

# Cloud Computing - The Effect of Generalized Spring Tensor Algorithm on Load Balancing

Shahrzad Aslazzandeh

University of Technology Sydney  
Center of Real-Time Information  
Network (CRIN)  
Australia, Sydney

[Shahrzad.Aslanzadeh@uts.edu.au](mailto:Shahrzad.Aslanzadeh@uts.edu.au)

Zenon Chaczko

University of Technology Sydney  
Center of Real-Time Information  
Network (CRIN)  
Australia, Sydney

[Zenon.Chaczko@uts.edu.au](mailto:Zenon.Chaczko@uts.edu.au)

Christopher Chiu

University of Technology Sydney  
Center of Real-Time Information  
Network (CRIN)  
Australia, Sydney

[Christopher.Chiu@uts.edu.au](mailto:Christopher.Chiu@uts.edu.au)

**Abstract**—In business world, competitors use innovative approaches to improve their performance and profits. Cloud computing is one of these creative concepts that allowed companies to further taking advantage of their potential.

Cloud computing is assisting companies to execute their business plans more efficiently. As cloud computing has multi-tenancy structure, availability and efficiency of the resources is essential foundation of the cloud architecture. Recent studies showed that, optimized cloud computing could be seen as an elastic network of resources that are interacting with each other, to minimize the waiting time and utilize the throughput. Therefore load balancing and resource management can be highlighted as the main concerns in cloud computing as they are impacting the network performance directly.

This research aims to discuss the current challenges existing in load balancing algorithms. Different metrics and policies of the relevant load balancer algorithms have been investigated and as a result, collective behavior has been proposed as a new policy for classification of elasticity mechanism in load balancing.

**Keywords**—Cloud Computing, Elasticity; Algorithm policies, Collective behavior

## I. INTRODUCTION

In cloud computing environment, tasks can be scheduled from anywhere at any time. Therefore having an optimal resource allocation policy to secure the resource utilization is essential [1].

Load balancing is one of those mentioned policies that can optimize the network performance by maximizing the throughput of the system and minimizing the response time [2]. This research offers a comprehensive study of available load balancing algorithms with their benefits and challenges. Moreover different policies for categorizing the behavior of the load balancing algorithms, from elasticity perspective, have been investigated and hence a new approach for architecting an optimized algorithm is suggested.

## II. ELASTIC CLOUD COMPUTING

### A. Necessity of Elasticity in cloud computing

In cloud computing, elasticity is focusing on identifying the optimized controlling manner of the resources in terms of dealing with future changes [3]. In some research studies, elasticity and scalability are used interchangeably.

However they are acting differently in cloud computing. The former is referring to competency of the cloud system to grow base on requirements, while the latter presents the capability of the system to allow customers request and release the resources on demands [4-5].

Elasticity improves the resource management policies which results in better resource provisioning and robust power consumption management. Amazon EC2, is one of the successful web services in cloud that set up elasticity as the foundation of their cloud structure [6-7].

### B. Related works: load balancing in cloud computing

Cloud computing brings the opportunity of handling the shared resources, which reduces the associated expenses and enhances the quality of services.

Based on scalability characteristic of cloud computing, businesses can easily add more nodes to their networks [8]. Brynjolfsson, et al [9], however; argued that although scalability enables the enterprises to add more resources to fulfill their requirements, it does not guarantee that adding or removing extra resources will result in better services qualities.

Chunlan et al[10] uses the water and electricity resources of a country as an example to elaborate the "availability" concept of cloud computing. Occasional outages can result in interruption of supplied water or electricity to the residents. Likewise, resources provided by cloud computing, cannot be completely relied upon, as technical issues may cause unavailability of such resources.

In 2009, due to some technical issues, Amazon lost the availability of its resources for about 6 hours, which caused the customers to suffer from that issue even after two days [7].

Furthermore, problems with networks and Internet connections can also affect resource availability in cloud, which results in losing accessibility to the information. The solutions given by different authors are mainly focus on optimization of load balancing techniques that enable the businesses, increase the percentage of their resource availability and guarantee on-time service delivery to their clients. [11-12].

### C. Load balancing metrics

Research studies present different metrics of designing an optimized load balancing algorithm. Below is describing these metrics in details [12-13]:

- *Network's Speed*: Based on bandwidth, network's speed helps the algorithm to distribute the load more efficiently.
- *Complexity*: Complex load balancing algorithm will cause performance issues. Although complex algorithm has better management over the network, it causes delays on the system. Thus it will have a negative impact on performance.
- *Fault tolerance*: The algorithm should be able to recover from errors more quickly. The structure of the network is playing an important role in terms of defining fault tolerance threshold.
- *Scalability*: The algorithm should be able to scale up or scale down base on the new requirements that will be added to the system.
- *Response time*: The gap between receiving the load, and load distribution, is called response time or waiting time. This attribute has direct impact on performance of the system.
- *Elasticity*: As it was described before, elasticity can optimize the behavior of the algorithm in terms of resource provisioning.

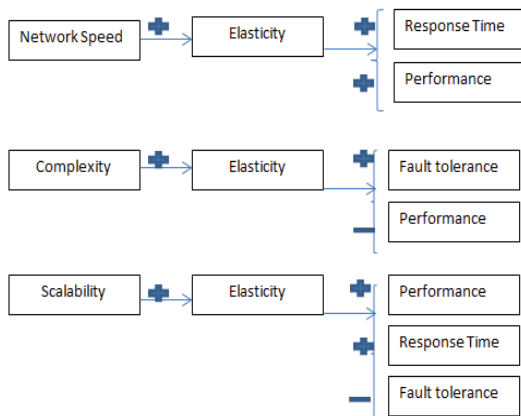


Figure 1- Relationship between Load balancing metrics and elasticity

Considering the above mentioned attributes, it is clear that there is direct relationship between load balancing metrics and performance of the system which is highlighted in "Fig.1".

The combination of the network speed and elasticity can have a positive impact on response time and performance which is depicted with "plus" sign. Increasing the complexity will amplify the fault tolerance. Scalability will optimize the performance and response time, but it has negative impact on fault tolerance that is reflected with "minus" flag.

Applying these metrics, the expected load balancing algorithm should minimize the energy consumption and should enhance the fault tolerance and resource provisioning ability [15].

### III. COMPARISON BETWEEN LOAD BALANCER PROVIDERS AND LOAD BALANCING ALGORITHM

In order to identify the gaps between different load balancing techniques, a comprehensive survey has been done to compare the popular load balancing algorithms. Based on load balancing characteristics, the algorithms are grouped in two categories:

#### A. Static Algorithm

- Round Robin algorithm is assigning the first received task to the first processor and second task to the second processor and so on. The problem with this algorithm is that, the load will not be distributed fairly as the timing of each task is the key value of the resource allocation in Round Robin algorithm [14].

- Randomized algorithm operates based on the probability of accessing the resources in the system. The location of the nodes doesn't have any impact on functionality of this algorithm. However; complex computational process of this algorithm is considered as the negative behavior of the randomized algorithm [12-13].

- Map Reduce algorithm is based on two main goals, mapping the tasks and reducing the load. At first stage, tasks sit in groups, using hash function, then each group assigns to the relevant resources and the load will be reduced. Overloading is the main issue in this algorithm. As this algorithm has the ability to map and reduce the load in parallel structure, therefore tasks overloading may occur in the system [12].

#### B. Dynamic Algorithm

- Token Routing: In this algorithm a token is moving between different nodes in different timing slots. The problem with this algorithm is that, the bandwidth and the essential requirements for execution of the tasks will not be analyzed. Therefore the tasks might be allocated to a node without initial requirements for execution [15].

- Central load balancing decision model: This algorithm is based on centralized model. There is central load manager which manages the tasks based on their priorities in the queue. The main challenge in this algorithm is related to the structure of the algorithm. If the central processor fails, then the whole system will fail as well [13].

- Weighted least connection: The key feature of this algorithm is the number of the connections to each server. Tasks will be assigned to the server with the least connections.

- Biased random sampling: This algorithm works well in scalable environments. It usually uses for self-regulating the network load. However; if the populations of the nodes increase the performance will be affected [14].

As described above static algorithms are focusing on rule based structure and predefined thresholds. Recent

studies showed that cloud networks are evolving in size and complexity. Consequently, network load will increase and will revoke the rule based conditions for balancing the load, throughout the whole network [16]. Therefore dynamic load balancing algorithms are playing critical role in terms of load managements in cloud networks.

#### IV. NEW ELASTIC MECHANISM POLICY

Galante [17] categories the elasticity policies into two groups:

1. Manual elasticity is referring to designing specific application interface (APIs), to enable the users request and release resources by themselves. This approach will help the customers to expand their networks easier and manage the resources based on their needs. However, dealing with the interface that is provided by the cloud providers is the major barrier in this matter.

2. Automatic elasticity: This method is structured based on two models: Reactive and predictive approach. Reactive model is focusing on rule based commands and system's behavior is dependent on rules and thresholds. In predictive approach, based on mathematical analysis, the system will forecast a suitable behavior to scale up/down automatically.

Although the above mentioned policies have been applied in different elasticity algorithms, still there is a need for more efficient policy to manage the usage of the resources in elastic network. The algorithm should be able to manage relational components and recognize anisotropic movement of the resource.

In this study, we explored another category of elasticity policy that is called "Collective behavior" methodology. This approach proves that the utilized resource provisioning is dependent on collection of fluctuation and magnitude motion of the nodes [18-19].

Considering the collective behavior policy, "generalized spring Tensor", STeM, can be the target algorithm for load balancing in cloud computing.

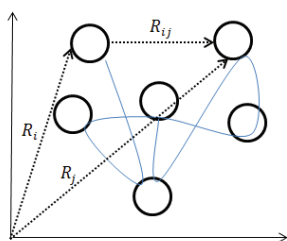


Figure 2- Modeling of the proteins in elastic network

STeM algorithm predicts the anisotropic movement of the load in the cloud network by using Force constant Hessian matrixes.

STeM algorithm has been applied in biological science field, describing the interaction between different proteins. This model simplified the structure of the

biological macromolecules by presenting them in an elastic mass spring network.

"Fig. 2" is simulating an elastic network of the nodes in a cloud network.  $i$  and  $j$  are the sample nodes with the position of  $R_i^0$  and  $R_j^0$  with the distance of  $R_{ij}^0$  [20].

Combining STeM algorithm with collective behavior policy of elasticity the network will remain stable as the fluctuation and magnitude of the upcoming load is predictable. Therefore, visualizing the optimized pattern of nodes collective behavior will be the focus of the future work of this research.

#### V. CONCLUSION AND FUTURE WORK

This research studied the different static and dynamic algorithms to highlight the existing challenges in load balancing.

Different load balancing criteria have been investigated and amongst all of those mentioned principles, elasticity highlighted as the most important specification in load balancing. Also collective behavior methodology has been proposed as the new categorization of elasticity. Applying collective behavior policy on cloud computing, we found that generalized spring tensor algorithm can be used as a novel load balancing algorithm which is cable of analyzing the fluctuation and magnitude of load in cloud. As a future work, the proposed idea will be implemented by using Matlab and java code to visualize the behavior of the load in terms of magnitude and direction on cloud network.

#### REFERENCES

- [1] Creeger, M. 2009, CTO Roundtable: Cloud Computing, Communications of the ACM, Vol. 52, No. 8, pp. 50-56.
- [2] Gupta, A. Cloud computing growing interest and related concerns. Computer Technology and Development (ICCTD), 2010 2nd International Conference on, 2-4 Nov. 2010. pp. 462-465.
- [3] Rimal, B., E. Choi, et al. 2010, A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems. Cloud Computing. N. Antonopoulos and L. Gillam, Springer London: 21-46.
- [4] Chaczko, Z. and S. Aslanzadeh 2011, C2EN: Anisotropic Model of Cloud Computing. 21st International Conference on Systems Engineering (ICSEng).
- [5] L. Bager, R.patt-Corner, and J. Voas, 2010, "Draft could computing synopsis and recommendations recommendations of the national insitute of standards and technology" Nist special publication, vol.146.
- [6] Nuaimi, K. A., N. Mohamed, et al. 2012, A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms. Network Cloud Computing and Applications (NCCA), Second Symposium.
- [7] Amazon Web Services, 2013, Amazon Elastic Compute Cloud (AmazonEC2)
- [8] Amrani, C. E., K. B. Filali, et al. 2012, A Compative Study of Cloud Computing Middleware. Cluster, 12th IEEE/ACM International Symposium on Cloud and Grid Computing (CCGrid)
- [9] Brynjolfsson, E., Hofmann, P. & Jordan, J. 2010. Cloud computing and electricity: beyond the utility model. Commun. ACM, vol.53, pp.32-34.
- [10] Chunlan, L. & Zhonghua, D. Value of Cloud Computing by the View of Information Resources. Network Computing and Information Security (NCIS), 2011 International Conference on, 14-15 May 2011. pp. 108-112.

- [11] Takabi, H., Joshi, J. B. D. & Ahn, G. 2010. Security and Privacy Challenges in Cloud Computing Environments. *Security & Privacy, IEEE*, 8, pp. 24-31.
- [12] Khiyaita, A., M. Zbakh, et al. 2012, Load balancing cloud computing: State of art. *National Days of Network Security and Systems (JNS2)*.
- [13] Ray, Soumya; De Sarkar, Ajanta *International Journal on Cloud Computing: Services & Architectur*; Oct2012, Vol. 2 Issue 5, p1
- [14] Paulin A Florence and V Shanthi. Article: Intelligent Dynamic Load Balancing Approach for Computational Cloud. *IJCA Proceedings on National Conference on Recent Trends in Computer Applications NCRCA:15-18*, October 2013. Published by Foundation of Computer Science, New York, USA
- [15] Nigni Jain Kansal, Inderveer Chana. Existing load balancing techniques in cloud computing: A systematic review. *Journal of Information Systems and Communication*, Vol. 3, 2012, Issue 1, 87-91
- [16] Prakash Kumar, Pradeep Kumar and Vikas Kumar. Article: Load Balancing Strategy for Computational Grid System using an Improved Algorithmic Approach. *International Journal of Computer Applications* 79(8):46-50, October 2013. Published by Foundation of Computer Science, New York, USA
- [17] Galante, G. and L. C. E. de Bona 2012, A Survey on Cloud Computing Elasticity, *IEEE Fifth International Conference on Utility and Cloud Computing (UCC)*.
- [18] Neeraj Pandey, Shashi Kant Verma and Vivek Kumar Tamta. Article: Load Balancing Approach for Scheduling Sequential Task in Grid Computing Environment. *International Journal of Computer Applications* 78(1):42-48, September 2013. Published by Foundation of Computer Science, New York, USA
- [19] Poornima Bhardwaj and Vinod Kumar. Article: An Effective Load Balancing Task Allocation Algorithm using Task Clustering. *International Journal of Computer Applications* 77(7):32-39, September 2013. Published by Foundation of Computer Science, New York, USA
- [20] I. Bahar, A. R. Atilgan, and B. Erman ,Direct evaluation of thermal fluctuations in protein using a single parameter harmonic potential, *Folding & Design* 2, 173-181, 1997.