ISSN 2319–7757
EISSN 2319–7765

Indian Journal of Engineering

A Survey on Task scheduling Algorithm and Quality of Service for Resource Allocation in cloud environment

Nihar Ranjan Nayak[1], Bhuvaneshwari S[2]

1. Doctoral Research Scholar, Department of Computer Science and Engineering, Pondicherry University, Karaikal campus, Email: nayak.niharranjan0@gmail.com.
2. Associate Professor, Registrar, Department of Computer Science and Engineering, Central University Tamil Nadu, Thiruvarur, Email: booni_67@yahoo.co.in

Publication History
Received: 09 November 2016
Accepted: 03 December 2016
Published: January-March 2017

Citation
Nihar Ranjan Nayak, Bhuvaneshwari S. A Survey on Task scheduling Algorithm and Quality of Service for Resource Allocation in cloud environment. *Indian Journal of Engineering*, 2017, 14(35), 63-70

ABSTRACT

Cloud computing is a network of server technology, use of computing resources (hardware and software).That are delivered as a service over the internet as a pay for use go model. A significant issue faced by Infrastructure as a service is task scheduling. Based on the user requirement quality of service resource allocation is effective. Although, there are different algorithms and quality of service based on the Existing system problem is solved, none of these can be prolonged. So this model has been verified through the UML diagram and determines that effective user interaction and customer satisfaction.

Key Words: Cloud Computing, Resource Allocation, Task Scheduling, QoS (Quality of Service).

1. INTRODUCTION

Cloud computing is virtualize and a network of server's technology that relies on-demand sharing computing resources rather than the personal device to handle application of the internet on pay-per-use basic system and provide reliable, customized and QoS (Quality of Service).

Cloud computing can be divided into the following types:

- Public clouds which can be register by anyone and the services they may use.
- Private clouds whose data can be managed within the enterprise and access without the restrictions of the network bandwidth, security etc.
- Hybrid clouds are a combination of the private/internal clouds and the external cloud resources.
- Community cloud the cloud infrastructure is shared between the organizations with similar interests and requirements whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud).
- Cloud service providers offer services that can be classified into the following three categories:
- Infrastructure as a Service (IaaS): Which allows customers to use hardware computing resources such as CPU, memory and processing power.
- Platform as a Service (PaaS): It is a development platform that support full "Software Lifecycle Process" that allows customer to develop cloud services.
- Software as a Service (SaaS): it providing software and application that is remotely available by consumers.

There are different algorithms for resource allocation and task scheduling in cloud differs according to the task accessed in environment. Various algorithms are survey in following section.

Scheduling is very important problem in cloud computing its aim is mapping between appropriate task or job to the appropriate available resources, by considering checks such as cost, deadline, Quality of service (QoS), high throughput, etc. Without effective scheduling resources are not utilize properly and decrease the availability, scalability etc. There are difference between Task and workflow scheduling but both are NP-hard problems, Task scheduling is much easier than workflow scheduling because scheduler only contains pool of tasks which do not have any interdependency and execute them in an arbitrary order. In other workflow scheduling is much more complicated because a workflow generally consists of a set of dependent tasks communicating with each other and scheduler should map the workflow tasks to the VMs by consider in their dependencies. The algorithms for task scheduling can be classified a static scheduling and dynamic scheduling algorithms where in static scheduling, the number of tasks or machine sets for scheduling and in dynamic scheduling, they are not fixed.

Resource allocation algorithm focuses the hardware capability and functionality of cloud computation. To increase flexibility, cloud allocates the resources according to their demand. And the major problem in task scheduling is load balancing, reliability, performance, scalability and reallocation of resources in dynamically. Actually task scheduling algorithm aim is to minimizing the execution task and maximize resource usage efficiently. The resource allocation strategy avoids Resource contention, Scarcity of resources, Resource fragmentation, Over-provisioning, Under-provisioning. And the quality of service based on resource allocation technique is efficient in reducing time and reducing cost. Basically quality of service parameters are network bandwidth, customer confident level, availability etc.

2. LITERATURE REVIEW

2.1 Scheduling and Resource allocation algorithm

2.1.1 Modified Round Robbin algorithm

It is one of the easiest scheduling techniques that utilize the principle of time slices. Here the time is divided into multiple slices and each node is given a particular time slice or time interval i.e. it utilizes the principle of time scheduling. Each node is given a quantum and its operation. The service provider provides the resources to the client on the basis to the time slice.

This Modified Round Robbin algorithm starts with the time of first request, which changes after the end of first request. If any new request is ready in the queue, the algorithm calculates the average of sum of the times of requests found in the ready queue including the new arrival request.

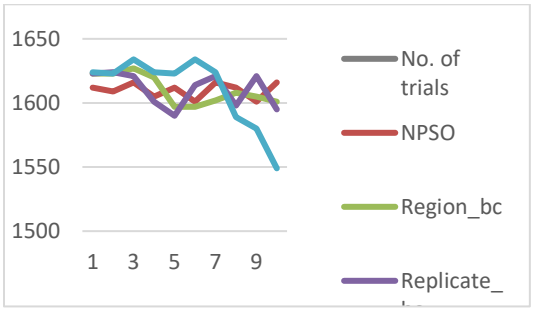
Nihar Ranjan Nayak and Bhuvaneshwari S,
A Survey on Task scheduling Algorithm and Quality of Service for Resource Allocation in cloud environment,
Indian Journal of Engineering, 2017, 14(35), 63-70,

This needs two registers -SR: To store the sum of the remaining burst time in the ready queue-AR: To store avg. of the burst times by dividing the value found in the SR by the count of requests found in the ready queue. After execution, if request finishes its burst time, then it will be removed from ready queue or else it will move to the end of the ready queue. SR will be updated by subtracting the time consumed by this request. AR will be updated according to the new data.

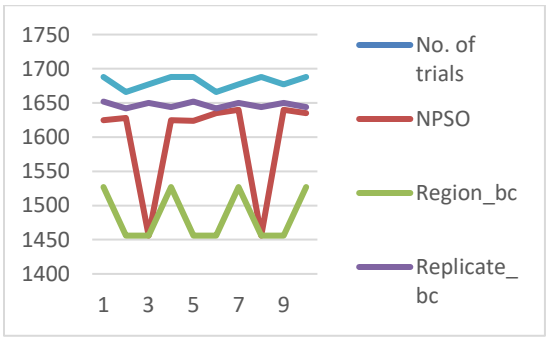
2.1.2 PSO (Particle swarm optimization)

The advantages of PSO such as fast search speed, it suffers from problems such as local optima, low convergence rate, etc. In the PSO-based workflow scheduling, the dimension of the particles is the number of tasks and each position of a particle indicates a mapping between the virtual machines and tasks.

To solve task scheduling and resource allocation in cloud computing, a PSO based algorithm is to assign each subtask to an appropriate resource (routing problem) and to sequence the subtasks on the resources (sequencing problem) in order to achieve the objectives of this scheme. To formulate the problem, cloud user tasks can denote the set of independent jobs, and each subtask is allowed to be processed on any given available resources. A subtask is processed on one resource at a time and the given resources are available continuously. This scheme shows that the PSO based fitness function is more effective and efficient with shorter completion time and lower cost. The below diagram I have taken 40 particle and 1000 iteration and calculated the graph of best and worst case.



(Best fitness)



(Worst fitness)

2.1.3 Bee Algorithm

Bee's algorithm in nature tracks the actions of bee to get their foodstuff. Initially they pick scout bee to a search a food areas, if that bee find the area with large food stuffs informs the place and direction to the other bees to find the area .Some other selected bee's and scout bee's collected honey as a foodstuff from different places. Identically some other set of scout bees inform the location of foodstuffs from different direction.

Select: Here scheduler using for find a task which has I/O, storage, memory is required to complete their task which will act as scout bees to find the area. ***Fitness:*** The main thing is a scout bee is identifies the location by using a fitness function which run that

task in particular resources. Actually, fitness refers to how much progress each task is making with their allocated resources compared to the same task executing on the entire group. ***Waggle:*** By identifying the location resources the scout's returns to scheduler and does the waggle function. Waggle function segregates the task present in scheduler based upon scout's information such as cost, memory and processor requirements. The combination takes place in such a way that the memory oriented information is passed to the memory oriented task for execution with fixed capacity. If a task exceeds capacity then the task has to wait until the scout task finds another resource available adjacent location.

2.1.4 Ant colony optimization algorithm

Ants are temporary form a group to perform a task of collecting food with reliability and consistency. Like ants cloud computing performs a complicated task providing resource optimally to customers to solve the problem of in an efficient a manner. Basic principles of ACO algorithm depending on the species ant behaviour when ant moving from food to nest or from nest to food or in both direction. Cloud Resources Scheduling Based Ant Colony Optimization algorithm is a random search algorithm, in *Travelling Sales Man* (TSP) problem given in n Place and the person starts from the point again reach the same point in shortest path. The main objective of this paper is to develop an effective load balancing algorithm using Ant colony optimization technique to maximize or minimize different performance parameters like CPU load, Memory bandwidth, and Delay or network load including various features such as scalability, reliability and autonomy for the different cloud environment. Similarly, a heuristic algorithm based on ant colony optimization has been proposed to initiate the service load distribution under cloud computing architecture and also update mechanism has been proved as an efficient and effective tool to balance the load.

2.2. Quality of Service

2.2.1 Workload identification

The amount of work to be done, especially by a particular person or machine in a period of time is called as workload. The cloud workload contains web server, application server, file server, transactional server etc. and adding the quality attribute for the workload. These are the example of workload identification.

- Websites: Website is a social networking site and freely available all the information oriented website for number of cloud consumer. The quality attribute contains for this workload is storage, Network Bandwidth, Availability etc.
- Online transaction: It includes online internet banking for transaction purpose only and contains online insurance policies. The quality attribute contains for this workload is Security, Internet Availability, accessibility etc.
- E-commerce: It includes all the malls and supermarket. The quality attribute contains for this workload is computing load and customizability etc.
- Financial Services: It includes banking and insurance system. The quality attribute contains for this workload is Security, availability and Integrity etc.
- Software Testing: It includes simulation based testing and software development application. The quality attribute contains for this workload is flexibility, testing time, computing capacity and self-service rate etc.

2.2.2 Workload Analysis

Cloud based workload patterns are used specify the type of application that user wants to execute. With the QoS and other requirement of workload it reduce the complexity of cloud. Above we seen how to identified the workload and it analyse through the process of workload pattern. Some process following the workload analysis are given below.

- Web service interface and API: It contains uncovering abilities of workload through user interface and web services. For example: Corporation constructing digital strength administration resolution showing Application Program Interfaces to other web services.
- Cloud deployment: Applications are deploying with QoS requirements like high availability and dynamic scalability. For example: Wholesaler store using web portal to scale down automatically when usage go beyond threshold and scale down as required.
- Storage base system: It deals with storage of unstructured data of large quantity. For example: Corporation keeping reports of authorized obedience in backup store.
- Instant service management: It includes the functions to start, suspend and stop cloud based applications and management of configuration of service. For example: An administrator of web application managing state of a Cloud based application through service portal.

Nihar Ranjan Nayak and Bhuvaneshwari S,
A Survey on Task scheduling Algorithm and Quality of Service for Resource Allocation in cloud environment,
Indian Journal of Engineering, 2017, 14(35), 63-70,

- Design for operation: How to develop a cloud based application which provides the function of logging and health status. For example: Design cloud based application which is user friendly through efficient Graphical User Interface (GUI).

2.2.3 Pattern Identification

- Websites –web service interface and API
- Online transaction- cloud deployment
- Ecommerce-Storage base system
- Financial Services-Instant service management
- Software Testing-design

2.2.4 Requirement of QoS

- Reliability: It can be formulated as follows:
- Reliability =MTBF = MTTF + MTTR
- Testing time: It can be formulated as follows: Testing Time = Time to prepare test environment +Time to execute Test Suite for a Cloud workload .(Test Suite is collection of test cases)
- Availability: It can be formulated as follows: Availability = MTTF/MTBF
- Network bandwidth: The network bandwidth can be calculated as number of bits transferred/received in a particular workload in one second. It can be formulated as follows: Network Bandwidth = Bits/second(b/s)
- Computing capacity: It can be formulated as follows: Computing Capacity = Actual Usage time of the Resource/Expected Usage time of the Resource.

2.2.5 Calculate metrics, weight and q-value

Conversion metrics

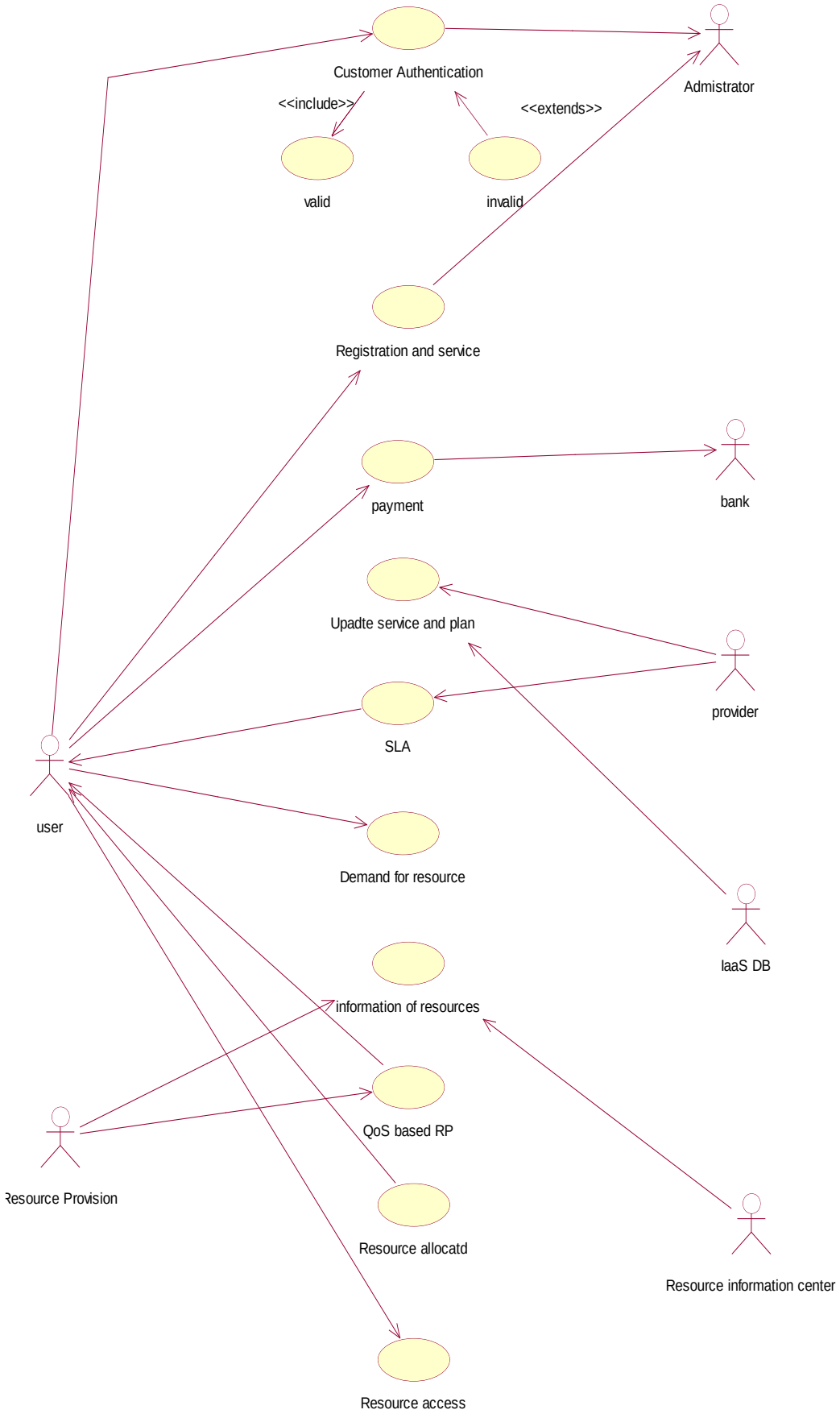
Weight (%)	weight
0-20	**1**
20-40	**2**
40-60	**3**
60-80	**4**
50-100	**5**

2.2.6 Measurement of quality attribute

Measurement of Q attributes	q-Value
Low	**1**
Medium	**3**
High	**5**

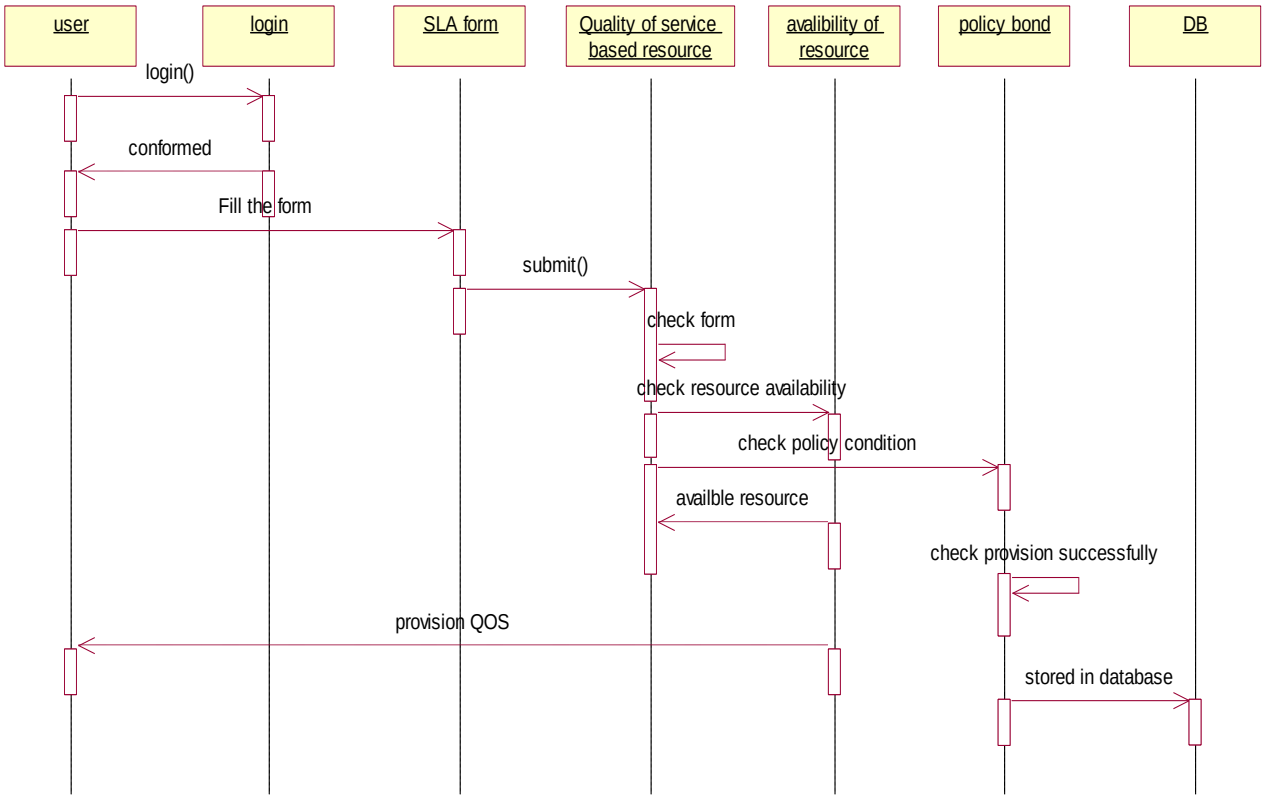
3. ARCHITECTURE DIAGRAM

3.1 Use case Diagram



(Use case diagram for Quality of Service based on Resource provision and allocate)

3.2 Sequence diagram



(Sequence diagram for Quality of service based policy)

4. CONCLUSION

Cloud computing is a prominent trend as an important platform for business, hosting the large computing system and service. It allots on demand dynamic resource allocation for providing quality of service to the consumer based on pay-as- you-use model to public. In this paper discuss about various task scheduling algorithm and the step of quality of service based resources allocation. Although, there are different algorithms and quality of service based on the Existing system problem is solved, none of these can be prolonged. All above discussed algorithm used for resources allocation depend on the task to be scheduled and quality of service based resources allocation is solved the problem of starvation. Depending on surveying the various algorithm and quality of Service it conclude that make span can reduce and because of Quality of service resource can be utilize more. In future cloud size increases, there is a need for better task scheduling algorithm and Quality Assurance develops.

REFERENCES

1. W. Zhao, K.Ramamritham, and J.A.Stankovic, "Pre-emptive scheduling under time and resource constraints", in: IEEE Transactions on Computers C-36 (8) (1987)949–960.
2. Alex King Yeung Cheung and Hans-Arno Jacobsen, "Green Resource Allocation Algorithms for Publish/Subscribe Systems", In: the 31th IEEE International Conference on Distributed Computing Systems (ICDCS), 2011.
3. Chandrasekhar S. Pawar and R.B.Wagh, "A review of resource allocation policies in cloud computing", IN: World Journal of Science and Technology (WJST) Vol3,PP165-167(2012).
4. Rerngvit Yanggratoke, Fetahi Wuhib and RolfStadler, "Gossip-based resource allocation for green computing in Large Clouds", 7th International conference on network and service management, Paris,France,24-28October,2011.
5. J. Gu, J. Hu, T. Zhao, G. Sun, "A new resource scheduling strategy based on genetic algorithm in cloud computing environment", J. Comput. 7 (2012) 42–52.
6. Linan Zhu, Qingshui Li， and Lingna He, "StudyonCloud Computing Resource Scheduling Strategy Based on the

Nihar Ranjan Nayak and Bhuvaneshwari S,
A Survey on Task scheduling Algorithm and Quality of Service for Resource Allocation in cloud environment,
Indian Journal of Engineering, 2017, 14(35), 63-70,

AntColony Optimization Algorithm",In: International Journal of Computer Science (IJCSI-2012) Vol 9, PP 1694-0814(2012).
7. Ratan Mishra and Anant Jaiswal, "Ant colony Optimization: A Solution of Load balancing in Cloud", in: International Journal of Web & Semantic Technology (IJWesT-2012) Vol 3, PP 33-50 (2012). DOI:10.5121/ijwest.2012.3203.
8. Hao Yuan, Changbing Li and Mankang Du, "Resource Scheduling of Cloud Computing for Node of Wireless Sensor Network Based on Ant colony Algorithm",In: Information Technology journal (ITJ-2012) Vol 11, PP 1638-1643 (2012) DOI:10.3923/itj.2012.1638.1643.
9. K C Gouda, Radhika T V, Akshatha M, "Priority based resource allocation model for cloud computing", Volume2,Issue 1, January 2013, International Journal of Science, Engineering and Technology Research (IJSETR).
10. Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu, "A Multi-dimensional Resource Allocation Algorithm in Cloud.
11. Zhang Qi, Boutaba Raouf. Dynamic workload management in heterogeneous Cloud computing environments. In: 2014 IEEE, network operations and management symposium (NOMS). IEEE; 2014. p. 1–7.
12. Son Seokho, Jung Gihun, Jun Sung Chan. An SLA-based cloud computing that facilitates resource allocation in the distributed data centers of a cloud provider. J Supercomput 2013; 64(2):606–37.
13. LaCurts Katrina Leigh. Application workload prediction and placement in cloud computing systems [PhD Dissertation]. Massachusetts Institute of Technology; 2014.
14. Chang Yao-Chung, Chang Ruay-Shiung, Chuang Feng-Wei. A predictive method for workload forecasting in the cloud environment. In: Advanced technologies, embedded and multimedia for human-centric computing. Lecture notes in electrical engineering, vol. 260. Netherlands: Springer; 2014.p. 577–85.
15. Quiroz Andres, Kim Hyunjoo, Parashar Manish, Gnanasambandam Nathan, Sharma Naveen. Towards autonomic workload provisioning for enterprise grids and clouds. In: 2009 10th IEEE/ACM international conference on, grid computing. IEEE; 2009. p. 50–7.
16. Nguyen Van Hien, Dang Tran Frederic, Menaud Jean-Marc. Autonomic virtual resource management for service hosting platforms. In: Proceedings of the 2009 ICSE workshop on software engineering challenges of cloud computing. IEEE Computer Society; 2009. p. 1–8.
17. Silva João Nuno, Veiga Luís, Ferreira Paulo. Heuristic for resources allocation on utility computing infrastructures. In: Proceedings of the 6th international workshop on Middleware for grid computing. ACM; 2008. p. 1–9.
18. Caron Eddy, Desprez Frédéric, Muresan Adrian. Forecasting for grid and cloud computing on-demand resources based on pattern matching. In: IEEE second international conference on cloud computing technology and science (CloudCom); 2010. p. 456–63.
19. Chaisiri Sivadon, Lee Bu-Sung, Niyato Dusit. Optimization of resource provisioning cost in cloud computing. IEEE Trans Ser Comput 2012; 5(2):164–77.
20. Vecchiola Christian, Calheiros Rodrigo N, Karunamoorthy Dileban, Buyya Rajkumar. Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka. Future Gener Computer System 2012; 28(1):58–65.
21. Feng Guofu, Garg Saurabh, Buyya Rajkumar, Li Wenzhong. Revenue maximization using adaptive resource provisioning in cloud computing environments. In: Proceedings of the 2012 ACM/IEEE 13th international conference on grid computing. IEEE Computer Society; 2012. p. 192–200.
22. Lua Kuan, Yahyapoura Ramin, Wiedera Philipp, Yaquba Edwin, Jehangiria Ali Imran. QoS-based resource allocation framework for multidomain SLA management in clouds. Int J Cloud Comput 2013;1(1) [ISSN 2326-7550].
23. Wu Linlin, Garg Saurabh Kumar, Buyya Rajkumar. Sla-based resource allocation for software as a service provider (saas) in cloud computing environments. In: 2011 11th IEEE/ACM international symposium on, cluster, cloud and grid computing (CCGrid). IEEE; 2011. p. 195–204.
24. S. Singh, I. Chana / Computers and Electrical Engineering 47 (2015) 138–160 159
25. Li Qiang, Hao Qinfen, Xiao Limin, Li Zhoujun. Adaptive management of virtualized resources in cloud computing using feedback control. In: 2009 1st international conference on, information science and engineering (ICISE). IEEE; 2009. p. 99–102.
26. Chieu Trieu C, Mohindra Ajay, Karve Alexei A, Segal Alla. Dynamic scaling of web applications in a virtualized cloud computing environment. In: ICEBE'09. IEEE international conference on, e-business engineering, 2009. IEEE; 2009. p. 281–6.
27. Herbst Nikolas Roman, Huber Nikolaus, Kounev Samuel, Amrehn Erich. Self-adaptive workload classification and forecasting for proactive resource provisioning. Concurr Comput: Pract Exp 2014;26(12):2053–78.