Project Proposal

Survey on Load Balancing in Distributed Systems.

CS550 – Advanced Operating Systems

September 10th 2020

Amit Nikam

1. Introduction

Distributed systems make it possible to share and make use of different resources such as computers, storage systems and other specialized devices. These resources are distributed and often are owned by different organization, agents or individuals. The behavior of user in a distributed system is difficult to characterize and so the management of resources and applications is a very complex task. When the demand for computing power in the system increases, it becomes important to balance the load in the system properly.

Load balancing is the process of distributing the load among various nodes of a distributed system to improve both job response time and resource utilization while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or lightly loaded.

2. Background Information

In distributed systems, load balancing is implemented at node level as well as the network level. These networked nodes can either be close to each other or around the world. The nodes within the system may try to balance out the load depending on the policies and algorithm defined for load balancing. Policies largely affect the way load balancing is achieved.

Among some of the algorithms which are used, static load balancing are known for their simplicity and they often works well for certain applications and environments. If all the information and resources related to a system are known, optimal or a sub-optimal static load balancing can be implemented which makes it possible to increase throughput of a system and to maximize the use of the resources by this method.

In most of the cases, this information is not available. Systems today are getting more complicated day by day and so it is clear, the potential of static load balancing are limited by the fact that they do no react to the current system state. In this scenario, dynamic load balancing algorithms and policies can be implemented to assign the job at the runtime.

Even after implementing dynamic load balancing, it might happen that the system is not at its optimal performance, i.e. some of the nodes are overworked while the others might be idle. This could be happening as a result of the different nature of nodes, jobs and heterogeneous nature of distributed networks. In this situation it might prove to be optimal to implement the adaptive load balancing algorithms and policies.

So, there is no perfect load balancing algorithm or policy in the distributed systems, but it is the goal of the researcher to achieve optimum performance from the system.

3. Problem Statement

In a distributed system, with multiple nodes there is a very high chance that some nodes will be idle while the other will be over loaded. A poorly balanced system can be slow, not utilizing resources properly and expensive to maintain. The goal of the load balancing algorithms and policies is to maintain the load to each processing element such that all the processing elements become neither overloaded nor idle.

As a distributed systems researcher or developer, it may become difficult to decide which load balancing policies and algorithms to implement since optimal implementation is completely dependant on the architecture and nature of the distributed system. It is challenging to achieve a optimal balance. If not implemented properly, improving the system wide load balance comes at additional cost, overheads and performance issues. Thus it becomes very important to have lucid understanding of load balancing.

4. Proposed Outcome

In this survey we intent to understand load balancing policies and algorithms through a comparative study. We would be comparing among the different policies and understanding their advantages over the others.

Additionally we will compare the static load balancing algorithms among themselves as well as with the dynamic and adaptive load balancing algorithms. This comparative study will involve time, complexity, utilization of resources, cost, overhead as some of the basic parameters for comparison.

Furthermore we intend to study a few specific use-cases and implementations which will give us a deeper understand of the topic.

5. Conclusions

In this study we intend to understand the load balancing strategies in detail. Load balancing in distribute systems is a popular area of research today as the demand for heterogeneous computing is increasing due to wide use of internet based solutions. Through this comparative study we also understand the differences between the algorithms; and although there is no prefect algorithm for a distributed system, we would be understanding which algorithms to implement in our solutions in order to achieve optimal performance. This study would offer a guideline for the researchers and developers intending to implement load balancing in their distributed computing systems.

6. Miscellaneous

a) Timeline

Week	Task
1	Study relevant papers, gather more info
2	Study papers related to policies and static load balancing
3	Write the initial findings as a report
4	Study papers related to dynamic and adaptive load balancing
5	Write the findings as continuation of report
6	Find papers related to further research in the field for specific use-cases and applications
7	Compile and write the final report
8	Write the final presentation

b) Deliverables

- One final report in PDF form.
- One final Powerpoint presentation.

References

[1] D.L. Eager, E. D. Lazowska and J. Zahorjan, "Adaptive Load Sharing in Homogeneous Distributed Systems," IEEE Trans. on Software Eng. ,vol. SE-12, pp. 662-675, May 1986.

[2] Penmatsa, Satish & Chronopoulos, A.. (2007). “Dynamic multi-user load balancing in distributed systems.” 1-10. 10.1109/IPDPS.2007.370312.

[3] Archer C.J., Mullins T.J, Sidelnik A. and Smith B.E. “Parallel Computing System Using Coordinator and Master Nodes for Load Balancing and Distributing Work”. United State Patent, 2010.

[4] Md. Firoj Ali and Rafiqul Zaman Khan. “The Study On Load Balancing Strategies In Distributed Computing System”. International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.2, April 2012.

[5] Berenbrink P., Friedetzky T. and Steger A. “Randomized and Adversarial Load Balancing”. CiteSeerx, 1997.

[6] Bernard G., Steve D. and Simatic M. “A Survey of Load Sharing in Networks of Workstations”. The British Computerm Society, The Institute of Electrical Engineers and IOP Publishing Ltd, 75-86,1993.

[7] Hamdi M. and Lin C.K. “Dynamic Load Balancing of Data Parallel Applications on a Distributed Network”. In 9th International Conference on Supercomputing, ACM, 170-179, 1995.

[8] Kabalan K.Y., Smari W.W. and Hakimian J.Y. “Adaptive load Sharing in Heterogeneous system: Policies, Modifications and Simulation”. CiteSeerx, 2008.

[9] Weinrib and Shenker S. “Greed is not Enough: Adaptive Load Sharing in Large Heterogeneous Systems”. INFOCOM, 986-994, 1988