Received: 18 April 2021 | Revised: 21 June 2021 | Accepted: 25 June 2021

DOI: 10.1002/cpe.6496

RESEARCH ARTICLE

WILEY

# A deadline aware load balancing strategy for cloud computing

Raza A. Haidri[1] | Mahfooz Alam[2] | Mohammad Shahid[3] | Shiv Prakash[4] | Mohammad Sajid[2]

[1]Department of Computer Science and Information Technology, Khwaja Moinuddin Chishti Language University, Lucknow, India

[2]Department of Computer Science, Aligarh Muslim University, Aligarh, India

[3]Department of Commerce, Aligarh Muslim University, Aligarh, India

[4]Department of Electronics and Communication, University of Allahabad, Allahabad, India

**Correspondence**
Mahfooz Alam, Department of Computer Science, Aligarh Muslim University, Aligarh, India.
Email: mahfoozalam.amu@gmail.com

## Abstract

The load balancing (LB) may be used at different levels to reduce overhead for the decision-making process. In the past decade, cloud computing has drawn a lot of attention from both the academic and commercial communities to get demanded resources (machines, platforms, data, storage, software, and so forth) as a service on rent economically. Generally, a situation may arise when requests are not meeting their deadlines and the cloud provider wants to finish the running application in minimum time. In this article, a receiver initiated deadline aware LB strategy (RDLBS2) has been proposed which attempts the migration of incoming cloudlets to appropriate virtual machines (VMs) where the deadlines of the cloudlets are met to optimize the turnaround time by exploiting the remaining processing capacities of VMs. A simulation study has been carried out by using Cloud-Sim as a simulator. A sensitivity analysis has been presented to analyze the effects on performance parameters by varying the number of cloudlets and the number of VMs while keeping the remaining input parameters fixed. The experimental evaluation and analysis suggest that RDLBS2 performs significantly better than its peers on objective parameters almost in all cases under study.

**KEYWORDS**

cloud computing, cloudlet migration, deadline meet, load balancing, QoS parameters, receiver initiated strategy, total gain, turnaround time

## 1 | INTRODUCTION

The load balancing (LB) is one of the most dominant and challenging issues in the modern era from the operating system (OS) level to high-performance computing (HPC) level to minimize the management effort in complex systems. But, sometime earlier, many organizations were not able to use HPC due to high cost and nonavailability. It has various real-time applications[1] such as those usually found in the perspective of multimedia, cloud computing (CC), microcontroller and microprocessors used for robotics, automation, real-time databases, and embedded systems attached with various real-time applications. These are categorized with a load which can be measured in million instructions (MI) per second associated with different applications.

The CC environment provides resource sharing and dynamic resource provisioning via virtualization.[2,3] The CC follows the pay per use model, that is, virtual resources can be taken on rent and payment is done as per the consumption.[4–6] There are three delivery models.[4,6–9] The first CC delivery model is infrastructure as a service (IaaS) which reduces infrastructure cost by allowing users to run any applications on CC provider's hardware, that is, contemporary applications can be transferred to/from company data-center. Common examples of IaaS are AWS EC2, Rackspace (hosting private cloud), and GoGrid (hybrid hosting, dedicated hosting). The second CC delivery model is platform as a service (PaaS) which permits users to create their cloud applications using cloud-specific tools and programming languages. The common examples of PaaS are GAE, Windows Azure, and Force.Com. Furthermore, the PaaS supports rapid development at a low cost.[10] The last one is software as a service (SaaS). It is the easiest

  