ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

A Study on Load Balancing Video Servers in Distributed System

Kunjal Surati[1], Vedashree Patil[2], Namoshi Roy[3], Ramakant singh[4], Smita Patil[5]

Students, Dept. of I.T., Atharva College of Engineering, University of Mumbai, Mumbai, India[1,2,3,4]

Professor, Dept. of I.T., Atharva College of Engineering, University of Mumbai, Mumbai, India[5]

ABSTRACT: Load Balancing effectively distributes the load among the available proxy servers. The goal of load balancing is to improve the performance of the system by balancing the loads among the computers. A Video-on-Demand system allows geographically distributed user to select and watch movies, videos as per their convenience.Video-on-demand (VoD) systems allow users to select and watch video content when they choose to, rather than having to watch at a specific broadcast time. In this paper, we have studied various load balancing policies. We have also studied different Video-on-Demand (VoD) system architectures and scheduling policies.

KEYWORDS: load balancing; Video-on-Demand, Round Robin, Cluster-based.

I. INTRODUCTION

As dynamic contents are changing traditional web environment, there is an increasing demand on high performance web servers, which leads to the use of cluster-based web servers. The growth of the Internet, along with the increasing popularity of dynamically generated content on the World Wide Web, has created the need for more and faster web servers that are capable of serving the over 100 million Internet users.

Load balancing is one of the most important problems in attaining high performance in distributed systems which may consist of many heterogeneous resources connected via one or more communication networks. Load balancing is the strategy by which load is redistributed among the computational elements (CE) of a heterogeneous network where they work cooperatively so that large loads can be distributed among them in a fair and effective manner.

Video on Demand are Systems which allow users to select and watch/listen to video content when they choose to, rather than having to watch at a specific broadcast time. Interactive Video-on-Demand systems (IVOD) is an extension of Video-on-Demand (VoD) .In order to improve the throughput, response speed and scalability of VoD server efficiently, many VoD server have thrown away single servers and turn into VoD server clusters, which centrally accepts all the coming requests and evenly dispatches them to the servers. In the paper [6], it proposes two efficient algorithms referred to as Rate-based Load Balancing via Virtual Routing (RLBVR) and Queue-based Load Balancing via Virtual Routing (QLBVR).A new Open Flow controller application for dynamic server load balancing is designed by continuous monitoring of load of each video server [7]. In the paper [1], a modified Round-robin load-balancing algorithm (MRR) for clustered based web system is proposed.

The remainder of the paper is organized as follows. In section 2, we have discussed about load balancing policies. Section 3 includes Video-on-Demand system architectures. Different Scheduling policies are described in Section 4.

II. LOAD BALANCING POLICIES

Load balancing for distributed computing system has been deeply studied for a long time. Generally there are two methods for load balancing in a distributed environment: i) A static load balancing approach which depends on static information such as memory space, CPU capacity etc in making load balancing decisions ii) Dynamic load balancing policies uses current state of the system to make load balancing decisions and therefore is able to further improve the system performance.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

A. Clustered Load Balancing Policy

In this policy the network is organized into N clusters [1]. Each cluster has a specific node which acts as the Cluster Manager and is responsible for managing the cluster load. Both the lower levels known as the worker nodes and N_{i-1} clusters send the load information to the Cluster Manager. The Clustered Load Balancing Policy starts with sending of load update messages by the lower levels to the Cluster Manager. After a particular time interval called estimation interval, each processor in the system calculates its load parameters. Load information parameters used to measure the load of a node are Response Time, Memory Utilization, CPU Utilization and Traffic intensity. This workload estimation is then send to the Cluster Manager. Using the workload estimation, the nodes are grouped into idle, balanced, medium loaded or highly loaded. Cluster Manager then decides whether the cluster is stable or not stable by calculating the stability and saturation of the cluster. Intra Cluster load balancing is done by transferring the task among clusters. This Intra Cluster task transfer is done by maintaining priority queues. Inter Cluster load balancing is done if the Cluster Manager fails to balance its load among its worker nodes.

B. Dynamic Feedback Algorithm Based On Neural network

In this policy, each real server dynamically collects the local load information with the help of parameters such as CPU utilization and memory usage of the system[5]. The client process in Information transfer module calls the load information collection module time to time to collect information about parameters by establishing UDP connection. If the server does not receive any packet from node i over a three time periods then that indicate node failure and its weight is set to 0 to isolate that node. When the cluster is idle Neural network training module trains the neural network and preserve the trained weights and thresholds in a file. Load calculation module with the schedule code block in the kernel achieves load-balancing scheduling. The new connections are allocate to each serve in proportion to the weights; which avoids large number of concurrent requests to arrive at one real server load information sampling period.

C. Queue-based Load Balancing via Virtual Routing

In Queue-based Load Balancing via Virtual Routing combines queue and rate adjustment policies[6]. QLBVR carries out two activities, first is coarse adjustments on job transferring and processing rates and second is fine adjustments on number of jobs in the queue. In Coarse adjustment on transfer of job and processing rates, at every instant of time Tn, an optimal solution is obtained by using a procedure for a single node. After a time interval Ts and according to the computed optimal solution each node adjusts its job transferring rates and job processing rate. In the second activity of Fine adjustment on queue length, there is a status exchange interval Ts which is divided into equal subintervals, denoted as estimation intervals Te. Now at this time instant Te, node i estimates and adjusts the earlier estimate of the queue lengths of its neighbouring nodes. It also has an accurate knowledge of its own queue length.

III. VIDEO-ON-DEMAND (VOD) SYSTEM ARCHITECTURES

A. Cluster-Based web server architecture

In cluster-based architecture web system as shown in fig 1, N server machines which have their own disk and operating system are connected through a high speed network [5]. This server machines are the nodes which responds to user requests. A cluster which includes large no of web servers utilizes only one host name and are virtual IP address because of which user get a single interface. For this a mechanism is used which controls the whole requests and hides the service distribution among all the servers. The load-balance controller is an important component which dispatches the incoming requests to all the server nodes and routes the load. The load-balance controller includes web switches which can be broadly classified as layer-4 and layer-7 web switches. Difference between these switches is that layer-4 web switches maintain a binding table so that each client TCP session is associated with the target server. It works at TCP/IP layer and utilizes content information blind distribution. On the other hand, layer-7 web switches establish a complete TCP connection with the client. It utilizes content distribution aware distribution, which checks HTTP request

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

content before deciding about dispatching. The load-balancing algorithm deployed in web switches has great significance to achieve good cluster performance.

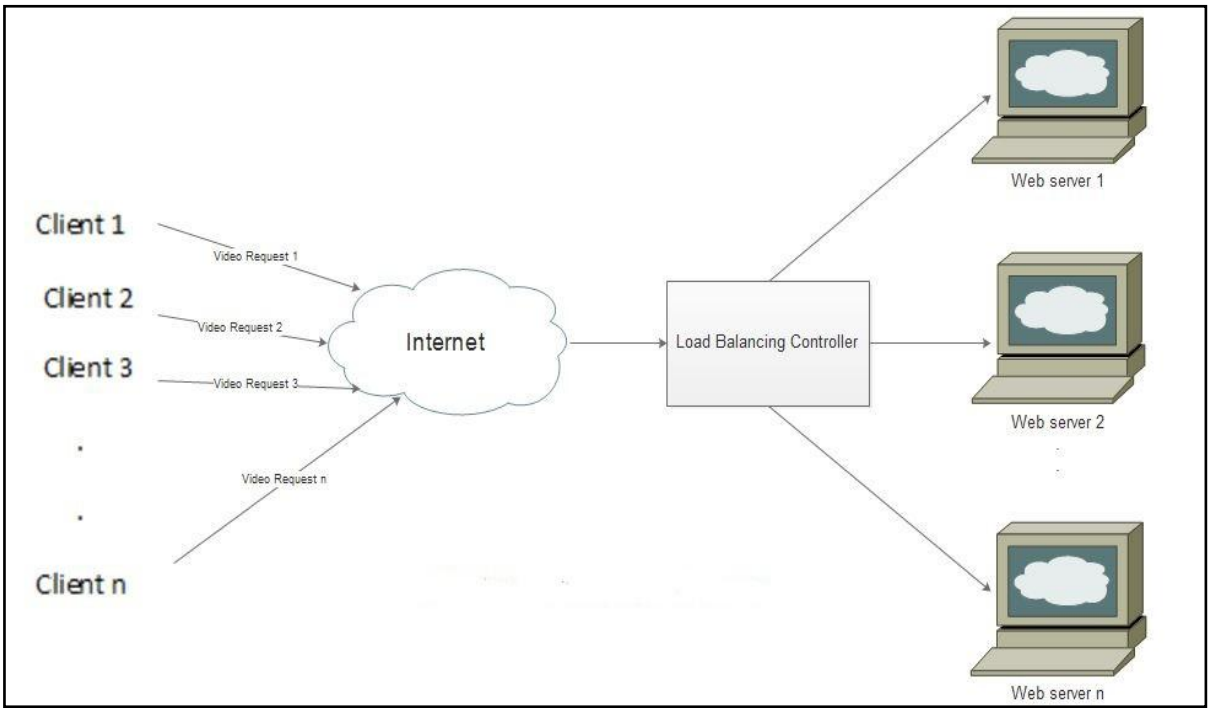

Fig 1: Cluster-Based web server architecture

B. Interactive NVOD system architecture

In Interactive NVOD architecture [3], the system consists of a VoD server, clients, and the network. Clients stream videos from the server using the network connecting them. The major components of the VoD server includes one queue for each video, a blocking and a waiting queue, a queuing manager, streams, a Split and Merge technique, I-Stream allocation module, a waiting scheduling policy, a blocking scheduling policy. The streams are divided into three categories: Full length multicast streams (B-Streams), Unicast interactive streams (I-Streams), Multicast patch streams (P-Streams). An I-Stream is a unicast dedicated stream which is used to service only the interactive requests, and therefore it cannot be shared with other requests. P-Streams and B-Streams are used to service the waiting customers and can also be shared with other requests to the same video. I-Streams are managed by split and merge technique. When delivering of the requested data is finished by the I-Stream, the server frees the I-Stream, which becomes available for use by other requests.

IV. SCHEDULING POLICES

A. Simple Round Robin Scheduling Policy

Simple round robin comes under the category of static policy since it does not consider system state information at the time of dispatching[2]. Simple round robin implements a circular queue where pointer is pointed at the last elected server, that is if server i was the last chosen node for request dispatching, the next request will be assigned to server i+1 ,where i=i+1 mod N here N is the no of servers. As web switch use highly complex and refined algorithm, Round Robin is more suites because of its simplicity which also helps in fast dispatching of requests. This avoids the web switches from becoming primary bottleneck of web cluster. However its disadvantage of using simple round robin

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

is that it's makes part assigned decisions. Also once host for process request allocation is chosen, it can't be changed during process execution to make any changes in the system load.

B. Modified Round Robin Policy

Modified Round Robin policy considers system parameter states while dispatching the request to servers, hence it come under dynamic polices[2]. Modified Round Robin policy maintains the circular queue in which pointer is points to last selected server. When a request is arrived at web switch, it calculates the load on all nodes and dispatch the request to the next server if its load is not biggest of all server else the request is assigned to next to next server. That is if server I is last server which was assigned a request ,then for next request, load on all servers will be calculated .If load on server i+1 is biggest ,then the request is assigned to server i+2.Else request is assigned to server i+1.

C. First Come First Serve Policy

In FCFS policy, the requests for all videos form a queue[4]. The first request in the queue is dispatched to next available server .First Come First Serve policy fairly distributes the request among the servers. FCFS-n policy is a type of FCFS policy in which a small part of server capacity is already reserved for n-hottest video requests and the cold video requests are scheduled according to FCFS policy.

V. CONCLUSION

In this paper, we have discussed various load balancing policies. We have also studied different Video-on-Demand (VoD) system architectures and scheduling policies. These policies can be used to effectively load the balance among video servers in distributed system. We deduce that cluster based load balancing with the help of round robin policy will effectively balance load among the video servers.

REFERENCES

1. Moumita Chatterjee, S K Setua,' A New Clustered Load Balancing Approach for Distributed Systems', Computer, Communication, Control and Information Technology (C3IT), 2015 Third International Conference , pp. 1-7, 7-8 Feb. 2015, Hooghly.
2. Xu Zongyu , Wang Xingxuan, 'A Modified Round-Robin Load-balancing Algorithm for Cluster-based Web Servers', Proceedings of the 33rd Chinese Control Conference, pp.3580-3584, July 28-30, 2014, Nanjing, China.
3. Kamal K. Nayfeh and Nabil 1. Sarhan 'Design and Analysis of Scalable and Interactive near Video-on-Demand Systems', Multimedia and Expo (ICME), IEEE International Conference, pp. 1-6, 15-19 July , 2013, San Jose, CA.
4. Sudhir N. Dhage, Smita K. Patil and B. B. Meshram, 'Survey on: Interactive Video-on-Demand (VoD) Systems', Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference, pp .435-440, 4-5 April 2014, Mumbai.
5. Yang Hu, Shanan Zhu, 'Load-balancing Cluster Based on Linux Virtual Server for Internet-based Laboratory' , Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference, pp. 2181-2185, 9-11 June 2014, Hangzhou.
6. M.Thejovathi, 'Dynamic Load Balancing Algorithms for Distributed Networks' , IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.2, February 2014.
7. Selin Yilmaz, A. Murat Tekalp, Bige D. Unluturk, 'Video Streaming Over Software Defined Networks With Server Load Balancing', 2015 International Conference on Computing, Networking and Communications, Internet Services and Applications Symposium, pp.722-726, 16-19 Feb 2015,Garden Grove, CA.