The QoGS Method Application for Selection of Computing Resources in Intercloud

R. Butkiene¹, G. Vilutis², I. Lagzdinyte-Budnike², D. Sandonavicius², K. Paulikas²
¹Department of Information Systems, Kaunas University of Technology,
Studentu St. 50–309, LT-51368 Kaunas, Lithuania
²Department of Computer Network, Kaunas University of Technology,
Studentu St. 50–416, LT-51368 Kaunas, Lithuania
gytis.vilutis@ktu.lt

Abstract—In this paper the selection of the right Cloud in Intercloud of computing services for fulfillment of the user task is analyzed. The main problem approached is how to select the right Cloud in Intercloud for the user task when there are no available or relevant resources in the private and public Cloud. In this case the Cloud, which gives the calculation results to the user in the shortest time including waiting time in queue, should be selected. For selection of the right Cloud the QoGS method has been chosen. The principle schema of QoGS method application in Intercloud has been presented. However, QoGS method works properly if the right set of weighted coefficients (SoWC) is selected. Existing methods are not suitable for selection of weighted coefficients because they require sending a lot of test tasks to Intercloud. Therefore, a new algorithm for selection of the best SoWC is presented. Results of the experiment show that it allows minimizing the workload of Intercloud by test tasks significantly.

Index Terms—Cloud, load flow control, computer networks.

I. INTRODUCTION

A Cloud computing technology becomes more and more popular. Therefore, users face the problem when they are selecting the Cloud for the task fulfillment. The time of task fulfillment in the Cloud is the main indicator of the computing Cloud quality and the main criterion in the Cloud selection. The time of the task fulfillment depends on many parameters of the Cloud: an amount of resources allocated for tasks, a data transmit rate, a size, complexity, and a number of the tasks in queue, etc. However, these parameters are not usually presented for a user. Even if the parameters are presented, the Cloud selection remains difficult if for different Clouds different sets of the parameters are presented. A unification of parameters describing the Cloud would decrease these difficulties. One way to unify the Cloud description is the use of ontology [1]. An application of ontology would allow to search and choose the right Cloud not only for the human-users but also for the software agents. Nevertheless, there are many parameters describing a Cloud. To assess all of them is time-consuming and requires the user’s experience. If we want to automate a selection of Cloud, a method, which is able to assess all parameters of Cloud and choose the best one, is needed.

The unification of Cloud’s parameters is not analyzed in this paper. The aim of this paper is to choose and apply the method for the right Cloud selection in Intercloud. Furthermore, this paper focuses on the Clouds providing services according SaaS (Software as a Service) model, for example graphics rendering services. The right Cloud selection is difficult when choosing from Clouds providing services according IaaS (Infrastructure as a Service) or PaaS (Platform as a Service) model. In these cases the preparation and adjustment of infrastructure, upload of image and so on can take considerable time.

II. INTERCLOUD STRUCTURE

The method for selection of the right Cloud for the user’s task fulfillment requires certain information about each computing Cloud; in other words, it requires the parameters about each Cloud. In this paper we refer to the Intercloud [2]–[6] (Fig. 1) where information about each Cloud in the Cloud network could be found, using services of the Intercloud Root. The Intercloud Root hosts the Cloud Computing Resource Catalogs for the Intercloud computing resources. The Intercloud Root and Intercloud Exchanges facilitate and mediate the initial Intercloud negotiating process among Clouds. Once the initial negotiating process is completed, each of these Cloud instances collaborates directly with each other.

The user accesses Private Cloud and receives its resources via Interface. Main function of this Interface is to describe the user’s requirements for the computing resources formally. When the specification of the user requirements is received, the Cloud Broker looks up for the proper resources in the Private Cloud. If the resources satisfying the user’s requirements are not found, the Cloud Broker accesses the Intercloud Root and looks up for the resources in the Public Cloud. The method for the resource search in the Intercloud is discussed in this paper.

III. RELATED WORKS

A method used by Cloud Broker (CB) for the resource selection plays an important role.

Manuscript received August 5, 2012; accepted April 19, 2013.
The role of the Cloud Broker is similar to the role of the Resource Broker, which is a common component of the computing Grid.

ERT method is mostly used resource selection method by Grid’s Resource Broker [7]. However, many researchers state that the Resource Broker working on ERT method poorly divides tasks in the network [8], [9]. The reason is that ERT method evaluates only two parameters of the resource: central processor speed C and job queue length L. Other methods, such as GSQN [10], Q-QoSM [11], [12] and QoS-based [13], evaluate three parameters of the resource. Each of them uses different sets of the parameters: GSQN evaluates the resource availability, a total time to delivery, and a cost; Q-QoSM – a data transfer rate, latency, and a cost; and QoS-based – central processor speed C, a storage size, and a RAM memory size. Whereas the QoS resource selection method [14], [15] evaluates all parameters of the resource, which are evaluated in the methods mentioned above, and more. A possibility to evaluate more parameters allows the user determining the right resource for his/her task fulfillment more precisely. An experiment shows that the QoGS method chooses the resources better than other methods do [15]. Moreover, the QoGS method forecasts the task fulfilment more precisely. An experiment shows that the QoGS method is more sensitive for the weighted coefficients the user has been inputted. If the weighted coefficients are selected improperly, the QoGS method can even worsen the time of the task fulfilment, comparing to the right resource in the Intercloud. In this case the formula (1) should be treated in this way: $n$ should express a number of Clouds in the Intercloud, $k$ - index of the task sent to Cloud, $V_{ni}^{D}$ – $i$-th parameter value of $n$-th Cloud, $V_{nj}^{M}$ – $j$-th parameter value of $n$-th Cloud, $V_{i}^{D_{\text{max}}}$ – the highest value of the $i$-th in the Intercloud, $V_{j}^{M_{\text{max}}}$ and $V_{j}^{M_{\text{min}}}$ would be the highest and the lowest value of the parameter $j$-th in the Intercloud respectively. $\omega_{kl}^{D}$ and $\omega_{kj}^{M}$ – weights of $i$-th and $j$-th parameters, set by $k$-th task.

A good selection of resources in the QoGS method is ensured by the right selection of weighted coefficients of the parameters (WCoP). The weighted coefficients (WC) point out parameters of the resource, which influence the time of the task fulfillment in the Cloud most. It is important to note that the QoGS method is sensitive for the weighted coefficients the user has been inputted. If the weighted coefficients are selected improperly, the QoGS method can even worsen the time of the task fulfillment, comparing to other methods. Therefore, it is very important to determine the right set of weighted coefficients (SoWC).

Using the QoGS method, the determination of the set of the weighted coefficients is done by the user. However, for the inexperienced user such task is quite difficult [16]. The WCoP determination algorithm for the CB would be
helpful. If the user does not set the best SoWC in the description of his/her task requirements, the CB sets it, considering the results of the WCoP determination algorithm.

The weighted coefficients are widely used in the clustering and classification tasks [17]–[19], neural networks [20]–[24] and other. The neural networks for determination of the weighted coefficients use self-learning methods. Such methods could be applied for recognition of the similar task that comes to the Cloud. Though, in the case of the QoGS method, not the task but the right resource for that task is selected. The parameters of the resource, in regard to a new task that came into the Cloud, differ (especially the number of tasks that are queued in the resource row). Therefore, self-learning methods of the neural networks are not applied.

For the solving optimization tasks the weighted coefficients are calculated, using Simplex method [25]–[27]. However, Simplex method cannot be used in the case of the QoGS method because it is impossible to create an equation system for selection of the best SoWC; there is a lack of information about the time of the task fulfilment. Therefore, a new solution for determination of the WCoP used in the QoGS method is needed.

IV. AN APPLICATION OF QOGS METHOD IN THE INTECLOUD

The QoGS method in the Intercloud should be applied in case when the relevant resource in the Private Cloud is not found. A principle schema of the Intercloud resource selection is depicted in Fig. 2. The user sends a task to the Cloud Broker and specifies the requirements for resources of Cloud (sets the preferences, weighted coefficients and constraints on parameters of Cloud) via Interface. The CB receives the parameters of all Intercloud resources from the Intercloud Root and filters the resources satisfying the user’s requirements. If resources, satisfying the user’s requirements, are found, the CB applies the QoGS method for selection of the most relevant Intercloud resource from filtered ones. When a relevant resource of the Intercloud is selected, the task is sent to that resource. After the task fulfilment a result is sent to the user via Interface.

When the CB applies the QoGS method for selection of the relevant resource, it takes the SoWC from WC storage. The SoWCs of all Intercloud resources are calculated by and stored in the Intercloud Root. The Intercloud Root recalculates the SoWCs of the Cloud resources if the structure of the Cloud has been changed. The Cloud Broker receives the SoWCs from the Intercloud Root and stores them in its own storage. The WC storage makes the normal work of the CB possible when the connection to the Intercloud Root is lost.

V. AN ALGORITHM FOR DETERMINATION OF THE WCoP OF THE INTERCLOUD RESOURCE

The purpose of the algorithm is to determine the best set of the weighted coefficients of the parameters of the Intercloud resources. As was mentioned above, the WCoP are the essential factor in the selection of the Cloud.

For the further consideration, the four parameters of the Cloud, which influence the successful and the fastest fulfilment of the task most, have been chosen: a number of the processors $N$, the average speed of the processors $C$, the parameters $N$ and $C$ define a capacity of the Cloud. Knowing these parameters, it is possible to predict how quickly the user’s task will be fulfilled. Knowing parameters $L$ and $W$, it is possible to predict when the fulfilment of the user’s task will start if the queue of the tasks exists. The parameter $W$ is especially important when deciding which Cloud should be selected.

The easiest way to obtain the best SoWC is to generate all possible combinations of the weighted coefficients (SoWC) for the Cloud parameters in specified variation limits, to calculate the $Q_{sd}$ (see formula (1)) with each SoWC for each Cloud of the Intercloud, to select the Cloud with the highest $Q_{nk}$, to send the test tasks to the selected Cloud, and to observe fulfilment time of the test task. The test task loads the Cloud resources minimally. Its purpose is to inform the Cloud Broker that it got into the computing resource. The test task ends immediately when a message to the CB is sent. From the set of the received test task fulfilment times the shortest one is selected.

Fig. 2. A Principle schema of the cloud selection in the intercloud.
The SoWC, which pointed out the Cloud, which executed the test task fastest, is considered as the best SoWC. However, this way can be time-consuming.

A number of generated SoWCs grows when the number of parameters increases or the variation limits of the WC are expanded. For example, when the variation limits of the WC are set to [1..2], a variation step is 1, and the number of the parameters increases from 4 to 5, the number of the SoWC increases from 16 to 32. When the number of the parameters is 4 and the variation limits of the WC are expanded from [1..2] to [1..10], the number of the SoWC increases from 8 to 10000.

It is necessary to emphasize that different but reiterating SoWCs generated among all possible SoWC can be found. A reiterating SoWC is considered such set where all values of that set are multiples of the values of another SoWC. The example where the SoWC is reiterating is presented in the formula (2), where the SoWC defines four parameters of the Cloud

\[
Q_{\text{SoWC}}^\text{SoWC} = \begin{array}{l}
\omega_W = 1 \\
\omega_L = 1 \\
\omega_C = 2 \\
\omega_N = 4 \\
\end{array}
\]

where \(\omega_W\) – WC for the average time of the tasks fulfillment \(W\); \(\omega_L\) – WC for the number of the tasks in the queue (or queue length) \(L\); \(\omega_C\) – WC for the average processor speed of the Cloud \(C\); \(\omega_N\) – WC for the number of the processors \(N\).

Different but reiterating SoWCs calculate the quality \(Q\) of the Cloud equally and do not give any benefit in the selection of the relevant Cloud. Thus, the reiterating SoWC could be rejected.

The algorithm (Fig. 3) described in this section does not embody shortcomings mentioned above. In the beginning of the algorithm an initial SoWC must be set. If the algorithm is executed for the first time, the value 1 is set to all WC in the initial SoWC. The variation limits of the WCoP are set to [1..2]. Then all possible SoWC are generated. Each generated SoWC is tested by sending the test tasks to the Intercloud for fulfillment. When the data about the fulfilled test tasks is received, the best SoWC of the iteration is selected. The best SoWC is considered a set, which pointed the task to the resource where fulfillment time of the task was the shortest. The best SoWC selected is compared to the best SoWC selected in the previous iteration. If the best SoWC selected in the current iteration is equal to the best SoWC selected in the previous iteration multiplied by 2, then the best SoWC selected in the current iteration multiplied by 2 and saved by the Intercloud Root. The algorithm completes. Otherwise, all values of the best SoWC selected in the current iteration are multiplied by 2 and for each parameter the variation limits of the weighted coefficients are changed in this way: the lower variation limit of the WC is obtained by decreasing the value of the WC by 1, and the upper variation limit of WC is obtained by increasing the value of the WC by 1. For example, if WC of certain parameter is 4, then the variation limits of this parameter will be [3..5]. When the new variation limits for the WCoP are set, the next iteration for selection of the best SoWC starts.

If the algorithm was executed once, then the best SoWC determined in the previous calculations is set as initial SoWC. The values of the initial SoWC are multiplied by 2 and the variation limits of the WCoP are set in the way described in the previous paragraph. Then the iteration for selection of the best SoWC starts. When the best SoWC are determined, it is recorded into the WC storage and used in the QoGS method. However, if the structure of the Cloud changes, it is necessary to recalculate the SoWC. The best SoWC received in the next iterations of the algorithm are more and more precise – it allows selecting more relevant Cloud.

However, the search of the most precise SoWC can take a lot of time.

---

**Fig. 3. An Algorithm for determination of WCoP.**
Therefore, a possibility to terminate the algorithm is provided. If the precision obtained (task fulfilment time) meets the user’s requirements, he/she can terminate the algorithm.

VI. EXPERIMENT

The aim of the experiment was twofold:
1) To determine how many times the number of the test tasks sent to Intercloud is decreased when the proposed algorithm is used comparing to the number of the test tasks sent to the Intercloud when all possible combinations of the SoWC are generated;
2) To determine when the execution of the proposed algorithm should be stopped.

The simulation model was created to do an experiment. The model simulates work of Cloud Broker in the Intercloud. The Cloud Broker applies the QoGS method and the proposed algorithm for the WCoP determination. Static parameters of the Clouds are presented in Table I. These parameters were used in simulation when, according to the method described in this paper, the load balancing of the flow of simulated tasks was performed. The first column “No.” identifies the Cloud. A number of processors \( N \) and average processors speed \( C \) influenced dynamic parameters (a number of the tasks in the queue \( L \), and the average time of the task fulfilment \( W \)) of the Clouds directly.

The structure of the Intercloud is similar to the structure of Grid network including services offered by them. But the structure of the Intercloud is underdevelopment now, so there is no possibility to acquire statistical data about operation of the Intercloud. Therefore the simulation model was validated comparing its operation with the work of BalticGrid network [28]. The simulation model comparing with BalticGrid network works with the average square error 0.7249.

During simulation the WC\(^s\) \((\omega_W, \omega_L, \omega_C, \omega_N)\) were chosen for 4 parameters: \( W, L, C, N \) (see the description of parameters in the previous section). The parameters of the tasks generated by the stream were:
1) The required CPU speed – 1 GHz;
2) The required number of the processors – 10-19;
3) The duration of the task performance – 10-100 min on the 1GHz processor.

The results of the experiment are presented in Table II. The data of the experiment is compared with the data, which would be obtained if all possible SoWC are generated under the given WC variation limits. An increase of the number of the SoWC is exponential if all possible combinations of the SoWC are generated. The results show that the proposed algorithm allows minimizing the number of the test task sent to the Intercloud from 6.25 % to 99.98 %.

An expansion of the WC variation limits and a correction of the WC of the best set did not give a reasonable effect. When the WC variation limits has been expanded up to \([1..10]\), the best set of the WC of four parameters was found. Next expansions gave the same results, which is the same average time of the task fulfilment.

In order to ascertain the practical benefit of the selected best SoWC, the stream of 1,000,000 tasks has been generated in the simulation model. The average time of fulfillment of these tasks is presented in Table II.

The experiment results show that there is no reasonable reason to continue the algorithm when after two last iterations the average time of the task fulfilment is received the same, or difference is not significant (the user’s needs are satisfied).

<table>
<thead>
<tr>
<th>WC variation limits</th>
<th>A number of all possible SoWC</th>
<th>A number of SoWC generated by the proposed algorithm</th>
<th>The best SoWC (including reiterating ones)</th>
<th>The average time of task fulfilment, min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>16</td>
<td>15</td>
<td>(\omega_W = 1; \omega_L = 2; \omega_C = 1; \omega_N = 1)</td>
<td>27.25991</td>
</tr>
<tr>
<td>1-5</td>
<td>625</td>
<td>96</td>
<td>(\omega_W = 1; \omega_L = 3; \omega_C = 1; \omega_N = 3)</td>
<td>26.72516</td>
</tr>
<tr>
<td>1-10</td>
<td>10000</td>
<td>177</td>
<td>(\omega_W = 1; \omega_L = 6; \omega_C = 2; \omega_N = 7)</td>
<td>26.70401</td>
</tr>
<tr>
<td>1-20</td>
<td>160000</td>
<td>258</td>
<td>(\omega_W = 1; \omega_L = 11; \omega_C = 4; \omega_N = 13)</td>
<td>26.69552</td>
</tr>
<tr>
<td>1-40</td>
<td>2560000</td>
<td>339</td>
<td>(\omega_W = 1; \omega_L = 11; \omega_C = 4; \omega_N = 13) or (\omega_W = 2; \omega_L = 22; \omega_C = 8; \omega_N = 26)</td>
<td>26.69552</td>
</tr>
<tr>
<td>1-80</td>
<td>409600000</td>
<td>420</td>
<td>(\omega_W = 1; \omega_L = 11; \omega_C = 4; \omega_N = 13) or (\omega_W = 2; \omega_L = 22; \omega_C = 8; \omega_N = 26) or (\omega_W = 4; \omega_L = 44; \omega_C = 16; \omega_N = 52)</td>
<td>26.69552</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>(C, GHz)</th>
<th>(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>3.</td>
<td>1.5</td>
<td>45</td>
</tr>
<tr>
<td>4.</td>
<td>1.5</td>
<td>55</td>
</tr>
<tr>
<td>5.</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>6.</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>7.</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>8.</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>9.</td>
<td>2.5</td>
<td>50</td>
</tr>
<tr>
<td>10.</td>
<td>2.5</td>
<td>55</td>
</tr>
<tr>
<td>11.</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>12.</td>
<td>3</td>
<td>60</td>
</tr>
</tbody>
</table>

TABLE I. THE STATIC PARAMETERS OF THE CLOUDS.

TABLE II. THE RESULTS OF THE EXPERIMENT.
The best average time of the task fulfilment received applying the QoGS method and proposed algorithm was 26.695 min. In order to ascertain that the obtained result is better than the results obtained applying another method, the Estimate Response Time algorithm was used in the same conditions. The average time of the task fulfilment received applying Estimate Response Time algorithm was 29.357 min.

VII. CONCLUSIONS

In this paper the QoGS method has been chosen for the selection of the right Cloud in the Intercloud. The principle schema of the QoGS method application in the Intercloud has been presented. Since the QoGS method is sensitive to the weighted coefficients of the Cloud's parameters, the algorithm for selection of the right set of WCoP has been proposed.

The proposed algorithm has been tested experimentally. The results of the experiment show that:
1) It allows minimizing the number of the test tasks sent to the Intercloud significantly. Thus, the proposed algorithm is more efficient than other algorithms;
2) The situation during execution of the algorithm when after two last iterations the average time of the task fulfilment is received the same, or difference is not significant from the viewpoint of the user, is an indication that the algorithm should be stopped;
3) The considerable expansion of the WC variation limits does not give a reasonable effect. The results of the experiment show that the WC variation limits [1..10] are enough for selection of the best SoWC.
4) The average time of the task fulfilment received applying the QoGS method is 9.067% shorter than the average time applying the Estimate Response Time algorithm.

REFERENCES