Engineering Science and Technology, an International Journal 20 (2017) 1249–1259

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Full Length Article

Simulated annealing based VM placement strategy to maximize the profit for Cloud Service Providers

Sourav Kanti Addya [a,*], Ashok Kumar Turuk [a], Bibhudatta Sahoo [a], Mahasweta Sarkar [b], Sanjay Kumar Biswash [c]

[a] Department of Computer Science and Engineering, National Institute of Technology, Rourkela, 769008 India
[b] Department of Electrical and Computer Engineering, San Diego State University, CA, USA
[c] Department of Software Engineering, Institute of Cybernetics, National Research Tomsk Polytechnic University, Tomsk 634050, Russia

ARTICLE INFO

Article history:
Received 5 March 2017
Revised 25 July 2017
Accepted 18 September 2017
Available online 30 September 2017

Keywords:
Cloud computing
Data center
Virtual machine
Energy efficiency
Revenue
Simulated annealing

ABSTRACT

Virtual machine (VM) placement strategies reported in the literature focuses mainly on minimization of power consumption and maximization of placed VMs. The revenue earned by a cloud service provider (CSP) depends on the number of VMs placed. Increasing the number of VMs placed by a CSP not only increases the power consumption but also decreases the profit margin of the CSP. In this paper, we propose a technique called maximum VM placement with minimum power consumption (MVMP) to maximize the profit earned by a CSP. The proposed technique attempts to maximize the revenue and minimize the power budget. It is formulated as a bi-objective optimization problem, and is solved using simulated annealing (SA) technique. To reach a sub-optimal solution more randomness is applied to SA. Our MVMP algorithm is compared to five state of the art algorithms in the realm of strategic VM placement, namely Marotta and Avallone (MA) approach, Hybrid genetic algorithm (HGA), Modified Best-Fit decreasing (MBFD), First-Fit decreasing (FFD) and Random deployment. We observe that MVMP performs better than Marotta and Avallone (MA) approach, HGA, MBFD, FFD and Random placement in terms of number of servers used, energy consumption, profit and execution time. Scalability of MVMP is verified using two different scenarios: (i) fixed number of VMs and, (ii) fixed number of servers. It is observed that MVMP is scalable too.

1. Introduction

Cloud computing is an emerging technology that uses Internet and centrally located remote servers for maintenance of data and applications. It allows businesses and consumers to use applications without installation and maintenance hassles and allows users to access their personal files at any location. The key technologies that made cloud computing feasible are virtualization, cyber-infrastructure and service oriented architecture [1]. Virtualization enables dynamic sharing of physical resources on cloud environment. It partitions the resources such as CPU, memory, bandwidth of one or more machines into multiple execution units called virtual machine (VM). Each VM acts as a complete system and is logically isolated from each other [2]. VMs are hosted on physical machines (PMs) also referred to as hosts. The process of mapping a VM onto a PM is known as VM placement.

Data center is a huge computational facility with many PMs at its disposal. A typical data center is a collection of high-end tightly-coupled servers or PMs. The expenses incurred in running a data center includes manpower, software licenses, depreciation of computing equipments, space to host the physical infrastructure, power budget etc. Reduction in power consumption is a major challenge in data center management. Servers and cooling systems in data centers consumes maximum power. However, by judiciously placing the VMs onto PMs it is possible to minimize the power consumption of servers and thereby increase the CSPs revenue.

In order to maximize revenue, a CSP will try to place as many VMs as possible. Increasing the number of placed VMs also increases the number of servers used which consequently increases the power consumption. Therefore, maximizing the number of VMs placed and simultaneously minimizing the energy consumption are contradictory parameters. Therefore, a VM

* Corresponding author.
E-mail addresses: kanti.sourav@gmail.com (S.K. Addya), akturuk@nitrkl.ac.in (A.K. Turuk), bdsahu@nitrkl.ac.in (B. Sahoo), msarkar2@mail.sdsu.edu (M. Sarkar), skbismu@gmail.com (S.K. Biswash).

Peer review under responsibility of Karabuk University.

https://doi.org/10.1016/j.jestch.2017.09.003
2215-0986/

placement technique that can place more VMs onto a lesser number of servers will maximize the CSP's profit.

In this paper, we propose a mechanism to maximize the profit for CSPs. The problem is formulated as a bi-objective optimization problem which simultaneously minimizes the energy consumption and maximizes the earned revenue. Simulated annealing (SA) technique is used to solve the optimization problem. This is because, for a small solution space it converges to a better sub-optimal solution than other meta- heuristic approaches [3]. A trade-off between energy consumption and VM placement is attempted to maximize the profit. The proposed bi-objective problem is compared with Marrota and Avallone (MA) approach [4], Hybrid Genetic Algorithm (HGA) [5], Modified Best-Fit decreasing (MBFD) [6], First-Fit Decreasing (FFD) [7–9] and, Random Placement. It is observed that the proposed mechanism outperforms all five benchmark algorithms.

Rest of the paper is organized as follows. In Section 2, some of the techniques reported in the literature, for VM placement is discussed. Cloud data center model, and the proposed MVMP technique is explained in Section 3. A simulated annealing based heuristic to solve the VM placement problem is explained in Section 4. Section 5 discusses the results. We conclude the paper in Section 6.

2. Related work

Most of the work reported in the literature, deals with either optimal VM placement or minimizing power consumption in a data center. Different techniques have been used to solve the VM placement problem. Work reported in [7,10] focuses on resource management in cloud data centers. The recent focus of researchers is on energy consumption and power management in data centers [11,12,5]. Some of the work discusses techniques to optimize the cost in data center management [13,14]. A few researchers have attempted to reduce the carbon footprint subject to essential QoS constraints [15]. In literature, many authors have tried to balance different single objectives functions to achieve a multi-objective solution [16–18]. In this section, we present a brief survey of different VM placement strategies in data centers with their relative merits and demerits. Bin packing is a well known technique to solve optimization problems. It is used by several authors for VM placement with different heuristics. First Fit Decreasing (FFD) is the most popular among these heuristics [8,9]. However, FFD has its own limitations. In FFD, all bins are in one dimension, where as servers in data centers are multi-dimensional.

Several authors have proposed genetic algorithm based optimization for VM placement in data centers. Mi et al. [11] have proposed an online self-reconfiguration approach for reallocation of VMs in large-scale data centers. Their technique attempts to optimize energy consumption by predicting future workloads. They have assumed, that applications that are deployed on VMs are independent of each other. Authors have not addressed the deployment issues when applications are dependent on each other.

Xu et al. [17] have proposed a two-level control system for mapping workloads to VMs, and VMs to physical resources. Their work aims at minimizing: (i) Total wastage of resources, (ii) Power consumption, and (iii) Thermal dissipation costs. Author's have also proposed a modified genetic algorithm with fuzzy multi-objective approach to evaluate their technique. They have shown that their work is superior to the well-known bin-packing algorithm-, and single-objective approaches.

Hybrid Genetic Algorithm (HGA) is also used for energy efficient VM placement. Tang et al. [5] have proposed a technique to optimize energy consumption in communication networks deployed in data centers. They have compared their proposed HGA with GA, and have shown that HGA is better then the traditional genetic algorithm (GA). However, authors have not addressed how energy consumption will be optimized by HGA in communication networks during VM migration.

Kantarci et al. [12] have proposed a mixed integer linear program (MILP) formulation for large-scale cloud systems where the cloud services are provisioned by many interconnected data centers. Their proposed MILP formulation, aims at virtualizing the backbone topology and placing the VMs on data centers with an objective to minimize power consumption. They have compared their proposed solution with a benchmark MILP model.

Chaisiri et al. [13] have proposed an optimal VM placement (OVMP) algorithm. The objective of their algorithm is to minimize the cost spent on each plan for hosting VMs in a multiple cloud provider environment under future demand and price precariousness.

Speitkamp et al. [14] have described a linear programming formulation for server consolidation problem. They have considered both on-demand payment plan and reservation, based approaches for hosting VMs. Their algorithm determines an optimal allocation which minimizes the cost spent on. The authors have not specified how their model works when workload changes over time.

Hermenier et al. [18] have modelled the VM provisioning and placement problem as two constraint satisfaction problem. They have proposed entropy resource manager, which dynamically consolidates the VMs on the basis of constraint programming and takes into account the problem of VM allocation to available nodes and the migration of VMs to these nodes.

Hamed et al. [15] have developed an optimization based framework, where the objective function ranges from energy minimization to carbon footprint minimization subject to the essential QoS constraints. They have not considered the energy costs during migration and dynamic placement. Also, they have not considered the deployment cost for initial VM placement.

Gao et al. [16] have proposed an ant-colony based optimization that simultaneously operates on total power consumption and resource wastage minimization. They have compared their technique with an existing multi-objective genetic algorithm, and two single-objective algorithms; a well-known bin packing algorithm and a MaxMin ant system (MMAS) algorithm.

3. Maximum VM placement with Minimum Power Consumption (MVMP)

To run a large data center huge amount of power is required. Most of the power in a data center is consumed by the servers and cooling systems [19]. Power consumed by cooling systems is difficult to minimize. However, a proper strategy with an appropriate technique can minimize the energy consumed in data centers by minimizing the number of servers used in a particular time frame. Minimization of the number of active servers also leads to minimization of energy consumption. The objective of a CSP is to maximize his profit.

The total profit earned by a CSP can be expressed as follows.

$$Profit = Revenue - Expenditure \tag{1}$$

CSPs will always try to maximize their profit. The expenses incurred by a CSP in running a data center can be computed as follows [20].

$$Expenditure = Exp_{space} + Exp_{cooling} + Exp_{op} + Exp_{phardware} \tag{2}$$

where, Exp_{space} is the cost of hosting the data center building and its physical infrastructure, Exp_{op} signifies manpower, software licenses, computing equipment depreciation cost etc., $Exp_{cooling}$ is the expense towards cooling systems deployed in a data center, $Exp_{phardware}$ is the