Indian Journal of Science and Technology, *Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/85054, December 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

QoS Based Cloud Service Selection to Handle Large Volume of Concurrent Requests

Jyothi Shetty[1]* and Demian Antony D'Mello[2]

[1]Department of Computer Science and Engineering, Srinivas School of Engineering, Mukka, Mangaluru - 575025, Karnataka India; jyothi_kishore@yahoo.com
[2]Department of Computer Science and Engineering, Canara Engineering College, Mangaluru, India; demian.antony@gmail.com

Abstract

Background/Objectives: With the exponential growth of cloud providers and users, the possibility of arrival of large number of concurrent requests within a small interval, may reduce the performance of broker. This paper presents a cloud service selection mechanism with service grouping. The main objective of this paper to show the performance difference between the selection mechanism with grouping and without grouping. **Methods/Statistical Analysis**: For the concurrent services groups and sub groups are created using functional and QoS parameters. Analytical hierarchy based selection algorithm is proposed. Performance of selection algorithm with grouping is evaluated. Performance of selection algorithm with grouping is compared with algorithm without grouping and the sufficient reduction in execution time is noted. **Findings:** AHP based selection algorithm with grouping improves the performance by reducing the computation time and reducing the repetitive repository access. **Applications/Improvements:** The limitation of AHP based algorithm is the rank reversal problem. The improved AHP based algorithm can be used to eliminate the rank reversal problem.

Keywords: AHP Algorithm, Concurrent Request, Cloud Service, Quality of Service, Service Selection

1. Introduction

Cloud computing is the latest trend in information technology used by IT organizations. Cost of storage service may be cheap for some cloud providers and may be expensive for computation service. For example, Amazon Elastic Compute Cloud (EC2)[1] provisioning have region based pricing for their IaaS services of some computation capabilities. The challenge for cloud customers is to discover the fact that among the diversity of cloud service providers who are providers that can meet their approximate requirement. There may be different perspective in between different cloud providers regarding functional and non functional requirements. This implies difficulty in order to assess service levels of variety of cloud providers by taking the considerations of reliability, security and quality. Therefore, we need not only to discover multiple cloud service providers but also to select the suitable cloud service provider. Decision making involves comparison of each cloud service based on the parameters. So some researchers have assumed selection problem as Multi Criteria Decision Making (MCDM) problem[2,3]. This framework let cloud users to compare and rank among diversity of cloud providers based on their priority and different dimensions and to select appropriate cloud providers

*Author for correspondence

based on their requirement. In cloud computing environment there is a possibility that large number of users give their request within a small amount of time. Handling such requests separately is time consuming process. In order to handle this situation efficiently, selection is applied on the grouped services. In this paper we have proposed an approach, in which services are grouped and selection algorithm is applied on the groups created.

2. QOS Model

Proposed system makes use of QoS parameters and user priority for service selection. QoS parameter metrics are built in the form of Service Measurement Index (SMI)[4]. Cloud consumers make use of these values and compare variety of available cloud services. By having properly defined cloud computing service characteristics, the proposed framework based on user requirement can provide comparison among available cloud service provider. This framework provides user the capability of comparing variety of available cloud offerings depending on their priority as well as several dimensions and helps them for selection of appropriate cloud service provider based on their need. A QoS based selection model discussed in[5], included the properties like accountability, agility, assurance, usability, security and privacy. In our proposed work, cloud service is selected using the following Quality of Service parameters.

1. Cost
2. Performance
3. Availability
4. Security

2.1 Cost

Before availing cloud computing service, the organization may have questions related to cost measures regarding whether or not it is cost efficient. Therefore, cost is considered as one of the important attribute for IT organization and their business. Cost is most quantifiable metric but there is a need to express characteristics of cost which are relevant to IT or business organization. During the evaluation of ongoing cost there is a need to examine the routine in which billing is determined. The chosen billing basis have dependency on the way of keeping track of costs with actual resource utilization, what are measures of predictions of invoices from month to month and what are mechanisms that are needed to provide proper authorization of consumption change. Cost are not limited to capacity based, usage based, seat based, user based, instance based, concurrent based, fixed monthly fee or profit sharing based and combination of several available multiple options. Proposed system defines cost in terms of Rs (Rupees).

2.2 Performance

Cloud service provider offers different kinds of solution for addressing the needs of different IT organization. These solutions provided by cloud service provider has different performance in terms of service response time, functionality and accuracy. There is a need for organizations to have insights about performance of application on several different cloud service providers in order to determine whether deployed application meet expected QoS requirement. Cloud consumers may have varying or challenging requirement for their time sensitive or critical system. Consumer must have knowledge about how these systems need to respond to specific input. Response time for this critical or time sensitive system may be associated with SLA which needs results instantaneously from system. The failure may result clients in serious consequences.

Service response time computes time in terms of:

- Interval of time in the unit of average length that a specific action required by system to complete.
- Assurance level defined by cloud service provider that response time will actually meet clients QoS requirement.
- Whether personnel resources and proper technology are provisioned for supporting committed response time.
- Whether a record details for time period of 6 months for successful meeting of response time is available. The efficiency of availability of service is measured in the form of response time. For example, in IaaS provider service response time is how quickly the service is available for use. If a consumer requests a virtual machine resource from cloud service provider, then service response time is defined as time period for provider in order to serve the client request. Service

response time includes application deployment, booting virtual machine, IP address assignment, and provisioning virtual machine. Service response time have dependency upon several sub factors such as average response time and maximum response time as promised by cloud service provider and time interval in which level of response time is missed. Maximum response time is defined as cloud service providers promised maximum response time. Proposed system defines service response time in terms of milliseconds.

2.3 Availability

This feature defines probable performance of cloud computing service as expected or as promised in SLA. Since every organization always wants to expand their business by providing best services to consumer they need to provide high level of availability. cloud clients who consume cloud service have requests which varies over time. Therefore, services must have availability feature based on changing business demands. There are multiple perspectives for availability feature:

- Service provider who promise to keep service to be up and running must of appropriate time interval. This is called as availability window.
- There must exist efficient personnel resources and proper technology so that promised service can be provided for defined availability window of appropriate length.
- The assurance level promised by service provider regarding availability targets that are to be met in realistic.
- Whether or not existence of data history which has track of records for meeting promised availability. Availability is defined as time in percentage that customer access the service provided by cloud provider.

Availability = (total service time) - (total time in which service is unavailable) / total service time

Proposed system defines service availability in terms of percentage.

2.4 Security

Protection of critical data and preserving privacy of sensitive data are important for IT organization. Housing and running business related data under the control of another organization is considered as critical issue since it requires strong security policy to be provisioned by cloud service provider. If we consider financial business organization, they need strong policies related to privacy and data integrity. A consumer of cloud computing service usually requires confidence that their system and business data cannot be viewed or modified by unauthorized third party. While hosting the business related data clients need to be provided with assurance that their data are not lost, misused, damaged, or stolen. Different clients may require different level of security which may be higher or lower based on their tolerance towards risk and they may need different level of security for different business application which may be used by single customer. For security proposed system uses point score from 0 to 10 and decimal values are not allowed within this range. 0 to 10 point scores are end points and scores 1 to 9 differentiates different alternatives.

0: No compliance - requirement described are not met and provider does not use any practices described by measure for evaluating services.

5: Partial compliance - only some kind of requirement and practices which are described are implemented by service provider and some of requirement and practices are not implemented thoroughly.

10: Complete compliance - each described requirement are properly met and defined practices are enforced by service provider.

3. System Architecture

This paper mainly deals with selection of cloud services. The system architecture is as shown in Figure 1. Broker is responsible for selection of cloud services, when large number of cloud services is involved. This framework act as a provision model for service selection based on consumer's priority and ranking of cloud services based on cloud service's QoS parameters.

3.1 Cloud Broker

This cloud broker component provides interaction of cloud consumers according to their application needs. This component helps in collecting users QoS requirement and performs ranking of cloud services based on analytic hierarchy process. It consists of 2 modules, Group Manager and Ranking System. Group

Manager groups the concurrent requests. Ranking system first creates AND/OR tree[6] for the requirements given and thereby finding the list of services that satisfy the given request. Then this information is given to the Group manager to create groups and sub groups. Ranking system ranks available cloud services according to user needs. Service Repository contains information about the services, including semantic details.

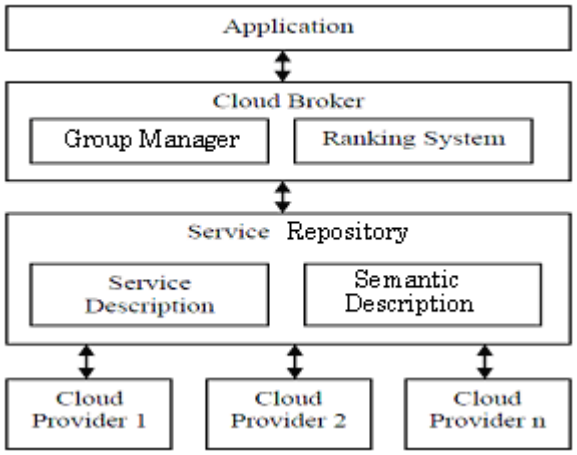


Figure 1. System architecture.

4. Service Selection

The service selection has to select the best service based on the non functional properties. The Phases involved in the selection process is as shown in Figure 2.

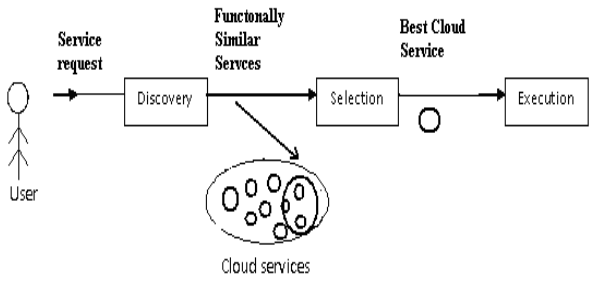


Figure 2. Phases of selection process.

The user request is given to the discovery module which identifies functionally similar services. Several types of discovery mechanisms[7,8] can be used to identify the services, but we have used syntactic and low level semantic based cloud service discovery mechanism[9] The best service out of these services is selected using the selection module. In this paper Analytic Hierarchy Process (AHP) based selection algorithm is used. The AHP[10,11] is a structured technique for complex decisions for proper organization and analyzing process. It was developed by Thomas L. Saaty. AHP and ANP (Analytic Network Process)[12] are mechanisms for solving problems related to Multi Criteria Decision Making (MCDM). ANP is used to solve the complex problems. AHP is one of the most widely used a multi-criteria decision making approach that simplifies complex and unstructured problems by arranging the decision factors in a hierarchical structure. AHP provides appropriate mechanism for consistency check of evaluating alternatives and measures and it also reduces bias in decision making problem. Different phases of AHP based selection process is as follows,

Phase 1: Hierarchical structure for cloud computing service based on QoS values

Phase 2: Relative weight computation for each QoS and service

Phase 3: relative value based weights for cloud computing service ranking.

Phase 4: Relative ranking aggregation for each attribute.

AHP based cloud service selection is proposed in the next section.

4.1 Service Selection Algorithm

Algorithm AHPbasedservice_selection
Input:
1. *Set of m attributes* $\{A_1, A_2 \ldots A_m\}$
2. *Set of services their corresponding values v for each attribute* A_k.
3. *m * 1 matrix UW,, which contains weights given by the user for each attribute*

Out put:
Ranked services

begin

for each A_k *do /* k= 1 to m*
/ Obtain n * n matrix P */*
begin
for each service S_i

begin
for each service S_j
begin
if higer value is better then
$P^k[i][j] = v_i/v_j$
else
$P^k[i][j] = v_i/v_j$
end
end
end

for k= 1 to m do
for each i do / i= 1to m */*
begin

$$E^k[i][j] = \frac{1}{n}\left[\sum_{j=1}^{n} P[i\,|\,j] * \frac{1}{\sum_{j=1}^{n} P[j\,|\,i]}\right]$$

end

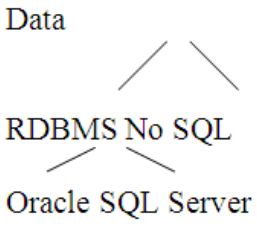
for k=1 to m do
for j= 1 to n do
$SE[j][k] = E^k[1][j]$ */* obtain n * m matrix*

/ obtain n* 1 resultant matrix*
Resultant matrix =Multiply(SE, UW)

end

4.2 Handling Concurrent Service Requests

When large number of requesters gives the service requests at small interval and if broker tries to handle that individually then performance degraded. In order to solve this problem efficiently the functionally similar services are grouped together. The semantic repository[9] is used to group the functionally similar services. The semantic repository stores the service information using synsets of the form (synset_id, lex_id, refs). Each service is linked to its hypernym and hyponym, which gives a semantic knowledge tree as follows:



For example if a user requests are, RDBMS, Oracle and SQL server these services are functionally similar services and these are grouped together. Given a set of semantically similar service requests, multiple requesters might have given different QoS constraints. To rank the services in the same group QoS constraints need to be combined. The users constraints are represented using AND/ OR tree. For example, the constraints given by the requesters R1, R2, R3 is represented as C1, C2 and C3 in Figure 3. This figure also shows the services selected based on the conditions given.

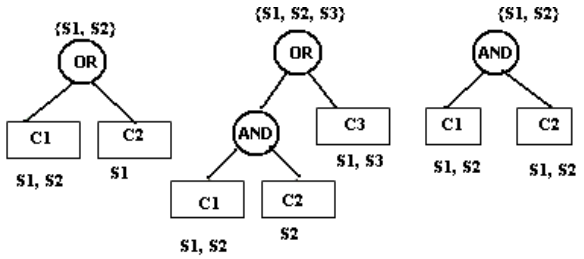


Figure 3. Requesters constraint in the form of AND/OR tree.

The AND/OR tree is evaluated based on the constraints given and Final selected services will be stored in the *requested service* array. The requesters array for the AND/OR trees given in figure 2 is represented in Figure 4.

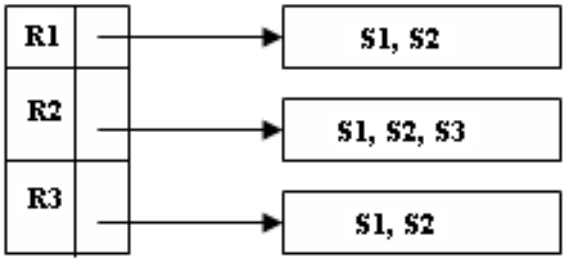


Figure 4. Requesters service array.

Set of concurrent services is obtained using,

$$S = \bigcup_{i=1}^{n} S[i]$$

In Figure 4, it can be observed that requests R_1 and R_2 have the same set of services, from which selection has to be done. So subgroups are created with the functionally similar services. On these grouped services ranking and

selection algorithm is applied. The selection algorithm ranks the services; the best services can be given to all the requests within that group. Suppose many customers are interested to have virtual machines and if provider has the ability to host all the virtual machines then same provider can satisfy the requests of all the customers otherwise two or more providers must be used to satisfy the requester's requirements. In our proposed work based on the ranked service list, the first provider is not able to satisfy all the requests then next ranked services are used to satisfy the request. The requesters are selected in First come first serve basis.

The complete sequence of activities of the Broker for service selection requests is summarized as follows.

a) The broker collects service selection requests for a time slice t_1.
b) The broker categorizes the service selection requests based on the semantic relationship of the services, as stored in the semantic repository.
c) For a group with the multiple requests obtain all relevant cloud services from the repository.
d) Create set groups of the cloud services, depending on the functional requirements. Cloud services with the same functional requirement are assigned to sub groups.
e) For each sub group, with several non functional requirement is categorized into different groups, consisting of similar non functional requirements.
f) For each sub groups created, apply AHP based selection algorithm to find the best cloud service for that group of requests.

5. Experiment and Results

This project provides framework for cloud consumers in order to resolve the problem of discovery of appropriate cloud providers who can meet consumer QoS requirements. Since there exists diversity of cloud service providers, time for executing AHP over service provider data vary since each cloud service have wide range of service attributes. The graph given in Figure 5 shows, execution time for AHP with respect to input size. Several experiments are conducted to support the theoretical concept of group selection algorithm. It as observed that the prototype developed for group selection algorithm effectively groups the semantically similar services, within that functionally similar service, further QOS similar services.

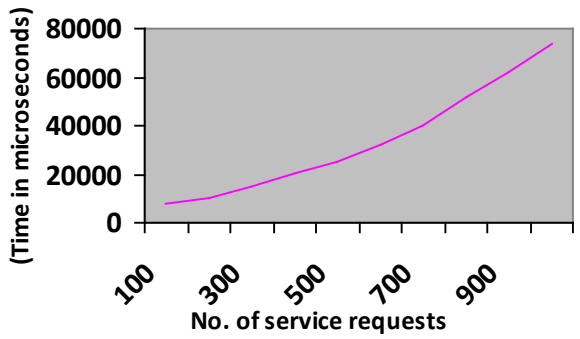


Figure 5. Input size vs. execution time.

Figure 6 shows performance comparison of selection algorithm, with grouping and without grouping. Since selection algorithm, without grouping requires repetitive execution of the requests compared to with grouping, the execution time required for execution of selection with grouping in less compared to without grouping. In case of proposed grouped selection algorithm, even redundant access of repository is reduced, thereby reducing the communication cost

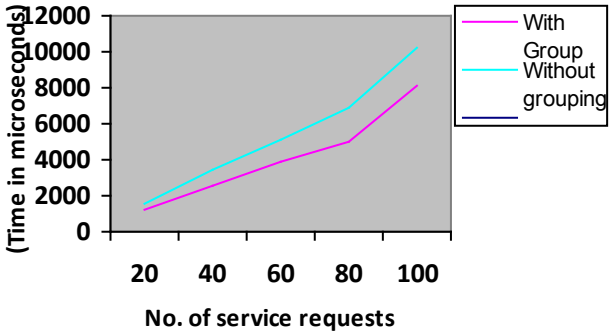


Figure 6. Performance comparison of selection algorithm.

6. Conclusion

This paper makes use of AHP process which decides the best cloud provider according to costumer needs. The limitation of this AHP based algorithm is that it suffers from rank reversal problem. In this paper we have also dealt with concurrent services and also tested the working of the selection algorithm on the groups and subgroups created in case of large number of concurrent services. So according to the groups drawn we can conclude that significant reduction in the execution time for the proposed selection algorithm compared to repetitive selection algorithm.

7. References

1. Amazon EC2 - Virtual server hosting [Internet]. [cited 2015 Mar 12]. Available from: https://aws.amazon.com/ec2/.
2. Zheng Z, Wu X, Zhang Y, Lyu MR, Wang J. QoS ranking prediction for cloud services. IEEE Transactions on Parallel and Distributed Systems. 2013; 24(6):1213–222.
3. Yu R, Yang X, Huangy J, Duanz Q, Ma Y, Tanaka Y. QoS-aware service selection in virtualization-based cloud computing. 2012 Network Operations and Management Symposium (APNOMS), Seoul; 2012. p. 1–8.
4. Service measurement index framework version 2.1, CSMIC Carnegie Mellon University Silicon Valley: Moffett Field, CA USA; 2014.
5. Mahalingam SK, Sengottaiyan N. A QoS guaranteed selection of efficient cloud services. Indian Journal of Science and Technology. 2015 May; 8(S9):103–10 .
6. D'Mello DA, Ananthanarayana VS. A QoS broker based architecture for dynamic web service selection. Proceedings of the 2nd Asia International Conference on Modelling and Simulation (AMS), Kuala Lumpur; 2008. p. 101–6 .
7. Suchithra M, Ramakrishnan M. A survey on different web service discovery techniques. Indian Journal of Science and Technology. 2015 Jul; 8(15):1–5.
8. Shetty J, D'Mello DA. Repository design strategies and discovery techniques for cloud computing. 2013 Proceedings of IEEE International Conferenceon Green Computing, Communication and Conservation Energy, ICGCE'2013, Chennai; 2013. p. 761–6.
9. Shetty J, D'Mello DA. An XML based data representation model to discover infrastructure services. 2015 Proceedings of IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy Materials, (ICSTM), Chennai, India; 2015. p. 119–25.
10. Saaty TL. Decision making with the analytic hierarchy process. International Journal of Services Sciences. 2008; 1(1):83–98.
11. Garg SK, Versteeg S, Buyya R. SMICloud: A framework for comparing and ranking cloud services. Proceedings of 2011 Fourth IEEE International Conference on Utility and Cloud Computing, (UCC), Victoria, NSW; 2011. p. 210–18.
12. Chung BD, Kwang-Kyu S. A cloud service selection model based on analytic network process. Indian Journal of Science and Technology. 2015; 8(18):1–5.