bandwidth links which affects it response times. This framework works as a forecasting engine which manages SLA by applying time series models and use bandwidth thresholds with varied time among multiple cloud providers' workloads. It was tested in Open Stack in perspective of checking bandwidth reliability for VM allocation and avoiding SLA violations for different customer requests.

The resource utilization in network virtualization for heterogeneous resources is effectively managed by hybrid virtualization provisioning in [34] by converting the complete mechanism to Integer linear programming for minimizing provisional cost. Two set of algorithms namely decomposition based and spectral clustering in [34] were used to enhance the effectiveness of provisioning for heterogeneous networks. The reliability of this provisioning was not tested for physical failures of systems in virtualized networks. In federated cloud data management in underlying operational environment leads to frequent SLA violations due to lack of QoS maintenance. This problem was dealt in [35] by developing policy based autonomic middleware by self adaptive nature in federated cloud. This adaptive middleware was tested in SaaS application environment and validated good results in terms of dynamic data management without any intervention of operators leading to less SLA violations.

The resource allocation model for federated geo-distributed clouds was proposed in [36] to manage globally shared resources among multiple cloud providers by establishing a resource contract apriori time interval for a 24 hrs time period. Each cloud service provider with pre-contracts of resources uses job scheduling and provisioning algorithms to achieve job completion within response times and effectively manage job allocations globally and locally having fair shared profit among multiple cloud providers. The service selection of resources among multiple cloud providers in federated clouds does not follow fairness while serving the resource request from cloud consumers. To deal this issue a fair service matching agent was proposed in [37] to distribute resource request fairly among cloud providers. This mechanism tries to address the availability problem of resources for different cloud consumers for satisfying their dynamic service requests instantly. The sharing of profit based on the dynamic allocation of resources request made by consumers among cloud providers have been not implemented in this mechanism.

The cloud elasticity still remains a challenging task with single cloud provider which can efficiently dealt in federated clouds by performing inter operations among other cloud providers. A residue based approach for resource provisioning was proposed in [38] by providing horizontal scaling of resources instantly for satisfying single request of resources. The greedy technique is also used to rank the clouds while monitoring the availability of resources in this mechanism. Managing QoS violations while providing computing and networking resources simultaneously in federated cloud, this problems is called as virtual network embedding problem in [39] is dealt by mixed integer linear programming model in two phase approach first splitting the virtual network request among multiple clouds and then perform intra cloud segment mapping for allocating virtual network resources for having less SLA violations. Autonomic resource management is a critical task in federated cloud which was managed in [40] by Connected Dominating sets by hierarchical approach fixing set of system components to work as Autonomic managers at data centers and share resources by collaboration with Global Managers for effective resource management among multiple cloud providers by generating dominating sets among the cloud providers to handle resource provisioning issue without QoS violations.

IV. RESEARCH ISSUES AND CHALLENGES OF QOS IN FEDERATED CLOUDS

Cloud service brokerage plays a prominent role in federated clouds during resource provisioning. Most of the research challenges involved with cloud broker were highlighted in [27] in which four major challenges were given more importance like legal rule compliance checking, legislation dynamic management, seamless service migration and QoS monitoring.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Much of the challenges have been resolved they are dealt with by isolated solutions the integration of all these solutions will be an added advantage for federated clouds in performing cloud brokerage job effectively. QoS is realized by QoS monitoring and analysis. QoS monitoring techniques are at its infancy stage and the job of monitoring has become complex because the trend is shifted to developing applications for micro services running as containerized processes while execution. QoS assessment is still an open issue for analyzing key cloud properties like elasticity and consistency.

As per [28] QoS Management for different cloud applications is still a research issue in terms of providing the guarantee of a service level for dimensions such as performance, availability and reliability. Much of the work towards traffic forecasting, resource consumption estimation and anomaly detection has been not provided validation in a cloud environment and became a robust challenge for the multi-tenant or federated cloud environment. Integrating workload characterization, system models and resource management solutions without QoS degradation is a big challenge and while building research prototype into a commercial solution is still an open point.

Interconnected clouds are really facing a big challenge as given in [29] for avoiding vendor lock-in problems in federated cloud environments. Developing standard interfaces for integrating different cloud provider's services to avoid cloud interoperability issues is still an unsolvable research issue. Meeting legal regulations in interconnected cloud environments to support legal agreements made for validating SLA is necessary. Implementing interconnected market places for taking care of economical aspects of different cloud providers while collaboration for pricing and billing is needed to study for federated clouds.

In [30] the importance of QoS differentiated resource pricing was highlighted for federated clouds. The semantic description of resources should be standardized for ensuring interoperability. An autonomous resource monitoring tool needs to be used for validation and performance measuring for a heterogeneous application deployed in the federated cloud environment.VM behavior modeling and workload of a federated cloud environment should be evaluated to realize the workloads of VM at different instances.

The Table1 list the performance metrics of QoS ware resource provisioning with specified implementation models, analyzing QoS parameters. These all implementation models along with their specific importance were highlighted in this section.

V.PROPOSED LAYERED AGENT BASED MODEL FOR ANALYZING FEDERATED CLOUD

Layered Agent Based Model consist of layering the agents to simulate federated cloud nature for analyzing different aspects like maximizing resource usability and providing optimal price option for consumers agents by considering different QoS factors. The Layered agent based model considers four layers for analyzing communication to satisfy consumers getting serviced for optimal price. The four layers in Figure 2 are Provider agent layer, Collated provider agent layer, Broker agent Layer and Cloud consumer agent layer. Simulating federated nature of clouds is possible by using this layered approach. Each layer has its own significance task to handle the functionality of federated nature to make it possible for analyzing resource utilization of cloud consumers for optimal service time.

Provider agent layer will form collation with other provider agents to make heterogeneous resources available at any instant. Collation formation is a tedious task which involves FLA (Federation Level Agreement) to share resources for certain QoS constraints and price. Each cloud providers will initiate the collation formation by communicating this FLA in order to generate maximum profit by collaboration. The next layer Collated provider agents will provide all set of collaborated resources for brokers to allocate for cloud consumers. The Broker agent layer performs major task in generation of service classes for different set of collated provider agent resources. This service classes generation can be used for QoS aware differential pricing for providing different prices for cloud consumers to meet their QoS demands along with resource requests. The cloud consumer agents will have a choice of selection of these service classes to get serviced his resource requests for an optimal price.

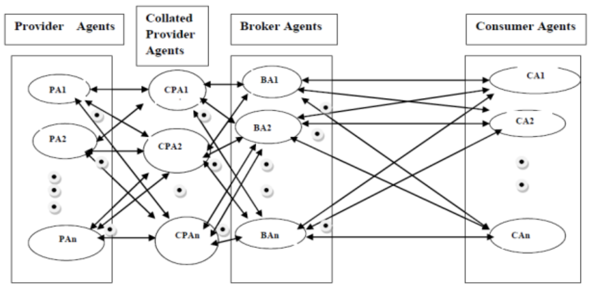


Fig 2 Layered Agent Based Model

.System Model

The layered agent based model provides scope for analyzing resources utilization and pricing options for federated clouds consumers using mathematical approach. Let us consider the provider agents $(PA_1, PA_2, \ldots, PA_n)$, collated providers agents $(CPA_1, CPA_2, \ldots, CPA_n)$, Broker agents as $(BA_1, BA_2, \ldots, BA_n)$ and Consumer agents as $(CA_1, CA_2, \ldots, CA_n)$ .The resource request made by consumer agent is treated as vector of VM instances through which they specify the resources required REQV$(VM_1, VM_2, \ldots, VM_n)$ Each VM instance will have its specified resources to get serviced by broker agent. Along with resource request vector of each VM instance will have equivalent QoS factors to meet while getting serviced .The QoS factor response time is considered for each VM instance is QOSV$(RT_1, RT_2, \ldots, RT_n)$ .The resource utilization is measured using $RUtil_\theta$(i) where θ is the user instance for particular time period T and 'i' is the service class selection to utilize maximum resources of that VM instance by user θ.

$$Tpr_j = Acpr_i + Srt_i * CVM_i \quad (1)$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Where Tpr_j is the total pricing function of 'j' cloud consumer, $Acpr_i$ is the acceptable price by cloud consumer from broker service class–i for utilizing those VM instance resources, Srt_i is the service time of service class-i and CVM_i is the capacity of resources of service class-i VM instances for utilization.

$CVM_i = \sum_{i=1}^{N} Nr_i * Pr_i$ (2)

Where Nr_i no of resources of that VM instance of service class-i Pr_i is the price associated with that resource of service class-I for that particular instance.

The resource utilization of 'j' cloud consumer is given by

$RUtil_\theta(j) = Ut_{Max} - Tpr_j - \theta\ QFunc(N_i, CVM_i)$ (3)

U_{Max} – Maximum utilized resources made by θ user's from service class-i

Tpr_j – Total pricing function of the j[th] user

$QFunc(N_i, CVM_i)$ - congestion function which measures the influence of violation of not satisfying that QOS factors while utilizing those resources for time period t.

Choosing a service class i is possible maximizing this $RUtil_\theta(j)$ for particular time period t for j[th] user

$i = \arg Max_{j \in \{1,2 \ldots m\}} RUtil_\theta(j)$

or else not to select the service class- i

if $RUtil_\theta(j) < 0$. (4)

$QFunc(N_i, CVM_i)$ is a critical function of considering QoS factors affecting the resource utilization of that service class-i for a no of users N_i while allocating its capacity of resources of VM instances CVM_i.Let us consider the QoS evaluations made for a particular service class by N users in a time slot T is E_i.

$E_i = \{e_{i,1}, e_{i,2}, \ldots\ldots, e_{i,N}\}$ where $e_{i,1}$ is the QoS evaluation made by first user of that service class- i,

Consider the mean value for QoS evaluations of VM instances.

$$Em_i = \frac{1}{N}\sum_{k=1}^{N} e_{i,k} \quad (5)$$

$$QFunc(\ ,\ CV) = \frac{1}{N_i} \sum_{1=1}^{N} e_{i,k} * CV \quad (6)$$

VI. RESULTS AND ANALYSIS

Let us consider the four service classes with set of resources as in table 3 and table 4 gives prices of resources and service time .Following data is taken with reference to AmazonEC2 prices to analyze mathematically for above algorithm. The figure 3a gives the comparison of maximum resource utility made by four consumer agents without QoS and with QoS and figure 3b gives the comparison of pricing payed by four consumer agents without QoS and with QoS parameters in to consideration.

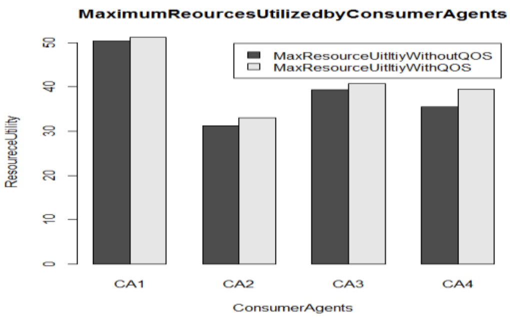
The results in figure 3a shows that 5-10% increase in resource utilization because of considering the QoS factor response time of service classes as evaluated by consumer agent. The results in figure 3b shows that 10-20% decrease in pricing of resources because of considering the QoS factor response time of service classes and QoS values effect is measured by QoS evaluation function at cloud broker

Table 2: Service Classes with Resources

Service classes	Small			Medium			Large		
	CPU	RAM	HD	CPU	RAM	HD	CPU	RAM	HD
Class 1	4	4	20	10	8	120	20	16	200
Class 2	5	4	30	12	8	150	17	16	250
Class 3	2	4	40	15	8	110	25	16	240
Class 4	4	4	20	11	8	150	22	16	270

Table 3: Service Classes with Price and Service time

Service classes	Prices in $			Service time (msec)
	Per CPU	Per GB	Per TB	
Class1	0.12	0.10	0.05	0.23
Class2	0.15	0.20	0.08	0.5
Class3	0.17	0.08	0.02	0.25
Class4	0.20	0.15	0.09	0.13



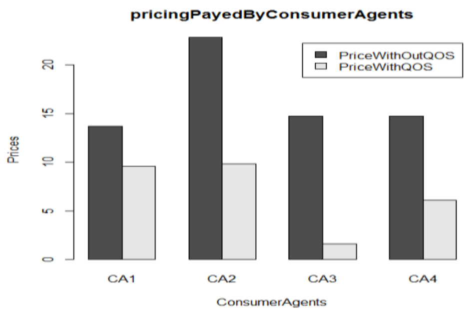


Fig 3 a: Comparison Of Maximum Resource Utility With QoS And Without QoS

Fig 3 b: Comparison Of Price Without QoS And With QoS

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VII. CONCLUSION

This study reviews the resource provisioning perspective through QoS effects in federated cloud environments. It highlights the facts of models which are implemented for creating collaboration among cloud providers to provide resources without violation of SLA. The responsibilities of cloud broker were made clear in managing the communication between collated cloud providers and cloud consumers. Many research issues are available federated cloud in terms of creating common middleware API's to manage the communication between broker and providers. The importance of QoS management for resource provisioning in terms of legal compliance rules and regulations among federated cloud providers are still open research issue for resolving. Better to build intelligent systems into a federated cloud environment to asses most of the QoS factors aspects and predict different functionalities of resource management.

The future work is to propose a layered agent-based model to build the clear communications API's among cloud providers in order to form collation and manage fault tolerance whenever there is breakage of collation and formation of new collation. Implemented a prediction model to analyze the resource utilization made by cloud users at different instances and builds a proactive way of avoiding resource contention of requests at cloud broker. Enabling QoS differentiated services at cloud broker to provision resources with effective QoS feedback from cloud users while utilizing their resources from particular collated cloud providers.

The result shows that the comparison of resource utilization and pricing with QoS and non QoS and it clearly states that resource utilization increases and pricing of resources will decrease because of QoS evaluation function at cloud broker. More fine results need to be tracked by increasing no of consumer agents and taking more QoS factors for QoS evaluation function.

REFERENCES

1. Feng, Ye, Wang Zhijian, and Huang Qian. "A novel QoS-aware mechanism for provisioning of virtual machine resource in the cloud ". Journal of Algorithms & Computational Technology. vol.10, pp. 169-175, 2016.
2. Li, Zengxiang, Long Wang, Yu Zhang, Tram Truong-Huu, En Sheng Lim, Purnima Murali Mohan, Shibin Chen et al. "Integrated QoS-aware Resource Provisioning for Parallel and Distributed Applications ".IEEE. In Proceedings of the 19th International Symposium on Distributed Simulation and Real-Time Applications, Chengdu, Sichuan, China, pp. 171-178 October 14-16, 2015.
3. Mollamotalebi, Mahdi, and Shahnaz Hajireza. "Multi-objective dynamic management of virtual machines in cloud environments." Journal of Cloud Computing, vol. 6, pp. 16, 2017.
4. Ma, Hua, Haibin Zhu, Zhigang Hu, Wensheng Tang, and Pingping Dong. "Multi-valued collaborative QoS prediction for cloud service via time series analysis ". Future Generation Computer Systems , vol.68 ,pp. 275-288, 2017.
5. Singh, Sukhpal, Inderveer Chana, and Rajkumar Buyya. "STAR: SLA-aware autonomic management of cloud resources." IEEE Transactions on Cloud Computing, 2017.
6. Alsarhan, Ayoub, Awni Itradat, Ahmed Y. Al-Dubai, Albert Y. Zomaya, and Geyong Min. "Adaptive resource allocation and provisioning in multi-service cloud environments." IEEE Transactions on Parallel and Distributed Systems, vol.29, pp.31-42, 2018.
7. Nikravesh, Ali Yadavar, Samuel A. Ajila, and Chung-Horng Lung. "An autonomic prediction suite for cloud resource provisioning." Journal of Cloud Computing , vol.6, pp.3,2017.
8. Anastasi, Gaetano F., Emanuele Carlini, Massimo Coppola, and Patrizio Dazzi. "QoS-aware genetic cloud brokering". Future Generation Computer Systems, vol.75, pp.1-13, 2017.
9. Reddy, K. Hemant Kumar, Geetika Mudali, and Diptendu Sinha Roy. "A novel coordinated resource provisioning approach for cooperative cloud market" . Journal of Cloud Computing vol. 6, pp.8,2017.
10. Javadi, Bahman, Jemal Abawajy, and Rajkumar Buyya. "Failure-aware resource provisioning for hybrid Cloud infrastructure".Journal of parallel and distributed computing, vol.72, pp. 1318-1331, 2012.
11. He, Hong. "A Resource Reservation based Framework for QoS-aware Resource Provision in Cloud Computing". International Journal of Grid and Distributed Computing, vol. 9, pp.193-2014, 2016.
12. Chang, Ben-Jye, Yu-Wei Lee, and Ying-Hsin Liang. "Reward-based Markov chain analysis adaptive global resource management for inter-cloud computing". Future Generation Computer Systems, vol. 79, pp.588-603, 2018.
13. Assis, Marcio RM, Luiz F. Bittencourt, and Rafael Tolosana-Calasanz. "Cloud federation: Characterisation and conceptual model ". IEEE.2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, IEEE, London, UK, pp. 585-590. December 8-11,2014.
14. Chudasama, Vipul, Dhaval Tilala, and Madhuri Bhavsar. "SLA Management in Cloud Federation". Springer, Cham, In International Conference on Information and Communication Technology for Intelligent Systems, pp. 397-404, 2017.
15. Manno, Giuliano, Waleed W. Smari, and Luca Spalazzi. "Fcfa: A semantic-based federated cloud framework architecture".IEEE, In High-Performance Computing and Simulation (HPCS), 2012 International Conference on, Madrid, Spain, pp. 42-52,July 2-6 2012
16. Al-Hazmi, Yahya, Konrad Campowsky, and Thomas Magedanz. "A monitoring system for federated clouds". IEEE, In Cloud Networking (CLOUDNET), 2012 IEEE 1st International Conference on, Paris, France, pp. 68-74.November 28-30 2012.
17. Kemchi, Sofiane, Abdelhafid Zitouni, and Mahieddine Djoudi. "AMACE: agent-based multi-criterions adaptation in a cloud environment." Human-centric Computing and Information Sciences, vol.8, pp.26, 2018.
18. Shu, Jian, Changyong Liang, Binyou Wang, and Jian Xu. "Building the federation of cloud service for big data.”. IEEE, In Big Data Analysis (ICBDA), 2018 IEEE 3rd International Conference on, Shanghai, China, pp. 166-169.March 9-12, 2018.
19. Pacholczyk, Milosz, and Krzysztof Rzadca. "Fair non-monetary scheduling in federated clouds." *arXiv preprint arXiv:1803.06178* ,2018.
20. Hou, Zhengxiong, Yunlan Wang, Yulei Sui, Jianhua Gu, Tianhai Zhao, and Xingshe Zhou. "Managing high-performance computing applications as an on-demand service on federated clouds." Computers & Electrical Engineering, vol. 67, pp. 579-595,2018.
21. Kohne, Andreas, Damian Pasternak, Lars Nagel, and Olaf Spinczyk. "Evaluation of SLA-based decision strategies for VM scheduling in cloud data centres." .ACM,In Proceedings of the 3rd Workshop on CrossCloud Infrastructures & Platforms, p. 6. April 18-21, 2016.
22. Halabi, Talal, and Martine Bellaiche. "Service Assignment in Federated Cloud Environments Based on Multi-objective Optimization of Security". IEEE, In Future Internet of Things and Cloud (FiCloud), 2017 IEEE 5th International Conference on, Prague, Czech Republic, pp. 39-46.August 21-23, 2017.
23. Yazdi, Nazi Tabatabaei, and Chan Huah Yong. "Simulation of Multi-Agent Approach in Multi-Cloud Environment Using Matlab ". IEEE. In Computational Intelligence, Modeling and Simulation (CIMSim), 2015 Seventh International Conference on, Kuantan, Malaysia ,pp. 77-79. July 27-29, 2015.
24. Anastasi, Gaetano F., Emanuele Carlini, and Patrizio Dazzi. "Smart cloud federation simulations with cloudsim.".ACM, In Proceedings of the first ACM workshop on Optimization techniques for resources management in clouds, pp. 9-16. 2013.
25. Kohne, Andreas, and Olaf Spinczyk. "Model for SLA-Based VM Scheduling in Federated Cloud Environments". Journal of Integrated Design and Process Science, vol. 18, pp.39-42, 2014.
26. Wen, Zhenyu, Rawaa Qasha, Zequn Li, Rajiv Ranjan, Paul Watson, and Alexander Romanovsky. "Dynamically partitioning workflow over federated clouds for optimizing the monetary cost and handling run-time failures.", IEEE Transactions on Cloud Computing, vol.1,pp.1-1,2016.
27. Casalicchio, Emiliano, Valeria Cardellini, Gianluca Interino, and Monica Palmirani. "Research challenges in legal-rule and QoS-aware cloud service brokerage." Future Generation Computer Systems, vol.78, pp.211-223, 2018.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

28. Ardagna, Danilo, Giuliano Casale, Michele Ciavotta, Juan F. Pérez, and Weikun Wang. "Quality-of-service in cloud computing: modelling techniques and their applications." Journal of Internet Services and Applications 5, vol.1, p.11, 2014.
29. Toosi, Adel Nadjaran, Rodrigo N. Calheiros, and Rajkumar Buyya. "Interconnected cloud computing environments: Challenges, taxonomy, and survey." ACM Computing Surveys (CSUR), vol.47, p.7, 2014.
30. Liaqat, Misbah, Victor Chang, Abdullah Gani, Siti Hafizah Ab Hamid, Muhammad Toseef,Umar Shoaib, and Rana Liaqat Ali. "Federated cloud resource management: Review and discussion.". Journal of Network and Computer Applications, vol. 77,pp.87-105,2017.
31. Kertesz, Attila, Gabor Kecskemeti, Zsolt Nemeth, Marc Oriol, and Xavier Franch. "A holistic service provisioning solution for federated cloud infrastructures.".ACM, In Proceedings of the First International Workshop on European Software Services and Systems Research: Results and Challenges, Zurich, Switzerland, pp. 25-26., June 05 2012.
32. Bardhan, Shoumen, and Dejan Milojicic. "A mechanism to measure quality-of-service in a federated cloud environment.”. ACM, In Proceedings of the 2012 workshop on Cloud services, federation, and the 8th open cirrus summit, San Jose, California, USA, pp. 19-24, September 21 2012.
33. Adegboyega, Abiola. "An adaptive resource provisioning scheme for effective QoS maintenance in the IaaS cloud.", ACM, In Proceedings of the International workshop on Virtualization Technologies, Vancouver, BC, Canada, p. 2. December 07 - 11, 2015.
34. Sun, Gang, Dan Liao, Dongcheng Zhao, Zhili Sun, and Victor Chang. "Towards provisioning hybrid virtual networks in federated cloud data centers." Future Generation Computer Systems, vol. 87,pp.457-469.2018.
35. Rafique, Ansar, Dimitri Van Landuyt, Eddy Truyen, Vincent Reniers, and Wouter Joosen. "SCOPE: self-adaptive and policy-based data management middleware for federated clouds”. Journal of Internet Services and Applications , vol.10,p.2,2019.
36. Xu, Jinlai, and Balaji Palanisamy. "Optimized Contract-based Model for Resource Allocation in Federated Geo-distributed Clouds." IEEE Transactions on Services Computing, 2018.
37. Sudhakar, S., N. S. Nithya, and B. L. Radhakrishnan. "Fair service matching agent for federated cloud." Computers & Electrical Engineering, vol. 76, pp. 13-23, 2019.
38. Kirthica, S., and Rajeswari Sridhar. "A residue-based approach for resource provisioning by horizontal scaling across heterogeneous clouds." International Journal of Approximate Reasoning, vol.101, pp.88-106, 2018.
39. Diallo, Marieme, Alejandro Quintero, and Samuel Pierre. "A QoS-based splitting strategy for a resource embedding across multiple cloud providers." IEEE Transactions on Services Computing, 2018.
40. Martin, John Paul, A. Kandasamy, and K. Chandrasekaran. "Toward Efficient Autonomic Management of Clouds: A CDS-Based Hierarchical Approach.".Springer, In Advanced Computing and Systems for Security, Springer, Singapore, pp. 61-73, 2018.

AUTHORS PROFILE

Dr. Kolla Bhanu Prakash, working as Professor in the Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India. His research interests include Wireless Network Communication, IoT.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*