A Novel Approach of Load Balancing and Task Scheduling Using Ant Colony Optimization Algorithm

Selvakumar A., Manonmaniam Sundaranar University, Tirunelveli, India
Gunasekaran G., J.N.N Institute of Engineering, Chennai, India

ABSTRACT

Cloud computing is a model for conveying data innovation benefits in which assets are recovered from the web through online devices and applications, instead of an immediate association with a server. Clients can set up and boot the required assets and they need to pay just for the required assets. Subsequently, later on giving a component to a productive asset administration and the task will be a vital target of Cloud computing. Load balancing is one of the major concerns in cloud computing, and the main purpose of it is to satisfy the requirements of users by distributing the load evenly among all servers in the cloud to maximize the utilization of resources, to increase throughput, provide good response time and to reduce energy consumption. To optimize resource allocation and ensure the quality of service, this article proposes a novel approach for load-balancing based on the enhanced ant colony optimization.

KEYWORDS
Ant Colony Algorithm, Cloud Computing, Load Balancing, Optimization, Virtual Machine

1. INTRODUCTION

Cloud computing refers to the delivery of computing and storage capacity as a service to a heterogeneous community of end-recipients. Cloud computing is an internet technology that utilizes both central remote servers and internet to manage the data and applications. This technology allows many businesses and users to use the data and application without an installation. Users and businesses can access the information and files at any computer system having an internet connection (Soundararajan et al., 2009). Cloud computing provides much more effective computing by centralized memory, processing, storage and bandwidth.

The cloud service provider (CSP) provides on demand provisioning of Hardware like processing power, I/O, large amounts of storage, etc. By utilizing virtualization, each user accesses the services of cloud through a virtual machine, where number of virtual machines shares a single physical server. Cloud computing provides a service-oriented platform for cloud users (Sharma, Sharma, & Sharma, 2012). As the cloud is dynamic, where number of tasks are to be scheduled dynamically

DOI: 10.4018/IJSI.2019040102
by the available virtual machines. The conventional algorithms like FIFO, Round Robin may not distribute the load efficiently. The load balancing algorithm must be dynamic i.e. the load on each resource must be updated periodically and check the availability of the resources whenever there is a need to allocate a resource to the task (Katyal & Mishra, 2013). Cloud is heterogeneous in nature; servers of different capacity may be available at data center. The following challenges need to be considered for efficient load balancing in the cloud.

The distributed dynamic load balancing algorithm is executed by all nodes present in the system and the task of load balancing is shared among them. The interaction among nodes to achieve load balancing can take two forms: cooperative and non-cooperative (Vouk, 2008). In the first one, the nodes work side-by-side to achieve a common objective, for example, to improve the overall response time, etc. In the second form, each node works independently toward a goal local to it, for example, to improve the response time of a local task. Dynamic load balancing algorithms of a distributed nature usually generate more messages than the non-distributed ones because, each of the nodes in the system needs to interact with every other node.

A benefit, of this is that even if one or more nodes in the system fail, it will not cause the total load balancing process to halt; it instead would affect the system performance to some extent. Distributed dynamic load balancing can introduce immense stress on a system in which each node needs to interchange status information with every other node in the system. It is more advantageous when most of the nodes act individually with very few interactions with others.

In non-distributed type, either one node or a group of nodes do the task of load balancing. Non-distributed dynamic load balancing algorithms can take two forms: centralized and semi distributed. In the first form, the load balancing algorithm is executed only by a single node in the whole system: the central node. This node is solely responsible for load balancing of the whole system. The other nodes interact only with the central node. In semi-distributed form, nodes of the system are partitioned into clusters, where the load balancing in each cluster is of centralized form (Alakeel, 2010). A central node is elected in each cluster by appropriate election technique which takes care of load balancing within that cluster. Centralized dynamic load balancing takes fewer messages to reach a decision, as the number of overall interactions in the system decreases drastically as compared to the semi distributed case.

However, centralized algorithms can cause a bottleneck in the system at the central node and also the load balancing process is rendered useless once the central node crashes. Therefore, this algorithm is most suited for networks with small size.

Figure 1. Load balancing in cloud computing
A Novel Key Management Scheme for Next Generation Internet: An Attack Resistant and Scalable Approach  
*International Journal of Information System Modeling and Design* (pp. 92-121).  
[www.igi-global.com/article/a-novel-key-management-scheme-for-next-generation-internet-an-attack-resistant-and-scalable-approach/208641?camid=4v1a](www.igi-global.com/article/a-novel-key-management-scheme-for-next-generation-internet-an-attack-resistant-and-scalable-approach/208641?camid=4v1a)

Mobile Agents for a Mobile World  
[www.igi-global.com/chapter/mobile-agents-for-a-mobile-world/117944?camid=4v1a](www.igi-global.com/chapter/mobile-agents-for-a-mobile-world/117944?camid=4v1a)