Scientific
Research

Proxy Server Experiment and Network Security with Changing Nature of the Web

Olatunde Abiona[1], Adeniran Oluwaranti[2], Ayodeji Oluwatope[2], Surura Bello[2], Clement Onime[3], Mistura Sanni[2], Lawrence Kehinde[4]

[1]Department of Computer Information Systems, Indiana University Northwest, Garry, USA
[2]Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
[3]Information and Communication Technology Section, Abdus Salam International Centre for Theoretical Physics, Trieste, Italy
[4]Department of Electrical and Electronic Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
Email: oabiona@iun.edu, aranti@oauife.edu.ng, aoluwato@oauife.edu.ng, apinkebello@yahoo.com, onime@ictp.it, misturasanni@gmail.com, lokehinde@oauife.edu.ng

Abstract

The total reliance on internet connectivity and World Wide Web (WWW) based services is forcing many organizations to look for alternative solutions for providing adequate access and response time to the demand of their ever increasing users. A typical solution is to increase the bandwidth; this can be achieved with additional cost, but this solution does not scale nor decrease users perceived response time. Another concern is the security of their network. An alternative scalable solution is to deploy a proxy server to provide adequate access and improve response time as well as provide some level of security for clients using the network. While some studies have reported performance increase due to the use of proxy servers, one study has reported performance decrease due to proxy server. We then conducted a six-month proxy server experiment. During this period, we collected access logs from three different proxy servers and analyzed these logs with Webalizer a web server log file analysis program. After a few years, in September 2010, we collected log files from another proxy server, analyzed the logs using Webalizer and compared our results. The result of the analysis showed that the hit rate of the proxy servers ranged between 21% - 39% and over 70% of web pages were dynamic. Furthermore clients accessing the internet through a proxy server are more secured. We then conclude that although the nature of the web is changing, the proxy server is still capable of improving performance by decreasing response time perceived by web clients and improved network security.

Keywords

Proxy Server, Network Security, Hit Ratio, Webalizer, Proxy Log Analysis

1. Introduction

Many organizations today rely heavily on the use of the internet and the WWW; this has open doors for network administrators to acquire skills to manage the ever growing demand for access and good response time. A typical solution to providing access and good response time is to increase the bandwidth; this is not a scalable option. An alternative solution is to deploy proxy servers to service the ever increasing request of users.

A proxy server is a server that sits between a client application, such as a web browser, and a real server. It intercepts all requests to the real server to see if it can fulfill the requests itself. If not, it forwards the request to the real server. A proxy server can improve network performance by functioning as a caching server. Most Internet Service Provider (ISP) and organizations have been installing proxy caches to reduce bandwidth and decrease the latency to their users [1]-[5]. The performance increase due to proxy servers has been widely reported; however, a study reports that proxy servers actually decrease performance [6]. A pertinent question that comes to our mind is that since the web is evolving from static to dynamic information repository, is there a future for the caching proxy server?

In order to further understand the nature of proxy server and how it can be used to provide improved access and response time to a large number of users requesting same object from the cache, we conducted a proxy server experiment. A non-intrusive network traffic monitoring system was setup in [7] to collect access logs from three proxy servers, for a period of five months to three years. These access logs were analyzed using Webalizer. The three proxy servers are institutional web proxy cache. Two of the proxy servers are on the academic network, the Obafemi Awolowo University, Ile-Ife, Nigeria (OAU), the Indiana University Northwest Computer Networking Lab in HH226, Gary Indiana (IUN). The third proxy is on the Wide Area Network of the International Centre for Theoretical Physics, Trieste, Italy (ICTP).

The rest of the paper is organized as follows. In Section 2, we review related work, followed by data collection in Section 3. In Section 4, we perform access log analysis on raw data and reduced data. In Section 5, we present the results of our analysis for each caching proxy server and finally, in Section 6, we conclude the paper.

2. Related Work

Caching can be applied at several locations, namely at the web client, web server and within the network (proxy servers) [8]-[10]. Caching proxy server has gained popularity on the Internet, due to their ability to keep local copies of documents requested by web clients and using them to satisfy future request for same document. This can save bandwidth and reduce delays perceived by web users.

Several studies have reported performance increase due to proxy servers. One of the major functions of a caching proxy server is to decrease access time. The result of a study in [11] showed that the average response time of a hit may be five times smaller than a miss. A 20% to 25% improvement in user perceived response time was reported in [12] [13]. Research on the effectiveness of proxy caching is very active. A study at Virginia Tech has shown that hit rates of 30% to 50% can be achieved by a caching proxy [14]. Other studies gave a range from between 20% to 60% hit rate [9] [11] [15] [16] and [17] reported hit rate of between 10% to 40% for a three level caching hierarchy, and about 35% to 40% for a university-level web proxy cache.

However, a study conducted in [6] reported a hit rate of 4%, which shows a decrease in performance. The reason for this decrease in performance was traced to the changing nature of the web, *i.e.* the web is evolving from static nature to dynamic repository. Furthermore, research into the ability of proxy servers to cache video was reported in [6]. In the last few years there have been research efforts to improve multi-level proxy cache configuration [18]-[21]. Other factors that may improve proxy cache performance are the replacement polices used by the cache and the workload characteristics. The results of [17] showed that combining different replacement polices at different levels of the cache can improve the performance of a caching hierarchy. Finally the results of [22] showed that the cache replacement polices are sensitive to Zipf slope, temporal locality, and correlation between file size and popularity but relatively insensitive to one-timers, and heavy tail index.

3. Raw Data Collection

We collected access logs from three proxy servers located at three different locations. Two of the proxy servers are located at the Obafemi Awolowo University, Ile-Ife, Nigeria and Indiana University Northwest, Gary, Indiana computer networking lab. The third proxy server is located at the International Center for theoretical Physics,

Trieste, Italy; we refer to the proxies as follows:

- ASOJU used by the OAU academic network;
- IUN used by only the students in computer networking lab HH226;
- ICTP used by the ICTP network.

ASOJU continuously recorded access log on a daily basis for six months, details can be found in [7], The IUN records proxy logs during the academic year (August-December and January-May) for a period of three years, while ICTP proxy server had only one month of access log. Two of the proxies are institutional-level proxy servers while the third is only used by students in the networking lab HH226.

4. Access Log Analysis

4.1. Raw Data Analysis

Webalizer [23] is capable of generating reports on a monthly basis and also a summary report for the entire period. We have five months summary, from September 30, 2006 to February 28, 2007 for the first OAU proxy server which is referred to as ASOJU access log. The five-month ASOJU access log recorded a total of 153,125,959 requests in 107 days of activity. The access logs for 45 days were not available due to down time and power outages. Similarly, we have three years of access log from September 2010 to October 2013 for the IUN proxy server which is referred to as IUN access log. The Three years IUN access log recorded a total of 62,675,342 requests in 210 days of activity. The access log was only collected when students are using the lab during the semester. Hence the need to collect log files for a longer period since the lab is only in use three months in a year. The eight days ICTP access log referred to as ICTP recorded a total of 5,458,868 requests in 8 days.

Table 1 provides a summary of the access logs for the three proxy servers studied. ASOJU has the highest activity in terms of number of request per day and also the highest average volume of bytes transferred per day.

In this study we are interested in requests for the transfer of web documents. Hence we study the response code in the access logs for all web requests. The breakdown of the HTTP reply code as a percentage of the total request is shown in **Table 2**. Web proxy server can provide many possible responses to web client [24]. Here are some response code and their corresponding meaning: The 200 series response code means a valid document was made available to the client, 300 series means redirection, 400 series means client error and 500 series means server error.

4.2. Raw Data Reduction Analysis

The access log recorded the amount of data transferred regardless of the source (*i.e.* from proxy cache, another cache or origin server). To know the actual workload of a proxy server, we consider all requests resulting in the

Table 1. Summary of proxy access logs (raw data).

Item	ASOJU	IUN	ICTP
Access log duration	107 days	210 days	8 days
Start date	Sept 30 2006	Sept 2010	Feb 11 2007
End date	Feb 28 2007	Oct 2013	Feb 18 2007
Total request	153,129,674	62,675,324	5,458,868
Avg request/day	1,431,118	298,453	682,359
Total bytes transferred (GB)	1637.4	495.9	86.3
Avg bytes/day (GB)	15.30	2.36	10.79

Table 2. Breakdown of HTTP response code.

Response code	ASOJU (%)	IUN (%)	ICTP (%)
200 series	38.34	67.00	45.20
300 series	12.30	20.00	28.00
400 series	46.00	9.00	2.70
500 series	2.16	1.00	0.90
Undefined	1.20	3.00	23.20
Total	100	100	100

documents being accessed from the origin server without an intermediate proxy. The objective is to evaluate the effectiveness of proxy caching.

Suppose a client using a proxy makes requests r_1, r_2, $\cdots$, r_n to pages, if a page has F objects out of which C can be obtained from the cache and W from the origin server. Total request R will be:

$$R = \sum_{i=1}^{n} r_i.$$

But not all requests will bring back data. Hence, all requests that will result in data transfer will be,

$$F = \sum_{i=1}^{n} W_i + \sum_{i=1}^{n} C_i.$$

So we can compute the document hit ratio (DHR) and byte hit ratio (BHR) as,

$$\text{DHR} = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} W_i + \sum_{i=1}^{n} C_i}, \quad \text{and} \quad \text{BHR} = \frac{\text{Cache byte}}{\text{Total byte}}$$

Cache byte = the no of bytes transferred from the cache;
Total byte = the total no of bytes transferred.

For DHR we only considered 200 and 300 series of response, in order to consider only successful transfer of documents to requesting clients. For BHR we did not consider the 400 series (client error). **Table 3** summarizes the reduced access logs for the three proxies. Based on the average number of request seen by each proxy server per day, ASOJU has the highest activity while IUN and ICTP have about the same activity. The successful transfer accounted for 45% to 87% while the total bytes transferred accounted for 64% to 89% similar to the observation in [19]. Other values on the table were calculated. The two performance metrics used in this study to evaluate the performance of the proxy servers are DHR and BHR.

5. Results

We observe that the total requests in the reduced data for ASOJU is smaller, this is expected since about 46% of the total request are error due to client authentication see **Table 2**. This is possible because ASOJU runs proxy authentication. Again about 60% to 78% of the requests are for dynamic pages that cannot be serviced by the proxy server. These observations support the fact that the web is fast changing from static nature to dynamic information repository [6]. However, the DHR range from 21% to over 38% for the three proxy servers analyzed in the study, these results are similar to the results obtained in [9] [11] [14]-[16]. Similarly, the BHR range from 21% to 29% for the three proxy servers. This result is also comparable with [11]. Since The ICTP data was only collected for only 8 days in the month, we can only plot the graphs of the hit ratios for ASOJU and IUN using the reduced data for a six-month period.
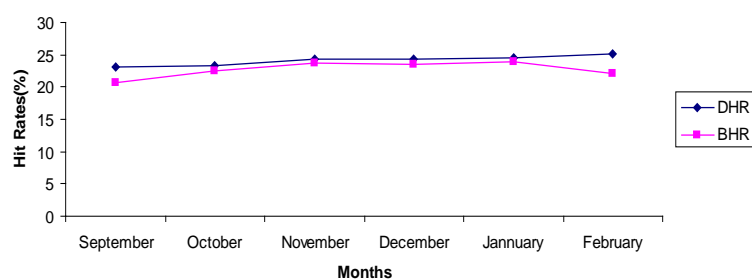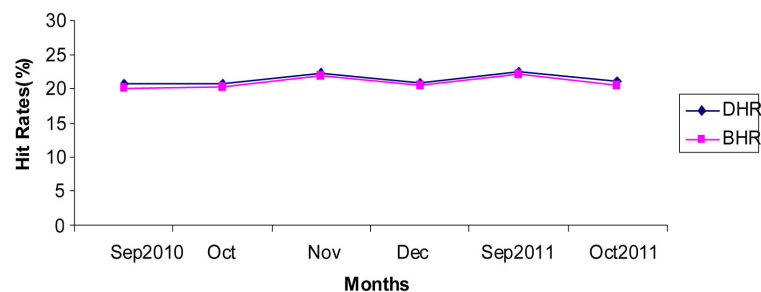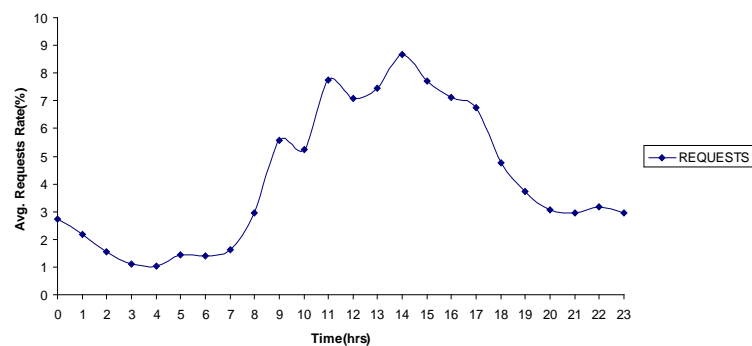
Figure 1 and **Figure 2** show that both hit ratios for ASOJU and IUN are not affected by the volume of the workloads across the months. We further study this observation on monthly hourly raw data. We are unable to generate the hourly reduced data, since the breakdown of the HTTP response used for generating the reduced data can only be obtained for monthly data. We study the monthly variations of the mean hit ratios across the hours of the day for the three proxy servers. The *y* error bars on the graph shows the variability of the hit ratios across the hours. We observe that our hit ratios in the following monthly graphs are relatively lower, varying in the 2% to 8% range. This is expected since the raw data contains client errors that were not removed. We also plot the mean monthly requests for the three proxy servers, in order to identify the peak periods of the day for the servers, since it varies.

Figure 3 shows the mean monthly hourly requests for ASOJU, the high usage periods (peak periods) are 9 hrs to 17 hrs and the low usage periods are 18 hrs to 8 hrs. This graph shows a typical work or social pattern in the environment. The traffic volume rises steadily with some deeps indicating break periods and fall steadily during the close of work for the day. It gives a representation of the user's access pattern. The graph shows that monthly hourly requests follow a normal distribution.

Figure 4 and **Figure 5** show the variations in the monthly average hit ratios for ASOJU. Both hit ratios

Table 3. Summary of proxy access logs (reduced data).

Item	ASOJU	IUN	ICTP
Total request	77,580,469	88,922,534	3,998,971
Avg request/day	725,051	499,565	499,871
Total bytes transferred (GB)	884.2	878.0	84.0
Avg bytes/day (GB)	8.26	4.93	10.5
Total internet requests	58,707,185	69,945,197	2,469,348
Total cache requests	18,873,284	18,977,337	1,529,623
Total cache byte (GB)	108.7	184.4	23.5
DHR (%)	24.3	21.7	38.3
Cache miss rate (%)	75.7	78.3	61.7
BHR (%)	22.8	21.4	28.8

Figure 1. ASOJU hit ratios for the reduced data.

Figure 2. IUN hit ratios for the reduced data.

Figure 3. Mean hourly requests for ASOJU.

ASOJU Hourly DHR (Oct.'06 to Feb.'07)

Figure 4. Variation in DHR for ASOJU.

ASOJU Hourly BHR (Oct.'06 to Feb.'07)

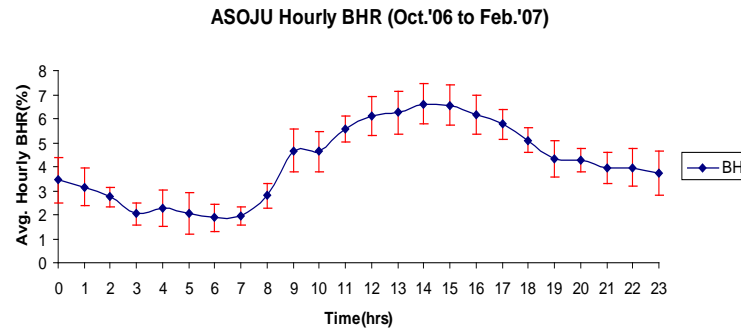**Figure 5.** Variation in BHR for ASOJU.

follow a similar trend, the standard deviation shown by the *y* error bars have a high dispersion for both ratios during the peak periods. This is expected since the traffic intensity increases during the peak periods. The variation of the DHR is higher; this is a reflection of the replacement algorithm and size of objects cached by the proxy server. This particular proxy is configured to cache small objects. Hence higher values of DHR, this will result in faster response time for the users.

Figure 6 shows the coefficient of variation (COV) for ASOJU hit ratios. The hit ratios show low variations during the peak periods (9 hrs to 15 hrs). This shows that neither ratio depend on traffic intensity.

Figure 7 shows the mean monthly hourly requests for IUN, the high usage periods (peak periods) are 0 hrs to 17 hrs and the low usage periods are 18 hrs to 20 hrs. This graph shows a typical access pattern for a student lab, the traffic volume is high for most time of the day with a small deep and then rise again. This pattern is however different from the access pattern of an academic network which has a high traffic during office hours (8 am - 5 pm).

Figure 8 and **Figure 9** show the variations in the monthly average hit ratios for IUN. Again, both hit ratios follow a similar trend, the standard deviation shown by the *y* error bars have a high dispersion for both ratios during the peak periods. This is expected since the traffic intensity increases during the peak periods. The variation of the BHR is higher; this is a reflection of the replacement algorithm and size of objects cached by the proxy server. This particular proxy is configured to cache large objects. Hence higher values of BHR, this will result in more bandwidth savings for the network.

Figure 10 shows the coefficient of variation for IUN hit ratios. Similarly, the hit ratios show low variations during the peak periods (0 hrs to 17 hrs). Again, this implies that neither ratio depend on traffic intensity.

Figure 11 shows the mean hourly requests for ICTP, the high usage periods (peak periods) are 8 hrs to 23 hrs and the low usage periods are 0 hrs to 7 hrs. The graph shows a typical social or work pattern. The traffic volume rises steadily with some deeps indicating break periods and fall slightly and remain high for the duration of the peak period. The graph shows the users access pattern.

Figure 12 shows the effect of traffic intensity on the ICTP hit ratios. Similarly, the hit ratios show low variations during the peak periods (8 hrs to 23 hrs). Again, this implies that neither ratio depend on traffic intensity. One may expect that when the number of client requests increases (peak periods) so does the number of hits. However, this is not the case since the peak hours user population has access patterns different from light load hours.
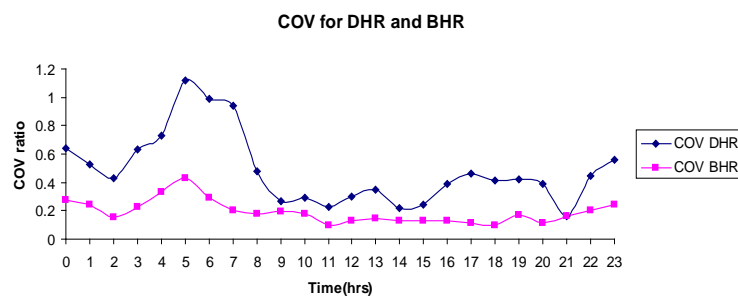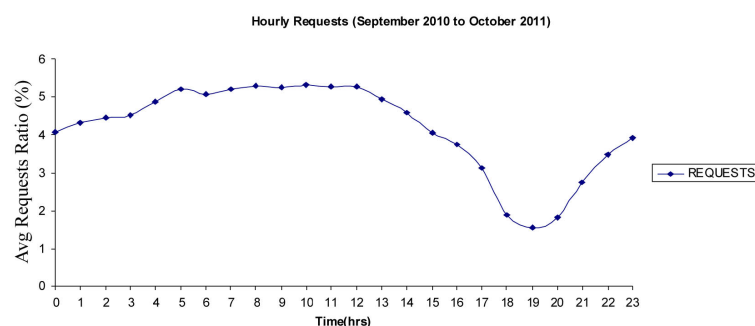
COV for DHR and BHR

Figure 6. COV for ASOJU hit ratios.

Hourly Requests (September 2010 to October 2011)

Figure 7. Mean hourly requests for IUN.

IUN Hourly DHR (Sept 2010 to Oct 2011)
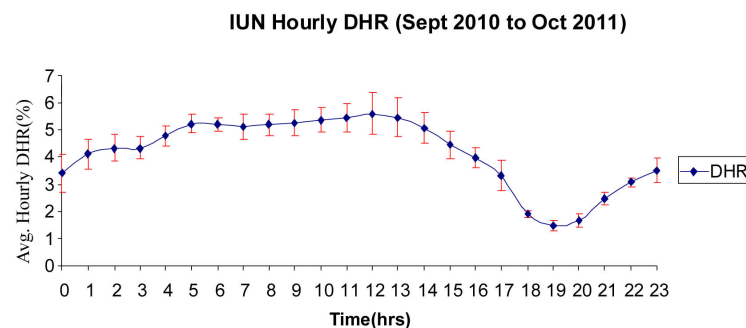

Figure 8. Variation in DHR for IUN.

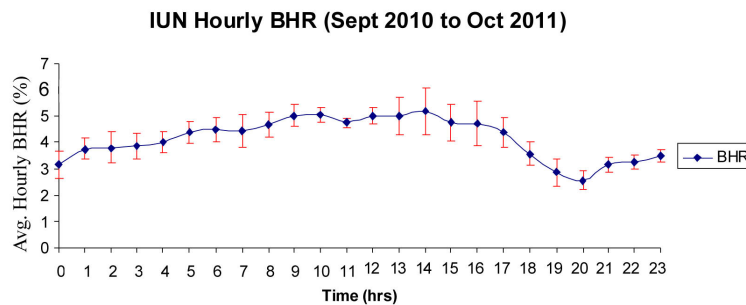**IUN Hourly BHR (Sept 2010 to Oct 2011)**

Figure 9. Variation in BHR for IUN.

The study also shows that the proxy server provides better security for clients accessing the internet using a proxy server. Most proxy server run dynamic host configuration protocol (DHCP), a service that provide clients with Internet protocol (IP) addresses that are private and using masquerading or network address translation, the clients can access the internet. From the outside, only the proxy server is visible. All clients using the proxy server are protected from attack, since they are not visible from outside the network. Only the proxy server
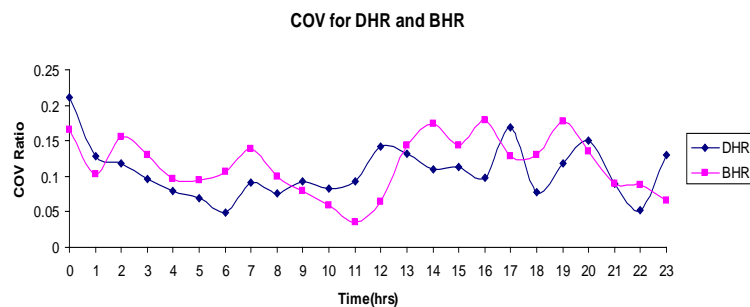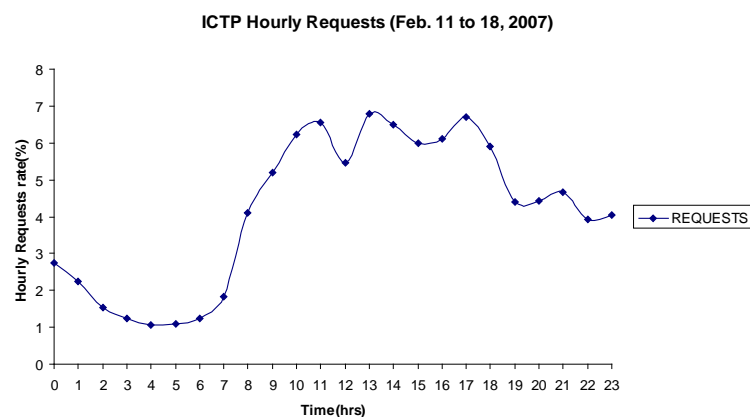
COV for DHR and BHR

Figure 10. COV for IUN hit ratios.

ICTP Hourly Requests (Feb. 11 to 18, 2007)

Figure 11. Mean hourly requests for ICTP.

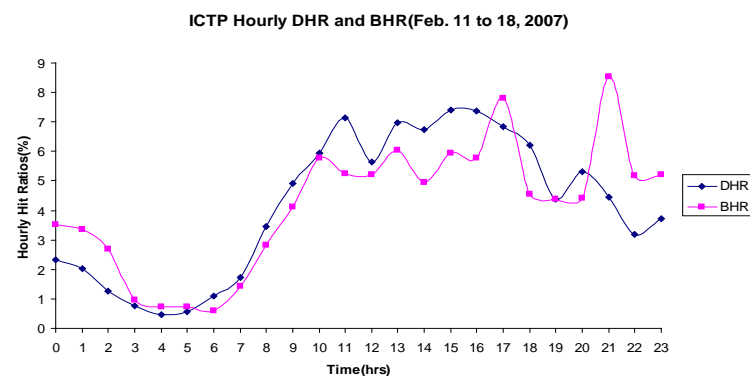**ICTP Hourly DHR and BHR(Feb. 11 to 18, 2007)**

Figure 12. Effect of traffic intensity on the hit ratios.

has a public IP address. We tried to attack clients behind the proxy server with no success. This technique shows that the proxy server provides a layer of security for clients accessing the internet using a proxy server.

6. Conclusion

This paper presents an experiment to determine the effectiveness of proxy servers and security provided by using proxy servers. We are also interested to know how the changing nature of the web has affected the performance of proxy servers and level of security provided by proxy server. We conducted a six-month proxy server experiment to know the performance of proxy servers. Access logs of varying durations were collected, from the three different proxy servers to see if it would have any effect on our results. We analyzed the logs using Webalizer. Two performance parameters—DHR and BHR—were used to evaluate the performance of proxy servers. We compute DHR and BHR for the duration of the study, and we also compute DHR and BHR for monthly and

hourly traffic to study the effect of traffic intensity on proxy server performance. The result shows a hit rate of about 21% to 38% and a byte rate of 21% to 28%, and the *y* error bar graphs show a high variation during the peak periods, while the COV graph shows a low or constant variation during the peak periods indicating that neither hit ratios depend on traffic load. The result shows that good performance can be achieved using proxy servers. Although the web is changing from the static nature to dynamic information repository, proxy servers actually improve performance and provide better security despite the changing nature of the web. In the future we hope to look into further enhancing security using honey pots and honey nets. We plan to investigate the cyclic multicast engine and proxy server as a possible technique to improve proxy server performance.

References

[1] Baentsch, M., Barum, L., Molter, G., Rothkugel, S. and Sturm, P. (1997) Enhancing the Web's Infrastructure: From Caching to Replication. *IEEE Internet Computing*, **1**, 18-27. http://dx.doi.org/10.1109/4236.601083

[2] Bestavros, A., Carter, R., Crovella, M., Cunha, C., Heddaya, A. and Mirdad, S. (1995) Application-Level Document Caching in the Internet. *Proceedings of the* 2*nd International Workshop on Services in Distributed and Networked Environments* (*SDNE*'95), Whistler, 166-173.

[3] Cohen, E., Krishnamurthy, B. and Rexford, J. (1998) Improving End-to-End Performance of the Web Using Server Volumes and Proxy Filters. *Proceedings of ACM SIGCOMM*'98 *Conference*, Vancouver, 241-253.

[4] Baentsch, M., Baum, L., Molter, G., Rothkugel, S. and Sturm, P. (1997) World Wide Web Caching: The Application-Level View of the Internet. *IEEE Communications Magazine*, **35**, 170-178. http://dx.doi.org/10.1109/35.587725

[5] Zhang, L., Floyd, S. and Jacobson, V. (1997) Adaptive Web Caching. *Proceedings of the NLANR Web Caching Workshop*, Boulder.

[6] Howard, R. and Jansen, B.J. (1998) A Proxy Server Experiment: An Indication of the Changing Nature of the Web. *Proceedings of the* 7*th International Conference on computer Communications and Networks* (*ICCCN*'98), Washington DC, 646-649.

[7] Abiona, O.O., Onime, C.E., Oluwaranti, A.I., Adagunodo, E.R., Kehinde, L.O. and Radicella, S.M. (2006) Development of a Non Intrusive Network Traffic Monitoring and Analysis System. *African Journal of Science and Technology*, **7**, 1-17.

[8] Rousskov, A. and Soloviev, V. (1998) On Performance of Caching Proxies. *Proceedings of the ACM SIGMETRICS Conference*.

[9] Caceres, R., Douglis, F., Feldmann, A., Glass, G. and Rabinovich, M. (1998) Web Proxy Caching: The Devil Is in the Details. *ACM Performance Evaluation Review*, **26**, 11-15. http://dx.doi.org/10.1145/306225.306230

[10] Abdulla, G., Fox, E., Abrams, M. and Williams, S. (1997) WWW Proxy Traffic Characterization with Application to Caching. Technical Report TR-97-03, Computer Science Department, Virginia Tech.

[11] Rousskov, A. and Soloviev, V. (1998) On Performance of Caching Proxies. *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, **26**, 272-273.

[12] DiDio, L. (1997) Proxy Servers Gain User Appeal. *Computerworld*, **31**, 16.

[13] Machlis, S. (1997) Planning Blunts Web Traffic Spikes. *Computerworld*, **31**, 6.

[14] Williams, S., Abrams, M., Standridge, C.R., Abdulla, G. and Fox, E.A. (1997) Removal Policies in Network Caches for World-Wide Web Documents. *Proceedings of the ACM SIGCOMM Computer Communication Review*, **26**, 293-305.

[15] Abrams, M., Stanbridge, C., Abdulla, G., Williams, S. and Fox, E. (1995) Caching Proxies: Limitations and Potentials. Boston. http://ei.cs.vt.edu/~succeed/WWW4/WWW4.html

[16] Glassman, S. (1994) A Caching Relay for the World-Wide Web. 1*st International World-Wide Web Conference*, Geneva, 25-27 May 1994, 69-76.

[17] Busari, M. and Williamson, C.L. (2001) Simulation Evaluation of a Heterogeneous Web Proxy Caching Hierarchy. *Proceedings of the IEEE MASCOTS*, Cincinnati, 15-18 August 2001, 379-388.

[18] Fan, L., Cao, P., Almeida, J. and Broder, A. (1998) Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol. *Proceedings of ACM SIGCOMM*'98 *Conference*, Vancouver, September 1998, 254-265.

[19] Mahanti, A., Williamson, C. and Eager, D. (2000) Traffic Analysis of a Web Proxy Caching Hierarchy. *IEEE Network*, **14**, 16-23. http://dx.doi.org/10.1109/65.844496

[20] Yu, P. and MacNair, E. (1998) Performance Study of a Collaborative Method for Hierarchical Caching in Proxy Servers. *Proceedings of World-Wide Web Conference*, Brisbane, 14-18 April 1998, 215-224.

[21] Tewari, R., Dahlin, M., Vin, H. and Kay, J. (1999) Beyond Hierarchies: Design Considerations for Distributed Caching on the Internet. *Proceedings of the* 19*th International Conference on Distributed Computing Systems*, Austin.

[22] Busari, M. and Williamson, C.L. (2001) On the Sensitivity of Web Proxy Cache Performance to Workload Characteristics. *Proceedings of the IEEE INFOCOM*, Anchorage, 22-26 April 2001, 1225-1234.

[23] The Webalizer Home Page. http://www.mrunix.net/webalizer/

[24] Vass, J., Harwell, J., Bharadvaj, H. and Joshi, A. (1998) The World Wide Web: Everything You (N)ever Wanted to Know about Its Servers. *IEEE Potentials*, **17**, 33-37. http://dx.doi.org/10.1109/45.721730

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.