Mingzhang Wu, Janne Koljonen, Timo Mantere*

Addressing Resource Allocation Issues in Cloud Computing Environment with Ant Colony Optimization

Abstract: Cloud computing is a fast growing and attractive paradigm in information technology, since it allows using resources on-demand wherever and whenever needed. The use of dynamic cloud resource allocation allows immediate accommodation to unpredictable demands and improvement in the return on investment as for the computing infrastructure. The cloud resources allocation optimization model is one of the core parts in cloud computing. However, despite the recent growth of the research in the cloud computing area, several problems with the process of resource allocation remain unaddressed. Cost and performance are two important but contradictive objectives in the cloud resources allocation process. Cost-performance trade-off constitutes a challenging multi-objective optimization problem in cloud resources allocation. In this paper, a new optimization model is proposed to solve this multi-objective optimization problem effectively. An ant colony optimization algorithm that optimizes the Quality of Service (QoS) and the response time in a simulated CloudSim environment that models five servers of varying characteristics. Experimental results demonstrate the effectiveness of the designed algorithms. Ant colony algorithm shows mostly higher performance than the round robin and greedy assignment algorithms that were used as benchmarks.

Keywords: ant colony optimization, cloud computing, CloudSim, cost-performance, resource allocation, trade-off problem

*Corresponding Author: Timo Mantere: University of Vaasa, School of Technology and Innovations, E-mail: timo.mantere@uva.fi

Mingzhang Wu: E-mail: mingzhang.wu@gmail.com

Janne Koljonen: University of Vaasa, School of Technology and Innovations, E-mail: janne.koljonen@uva.fi

1 Introduction

In the recent years, information technology (IT) has been integrated into our daily life more and more. The major applications are build up based on network and internet technologies. We are now in an era of “big data” with rapid growth on the number of transactions, information, and data. However, low cost, fast speed, and efficient computing are desired. The traditional network and local computation capacity are unable to meet these needs. Instead, distributed network technologies are developed to enable the utilization of distributed computing resources from the internet. How to integrate and distribute the resources, such as servers, over the internet give new research topics to be considered.

Cloud computing is a result of the convergence of several technologies, such as, (1) hardware, (2) internet technologies, (3) distributed computing, and (4) systems management. The main advantage of cloud computing is providing computing resources based on the public utility model (compare to water, electricity, gas, and telephony) to enhance reliability, scalability, and performance.

From the technical perspective, cloud computing is the integration of many aspects of technologies, such as (1) virtualization, (2) utility computing, and (3) distributed computing. From the business perspective, cloud computing is a new business model. It enables (1) sharing information among users, (2) buying resources on-demand without large investments, (3) selling capacity to many users, and subsequently (4) improving the return on investments due to better rates of capacity use. Furthermore, (5) investing on the latest, high-performance infrastructure should give a business advantage to the service provider.

Ant Colony Optimization (ACO)is inspired by the observation of the collaboration activities of ants searching for food. Ants can gradually find out the shortest path between a food source and the nest of the colony. ACO can be modeled as a graph, including resource assignment. However, in its original form, ACO modifies the edges of a graph, not the nodes as will be done in this paper.

2 Experimental setup

The role of ACO is tackle the Cost-Performance trade-off problem (CPTOP). First, the problem is transformed into a Travel Salesman Problem (TSP). Second, the separate objectives of cost and performance are integrated into a single objective function, which is then optimized in the CloudSim environment.

CloudSim provides many ways for managing and utilizing the resources, such as virtual machine (VMs), datacenter, etc. It offers: (1) support to modeling and simulation of a large cloud computing infrastructure, (2) a self-contained support data center, service agent, scheduling and allocation strategy platform.

As for the optimization, the objectives for the customers and the cloud service providers are different: The objective of the customer is to maximize the performance of resources with a fixed cost. For the cloud service provider, the total amount of resources is fixed, and the objective is to add as many customer requests to the cloud as possible.

The objective – an energy function – can be defined in terms of idle capacity, that is, unoccupied capacity that cannot be assigned to another VM due to limitations in some other resource type. The optimization then follows the principle of minimum energy.

3 Experiments and results

To test the algorithm, a small cluster scenario with five servers having different characteristics and a single type of virtual machine (VM) was used. To compare the ACO algorithm with other assignment schemes, the results with the round-robin and a customer greedy heuristic schemes were also tested.

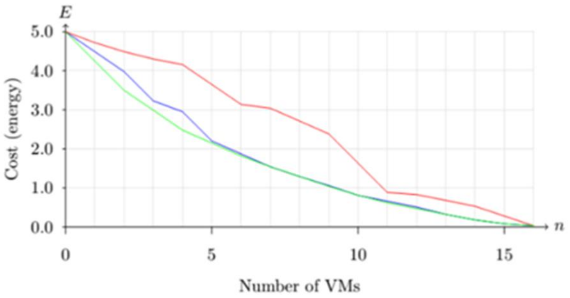

Figure 1. Comparison of the efficiency of the algorithms: round-robin (blue), greedy (red), and ACO (green).

Figure 1 shows the energy for each of the three assignment strategies for an increasing number of standard size VMs. The ACO algorithm described in this study (green line) has lower energy than the other two, although the round-robin strategy (blue line) is close to optimal. The greedy algorithm is evidently the worstof these three with practically any number of VMs.

4 Conclusions

Cloud computing enables a more efficient way to use utility computing resources and services. Users can access the computing resources with virtualized technologies and pay only for the resource accessed while getting the level of quality of service (QoS) wanted.

In this study, an optimal assignment was found by minimizing a combined energy function that measures cloud provider's costs. However, the test setup was relatively limited and simplified, due that a full-scale simulation of cloud computing services are complicated. The benefit for the cloud provider is to maximize the possibility to add further VMs to the existing cloud infrastructure without performance degradation or delays.

Three assignment policies were simulated and tested in CloudSim: round-robin, a customer greedy heuristic, and an optimized allocation implemented as an Ant Colony Optimization algorithm. When comparing the three assignment policies, the round-robin can be said to be both simple and efficient. The greedy assignment, where a customer can choose to allocate a VM to the host with the freest capacity was rather expensive. However, in this implementation, each processor had the same MIPS, so the results might be different in an environment with more versatile set of resources available.

The way that ACO uses resources differed in an optimal case from the round robin method only so that one extra VM was assigned to a different host. Nevertheless, this little change makes ACO to obtain the best energy function values. The further tests with varying numbers of VMs validated the mutual order of the three algorithms; ACO consistently outperforms the simple round-robin method slightly, while the round-robin method outperforms the simple greedy method significantly.

These experiments have some limitations. Actually, the implementation of ACO in CloudSim makes a solution 'all at once'. Therefore, it is recommended to develop an ACO variant that could find an optimal policy with a more dynamic situation, where VMs are created and terminated all the time.