

Chapter 7

Metaheuristic Approaches to Task Consolidation Problem in the Cloud

Sambit Kumar Mishra

National Institute of Technology Rourkela, India

Kshira Sagar Sahoo

National Institute of Technology Rourkela, India

Bibhudatta Sahoo

National Institute of Technology Rourkela, India

Sanjay Kumar Jena

National Institute of Technology Rourkela, India

ABSTRACT

The service (task) allocation problem in the distributed computing is one form of multidimensional knapsack problem which is one of the best examples of the combinatorial optimization problem. Nature-inspired techniques represent powerful mechanisms for addressing a large number of combinatorial optimization problems. Computation of getting an optimal solution for various industrial and scientific problems is usually intractable. The service request allocation problem in distributed computing belongs to a particular group of problems, i.e., NP-hard problem. The major portion of this chapter constitutes a survey of various mechanisms for service allocation problem with the availability of different cloud computing architecture. Here, there is a brief discussion towards the implementation issues of various metaheuristic techniques like Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), BAT algorithm, etc. with various environments for the service allocation problem in the cloud.

INTRODUCTION

In recent times, a large number of improvements have been done in distributed computing like flexibility, reliability, efficiency, etc. The cloud computing is one of the models which provides verities elastic services by using internet technologies. The term “elastic computing” refers to the ability of dynamically acquiring computing resources and supporting a variable workload. The cloud applications are normally based on the client-server model, and the applications are running on the local machines while the executions are conducted on the cloud. There are three actors (CSU, CSP, and CSN) in the cloud environment.

DOI: 10.4018/978-1-5225-1721-4.ch007

Metaheuristic Approaches to Task Consolidation Problem in the Cloud

The cloud service user (CSU) is a person or an organization that consumes delivered cloud services. The cloud service provide (CSP) is an organization that provides and maintains delivered cloud services. The cloud service partner (CSN) is a person or an organization that provides support to the building of the service offer by a CSP. The cloud user has to request the cloud service provider for the execution of their tasks, and an agreement has to make between them called service level agreement (SLA). According to the SLA, the CSP provides services to the user. Otherwise, the CSP may give a penalty for the violence of SLA. A cloud service provider maintains a massive infrastructure to support elastic services. Cloud computing is cost-effective due to resource multiplexing and one of the essential terminologies of the cloud computing is virtualization. Virtualization is the creation of a virtual component of different resources of physical computing devices or a single physical machine can run multiple operating systems concurrently, each in its own virtual machine.

The National Institute of Standards and Technology (NIST) has defined the cloud computing system as a model that provides on-demand services from a shared pool of computing resources like storage, servers, network, services, etc. and these services are provided to the user in a faster mode as well as stop the services rapidly (Mell & Grance, 2011). According to NIST definition, the cloud system model has five essential characteristics, four deployment models, and three service models as shown in Figure 1. The difference between the four deployment models is the comparison of scalability, reliability, security, and cost.

Essential Characteristics of Cloud Computing

1. **On-Demand Self-Service:** The cloud resources can be registered by the consumer for their usage without the necessity of human interaction with each CSP. Here, the computing resources include processing power, storage, virtual machines, etc.
2. **Broad Network Access:** Capabilities are accessible to the network system which can access in different ways that advance the system utilization in heterogeneous gadgets, as, for example, electronic devices like laptops, mobiles, and telephones.
3. **Resource Pooling:** The computing resources provided by the CSP serve numerous clients using a multi-tenant model, with various resources assigned and reassigned dynamically as per customer interest. Because of independent location feeling, the client has no knowledge of the exact location of different virtual machines.
4. **Rapid Elasticity:** Availabilities of services can be dynamically provisioned and released automatically by the scale out and scale in by discharging required resources on demand.
5. **Measured Service:** Utilization of cloud resources measured by monitoring CPU time, storage usage, bandwidth usage, etc. Different cloud service provider provides services to the users at a different level of abstraction.

Deployment Models of Cloud Computing

1. **Private Cloud:** This type of cloud infrastructure exclusively used by one organization or company or a group of clients. The name of this kind of infrastructure tells about the higher level of security, and it results in the increase in cost. This type of cloud may be owned, self-managed, and self-operated by any organization. The StACC (Collaborative Research in Cloud Computing) set

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/metaheuristic-approaches-to-task-consolidation-problem-in-the-cloud/171352

Related Content

Cloud-IoT-Based Healthcare Applications: Case Study I-Seizure Prediction

M. Abinaya (2025). *Revolutionizing Healthcare Systems Through Cloud Computing and IoT* (pp. 121-150).
www.irma-international.org/chapter/cloud-iot-based-healthcare-applications/359850

Performance Evaluation of Hypervisors for Cloud Computing

Sanjay P. Ahuja and Suganya Sridharan (2012). *International Journal of Cloud Applications and Computing* (pp. 26-67).
www.irma-international.org/article/performance-evaluation-hypervisors-cloud-computing/72326

A Load Balancing Algorithm for Enhanced Active Monitoring in a Cloud Environment

Sudhir Kumar Mohapatra, Tegegn Dita, Mesfin Abebe, Jasobanta Laha and Biswajit Tripathy (2023). *Privacy Preservation and Secured Data Storage in Cloud Computing* (pp. 244-255).
www.irma-international.org/chapter/a-load-balancing-algorithm-for-enhanced-active-monitoring-in-a-cloud-environment/333141

IPCRESS: Tracking Intellectual Property through Supply Chains in Clouds

Lee Gillam, Scott Notley, Simon Broome and Debbie Garside (2015). *Enterprise Management Strategies in the Era of Cloud Computing* (pp. 171-191).
www.irma-international.org/chapter/ipcress/129744

Integration of the Internet of Things and Cloud: Security Challenges and Solutions – A Review

Chellammal Surianarayanan and Pethuru Raj Chelliah (2023). *International Journal of Cloud Applications and Computing* (pp. 1-30).
www.irma-international.org/article/integration-of-the-internet-of-things-and-cloud/325624