A REVIEW-LOAD BALANCING OF WEB SERVER SYSTEM USING SERVICE QUEUE LENGTH

Brajendra Kumar, M.Tech (Scholor) LNCT,Bhopal 1; Dr. Vineet Richhariya, HOD(CSE)LNCT Bhopal 2

Abstract

In this paper, we describe on the load balancing system of the web server system. Web based service are used too many organization to support their customers and employee. An important issue in developing such services is ensuring the quality of service (QOS), that user experience is acceptable. Recent year have been more researcher try to it the maintained the quality of web server, with load balancer and to be used some techniques. In this paper work to use on that the weighted least connection (WLC) and Round Robin (RR) algorithm for the load balance of the web server with service queue length. These techniques are very effective to handle the dropping rate and performance throughput or delivery of packet to view the client side. Now we consider these parameters as for as memory load for the back end to supported by the server side, number of weighted active connection for at a time and service queue length of the web server. Based on these parameters, we are calculating the server load and also create rules for request allocation on the web server.

Keywords- web server system, load balancer, weighted least connection and round robin mechanism, service queue length with memory size.

1. Introduction

Load balancing to be provide a very useful to us for the web server system. Recently years to be more researchers to analyze do that effectively services to be provide for the web server system. Now here to mentioned that the web server load balancing system. The load balancing system has been found to provide an effective and scalable way of managing the ever- increasing web traffic. Although we try to get performance characteristics of the system, that user a load balancer. Typically a load balancing method or strategy is used to decide how the load balancers choose. Where to send the request there are many strategies available depending on the vender; however a few common ones are discussed this section.

With that Round robin method of load balancing, the load balancer will send traffic to each node in succession. This method will evenly distributed the traffic but does not take into account the current load or responsiveness of the nodes, and weighted Least connection like the least connection methods, these load balancing method selects pool methods or nodes based on the member of active connections. However the weighted least connection methods also based their selection on server capacity. Weighted least connection method work best in environments where the servers have different capacity. This method is more intelligent, but if the connection features is enabled, the weighted least connection methods do not include the connection in the calculation when selecting a pool member or node. The weighted least connection methods uses only achieve connection in their calculation.

All load balancer are capable of making traffic decision based on traditional OSI layer 2-3 information. More advance load balancer, however can make intelligent traffic management decision based on specific layers 4-7 information constrained with in the request issued by the client, such application is required in many application environments, including, those in which is a request for application data can only be met by a specific server or set of servers. Load balancing decision are made quickly, usually in less than on millisecond and high- performance load balancers can make millions of decision per second.

An administrator select algorithm implemented by the load balancer determines. The physical or virtual server and send the request. Once the request is received and processed, the application servers send its response to the client via. The load balancer manages all bi-directional traffic between the client and server, it maps each application response, and users receive the proper response. Load balancer can also configured

to guaranteed that subsequent request from the same user, and part of the same session, are directed to same server as the original request, called application that must maintain state.

2. Related work

In this section describe on the related work on the current content with respect to server load balancing of the web server system.

Rmana et al [2] in this paper the performance analysis of various distributed web server system load balancing algorithm based on different qualitative parameters, considering static and dynamic load balancing approaches are considered. The analysis indicates that these both static and dynamic types of algorithm can have advancements as well as weaknesses over each other. The main purpose of this paper is to help in the design of new algorithm in future by study of behavior and characteristic of various existing static and dynamic algorithm are more stable then dynamic algorithms but at a same time dynamic distributed algorithms are always considered better then static algorithm in distributed web server systems.

Sarabjeet et al.[3] in this paper, author present the performance analysis of various load balancing algorithms based on different parameters like fault tolerance, overload rejection, stability, and cooperative, resource utilization considering two typical load balancing approaches static and dynamic. The analysis indicates that both static and dynamic types of algorithm can have advancement as well as weakness over each other. Static load balancing depends on the current situation of the system. This analysis is useful for those who to develop a new load balancing algorithm for web server.

Zheng Wen, Lei Shi, Runjie Liu, Lin Qi and Lin Wei [4] according performance of load balancing scheduling polices in web server cluster system is greatly impacted by the characteristics of work load, based on the analysis of the load characteristics for scheduling algorithm, prediction-based adaptive load balancing model (RR_MMMCS_A_P) is proposed in this paper, the arrival rate the size of the fallow-up , request are predictive by RR_MMMCS_A_P and rapid adjustment of the corresponding parameter to balance the load between servers. Experiment have shown that

compared with CPU based and CPU scheduling strategy, RR_MMM_CS_A_P have better performance

in reducing average response time for both calculation-intensive and data_-intensive jobs.

Cardellini et al. [6] analyze various dispatching policies under realistic situation where state information needs to be estimated. They also that packet rewriting by the dispatcher presents problems because the dispatcher must rewrite incoming as well as outgoing packets, and outgoing packets typically out number incoming request packets.

Pao and Chen [7] presented the dispatcher-based load balancing architecture. This architecture is based on the decision-making based on the behavior of the server system. When the backend servers have different computation power, the system must use the real time loading information as the decision-making criteria. They proposed a dispatcher architecture that uses the remaining capacity of each backend servers to decide the most appropriate server. In addition to improve the performance, the system can also use the remaining capacity algorithm to reduce the cost project work. We have introduction a new performance parameter, queue size for selected the best appropriate web server to transfer new request.

S. Parkes, N. Gandhi, J.Hellerstein, D.Tilbury, T. Jayram, and J. BIgus, “Using control theory to achieve service level object design performance management”, 2002.[9] In this paper, we have demonstrated a methodology for constructing and analyzing closed-loop system using a statically approach to system identification. This approach is more generally applicable then the conventional first- principals approaches, and also has the potential to adapt to changes in the underlying system (such as a new software release). The fit for our models of a lotus notes, email server is quit good R^2 is no lower than 75%, and is as high as 98%.

In this paper we have restricted ourselves to a simple control law in order to demonstrate. The value of this approach. We plan to study more complex controllers to asses if control theory provides useful inserted in applying our methodology to other, service level management situations both to refine our methodology an to asses its value.

Anders Robertsson, Bjorn wittenmark and Maria Kihl, “Analysis and design of admission control in web server system”, IEEE, 2003, [10] this paper discussed from a control point of view the modeling of service control-node. The queue is assumed t be an M/G/1-

system and is modeled by a non linear flow model and a simplified discrete- time model is used in the analysis and design of the system.

An admission control system based on a PI- controller combined with an antirust windup feature is developed the stability of the admission control system is analyzed and the stability of the simplified queue model are verified through discrete- event simulation of the system.

We are trying to get the reduce load of the web server system regarding above mention problem of the web server load balancing and web server architecture. We consider another performance parameter of the web server that is memory load, number of active connection and queue length of the web server. Based on these parameters we are calculating the server load and also create some rules for request allocation on the web server. These rules help us to find the best web server to transfer the new requests.

3. PROPOSED METHOD

In this section we describe on the proposed methodology covered with that web server load balancing technique and algorithms.

3.1 web server load balancing algorithm

Load balancing of the web server system has been the area of research in the few decades. Load balancing algorithms always try to balance the load by transferring the load from heavily loaded server to lightly loaded servers. It is the process to remaining the load to the individual server to make resource utilization effective and better response time.

The main aspects to be considering when we are designing a load balancing algorithm are estimation of load, memory load, number of connection, behavior of the system. Depending on the current state of the system load balancing algorithm can be divided in to categories static and dynamic load balancing algorithms.

Such as the illustrate that and discuss below a different type of the web server load balancing technique and algorithm as for as the review with the study there.

A. Static load balancing algorithm
B. Dynamic load balancing algorithm

A. Static load balancing algorithm

Static load balancing is [2] the simplest form of the two types. Static load balancing is the selection and placement of a logical process on some processing element located on a distributed network. The selection of the processing element for some for some logical process is based upon some weighting factor for that process, or some kind of workload characterization of the processing elements or the network topology, possibly both. The process of selecting a processing element for a logical process requesting service may be done with two possible methods, a stateless method or a state –based method. With a stateless method the selection of a processing element without regards any knowledge of the state. A state-based method of the selecting a processing element of the implies that the selection of a processing element requires knowledge of the system state, either globally, or locally.

Globally knowledge implies that the state of the all components of the system is known and local knowledge implies that only partially knowledge is known. If the state of the system is needed then some kind of messaging or probing of network resource among the requesting processes, agent or processing elements is needed to determine their availability. Once a processing element is selected for a logical process, the logical process is executed on the selected processing element for a logical process lifetime. Two examples of stateless placement techniques for selecting a processing element for logical process are round robin and random placement. Round robin placement selects the next processing element from a predefined list. Random placement selects an element randomly from a set of processing element. Static load balancing algorithms are of the two type round robin load balancing algorithm. Weighted Round Robin load balancing algorithm and Random Allocation algorithm.

i. Round robin load balancing algorithm

Round Robin (RR) [2] algorithm distributed client requests evenly to all servers in round robin manner, meaning that server choosing is performed in series and will be back to the first server if the server has been reached. Servers choosing are performed locally on each server, independent of allocations of other server. Advantage of round robin algorithm is that it is simplest in nature. Round robin load balancing provides better performance for homogeneous environment where the entire server has same capacity. In generally cluster of the web server have different capacity. RR load balancing treats the entire server equally and provides request in series, so the higher capacity server works efficiently, but low capacity server does not provide better result for same number requests.

ii. Weighted round robin algorithm

In weighted Round Robin algorithm [2], this algorithm maintains a list weight and assigns a weight to the servers present in a cluster. These weight works as to the server. Higher weighted server has high priority and forwards new request in proportion to the weight of each server. This algorithm uses more computation times then Round Robin algorithm. However, the additional computation results in distributing the traffic more efficiently to the server that is most capable of handling the request.

iii. Random allocation algorithm

In a Random allocation, the HTTP requests are assigned to any server picked randomly among the cluster of servers. In such a case, one of the servers may be assigned many more requests to process, while the other servers are sitting idle. However, on average, each server gets its share of the load due to the random selection. This algorithm is easy to implement. This algorithm can lead to overloading of one server while other server is idle.

B. Dynamic load balancing algorithm

If differs from static algorithms in the workload is distributed among the servers at runtime. The load balancer assigns new requests to the server based on the information collected. In a distributed system, Dynamic load balancing can be done in two different ways: distributed and non-distributed [3]. In the distributed one, the dynamic algorithm is executed by all servers present in the cluster and the task of load balancing is shared among them.

In non-distributed type, either one server or clusters of servers do the task of load balancing. Non-distribution dynamic load balancing algorithm can take two forms: centralized and semi-distributed. In the first form, the load balancing algorithm is executed only by a single server in the whole system: the central server. This server is solely is responsible for load balancing of the whole cluster. The other servers interact only with the central server. in semi-distributed form, servers of the large cluster are portioned into cluster by appropriate selection technique, which takes care of load balancing within that cluster [3]. Hence the load balancing of the whole system is done via the central servers of each cluster. Centralized dynamic load server decreases drastically as compared to the semi-distributed case. However, centralized algorithm can render useless once the central server crashes. Therefore, this algorithm is most suitable for network with small size.

In this paper several related research about load balancing algorithm and web server architecture have been discussed. We have also discussed about dispatcher-based approach, admission control mechanism of a web server and discuss distributed computing and client-server model architecture.

1. Distributed computing

In general, distributed computing is any computing that involves multiple computers remote from each that has a role in a computation problem or information processing. In business enterprises, distributed computing generally has meant putting various steps in business processes at the most efficient place in a network of computers. In the typical transaction using the 3- tire model, user interface processing is done database access and processing is done in another computer that provides centralized access for many business processes. Typically, this rather distributed computing uses the client/server communication model.

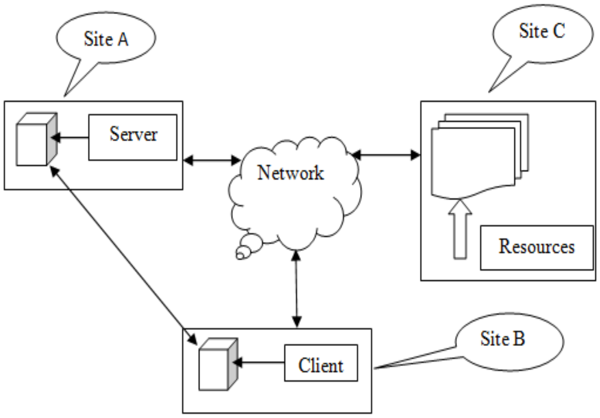

Figure:-1. Simple Distributed Architecture

2. Client server architecture

The client-server model is a distributed computing structure that partition takes or workloads between the provide of a resource or service, called server and service requesters, called clients. A client is a single- user workstation that provides presentation services and the appropriate computing, connectivity and the database services and the interface relevant to the memory providing computing, connectivity and the database services and the satisfies the business need by appropriately allocating the application processing between the client and the server processors. The protocol is the client requests the services from the server; the server processes the request and requests the result to the client. The communication mechanism is a message passing inters process communication (IPC) that enables distributed placement of the client and server processes.

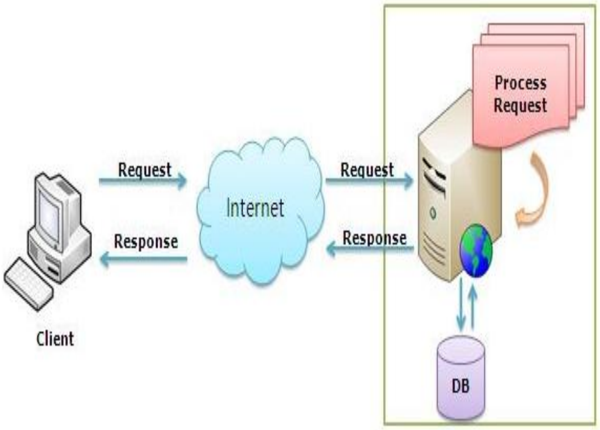

Figure:-2. Simple Client-Server Architecture

3. Dispatcher based Approach

To centralized request scheduling and fully control the client request routing. A dispatcher-based approach has been proposed in the paper. The dispatcher uniquely identifies each server in the system through a private address that can be a different protocol levels, depending on the architecture. Dispatching of request can be done using various technique Round Robin, lead connection etc.

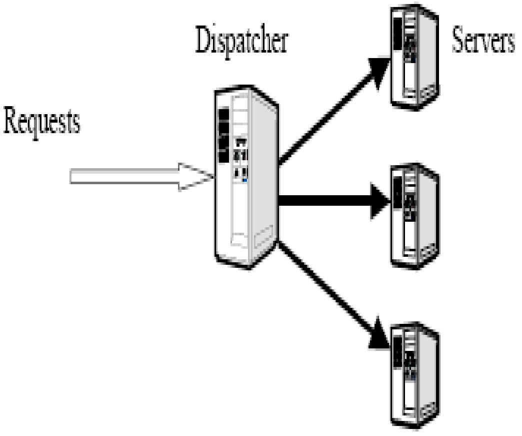

Figure: - 3. Dispatcher-based Server cluster

4. Admission control mechanism

Admission control mechanism is used to protected web server from overloaded and provide performance guarantee or differentiated service class of requests.

Kihl et at [8] have presented a control theoretic model of the web server. They use non linear control theory. Their model is based on the control theory. They developed a non-linear stochastic control theoretic model of GI/G1 system. They also show that their model can be used to design admission control mechanism that behaves well in the real system that is the queuing system. They also develop two type of controller PI and RST controller both commonly used in automatic control.

Conclusion

Load balancing algorithm on the Web server system has been used to improve the availability and reduce the overloading of the Web Servers. After the comparative analysis of the various classical approaches of the load balancing, these are the main problem to face on the throughput of the web server system and dropping rate of the request. Web server load balancing address several requirements that are becoming increasingly important in network such as increased scalability, high performance, high availability and disaster recovery. We have developed an efficient architecture of the Web server system and to reduce these are the problems.

In experimental analysis on the related works, we have observed that these are load balancing algorithm gives better result in homogeneous environment but heterogeneous environment to the increase the dropped the request and consistently down the throughput.

When the drop rate of any server is less, then it gives higher throughput and Web server processes more requests. Now we developed the new architecture and new PC load balancing algorithm for the load balancing of web server system and compare to the existing load balancing algorithm these are the round robin and weighted lest connection for the comparative analysis result a web server load balancing in resultant paper.

Reference

[1] Mikael Aderson, “Introduction to web server modeling and control Research”, Technical report, Octuber 28, 2005.

[2] K. Ramana, A. Subramanyam and A. Ananda Rao, “Comparative Analysis of distributed web server system load balancing algorithm using Qualitative parameters”, VSRD International Journal of Computer Science and Information Technology, Vol. 1(8), 2011, pp. 592-600.

[3] Sndeep Sharma, Sarabjit Singh,and Meenakshi Sharma, “ Performance analysis of load balancing algorithm”, World Academy of science, Engineering and Technology, vol. 38, 2008.

[4] Zheng Wen, Lei Shi, Lin Qi and Lin Wei, “A Predictive Adaptive load Balancing Model”, In proceeding of the 9th IEEE International Conference on fuzzy and knowledge discovery (FSKD 2012) Cihna.

[5] O. Damani, P. Chung, and C. Kintala, “One IP: Technique for hosting a service on a cluster of machines ”, In proceeding of 41st IEEE computing society International Conference, pp.85-92, April 1998.

[6] Cardellini, V. Colajanni, M., and Yu, “Dynamic load balancing in geographically distributed heterogeneous web server”, in proceeding of the IEEE 18th International Conference on distributed computing system, Amsterdam, the Netherlands pp.295-302, 1998.

[7] T.L.Pao, J.B.Chen, “the Scalability of heterogeneous Dispatcher based web server load balancing architecture”, proceeding of the 7th International Conference on parallel and distributed computing, application and technology, pp.213-216, 2006.

[8] M. Kihl, A. Robbertsson, “Performance modeling and control of server system using non- linear control theory”, in proceeding of 18th Internation Teltraffic Congress, 2003.

[9] S. Parkes, N. Gandhi, J.Hellerstein, D.Tilbury, T. Jayram, and J. BIgus, “Using control theory to achieve service level object design performance management”, 2002.

[10] Anders Robertsson, Bjorn wittenmark and Maria Kihl, “Analysis and design of admission control in web server system”, IEEE, 2003.

Biographies

BRAJENDRA KUMAR[1], Received the B.Tech. degree in Information technology, Engineering from the University of Mahatama Gandhi Chitrakoot Ghramodya viswavidhyalya, Satna, M.P., in 2008, the M. tech degree in software Engineering(pursuing) from the University of RGPV, Bhopal, M.P., in 2014, bsingh.lnct@gmail.com[1]

DR. VINEET RICHHARIYA[2], HOD (CSE), LNCT, Bhoapl (M.P.), vineetrich100@gmail.com[2]