Reverse Proxy Comparison Analysis for Optimization of Web Server Services

Abdul Aziz¹, Auliya Hanif¹

¹ Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta, Jl. Prof. DR. G.A. Siwabessy, Kampus UI, Depok, 16425, Indonesia

E-mail: abdul.aziz@tik.pnj.ac.id

Abstract - Web is one technology that is currently widely used in various needs, both used for media campaigns, information systems, multimedia services, social networking and others. Web services provider server often encounter difficulties in servicing very high service requests from clients. The need for a solution in improving a web service, one of the right solutions is to add a reverse proxy to improve the performance of a web server by previously making a design and testing in implementing a reverse proxy. In its implementation, the system design and configuration process requires configuring a Squid application on a proxy server that acts as a web cache request coming from the client, then building multiple web servers located behind the proxy server where the site used in the web server is wordpress. By making a reverse proxy for web server this is expect to be a useful solution in optimizing the performance of a web server.

keywords - proxy, reverse proxy, web server.

1. INTRODUCTION

The rapid development of technology and information is influence by the increasing use of internet in Indonesia. Based on the Indonesia Internet Statistics Data 2016 survey conducted by APJII (Association of Indonesian Internet Service Providers), the number of internet users in Indonesia in 2016 was 132.7 million users or about 51.5% of the total Indonesian population of 256, 2 million. Compared to internet usage in 2014 of 88.1 million users, it can be conclude there is an increase of 44.6 million users within 2 years (2014 - 2016) [1].

Along with the increasing use of the internet every year then the traffic data and internet network traffic is increasing drastically, so the server that accommodates the site in serving Internet users have a high enough process load, especially web server that accommodate popular sites that are often accessed by the user Internet. It may even be that the web server cannot afford to cater for more clients' access requests because of its high processing load. This can cause the web server to be slow due to overload, and eventually become down. If there is down on the server then it can affect client satisfaction in using the web server service.

In solving this problem, web service providers can upgrade server hardware to a higher specification to deliver higher performance. However, if one day the traffic from the client increases again and the server can no longer serve the high processing load, then the server should be upgraded to higher performance and specification again, so this solution will only solve the problem for the short term. Especially for the providers of local contents of newly developed would be a problem if you have to upgrade servers that require a high enough cost it is used reverse proxy that can ease the performance of the web server so more cost-effective.

Reverse proxy is a method to reduce the process load on web server by using web cache technique and become a bridge between web server with internet or client accessing web service. Reverse proxy server speed up the response and download time of web pages. This can be the right solution for newly developed local content providers because in addition to reducing the cost of hardware upgrades from servers and saving on bandwidth, the performance of web servers will increase.

The reverse proxy server service varies depending on the web server being use. On certain web servers in the required additional tools in the implementation of its reverse proxy, but on other web server is only required to activate the method only. Before the reverse proxy server will be the solution and implemented then there needs to be comparison analysis on some web servers that use reverse proxy.

In this work, we compare reverse proxy on web server what is more effective to use seen from side of traffic, memory and cpu utilization. The scope of this work is limit by several aspects.

(1) The system is implement in a virtual machine using Kvm.
(2) The Operating System used is Ubuntu Linux 16.04 (Xenial Xerus).
(3) The compared web servers are Nginx, Apache, and OpenLiteSpeed.
(4) The tools used for testing are ApacheBenchmark on Apache and Weighttp on the Nginx and OpenLiteSpeed web servers.
Reverse proxy is a method to reduce the process load on web server by using web cache technique and become a bridge between web server with internet or client accessing web service. Reverse proxy server speed up the response and download time of web pages. This can be the right solution for newly developed local content providers because in addition to reduce the cost of hardware upgrades from servers and saving on bandwidth, the performance of web servers will increase.

Based on the background of problems that have been described above, it can be formulated a problem that can be used as a reference, namely: what web server is a more efficient in terms of resources and network traffic by using a comparison between three web servers using reverse proxy.

2. RELATED WORK

In this section, to optimize the web server, we provide some related work in three aspects: reverse proxy, proxy, and web server itself.

2.1. Reverse Proxy

Reverse Proxy using an input-data cleansing algorithm that uses MD5 hashing to curb SQLIA that attempts to attack through entry forms and uses regular expression to curb SQLIA's efforts through a System URL using a sanitization app on a proxy server that will sanitize the request before being forward to the primary server and databases [2].

Reverse Proxy on HA Cluster web proxy server can improve the performance of web server by serving the number of request 0.77% more and service time 30.68% faster than a single web server [3]. Web servers with reverse proxy can serve more demand than a single web server, and can minimize resource usage on the web server itself [4].

Nginx as a dynamic reverse proxy. In addition, based on the dynamic Reverse proxy framework, three optimization approaches are propose for better elasticity optimization. Three methods are Batch Request Committing (BRC), Batch File Processing (BFP) and In Memory Configuration (IMC). This study compares the maximum Throughput between original and optimized Case, then discusses the relationship between batch size and response time [5].

The system proposes a hybrid web caching proxy architecture to improve computer network performance, by integrating forward proxy caching techniques and reverse proxy caching techniques. Proxy forward is use to forward clients from clients on the internal network to the origin server of the content. Use forward proxy caching to reduce network traffic, reduce page load time during normal loading and loading in "peak" conditions, reduce dropped packets [6].

2.2. Proxy

Proxy servers are very effective for managing limited-capacity bandwidth using ACL and cache on Squid. Squid is one of the GPL licensed GPL applications (GNU Public License) and is a proxy server implementation that can store objects in the cache [7].

Because the proxy works on the application layer, the proxy server can run on many applications including HTTP Proxy or Web Proxy for HTTP or Web protocol, FTP Proxy, SMTP Proxy or POP Proxy, NNTP Proxy for Newsgroup, Real Audio or Real Video Proxy for multimedia streaming, IRC Proxy for Internet Relay Chat, and more. Each will only receive forward or forward a filter on the packets generated by the corresponding service [4].

Proxy specific apps have many configuration options. For example, web proxies can be configure to deny access to certain websites at certain times. Other proxies also, for example configurable only allow FTP downloads and do not allow certain users to play RealAudio files, prevent access to email servers before a certain date, and more [4].

Proxies can be understand as third parties standing in the middle between the two interconnected parties and serve as intermediaries. That way the first party and the second party are not directly related, but each corresponds to an intermediary, the proxy [8].

Proxy server is one solution that can be use, the management company will be able to limit the uses of bandwidth internet, internet usage settings and reduce or minimize virus attacks, worms, spyware and DDOS [9].

The Main Functions and Usefulness of the Proxy is Cache, Caching here is defined as the internet object storage mechanism (image or web page) of a web server that has been accessed [2], because the proxy acts as an intermediary, then the proxy gets the object first from the source to then forwarded to the actual requestor. This caching process is invisible to the client (transparent), because it is not visible to the client who actually gave the object he requested, whether the proxy took from his cache or the web server. From the client side, all replies are directly from the web server. In the caching process, the proxy also stores the object for itself in the provided disk space. So at some point the client will request a service to the web server with the same object on the storage, the proxy will immediately be able to provide the requested object, without having to ask again to the web server. If the requested object does not exist, the new proxy will ask the web server and provide it to the client [4]. Content stored on the hard disk is call a cache object that will be uses if the client returns to visit the web server. In the next visit, the browser will check the validation of the stored content, validate is done by comparing the existing content header in the cache object with that on the web server, if the content has
not been expired then the content will be displayed back to the browser [5].

2.3. Web Server

The web server provides the requested information by the computer or other client hardware (hardware) and serves every request that comes from anywhere. Information will be send by the server and will be receive and read by the client through the browser program. The basic principle of design itself should consider several things, among others, how the hardware conditions will be used, what software is selected, server settings and optimization, maximum access and traffic and management facilities [10].

Web server is a software server that becomes the backbone of the World Wide Web (WWW). Web servers wait for requests from clients using browsers such as netscape navigator, Internet Explorer, mozilla firefox, and other browser programs. If there is a request from the browser, then the web server will process the request and then give the results of the process of the desired data back to the browser [4], the most important thing in the process of making a web server is choosing which software will be uses as our web server [8]. That so we need the following considerations:

1. License of software to be used (freeware, shareware or commercial).
2. Ease of installation.
3. Ease of configuring.
4. Ease to add or change peripherals.
5. The amount of space needed to store the minimal files needed for the software to function properly.
6. Prospect of the software in the future.
7. Performance and consumption of resources used the software.
8. What facilities can be support by the software
9. Technical support (have sites or mailing lists to ask if a problem occurs).
10. Support platform (any type of operating system that can run the software).

3. METHODS AND PROCEDURES

In this section, the system design process of the built of reverse proxy is described, this section contains a description of the system is the concept of the reverse proxy itself, the workings of the system that contains the flow diagram of how the testing work on the reverse proxy, as well as specs system that contain the specifications of hardware and the software used in this study.

Reverse proxy system is used to improve web server performance. The method by reverse proxy to improve web server performance is using web cache method. Web caching in this case is define as the mechanism of storing objects from the Internet in the form of web pages that contain pictures, audio, video, etc. By using this method, the reverse proxy server acts as an intermediary between the client and the web server accessed by the client. Therefore reverse proxy server will get the object first from the source (web server) and then forwarded to the actual requester (client). The process of web caching is not visible from the client side (transparent proxy), because for the client it will not appear who the object giver is asking, whether from reverse proxy server that takes from its cache storage or from web server. Because of the client side, all replies will be visible directly from the web server.

In the web caching process, the reverse proxy server also stores the requested object itself for the available storage. If at any time the client requests a service to a web server with the same object and is already on the reverse proxy server storage, then reverse proxy server will directly provide the requested object to the client, without having to request the object again to the web server. However, if the object requested by the client is not contained in the reverse proxy server storage, then the reverse proxy server will request the object on the web server and give it to the client.

![Block Diagram of Optimization of Web Server](image)

Reverse proxy is implemented on three different types of web servers, namely Nginx, OpenLiteSpeed and Apache web server for comparison analysis. Different reverse proxy tools are uses on each web server. On the Apache web server the Squid tool is use and configured to be a reverse proxy server. On the Nginx and OpenLiteSpeed web servers only need to enable the module and the rproxy.xml template.

Initially the web server will be test by the client using ApacheBenchmark. Then the bash file is create to run the ApacheBenchmark command. Then the next stressing web server will start client tester. On the web server that has been installed network monitoring tools that cacti, cacti will read the process of stressing and will release the results but if the web server was not strongly received the stressing done by the client or tester, then the web server will experience overload and the web server will be down.
There is a benchmark application on a vm client or tester that works for web server stressing, benchmark applications on the Apache web server using ApacheBenchmark tools and on Nginx and OpenLiteSpeed web servers using the Weighttp tool. On the web servers side there are monitoring applications that cacti to monitor the network to see the amount of traffic and resources used by the web server. The flow diagram in Figure 2 explains how the testing process works.

![Figure 2. Flow Diagram of Testing Process](image)

The first step required to install and configure a web server on a virtual machine is to install the required packages. Figure 3 is a command line for installing packages needed to create Apache web server, the packages include Apache web server itself, php5.6 as the programming language used, mysql-server as database server used, Phpmyadmin as admin-panel of the mysql database, and php 5.6-mysql as the connection between the programming language and the database server.

```bash
[root@auliyahanif ~]# apt-get install apache2
[root@auliyahanif ~]# apt-get install php5.6
[root@auliyahanif ~]# apt-get install libapache2-mod-auth-mysql
[root@auliyahanif ~]# apt-get install php5.6-mysql
[root@auliyahanif ~]# apt-get install mysql-server
[root@auliyahanif ~]# apt-get install phpmyadmin
```

**Figure 3. Installing Apache Dependencies Package**

## 4. RESULTS AND DISCUSSION

The test is performed by activating the service server and then accessed by the client using the benchmark AB (apache benchmark) tool, which is the default utility on the apache web server, so apache benchmark is automatically install when we install apache web server. In Figure 4 the benchmark apache tool used in the client to quantify the benchmark performance of the web server cluster.

```
[root@client# ab -n 1000 -c 100 http://domain.com
```

**Figure 4. Command For Do the Benchmark**

Information:
- `-c 100` = number of concurrent connections, as many as 100 connections simultaneously.
- `-n 1000` = number of requests sent, total 1000 requests.
- `http://domain.com` is the domain of the website to be tested.

Figure 5 shows the result of stressing the web server using ApacheBencmark.

```
HTTP/1.1 200 OK
Date: Sun, 26 Jun 2022 14:05:56 GMT
Server: Apache/2.4.25 (Ubuntu)
Content-Length: 70
Content-Type: text/html

Welcome to domain.com!
```

**Figure 5. Benchmark Result**

From the results of comparison testing has been do, by benchmarking web servers that use reverse proxy with those who do not use reverse proxy. By testing several major web pages. The first test result is benchmarking Apache2 web server using and not using reverse proxy. Use the following parameters:

- Request: 1000
- Concurrent: 100 – 1000
- Page: `http://domain.co.id/`

```
#ab -c 1000 -n 1000 http://domain.co.id
```

Table 1
The second test result is benchmarking Nginx web server that uses and which does not use reverse proxy. Use the following parameters:

- Request: 1000
- Concurrency: 100 - 1000
- Page: http://domain.co.id/

```bash
#ab -c 100 -n 1000 http://domain.co.id
```

The third test result is benchmarking OpenLiteSpeed web server that uses and which does not use reverse proxy. Use the following parameters:

- Request: 1000
- Concurrency: 100 - 1000
- Page: http://domain.co.id/

```bash
#ab -c 100 -n 1000 http://domain.co.id
```
5. CONCLUSION

From the data tests result that have been done can be analyzed that Apache web server is a web server with memory consumption, CPU consumption, and the smallest traffic compared with two other web server that is Nginx and OpenLiteSpeed. This is because Apache uses additional tools to reverse proxy is squid and squid service is place on the front-end server in front of the actual web server.

From the results of the analysis in testing reverse proxy server, it can be conclude several things:

(1) Web server with reverse proxy can serve more demand than without using reverse proxy.

(2) Apache web server is a web server with memory consumption, CPU consumption, and the smallest traffic compared to two other web servers namely Nginx and OpenLiteSpeed.

REFERENCES


