ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 14, May 2016

Distributed Secure Queuing in CDN Using Load Balancing and Automation with Prediction

Gowtham.C. B.Tech, M.E.,

J.Gift Lee Jones B.Tech.,M.E.,

***Abstract*— A content delivery network is a large network of interconnected servers and clients nodes that communicate with each other to transfer data between them. The main aim of a content delivery network is to provide the clients with service and data, that they request from the servers. A single server in a content delivery network can be connected with a number of servers simultaneously depending upon the number of clients in the network. And the servers in a content delivery network contain the same copy of data that is to be shared among the clients. The main aim of a content delivery network is to provide the clients with high availability of data and accessibility of data. In this paper, a content delivery network that includes the functions of queuing and load balancing is proposed to manage queue in a network and provide better service. The main area where a CDN fails in is providing security to the data in the servers. Hence in this proposal, a content delivery network that not only offers increased performance and network management, but also security to the data is put forth.**

***Index Terms*—Content delivery, load balance, queuing, security.**

I. Introduction

A content delivery network is a large network of interconnected servers and clients nodes that communicate with each other to transfer data between them [1]. The main aim of a content delivery network is to provide the clients with service and data that they request from the servers. A single server in a content delivery network can be connected with a number of clients simultaneously depending upon the number of clients in the network as shown in fig.1.1. And the servers in a content delivery network contain the same copy of data that is to be shared among the clients.

The main aim of a content delivery network is to provide the clients with high availability of data and accessibility of data. Although a content Delivery network provides high information availability and is aimed at reducing congestion in a network, it still suffers from congestion and low performance due to the increase in the request of information and the increase in the number of clients in the network. The poor performance may be due to the inability of the server to handle heavy traffic in the network and inability to serve a large number of clients at the same time. And though load balancing in used in the existing system of content delivery networks, it still fails to meet the request of the clients in a Network.

Although load balancing is used in the content delivery networks to reduce congestion in the available technology, it still does not help to avoid or reduce congestion in the network when the number of clients or the number of requests in a network increases. The reason for this is the increasing demand for data sharing in the internet where data and information is shed across various individuals from various distances. The access of data and the need of data from various sources across large distances in the internet has introduced a vital need for data than has to be transferred secure and efficiently.

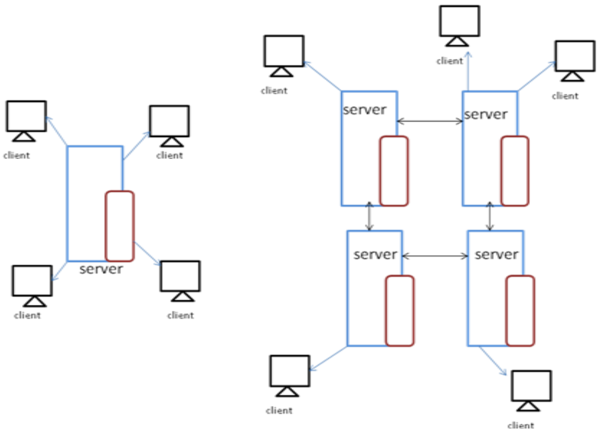

Fig 1: A simple network and a Content Delivery Network comparison.

All Rights Reserved © 2016 IJARBEST

In this work, a load balancing system that uses a queuing model to queue the requests from the clients in a content delivery network is proposed. The queuing models used are either rate based queuing or queue adjustment model. The load balancing algorithm used here is a dynamic load balancing algorithm which analyses the status of the servers and balances the load without any manual input. It is far more suitable in a network than a static load balancing algorithm which needs manual input to change between processors or devices for load balancing. The initial request from the client is sent to the server in the content delivery network. This request is received by the surrogate server that contains the load balancing unit and the queue.

II. LOAD BALANCING

Load balancing is a computer networking method for distributing workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units or disk drives. Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid of any one of the resources. Using multiple components with load balancing instead of a single component may increase reliability through redundancy. Load balancing is usually provided by dedicated software or hardware, such as a multilayer switch or a Domain Name System server process. One of the most commonly used applications of load balancing is to provide a single Internet service from multiple servers. Commonly, load-balanced systems include popular web sites, large Internet Relay Chat networks, high-bandwidth File Transfer Protocol sites, Network News Transfer Protocol (NNTP) servers and Domain Name System (DNS) servers.

In recent times, some load balancers have evolved to support database systems, these are called database load balancers. For Internet services, the load balancer is usually a software program that is listening on the port where external clients connect to access services. The load balancer forwards requests to one of the "backend" servers, which usually replies to the load balancer. This allows the load balancer to reply to the client without the client ever knowing about the internal separation of functions. It also prevents clients from contacting backend servers directly, which may have security benefits by hiding the structure of the internal network and preventing attacks on the kernel's network stack or unrelated services running on other ports.

a) Least loaded algorithm

b) Next neighbor load sharing.

In this work, the load balancing mechanism proposed is a combination of the least loaded algorithm and the next neighbour load sharing mechanism which combines the functions of both the algorithms.

The Least loaded algorithm finds the most least occupied server in the network and shares the load with that server.

In the next neighbor algorithm, the load balancer finds the next immediate server to the server that offers service and shares the load with the next server.

The load balancer module is placed in the surrogate server. The surrogate server is an intermediate server between the server and the client. The load balancer gets the status of the servers when the request is queued and sends the request to the least loaded server in the network.

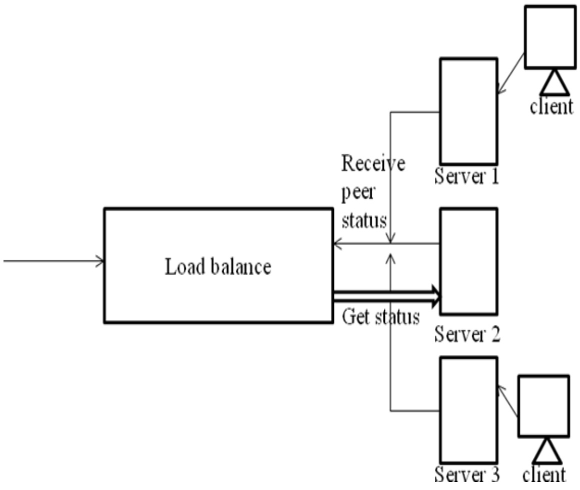

Fig 2 : Load balancing system

III. QUEUING

The Queuing models are implemented to aid in the process of load balancing. The main queues introduced in this method of managing the load by load balancing are[1] a) rate based queuing, b) queue adjustment model and c) hybrid queue model which is in theory the combination of rate and [2] queue adjustment

All Rights Reserved © 2016 IJARBEST

model. In rate based queuing the request from the clients is analysed and queued in a queue structure based on the rate of incoming requests. In queue adjustment model, the requests are organised in a [4] queue prior to sending them to the servers. And the hybrid queue adjustment model is a combination of both rate and queue adjustment. In this work, the rate adjustment model is implemented wherein the rate of incoming requests from the client to the server is analysed and based on that the requests are queued and sent to the servers. The incoming client requests are sent to the queue to control and reduce the rate so that it can be sent to the servers without any congestion, collision and packet dropping. The requests are queued in the surrogate server which contains the load balancer and the queue mechanism module. The load balancer retrieves the status of the servers in the content delivery network. Based on the statuses, the surrogate server establishes a connection with the least loaded node and sends the queued request to the server. By managing the incoming requests and sending them to the most free server, the clients can get better service and also this ensures that the servers can perform to their optimal capacity. Thus this helps to manage congestion better than conventional methods in a content delivery network.

IV. SECURITY

Here, the content delivery network is incorporated with a authentication module that aims at authenticating the various clients that request service from the servers. But since the process of authenticating each and every client every time they request service would be time consuming and in efficient to maintain the large amout of data about the clients, the system maintains the client details over a specific period of time window. This is such that is the if the duration of the request between the server and the client exceeds a certain time span, the server severs the ties with the client and requests new authentication from the client. Although the process of authenticating is time consuming comparatively, it does provide the feature of security in a content delivery network. Thus by a process of authentication and validation the security in a content delivery network can be improved.

The system architecture as shown in fig. 3 describes the architecture and the flow of control and data in the project elaborately. There are several servers that are interconnected with several nodes forming a content delivery network. Assuming, a client node requests for a data from a server. The request is first intercepted and received by a surrogate server. The load balancer is initiated first and it gathers the statuses of the servers in the network i.e., 0 or 1. Here, 0 is free and available to communicate and 1 is occupied and in service to a client in the network. Depending upon the status of the server as obtained by the load balancer, the request is scheduled in the queue to be sent to the available server say, server 3. Now the queued request is sent to the available server to get the required file from the server. In case if the server is not available or gets another client to be served during a communication then the client can be redirected to another free server in the network since all the servers contain the same copy of data that is to be shared. This type of content delivery networks are used in information portals and data hubs where the data is shared by a large number of clients in the network. By this the problem faced in data availability can be reduced or avoided in the content delivery network among the clients and the servers.

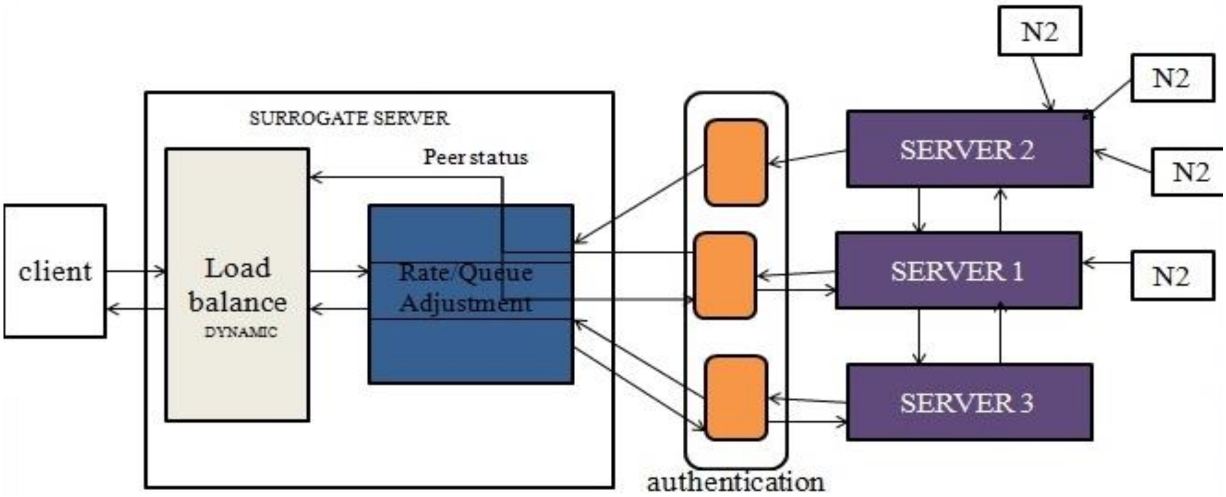

Fig 3: Architecture of the system

V. CONCLUSION

By using the methods of load balancing and network queues it is possible to thus regulate the congestion in a network and manage a content delivery network efficiently. This helps vastly in the current technology as a content delivery network plays a major role in being the backbone of the internet and data hubs. The queuing models can thus be used to manage the requests from the clients and help in accessing the servers much more efficiently. Although a model where the congestion in a content delivery network is avoided or reduced immensely can be implemented, there are still methods and ways to improve it.

Also by incorporating an energy saving mechanism to control the energy consumed by a content delivery mechanism, the energy consumption of a CDN can be reduced. This can be implemented by switching off the servers that are not in use at any given instance in the network.

Thus by incorporating a mechanism to improve the quality of security and by providing confidentiality to a content delivery network the work on a content delivery network can be improved.

REFERENCES

[1] Sabato Manfredi, Francesco Oliviero, Simon Pietro Romano . "A Distributed Control Law for Load Balancing in Content Delivery Networks".. IEEE transaction.Feb 2013

[2] Z. Zeng and B. Veeravalli . "Design and performance evaluation of queue-and-rate-adjustment dynamic load balancing policies for distributed networks". IEEE Transactions on computing, 2012

[3] Maggie Mashaly Paul J. Kühn . “Load Balancing in Cloud-based Content Delivery Networks using Adaptive Server Activation/Deactivation”.IEEE transactions on Networking, 2012.

[4] Abhinav Kamra, Huzur Saran, Sandeep Sen, Rajeev Shorey . ‘Fair adaptive bandwidth allocation: a rate control based active queue management discipline. IEEE transactions on Computer Networks 2012.

[5] James Aweya, michel ouellette, Delfin Y.Montuno . "A controlled theoretical approach to active queue management".. IEEE transaction Nov 2010.

[6] Ingmar Poese, Benjamin Frank, Bernhard Ager “Improving Content Delivery Using Provider- aided Distance Information”. IEEE transactions 2011.

[7] Vimal Mathew, Ramesh K. Sitaraman and Prashant Shenoy “Energy-aware load balancing in content delivery networks”.IEEE Transactions 2011.

[8] F. Oliviero, and S. P. Romano . "Distributed management for load balancing in content delivery networks.S. Manfredi, , IEEE GLOBECOM Workshop, Dec. 2010.

[9] Vimal Mathew, Ramesh K. Sitaraman and Prashant Shenoy .“Energy-aware load balancing in content delivery networks” ..IEEE Transactions 2011.

[10] Wentao Wang, Xiaozhong Geng , Qing Wang .“Design of a dynamic load balancing model for multiprocessor systems”, IEEE Transactions 2011.

[11] S.Manfredi, F. Oliviero, and S. P. Romano. "Distributed management for load balancing in content delivery networks., IEEE GLOBECOM Workshop, Dec. 2010

[12] H. Yin, X. Liu, G. Min, and C. Lin, ."Content delivery networks: A Bridge between emerging applications and future IP networks” .IEEE Transactions in computer networks, Aug. 2010.

[14] D. Cavendish, M. Gerla, and S. Mascolo. "A control theoretical approach to congestion control in packet networks". IEEE/ACM Transactions on Networking,October 2004

All Rights Reserved © 2016 IJARBEST