

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Load Balancing through Multipath Selection Mechanism for Big Data Centers in Clouds

A. Vijay Kumar¹, S. Vasundra²

M.Tech, Department of CSE, JNTUACE, Ananthapuramu, Andhra Pradesh, India¹

Professor, Department of CSE, JNTUACE, Ananthapuramu, Andhra Pradesh, India²

ABSTRACT: Cloud Storage has gone to immense usage these days making it more complex for frequent transfer of large amount of data in data centers. So, load balancing has become a key challenging issue in which large amount of data should be transferred through several thousands of connected servers in big data centers. Open Flow scheduling can balance the data flow through its programmatic traffic controller which makes it more reliable than other mechanism. Existing Open Flow scheduling schemes focus on selecting the transmission route for data before the actual data transfer begins. This results in failure of dynamical data transfer which is major challenging issue in present days. In this paper, multipath selection mechanism for big data centers in clouds is proposed in which when the data transmission route is overflowed with large amount of data then the path with lowest data transmission rate is selected. If the incoming data is beyond the bandwidth of route then data is equally distributed among all the available routes making it optimal load balanced route selection.

KEYWORDS: load balancing, Open Flow scheduling, big data centers, dynamic data transfer

I. INTRODUCTION

Big data centers are made up of large number of switches [1] which are used to transfer huge amount of data through servers. Large amount of data is continuously transferred among the data centers which causes lot of resource utilization and load balance problems. Load balanced scheduling is a method through which the data traffic is evenly distributed among all the network links available in data center which reduces the delay in data transfer. Different types of schemes have been proposed to overcome the load balancing problem [2][3], but all these schemes focus only on the initial route selection before the actual data transmission flow begins. However, possibility of network states and work load may change dynamically which results in failure of initial route selection. As a result, the existing schemes does not meet the requirements of dynamical load balancing. By observing all this possible failures and drawbacks, Load Balancing through Multipath Selection Mechanism for Big Data Centers in Clouds is proposed. In this system if a network route is overflowed with data transmission and the next data flow is not able to transfer the data at right time then we select the least load having network route for the transfer of that data flow.

II. RELATED WORK

Load Balancing is a process in which all resources in a system are shared equally by all tasks based on requirements. In precise load balancing optimizes [4] resource utilization and maximize the throughput, which decreases time delay of transmission. There are various scheduling mechanisms available. Among them static and dynamic scheduling are most used ones [5].

In static scheduling schemes such as Ratio-based, Round Robin and others, a predefined network route for data flow transmission is done. Network states and workload changing is not considered in static routing. This results in failure of data transmission when the network states change dynamically. Network path unavailability leads to such problems and data cannot flow to the assigned data center. So, in order to overcome such problem dynamic load balancing is implemented in network selection.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

In dynamic load balancing [6] data centers use hardware machines to overcome the workload balance. But these machines are very costly and cannot be used for longer periods of time. To reduce such problems load balancing through open flow [7] schemes came into use.

One of the most popular one is NOX [8] which can separate data packet control from data forward and it can be easily shifted to other network paths dynamically making it a load balanced path.

A. Motivational Scenario:

The main motivation for the proposed system is as depicted in the fig.1[9] below where there are 5 data flows ($f_1, f_2, \dots, f_6$), and there are 6 switches of the network. Initially each and every network path is assigned with data flows. When you look at the first network path the flow time for f_3 will be higher than the remaining data flows. So, in order to reduce the delay in time as well as better usage of bandwidth the f_3 is transferred to next network path which has lower load than the other one.

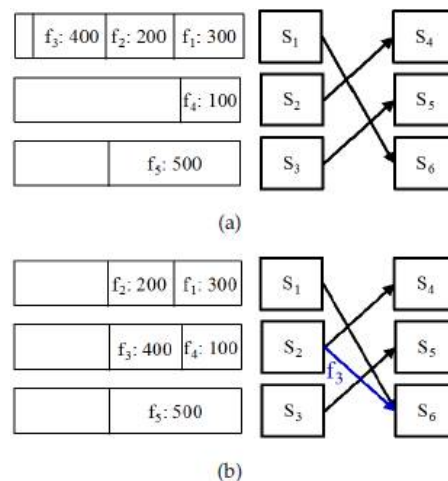


Fig.1 a) regular load balancing b) multipath selection mechanism

By using this scenario as a base for multipath selection mechanism the proposed system is implemented in order to maintain the load balance in network paths in data transfer. Fig. 1 a) represents the regular load balancing mechanism where the data transfer follows first come first serve type. Fig. 1 b) represents the motivational scenario for better load balancing dynamically which is the proposed system.

III. PROPOSED SYSTEM

In the proposed system, multipath selection mechanism is implemented which can select alternate path for data flow if the load is high on current network path. This reduces the network load and maximizes the resource utilization which ultimately reduces the time delay for data transfer.

Following architecture shows the implementation of multipath selection mechanism and its behavior.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

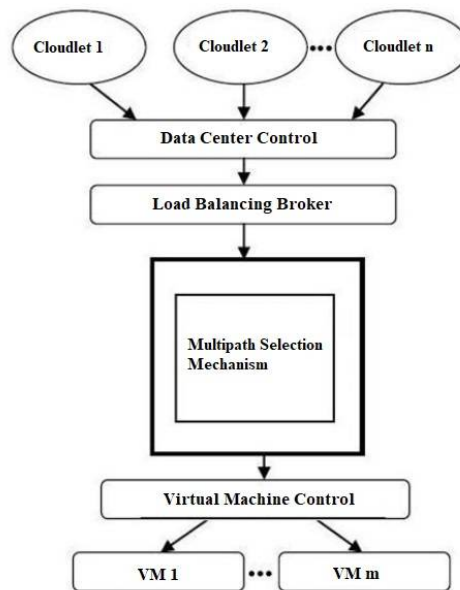


Fig. 2 Implementation modules of proposed work

The implementation modules in Fig. 2 involve mainly of Data center control, Load balancing broker, Multipath selection mechanism(MPM), and Virtual machine control. In this proposed work, the cloudlets are initially allocated to each data center and then they are processed by the broker who executes the load balancing through MPM. After this the data transfer is assigned to virtual machines by virtual machine control.

The modules of architecture are explained as follows

A. Data center control:

Data center control maintains all the detailed list of data centers such as data center name, data center id, host id, its ram and the bandwidth. This takes in the data flow from cloudlets and assigns from which data centers the transfer has to take.

B. Load Balancing Broker:

Load Balancing Broker acts as an intermediary between the Data center and the virtual machines in them. This implements the Multipath Selection Mechanism to assign work load to desired virtual machines.

C. Multipath Selection Mechanism:

Multipath selection mechanism is implemented in order to reduce the work load on network routes while transferring the data from one data center to another. In this mechanism, initially the flow data are assigned to empty network paths, and when each path is preoccupied and the next data flow incomes the mechanism checks if the current path is overloaded or not. If it is overloaded then the path with lowest flow load is taken into consideration and the data flow is transferred to that path.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Algorithm

```
Multipathselection() {  
  for each task in task queue  
    initflowtask = 0;  
    while (task is assigned to a vm) or (flowtask>threshold){  
      increment flowtask  
      if the load > other paths  
        find the lowest load path and transfer the flow  
        assign task to vm if indegree >0; decrement indegree  
    }  
    remove task from task queue  
  }  
  Process completed tasks(){  
    increment degree of vm assigned to the task  
  }  
}
```

D. Virtual Machine Control:

Virtual machine (VM) control controls the assignment of the tasks for each virtual machine present in the network. It shares host list with other vms. It processes cloudlets. This processing happens according to the algorithm, defined by the multipath selection mechanism. Each VM has an owner, which can submit cloudlets to the VM to be executed.

IV. PERFORMANCE RESULTS

The proposed work is executed and compared with two of the existing load balancing techniques, LOBUS and Round Robin. A generalized checking is done whether the proposed system can reduce the end to end delay in data transmission and it resulted in maintaining low delay than the other two.

The following figure Fig.3 represents the graph in which Multipath Selection Mechanism (MPM) reduces the time delay more efficiently than that of LOBUS and RR (Round Robin).

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

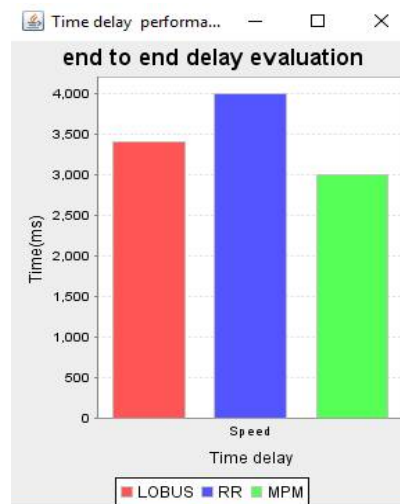


Fig. 3 End to End delay evaluation

End to End delay is most important factor for faster delivery of data. The MPM shows significant improvement when compared with other load balancing mechanisms. When end to end delay is low then there will be faster data transfer which decreases energy consumption by big data centers and also on time receive of data.

Bandwidth utilization of the proposed system is evaluated and it outperforms the existing LOBUS and RR.

The following figure Fig.4 represents the graph in which Multipath Selection Mechanism uses the maximum bandwidth and reduces the load on remaining network paths making it a well-balanced flow.

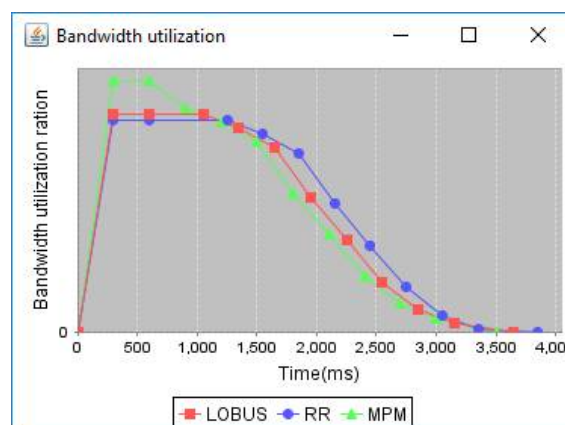


Fig. 4 Bandwidth utilization ration

Bandwidth is needed for any data transfer in big data centers. If more bandwidth is utilized then the faster the data transfer takes place. The MPM in above graph shows that the bandwidth is utilized to its maximum extent so that a faster delivery of data is taking place. This improves more dynamic data transfer along with load balance.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

V. CONCLUSION AND FUTURE WORK

Load balancing while transferring data in clouds is a complex and time taking so the proposed mechanism is implemented to reduce the load on the network path while transferring the data dynamically. It also maximizes the band width utilization and reduces the end to end delay. This mechanism can be used in large data centers where network links are inter connected and data transfer occurs continuously.

For the future enhancement, we will implement this mechanism in Open Flow networks in three different patterns, uniform, semi uniform and irregular data flows in network paths.

REFERENCES

- [1] Z.Z.Cao, M.Kodialam and T.V.Lakshman. Joint Static and Dynamic Traffic Scheduling in Data Center Networks. in Proceedings of IEEE INFOCOM 2014, pp.2445-2553.
- [2] N.Handigol, S.Seetharaman, M.Flajslik, N.McKeown, and R.Johari, Plug-n-Serve: Load-balancing web traffic using Open-Flow, Demo at ACM SIGCOMM, Aug. 2009.
- [3] M.Schlansker, Y.Turner, J.Tourrilhes, and A.Karp, Ensemble Routing for Datacenter Networks, In ACM ANCS, La Jolla, CA, 2010.
- [4] S. Vasundra, Load Balancing through shortest path in regular mesh topologies, in IJSSAN, pp:149-153, vol.4, issue 1, Jan 11 – June 2011
- [5] Z.Z.Cao, M.Kodialam and T.V.Lakshman. Joint Static and Dynamic Traffic Scheduling in Data Center Networks. in Proceedings of IEEE INFOCOM 2014, pp.2445-2553.
- [6] T.N.Anitha, R.Balakrishna, An Efficient and Scalable Content Based Dynamic Load Balancing Using Multiparameters on Load Aware Distributed Multi-Cluster Servers, IJEST, 2008.
- [7] R.Wang, D.Butnariu, and J. Rexford, OpenFlow-Based Server Load Balancing Gone Wild, in: Proceedings of Workshop on Hot-ICE, Mar. 2011.
- [8] N.Gude, T.Koponen, Justin Pettit et al., NOX: Towards an Operating System for Networks, SIGCOMM Comput. Commun. Rev., Vol.38, July 2008, pp. 105-110.
- [9] Feilong Tang, Laurence T. Yang, Can Tang, Jie Li and Minyi Guo. A Dynamical and Load-Balanced Flow Scheduling Approach for Big Data Centers in Clouds. in Proceedings of IEEE Transactions on Cloud computing 2016, Issue.99, March 2016.