

Load Balancing System for Multiple distributed computing service providers

Anit Kumar^a, Dr. Priyanka Chawla^b

^aResearch Scholar, Lovely Professional University, Phagwara- 144411, India

^bProfessor, Lovely Professional University, Phagwara- 144411, India

Abstract: Distributed computing is a term often used in the computer engineering and must be involved in any application when it comes to small to medium or large scale computation and supporting a minimum set to comparatively larger set of features. With distributed computing, an application is able to communicate on layers of different technologies. Users might not be restricted to a particular geographical region and users of the applications of different technologies are able to communicate with each other through a centralized distributed computing server. Since the inception of distributed computing in the center of many industrial applications, the number of the users increases day by day because it provides the mean of providing a platform of consuming a wide of variety of distributed services to exhibit accordingly variety of services to the end users. By the time to handle the traffic of enormous number of users on a single platform, there need a separate layer of load balancing system integrated on the application platform of supporting different distributed computing services. In this paper, we have represented an overview on the need and architectural construction of such a load balancing system.

Keywords – Distributed systems, Load balancing, Synchronization, Scheduling

1. Introduction

The Distributed computing is a very common term in the field of computer science and also a very general computing paradigm which is being used by almost all the scientific and industrial computing applications. All the information technology applications have stand on the base of distributed computing. It came on demand day by day with the evolution high speed internet. Computing services are generally deployed on a centralized server somewhere at a particular geographical region and its consumers are generally spread over the world wide. It makes possible only through establishment a communication channel between a consumer and the server through underling computer networks. It also facilitates the communication between an user to another user. Now a days, all mobile, radio and satellite [1] communication is using the concept of distributed computing to provide the services to its users which are globally spread. Cloud computing is also a very common example in the field of distributed computing and now a days used by many organizations to provide the services to its customers. Since the needs of customers increases day by day and they want as much as services on a single platform, so there comes up of accessing many distributed services on a single platform over some network. As soon as this kind of computer applications started to come into the commercial market, the number of its users started to increase exponentially over the time. By the time it became very difficult to handle and respond to its various users at the same. There is a very risk for the organizations of losing their customers, if they are unable to provide services based with a minimal possible response time. If an application has comparatively less [2] [3] feature but its response is quick, then it became a widely used product in the market. Hence almost all the organization invests a large amount of money yearly on the maintenance of its computer applications to boost its performance and response. Even though there are still many challenges in this field, since organization has a limited computing and networking resources to provide their services to the users are in millions in number. All these are not necessarily be in same geographical region and may belong to different geo- locations. Therefore we want to pursue our research on designing a load balancing system which can integrated with such an application which acts as a single platform for its users and consumes many distributed computing services to provide a variety of services to the users.

2. Resource Scheduling and load balancing system

Resource scheduling and load balancing is an one of the essential component [6][10] of every application which is comprising the different computing services and running on a single application platform, since it is giving the services to the enormous number of users which are having many real life needs and could be fulfilled by information technology.

we suggest to follow [11] [12] the below steps for resource scheduling and load balancing:

1. Users should be able to place a request for a service on the application platform
2. Users request should pass through the load balancing system which should be loosely coupled with the application platform
3. All distributed computing services should be loosely coupled with the load balancer and the common interface with the application platform
4. Only authenticate users requests should be processed by the load balancer otherwise throw unauthenticated user error
5. Synchronization and mutual exclusion of user requests over network resources must be implemented otherwise overlapping of user requests will result into invalid results and cause severe loss
6. As soon as an underlying distributed system fetches the response from its own abstract layer it should pass to the end user through the application platform load balancer
7. Load balancer should make a request pool to send again the same user request for a service and for it the user need not to make the same request again and again at their end

3. Scheduling classification

Followings are the two defined network resource scheduling methods:

1. **Static scheduling:** When the initial resource configuration and network topologies are known before initiating a user request and pre-defined parameters and metrics are used to implement the underlying computational system.
2. **Dynamic scheduling:** When the initial resource configuration and network topologies are unknown and dynamic variables and metrics are used in the algorithm to implement the underlying computational system.

4. Challenges and Issues in Load Balancing system

We have a wide number of computing applications are available and providing the same set of facilities to the end users. But few of them are able to fulfill the needs of customers [4][7] which want their request accomplishment in a minimum possible duration. Need not to add there is always a security concern of accessing a service over a network where each resource on the network must be synchronized [5][8][9] among its users and only authenticate users can access the services. For a load balancing schemes over a network for an application has many challenges and issues starting from its implementation to providing the services to an end user. Failing to do so over by any organization will lead to huge reputational and financial loss.

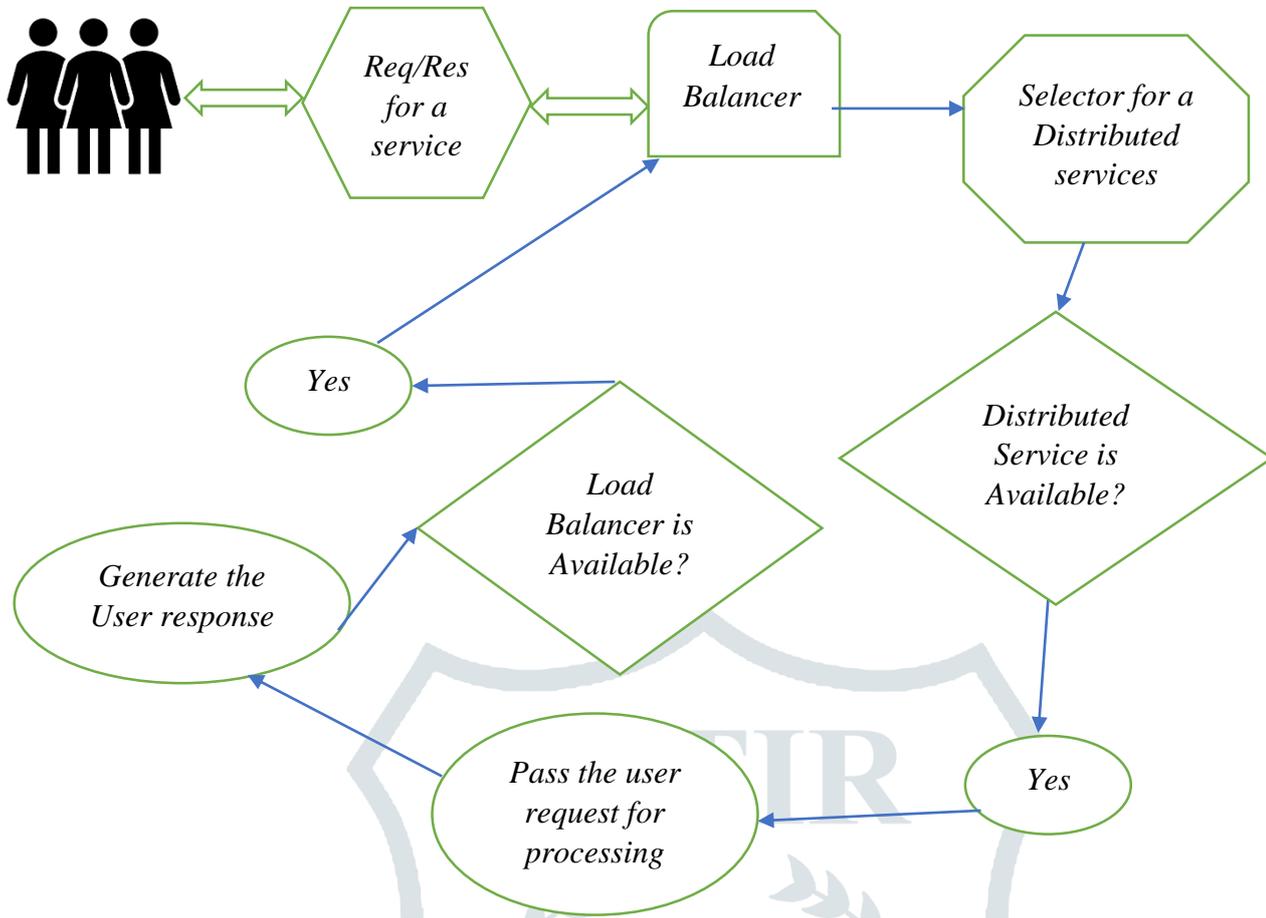


Figure- 1: Work-Flow diagram of a load balancing system

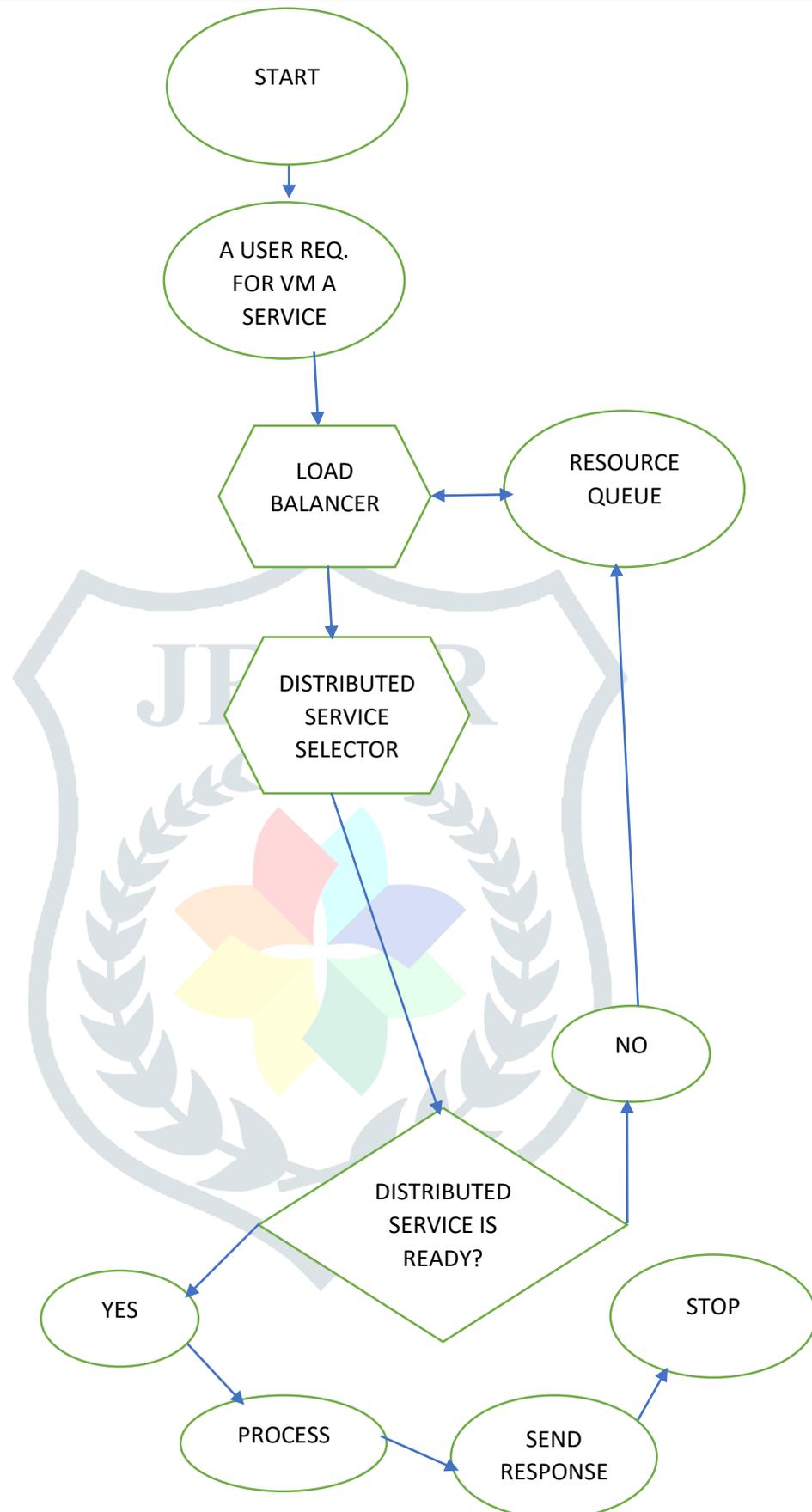


Figure- 2: Flow-Chart of a load balancing system

5. Conclusion

In this paper we have worked to give an overall need and architecture of a common load balancing system which may be the use by many applications consuming different distributed computing over a single platform. As for example many real world applications like PayTM, MakeMyTrip, Amazon and Flipkart etc are the general common example in real life having the same architectural behavior and layers uses a separate layer of load balancing system to give the services to their end users. These organizations spend large amount of moneys

just on maintenance and improvement of their load balancing system. But still in order to compete with their commercial counterpart organization, all these are going deep and deep to bring better system than others

REFERENCE

- [2] Tomoya Enokido, Ailixier Aikebaier, and Makoto Takizawa, Member, “Process Allocation Algorithms for Saving Power Consumption in Peer-to-Peer Systems”, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 58, NO. 6, JUNE 2011
- [3] Tao Hu, Julong Lan, Jianhui Zhang and Wei Zhao, “EASM: Efficiency-aware switch migration for balancing controller loads in software-defined networking”, Peer-to-Peer Networking and Applications, 2018
- [4] Tordsson J, Montero RS, Moreno-Vozmediano R, Llorente IM, “Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers”, Future Generation Computer Systems 2012; 28(2):358–367
- [5] Yu-Kwong Kwok and Lap-Sun Cheung, “A new fuzzy-decision based load balancing system for distributed object computing”, ELSEVIER- Journal of parallel and distributed computing, 2003.
- [6] Yi Lua, Qiaomin Xie, Gabriel Kliot, Alan Geller, James R. Larus and, Albert Greenberg, “Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services”, Performance Evaluation 68 (2011) 1056–1071
- [7] Deepak Nadig, Byrav Ramamurthy, Brian Bockelman and David Swanson, “APRIL: An Application-Aware, Predictive and Intelligent Load Balancing Solution for Data-Intensive Science”, National Science Foundation, 2011
- [8] Deepak Puthal, Rajiv Ranjan, Ashish Nanda, Priyadarsi Nanda, Prem Prakash Jayaraman, and Albert Y. Zomaya, "Secure Authentication and Load Balancing of Distributed Edge Datacenters", Secure authentication and load balancing of distributed edge datacenters, J. Parallel Distrib. Comput., 2018
- [9] Badr Eddine Sabira, Mohamed Youssfi, Omar Bouattane and Hakim Allali, “Authentication and load balancing scheme based on JSON Token For Multi-Agent Systems”, Second International Conference on Intelligent Computing in Data Sciences (ICDS 2018)
- [10] Varun Kumar Sharma and Mahesh Kumar, “Adaptive load distribution approach based on congestion control scheme in ad-hoc networks”, International Journal of Electronics, 2018