A Web Cluster Scheme using Distributed File Server in Internet Environments

Jun-Tak Han

Dept.of Computer Engineering
Hanzhong University, Kangwon, Korea

ABSTRACT

In this paper, we propose to new dispatcher method, which doesn't depend on an operating system of the server, and the direct routing method, by which a server answers a client's request at first hand. And, propose new web clustering scheme based on the contents on the web where web servers composed of cluster, with each different contents, answer client's request. The other purposes are to reduce overhead of the dispatcher through load balance, and to minimize the time to take in responding to a client's request. The performance of new web cluster scheme was improved by about 39% than that of the existing RR method. It was identified that the performance of the proposed web cluster method was extraordinary improved comparing with that of the existing RR method as a whole.

keywords : web cluster, dispatcher, load balancing, distributed file server

1. INTRODUCTION

The rapid growth of Internet on the WWW is increasing Internet traffic. Especially, it is verified that the importance of web on the internet is getting increased, considering that a sizable proportion of traffic is about a web service. With the increase of traffic, the way to raise the performance and availability of web server is further made much of and a lot of studies on the web clustering scheme, whose expandability and performance against price is superior, is proceeding to cope with the request for web service. The cluster scheme to improve a system's performance, beyond a limitation of parallel processing scheme, is to connect several servers connected to a network to one server. The cluster scheme based on the WWW service is called web server cluster. On the other hand, there are four methods to realize web cluster by the architecture; Broadcasting method[1], DNS(Round-Robin DNS) method[2], Direct Routing method[3] and Dispatcher method[4]. The existing structure of web cluster is composed of one dispatcher which transfers a client's request to a suitable web server and several server nodes with the same contents. It is a dispatcher way to deliver client's request for service from a dispatcher to the whole servers by a proper scheduling way,
not to be concentrated on a special web server. The above explanation is as the following Fig. 1.
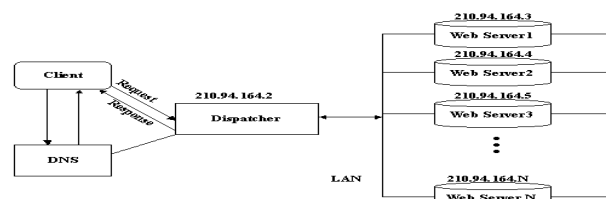

Fig. 1. The existing structure of dispatcher method

All client's requests can be controlled by dispatch method and they can be transferred to a web server through a processing of dispatch. In the other respect, it is too much of an expense to control and change all packets arriving at a dispatcher, and also a traffic jam can occur in the case that client's requests got increased[5]. In the case that a packet(HTML) requiring little data like a text information is transferred, it has little difficulty. However, in the case that packets requiring a great deal of multimedia data like moving picture file, music file and image file are transferred, the load of a special web server is increased and the load of the whole web servers gets unequal even though the load among the web servers is basically equal. Therefore, it spends much time to deal with the packets or any more services may not be possible in the serious case. According to, this paper proposes to new dispatcher method, which doesn't depend on an operating system of the server, and the direct routing method, by which a server answers a client's request at first hand.

2. THE PROPOSED NEW WEB SCHEME

The purpose of a web cluster scheme proposed in this paper is to reduce the overhead for traffic of dispatcher via load balancing or by providing each different content for each web server and to minimize the time to take to respond to a client's request, not focusing on effective load balancing from a dispatcher to each web server. The proposed structure of web server cluster is as the following Fig. 2.
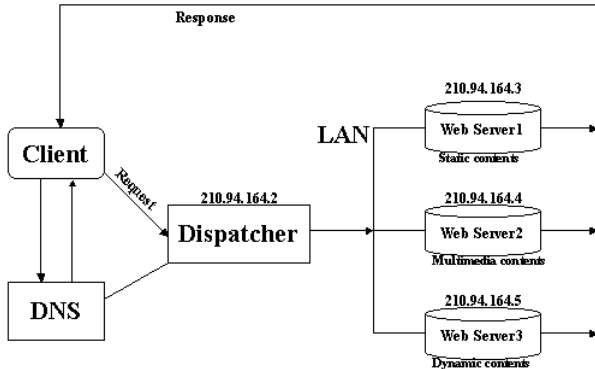

Fig. 2. The proposed structure of web server cluster

The proposed new web cluster scheme is composed of a dispatcher, which distributes all requests of clients to web servers with each corresponding content, and three web servers which respond to the request. As each server has its own IP address, clients recognize a dispatcher to be a server. They can access to a singe IP address of dispatcher and a web server itself gets hidden. The dispatcher processes by a unit of packet and delivers an IP packet transmitted from a client to the corresponding web server, when a destination address of IP packet header information is changed into an IP address of the real web server and the request is delivered to the real web server. When a requested web server responds to a client, it responds after its address has been converted into an address of dispatcher server. The process of this IP packet header information rewriting is as the following Fig. 3.
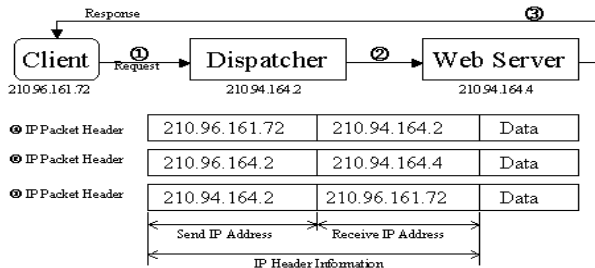

Fig. 3. The header information rewriting of IP Packet

On the other hand, as the dispatcher has an index file for the contents of web servers, it delivers the request from a client to the corresponding server. Monitoring web servers, the dispatcher server delivers the request to the content which the user wants, and each weighted web server delivers the load of special content to different web server via the dispatcher.

Three web servers have the individual different types of content. As Fig. 2 shows, Web Server 1 has a static content(HTML, Image, etc.) and Web Server 2 has a multimedia content(ASP, JSP, etc.). Web Server 3 has a dynamic content(ASP, JSP, etc.).

The corresponding web server responds to a client's request delivered to the dispatcher server according to each type. The web servers organized with cluster are composed of LAN.

2.1 The principal of operation

The operation of proposed web cluster is as the following Fig. 4. A client requests an address from DNS with URL including the name of dispatcher server and DNS answers the address of dispatcher server. The requested dispatcher server delivers the request to the web server with the corresponding content. The web server that has dealt with the request conceals its address and responds the dispatcher

address to the client. The client keeps requesting a service to the dispatcher, not knowing the real address of web server.
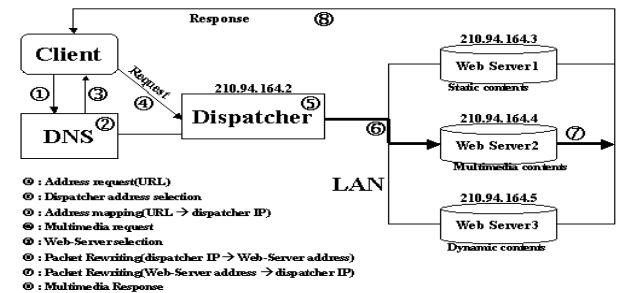

Fig. 4. The operating principal of new web server cluster

3. SIMULATION AND RESULT

One dispatcher and 3 web servers were installed to evaluate objectively the proposed new web cluster scheme. The specification of a system is as table 1. The network simulator of UC Berkely University, ns-2.1b8a in the operating system, WOW Linux 7.0 was used. The data set used to appraise its performance was sampled from the log files of web server of Hanzhong University. The performance of new web cluster scheme was compared to that of RR.

Table 1. The devices used for a test

Device Name	CPU	MEMORY	DISK
Dispatcher	Pentium 1.7GHz	256MB	80GB
Web Server1	Pentium 1.7GHz	256MB	80GB
Web Server2	Pentium 1.7GHz	256MB	80GB
Web Server3	Pentium 1.7GHz	256MB	80GB

3.1 The analysis of log files

To realize the new web cluster scheme proposed in this paper, the web server log files of Hanzhong University were analyzed. The number of total requests is 64,691 and the statistics of the requests by the content is as the following Fig. 5.
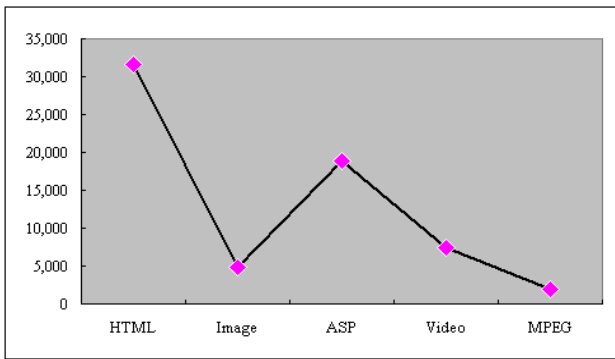
Fig. 5. The web server log analysis

Fig. 6. and Fig. 7. show the statistics having analyzed the pattern of contents and according to the statistics, 30,000, 40,000, 50,000 and 60,000 input data were made of.
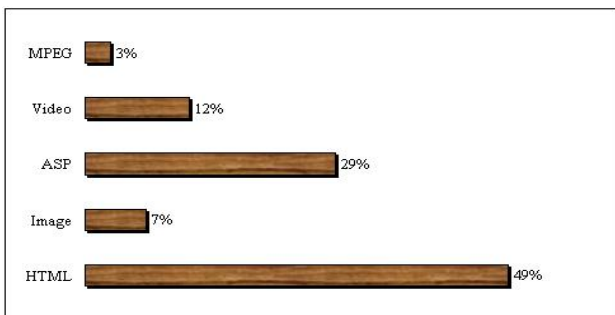

Fig. 6. The statistics of the web server request

As you can see the final analysis of the web servers of by the content in the Fig.7, they are analyzed as follows; The request for static contents (HTML, IMAGE etc) was 56%(The number of request, 36,396), for dynamic contents(ASP etc) was 29%(18,847), and for multimedia contents(Video, MPEG) was 15%(9,448).
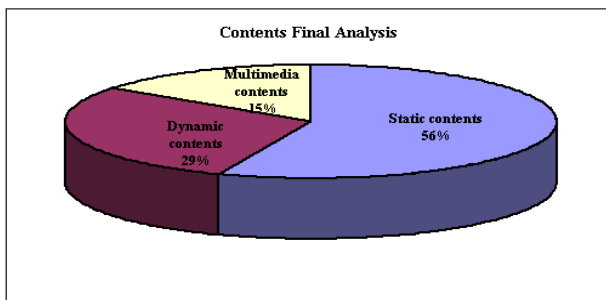

Fig. 7. The final analysis of the web servers by the contents

3.2 Performance evaluation

To compare new web cluster scheme with the existing RR method, the above mentioned log files, as input data, were compared and analyzed. The number of response per second for static contents, dynamic contents and multimedia contents

was observed for RR method and web cluster scheme when the number of requests was respectively 30,000, 40,000, 50,000 and 60,000 for their objectivity.
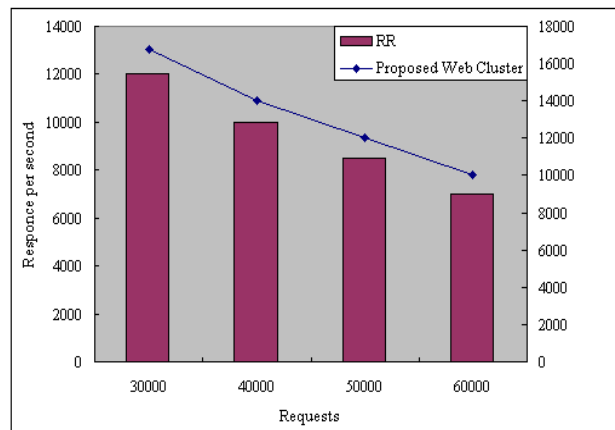

Fig. 8. The result of performance evaluation of RR method and the proposed web cluster scheme for static contents

As Fig. 8, Fig. 9 and Fig. 10 show, the performance of new web cluster scheme was improved by about 39% than that of the existing RR method.
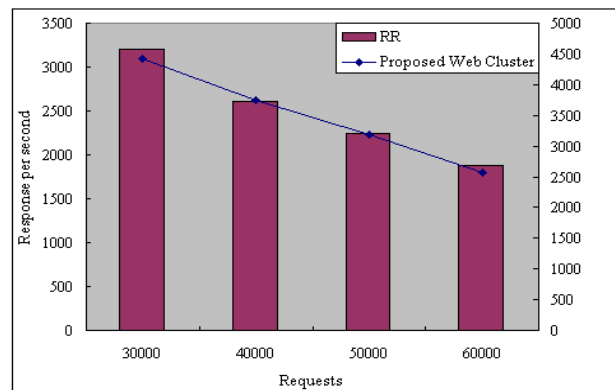

Fig. 9. The result of performance evaluation of RR method and the proposed web cluster scheme for dynamic contents

It was identified that the performance of the proposed web cluster method was extraordinary improved comparing to that of the existing RR method as a whole. That's why the short term solution for request had been achieved by the web cluster scheme, unlike the existing method. And also as the load was well distributed at a dispatcher, the increase
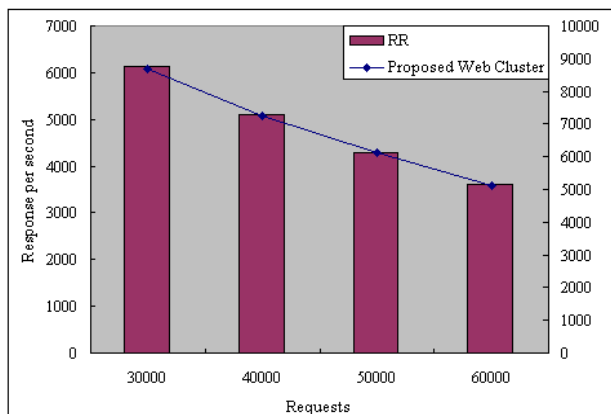of traffic of any special web servers could be prevented.

Fig. 10. The result of performance evaluation of RR method and the proposed web cluster scheme for multimedia contents

4. CONCLUSIONS

An exponential growth of Internet users results in overload of information supply servers. Besides, as the internet services including web become the important means for such business as electronic-commerce, the continuous service of internet and the fast service response for the clients' requests are getting important. While the existing method had focused on the study on load balancing to solve the problem, performance of the whole system could be lowered due to the complexity and irregularity of WWW if the load balancing web system was not properly organized. This paper proposes a model that a dispatcher delivers the client's request to the web server having the content suitable for it as the web servers composed of cluster have the individual different contents. The proposed web cluster method has the following virtues, compared to the existing solution.

Firstly, the web cluster scheme, as it has a function of load balancing, prevents the overload of a special web server and operates effectively the performance of the whole web cluster as its overhead for dispatcher is a little. Secondly, unlikely the service response of the existing dispatcher method, as it is delivered directly to a client regardless of dispatcher, the response time is reduced. Thirdly, as each web server has the individual different server, the client who wants a fast service can get a rapid response. The weakness of new web cluster scheme based on the contents on the web is that service can be paralyzed when a dispatcher is down. Therefore, the next goal of the study on new scheme will be what seeks for the scheme to realize the auxiliary dispatcher convertible when the main dispatcher is down.

REFERENCES

[1] J. Kangashafju, K. W. Ross, "A clustering structure for reliable multicasting" *Computer Communications and Networks*, 1999. Proceedings. Eight International Conference, pp.378-383, 1999.

[2] J. Kangasharju, K. W. Ross, "A replicated architecture for the Domain Name System" *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE computer and Communications Societies*, Vol.2, pp.660-669, 2000.

[3] Dongeun Kim, Cheol Ho Park, Deayeon Park, "Request rate adaptive dispatching architecture for scalable Internet server" *Cluster Computing, 2000. Proceedings. IEEE International Conference*, pp.289-296, 2000.

[4] M. Colajanni, P.S Yu, V. Cardellini, "Dynamic bad balancing in geographically distributed heterogeneous Web-servers" *Proc. of 18th IEEE Int'l,. Conf. on Distributed Computing Systems(ICDCS'98),*

[5] M. Colajanni, P.S Yu, V. Cardellini, "Dynamic bad balancing in geographically distributed heterogeneous Web-servers" *Proc. of 18th IEEE Int'l,. Conf. on Distributed Computing Systems(ICDCS'98),* pp.295-302, Amsterdam, The Netherlands, May 1998.

[6] Vivek S. Pai, Mohit Aron, Gaurav Banga, Michael Svendsen, Peter Druschel, Willy Zwaenepoel, Erich Nahum, "Locality-Aware Request Distribution in Cluster-based Network Services" *ACM 8th, ASPLOS* Oct. 1998.

[7] G. Hunt, E. Nahum, and J. Tracey, "Enabling content-based load distribution for scalable services" *Technical report, IBM T.J. Watson Research Center*, May. 2003.

[8] T. Schroeder, S. Goddard, and B. Ramamurthy, "Scalable Web server clustering technologies" *IEEE network*, pp.38-45, May. 2005.

[9] Rao, A., Lakshminarayanan, K., Surana, S., Karp, R., and Stoica, I, "Load Balancing in Structured P2P Systems" *Second International Workshop on Peer-to-Peer Systems*, 2003.

[10] J. Aspnes, J. Kirsch, and A. Krishnamurthy, "Load balancing and locality in range-queriable data structures" *In TwentyThird Annual ACM SIGACT-SIGOPS Symposium on PODC*, 2004.

[11] L. A. Barroso, J. Dean, and U. Hlzle, "Web Search for a Planet: The Google Cluster Architecture" *IEEE Micro*, 23(02):22–28, 2003.

Jun-Tak Han
He received the M.S. degree in Computer Science from Hanyang University, in 1994, and Ph.D. in Computer Engineering from Dongguk University, in 2003. Since 1995 he is an associate professor in Department of Computer Engineering, at Hanzhong University, Donghae, Republic of Korea.   From 2004 to 2006, he was chief at computer center of Hanzhong University. His main research interests include parallel/distributed processing, I/O disk system, and computer game engineering.