Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 832 – 841

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Autonomous Agent Based Load Balancing Algorithm in Cloud Computing

Aarti Singh[a], Dimple Juneja[b], Manisha Malhotra[a]*

[a]*Maharishi Markandeshwar University, Mullana, 133205, India*
[b]*Dronacharya Institute of Management and Technology, Kurukshetra, 136119, India*

Abstract

Cloud Computing revolves around internet based acquisition and release of resources from a data center. Being internet based dynamic computing; cloud computing also may suffer from overloading of requests. Load balancing is an important aspect which concerns with distribution of resources in such a manner that no overloading occurs at any machine and resources are optimally utilized. However this aspect of cloud computing has not been paid much attention yet. Although load balancing is being considered as an important aspect for other allied internet based computing environments such as distributed computing, parallel computing etc. Many algorithms had been proposed for finding the solution of load balancing problem in these fields. But very few algorithms are proposed for cloud computing environment. Since cloud computing is significantly different from these other types of environments, separate load balancing algorithm need to be proposed to cater its requirements. This work proposes an Autonomous Agent Based Load Balancing Algorithm (A2LB) which provides dynamic load balancing for cloud environment. The proposed mechanism has been implemented and found to provide satisfactory results.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Agents; Cloud Computing; Dynamic Load Balancing.

1. Introduction

* Corresponding author. Tel.: +91-8427465352
E-mail address: mmanishamalhotra@gmail.com

1877-0509 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.168

The cloud computing is a distributed internet based paradigm, designed for remote sharing and usage of different resources and services like storage, computational capabilities and applications etc. with high reliability over the large networks. However, due to dynamic incoming requests, dynamic resource allocation is required in it. This inherent dynamism in cloud computing requires efficient load balancing mechanisms. Load balancing concerns distribution of resources among the users or requests in uniform manner so that no node is overloaded or sitting idle. Like in, all other internet based distributed computing tasks, load balancing is an important aspect in cloud computing. In the absence of load balancing provision, efficiency of some overloaded nodes can sharply degrade at times, leading to violation of SLA. In traditional distributed computing, parallel computing and grid computing environments load balancing algorithms are categorized as static, dynamic or mixed scheduling algorithms based on their nature [6] where:

a)Static Load Balancing Algorithm is suitable for small distributed environments with high Internet speed and ignorable communication delays.

b)Dynamic Load Balancing Algorithm focuses on reducing communication delays and execution time and thus are suitable for large distributed environments.

c)Mixed Load Balancing Algorithm focuses on symmetrical distribution of assigned computing task and reducing communication cost of distributed computing nodes.

Based on above categorization, cloud computing clearly falls under the second category. It means balancing load in cloud computing environment requires focusing on dynamic load balancing algorithms. In traditional distributed environments process migration is less expensive due to small process granularity whereas in CC environment, process migration is expensive due to high granularity of data involved. Thus cloud computing environment requires a load balancing algorithm which could cater to dynamic service demands of users while providing optimized load balancing. Following parameters are available in literature for measuring efficiency of a load balancing algorithm in CC environment [4]:

i. *Reliability*: The algorithm must be reliable, since process failure while transferring job from one location to other may lead to increased waiting time and customer dissatisfaction.

ii. *Adaptability:* Algorithm must be capable of adapting the dynamically changing user requests and provide task allocation in minimal amount of time.

iii. *Fault Tolerance:* The algorithm must ensure fault tolerance, so that in case of a problem in the system complete load balancing mechanism does not stop working.

iv. *Throughput:* the algorithm must ensure increased throughput at minimal expense. If a load balancing algorithm doesn't increase system throughput, it defeats its own purpose.

v. *Waiting Time:* Algorithm should minimize wait time of a task for allocation of resources to it.

Next subsection elaborates major components of a dynamic load balancing algorithm.

1.1Components of Dynamic Load Balancing Algorithms

Literature review highlighted that a load balancing algorithm has five major components [6] as discussed below:

Table1: Components of Load Balancing

Sr. No.	Policy	Function
1	**Transfer Policy**	This policy is responsible to determine when a task should be transferred from one node to the other node.
2	**Selection Policy**	It focuses on selecting the processor for load transfer so that the overall response time may be improved.
3	**Location Policy**	It determines the availability of required resources for providing services and makes a selection based on location of resources.
4	**Information Policy**	This policy acquires workload related information about the system such as nature of workload and average load on each node. It is also responsible for exchanging the information from one node to another, along with method of exchange and amount of the information to be exchanged. For exchanging load information of a node one of the following three methods may be adopted: ***i. Broadcast Approach:*** If it is assumed a broadcast communication medium is available, then a load exchange is done whenever the load node changes. ***ii. Global System Load:*** Whenever a node does not acknowledges the revert from another node in a complementary stage, it presumes that all nodes are overloaded.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات