Research on Economical Management Model Based on Qos Constraint

Wenming Sun[1, a], Kai Du [2]

[1]Anhui Radio&TV University, Anhui, China

[2]Tianjin ChuangZhan Tongcheng Technology Corporation, Tianjin, China

[a] 34176556@qq.com

Abstract. This paper proposes an economical resource management model based on QoS constraint, analyzed the significance of introducing QoS and SLA, and discussed the relationship between QoS and SLA management. Finally, the paper realized the mapping of QoS from application layer to resources layer through SLA negotiation, and implemented the optimal allocation of virtual machine resources by using the minimal strategy.

Keywords: QoS constraint; resource management; economic management model.

1. Introduction

User adopts the demand model to obtain service from the cloud provider and cloud system structure design can use the producer-consumer model with the economy principle in human economy market to realize resource allocation and management in the process of cloud computing. From the perspective of the use of resources from the cloud users, the user, through the payment, buys the necessary resources from the service provider to get the appropriate service. It's natural to regard the commodity market as a prototype of cloud computing economic management. Therefore, cloud computing can be regarded as a product-oriented computing market environment. We can apply the different economic models to solve the problem of service and user's needs in the grid computing cloud economy market environment [1-2].

This paper proposes a fair grid resource scheduling strategy: on the condition of meeting the needs of users QoS, tasks are allocated to the most appropriate resources, and with the consideration of competition strategy of earliest time in the resources allocation, they will be ranked in accordance with the adjusted fair competition time. HPC4U and Assess Grid introduce the concept of SLA perception, using the resource management strategy of running time estimation to provide cooperative monitoring protocol, to support fault tolerance and realize QoS management.

The above work has studied the cloud resource allocation strategy from different aspects, but not deepens into the significance of QoS in the cloud computing architecture, the difference between QoS management and SLA management and the influence of QoS on resources allocation [3-5].

2. Economic resource management models of QoS constraints

2.1 Model description

The simplified model of cloud computing economic management based on QoS includes cloud users, brokers, cloud providers and cloud services market, as shown in Figure 1. QoS and SLA management mechanisms are introduced into the model.

QoS and SLA module are introduced to realize QoS management and SLA management in cloud computing economy model. It's bound to face QoS problems with the realization of cloud market resource monitoring, storage, network, virtual machine, service migration and fault tolerance and other functions. Service quality (QoS) provides service performance guarantee, availability guarantee, and other aspects of security, reliability, etc. QoS requires cloud service providers to be associated with cloud users, but when the cloud users obtain the required cloud services QoS through the market, they do not care how the cloud market and cloud service providers meet the service QoS. Therefore, SLA plays an important role in the decision of service providers and end users.

SLA Agreement (Service Level Agreement) is a protocol that is signed by users and service providers about how to provide service to the user. SLA includes technical and commercial legal

parts and the technical part is called the service level specification. The main functions of SLA management include SLA consulting, SLA creation, SLA management, default detection, contract compensation, penalty, etc,[6-8].

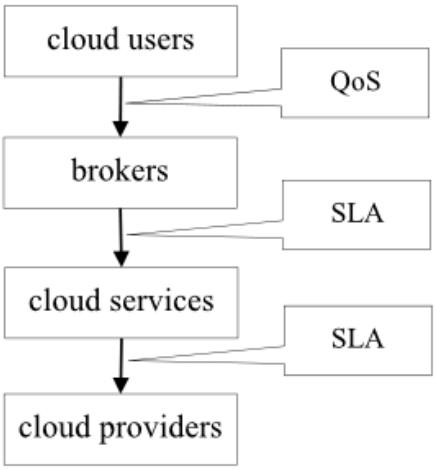

Figure 1. Simplified model of cloud computing economic management based on QoS

2.2 Mapping of QoS from application layer to virtual machine

QoS is a quality description one or more objects' aggregate behavior. It is clear that the QoS has different meanings in different application environments. In traditional network applications, multimedia applications, grid applications and cloud computing applications, its objects and descriptions are not the same. Because of the complexity, the limitation of expression ability, the QoS mechanism is divided into provision, control and management. QoS provision mechanisms include QoS mapping, resource reservation, QoS testing, and QoS negotiation / negotiation.

In order to ensure the QoS of cloud users, we need pay attention to the QoS provision mechanism of, and know how to achieve the mapping and negotiation of QoS through SLA. The research of QoS can be divided into 3 levels: application layer, system layer, resource layer, and it has general significance. At different levels, the QoS has different expressions and focuses.

The mapping transformation of QoS parameters is the process of establishing and maintaining different SLA. Because of the virtualization and distribution of cloud computing applications, it is not realistic to establish a SLA negotiation with all the resources needed to complete an application or service. Therefore, SLA can be decomposed to implemented collaborative management in different resource groups. This paper proposes a method of resource allocation based on QoS constraint for the secondary SLA. When the cloud users ask cloud market service requirements with QoS constraints through the broker, brokers and cloud market make the first SLA negotiation for each service and service is decomposed into different tasks which establish the corresponding TSLA. That is the realization of mapping from application to virtual resources. When QoS negotiation is successful, a set of TSLA mapping relationship is formed [10-13].

Assuming that cloud applications are a collection of cloud services, the relationship between cloud application P and service S is:

P=S1, S2,,, Si (1)

Each service Si corresponds to a TSLA negotiation process, including different QoS requirements, and then you can see the TSLA negotiation process as a mapping conversion from application service layer to cloud market virtual layer, and then the cloud services can be converted into a combination of QoS needs;

Si=Qi1, Qi2,,,Qij (2)

Thus, a cloud application can be expressed as a collection of QoS requirements:

$$P = U\{S_i\} = U\{U\{Q_{ij}\}\} \quad (3)$$

On the other hand, the resource device layer can be abstracted as a collection of a set of logical resources, i.e.:

R=R1, R2,,, Rm (4)

Although the resource layer is composed of physical resources, when resources device layer provides cloud market with logical resources, it needs to consider the overall service performance of the physical resources, the load of the resources, the resource sharing strategy, and the QoS parameters of the logical resource. In cloud computing environment, the physical machine provides logical resource through virtual technology. The virtual machine can be regarded as a collection of QoS requirements:

Rm=q1, q2,,, qk (5)

The ultimate goal of cloud resource management is to decompose the cloud service into a set of minimum scheduling units of the task set and to adjust the task according to a certain objective function until achieve the user's satisfaction. During the mapping from logic resources to physical resources, we need to make the second SLA negotiation, for example, to achieve the mapping from system QoS to network and environment, from logic resources QoS to physical resources, from security QoS to the local security policy module, from the availability of QoS to local availability management module and from billing QoS to the local billing module.

3. QoS constrained resource allocation algorithm

3.1 SLA coordination mechanism

According to the analysis of the second section, we can get the secondary SLA negotiation mechanism with QoS constraints, as shown in Figure 2.

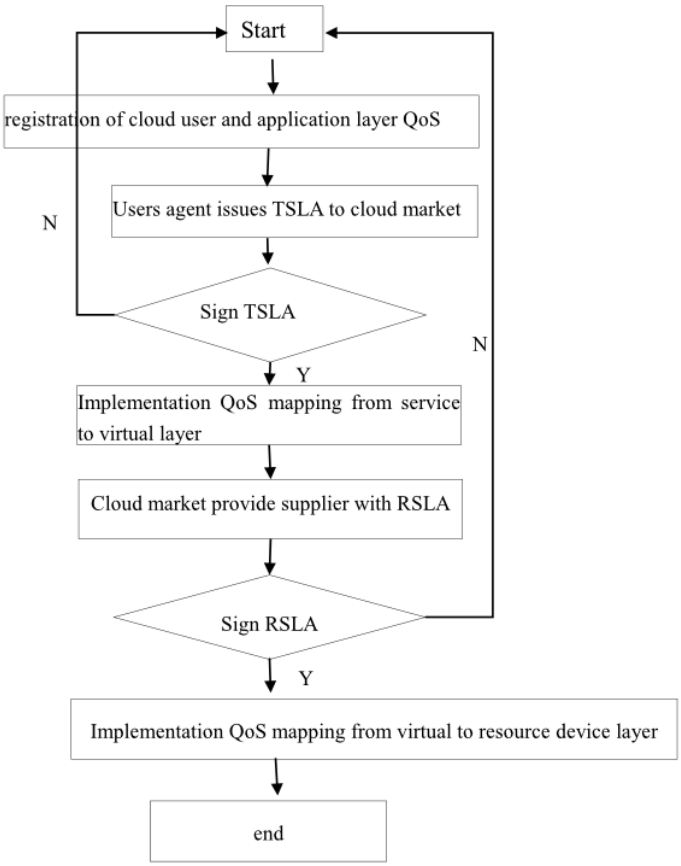

Figure 2 the secondary SLA negotiation mechanism with QOS constraint

Step 1 cloud users register in the cloud market registration and make a request for cloud services QoS.

Step 2 user agents issue cloud market a TS-LA contract containing a service request.

Step 3 cloud market analyzes TSLA contract and maps the QoS to a requested combination of QoS, oversights module and make TSLA negotiation through the cloud market management. If TSLA fails, the transaction ends.

Step 4 TSLA successes, achieving the mapping of QoS from service layer to virtual layer QoS, and realizing the decomposition from service to task.

Step 5 cloud market sends the supplier a RSLA contract containing the virtual layer QoS request, negotiating through the cloud market management and supervision module, if the RSLA negotiation fails, the transaction ends.

Step 6, RSLA successes, achieving the mapping of QoS from virtual layer to resource device layer and scheduling task to resources.

3.2 Virtual machine resource allocation based on utility function

From the above analysis, we can see that the process of the secondary SLA negotiation can be simplified as the QoS mapping from service to resource, and complete the scheduling of resources. Namely, from (3) to (5) and the matching strategy is:

$$\begin{Bmatrix} F_{map}=P\rightarrow R \\ s.t\phi\left(F_{map}\right)=Min \end{Bmatrix} \tag{6}$$

The essence of the resource scheduling algorithm based on RSLA is to find the matching function Fmap and optimize its objective function (Φ（F_{map}）). The utility function can be defined according to the different resource management strategy objectives, and the minimum execution time is used in this paper.

In cloud computing, the logic resource is the virtual machine one. T={T1, T2, Tn}, Ti is presented as MI, that's the instruction number. TimeT is the completing time by the users. The collection of virtual resources is V={VM1, VM2,, VMj,,,,,VMm}. VMj is presented as MIPS, which is the core of virtual machine, and generally n>m, .

The utility function is defined as: U=Min (TimeT). The optimization algorithm based on the utility function is shown as table 1.

Table 1 resource scheduling algorithm based on minimum running time

Algorithm: virtual machine allocation of minimum running time
Step1 to obtain the number of tasks in the task list
Step2 to obtain the number of virtual machines (m) in virtual machine list
Step3 DownSort (T) /to descend the task set according to MI
Step4 UpSort (V)/to ascend virtual machine resource collection according to MIPS
Step5 For (i=1; i<n; i+ +)
For (j=1; j<m; j+ +)
E (I, J) =Ti/Vj// T, V respectively stands for task set and virtual machine set after ranking
Step6 to start from the task of the matrix row number 0, each attempt to assign to the last column with the corresponding VM, if the virtual machine is the best, then complete the allocation; otherwise the task will be assigned to the current results of the optimal VM.
Step7 to repeat Step6 until Step7, until all tasks are assigned to the virtual machine.
Step8 End

It is worth knowing that the algorithm uses MIPS to describe the virtual machine, so the task execution time cannot be too small, but be higher than the minimum threshold, otherwise the optimal virtual machine cannot be found. In addition, the number of tasks should be far greater than the number of virtual machines, in order to fully reflect the significance and superiority of the algorithm.

4. Conclusions

The quality of cloud services is guaranteed by QoS constraint and SLA to realize the optimal scheduling of resources. The next step to research is to design QoS and SLA management module and to measure and evaluate the QoS parameter.

Reference

[1] Rimal B P,Jukan A,Katsaros D,et al.Architectural Require-ments for Cloud Computing System:An Enterprise Clound App-roach[J].Grid Computing,2010,9(1):3-26

[2] Armbrust M, Fox A, Griffith R, et al. Above the clouds: aBerkeley view of cloud computing [R].UCB/EECS-2009-28,E-lectrical Engineering and Computer Sciences,University of Cal-ifornia at Berkeley,2009

[3] Maurer M,Ivona Brandic V C E,et al.Cost-benefit analysis of an SLA mapping approach for defining standardized cloud com-puting goods [J].Future Generation Computer Systems, 2011, doi:10.1016/j.future.2011.05.023

[4] Buyya R,Abramson D,Giddy J.et al.Economic models for re-source management and scheduling in grid computing[J].Con-currency and Computing,2002,14:1507-1542

[5] Neumann D,Christof Weinhardt J S.Bridging the Adoption GapDeveloping a Roadmap for Trading in Grids[J].Electronic Mar-kets,2008,18(1):65-74

[6] Nimis J, Anandasivam A, Borissov N, et al. SORMA-BusinessCases for an Open Grid Market:Concept and Implementation[C]MProceedings of the 5th international workshop on Grid E-conomics and Business Models(GECON.08).2008:173-184

[7] Younge A J,von Laszewski G,Wang L-i zhe,et al.Efficient re-source management for Cloud computing environments[C]MGreen Computing International Conference.2010:357-364

[8] Doulamis N, Litke A D A, et al. Adjusted fair schedu-ling and non-linear workload prediction for QoS guarantees inGrid computing[J].Computer Communications,2007,30(3):499-515

[9] Battre D,Hovestadt M,Keller A,et al.Planning-based schedu-ling for SLA-awareness and Grid Integration [C]MProc. ofWorkshop of the UK Planning and Scheduling.Special InterestGroup, University of Paderborn.2007

[10] Sulistio A,Buyya R.A Time Optimization Algorithm for Sched-uling Bag-of-Task Applications in Auction-based ProportionalShare Systems[C]MProc.of the 17th International Symposiumon Computer Architecture and High Performance Computing(SBAC-PAD.05).2005:235-242

[11] Buyya R,Sudharshan Vazhkudai.Compute Power Market:To-wards a market-oriented grid[C]MProceedings of First IEEE/ACM International Symposium on Cluster Computing and the Grid.2001:574-581

[12] Lai K,Rasmusson L,Adar E,et al.Tycoon:An implementationof a distributed market-based resource allocation system[J].Multi agent and Grid Systems,2005(3):169-182

[13] Sabata B,Chatterjee S,Davis M,et al.Taxonomy for QoS spec-ifications[C]MProc of the 3rd Int.l Workshop on Object-Orien-ted Rea-l Time Dependable Systems.1997:100-107