International Journal of Swarm Intelligence Research
Volume 14 • Issue 1

Multi-Objective Binary Whale Optimization-Based Virtual Machine Allocation in Cloud Environments

Ankita Srivastava, Babasaheb Bhimrao Ambedkar University, India
Narander Kumar, Babasaheb Bhimrao Ambedkar University, India*
https://orcid.org/0000-0003-4633-1879

ABSTRACT

With the rising demands for the services provided by cloud computing, virtual machine allocation (VMA) has become a tedious task due to the dynamic nature of the cloud. Millions of virtual machines (VMs) are allocated and de-allocated at every instant, so an efficient VMA has been a significant concern to enhance resource utilization and depreciate its wastage. Encouraged by the prodigious performance of the nature-inspired algorithm, the binary whale optimization approach has been eventuated to get to grips with the VMA issue with the focus on minimizing the resource waste and volume of servers working actively. The deliberate approach's accomplishment is assessed against the literature's well-known algorithms for VMA issues. The comparison results showed that the least resource wastage fitness of 15.68, minimum active servers of 216, and effective CPU and memory utilization of 88.31% and 88.79%, respectively, have been achieved.

KEYWORDS

Binary Whale Optimization Algorithm, Cloud Computing, Metaheuristic Algorithm, Resource Allocation, Resource Wastage, Swarm Optimization, Virtual Machine Allocation, Virtualization

1. INTRODUCTION

The progression in technology has led to the emergence and evolvement of cloud computing from the basic computing paradigm of distributed, parallel, and grid computing. It has revolutionized the procedure for managing the data or information and the resources which have impacted human society socially and economically. The services are offered to the users in the resources form (e.g., storage, CPU, servers, network, and application). These resources are organized at one central point, the data center, from where accessing these resources comes at ease. The overall management of data centers is the responsibility borne by the service providers. The service providers serve the users' interests with three basic services via the internet IaaS for hardware resources, PaaS for runtime environment,

DOI: 10.4018/IJSIR.317111
*Corresponding Author

and SaaS for software resources. These former services are being accomplished with virtualization. Virtualization facilitates the creation and configuration of VMs possessing variant operating systems with varying resources (memory, CPU, and storage) that are then nailed on the host machine to furnish the services desired by users. It also expedites sharing of multiple VMs deployed on one distinct server and sharing the hardware resources. To accomplish the request or interest demanded by users, VMs are created and configured dynamically with variable configuration and resource demands. The key objective in this reference is to allocate resources in ways that they are effectively utilized, decreasing resource waste and resulting in low operational costs. Adopting an effective VMA algorithm is one way to accomplish this. VMA allows VMs to be placed on the servers such that computing resources must be efficiently utilized while also reducing the volume of active servers. With the decisive allocation techniques, there come challenges, including a reduction in QoS, reduction of performance with the curtailed energy usage, and then satisfying the user's QoS within the promised SLA. An effective VMs allocation narrows down the count of active and balanced multidimensional resource usage by servers, ultimately reducing the energy expenditure by the cloud. Figure 1 demonstrates the allocation of VMs in which all the VMs are allotted randomly to the server, while after optimization in figure 2, the VMs are consolidated in the first three servers, and the rest of the unutilized servers are turned off, thus saving the resources and energy. Identifying optimal VMs allocation belongs to the NP-complete problem (Lo, V. M. 1983), and achieving the optimum resolution of this is typically computationally infeasible, when the cloud involves multiple hosts and users (Widmer, T., Premm, M., & Karaenke, P. 2013). The issue can be addressed to minimize resource wastage and the number of active servers in the cloud data center. Various methods have been applied to resolve this issue in the literature. A theorem was given (Wolpert, D. H., & Macready, W. G. 1997), which stated that not a single meta-heuristic algorithm is available which can efficiently resolve all the optimization problems. In this regard, the WOA (Mirjalili & Lewis, 2016) has been successfully utilized for various optimization problems (Prakash, D.B. & Lakshminarayana, C.,2017; Sun, Wang, Chen, & Liu, 2018; Chen, H., Xu, Y., Wang, M., & Zhao, X., 2019; Too, J., Mafarja, M., & Mirjalili, S., 2021).

A meta-heuristic-based Genetic Algorithm (GA) has been used widely to resolve the allocation problem (Kaaouache, M. A., & Bouamama, S.,2018; Abohamama, A. S., & Hamouda, E., 2020). These methods are the victims of basically two issues. These algorithms converge when a large number of iterations are performed and claim for a large population size as opposed to other metaheuristic algorithms. Due to this, it involves a high processing time, and sometimes it also suffers from premature convergence. While other meta-heuristic algorithms like ant colony-based, grey-wolf-based algorithms converge with lower population size. This motivates exploiting different meta-heuristic algorithms that benefit from low population size and iterations or generations. The study introduces a novel state of the art for accomplishing VM allocation, which subsequently truncates the active servers' count and minimizes the undergoing resource wastage, thus enhancing the proportion of resource utilization.

Further, the study is assembled as section 2 background study for the associated work being performed, and section 3 demonstrates the proposed method. The algorithm for the given work is discussed in section 4, which is supported through simulation. The result analysis is done in section 5, and further discussion is performed in section 6, which is finally concluded in section 7.

2. LITERATURE REVIEW

Extensive advances in the cloud have led to a dramatic rise in the volume of data centers. Hence, energy consumption generated by these centers continues to upsurge the cost incurred by the cloud system. So, in this scenario mitigating the energy exhausted by these centers has now become a critical concern that needs urgent attention. Energy consumed through these data centers varies directly with resource usage, thus, to prevent energy consumption, one needs to be careful about resource consummation as it aids in reducing operating costs. The resources should be used proficiently, thus

Gene Expression Dataset Classification Using Artificial Neural Network and Clustering-Based Feature Selection

Audu Musa Mabu, Rajesh Prasadand Raghav Yadav (2020). *International Journal of Swarm Intelligence Research (pp. 65-86).*

www.irma-international.org/article/gene-expression-dataset-classification-using-artificial-neural-network-and-clustering-based-feature-selection/240630