

NEW FRAMEWORK FOR WEB PROXY SERVER THROUGH CLUSTER BASED PREFETCHING

Niklesh Rathore¹, Ms. Swati Tiwari²,

PG Scholar, Computer Science & Engineering Department, IPS Academy, Indore (M.P.)¹

Assistant Professor, Computer Science & Engineering Department, IPS Academy, Indore (M.P.)²

ABSTRACT

Web caching is the technique of storing a copy of the data that has been requested by the client. Web caching is used to reduce the network traffic by caching web pages at the proxy server level. Now a day's caching alone is not sufficient because of World Wide Web has evolved rapidly from a simple information-sharing mechanism. This mechanism offer only static text and images to a rich assortment of dynamic and interactive services, such as video/audio conferencing, e-commerce and distance learning. Web is demanding to improve the cache performance. If we use the prefetching technique with caching then the performance of cache is improved. Prefetching is the process of bringing data from the web server into the web cache before it is needed. When the client needs data, then instead of waiting for the responses from the memory, it can directly access the data from the cache, thus the response time of the user request is reduced. We have proposed technique Enhanced Dynamic Web Caching for Scalability & Metadata Management using association rules and markov model. Our proposed Prefetching techniques are useful to reduce the latency, control the network traffic and server load and get better the user approval. Markov model and prediction by incomplete match based approach exploit memory efficiently.

KEYWORDS: Proxy server, Caching, Prefetching, Markov Model.

I. INTRODUCTION

With the advent in technology, the web has grown as a rich collection of dynamic and interactive services. This huge growth of web has resulted into an increasing load on the web servers. Users often experience long and unpredictable delays when retrieving web pages from remote sites. Therefore, a clear solution to improve the class of web services would be the increase of bandwidth, but such type of solution increases the cost of system. However, higher bandwidth can resolve for the time being the problems since it would no difficulty the user to create more and more resource-

greedy applications for the network. In many research, researcher proved and suggest that a cache-based approach is very helpful to improve the performance of the web for lower cost. We notice that a single user again and again requests the same web object many times during a small interval of time and web object access is non-uniform over all web servers. Additional, different users many times request the same web object. If we can store commonly requested objects closer to web clients, users should see lower latency when browsing. Web caches are the systems that keep copies of frequently accessed objects close to clients. The development of web caching has spurred new research in many areas [1, 2].

There are three features of Web caching which make it attractive to all Web participants, including users, network managers, and content creators:

- Caching reduces network bandwidth usage.
- Caching reduces user-perceived delays.
- Caching reduces loads on the origin server.

Since web caching alone is not sufficient because of World Wide Web has evolved quickly from a simple information-sharing mechanism. There for the caching at proxy could be improved by pre-fetching the pages that may be requested by the users in near time and by following a web caching policy that synchronizes with the pre-fetching technique. Web prefetching is fetching web pages in advance by proxy server/client before a request is send by a client/proxy server. The major advantage of using web prefetching is reduced latency. When a client makes a request for web object, rather than sending request to the web server, it may be fetched from a pre-fetch area. The main factor for selecting a web prefetching algorithm is that its ability to predict the web objects to be pre-fetched in order to reduce latency. Web Prefetching exploits the spatial locality of web pages, i.e. Pages that are linked with current page will be accessed with higher probability than other pages. Web prefetching can be applied in a web environment as between clients and web server, between proxy servers and web server and between clients and proxy server [3]. If it is applied between clients and web server, it is helpful in reducing user perceived latency, but the problem is it will increases

network traffic. If it is applied between proxy server and web server, can reduce the bandwidth usage by prefetching only a specific number of hyper links. If it is applied between clients and proxy server, the proxy starts feeds pre-fetched web objects from its cache to the clients so there won't be extra internet traffic.

The process obtainable in this paper integrates the pre-fetching technique with the web caching scheme with the purpose of improving presentation of the proxy server. The integrated scheme would enlarge the performance of the proxy server in conditions of the Hit Ratio and the Byte Hit Ratio as opposite to a simple web caching approach.

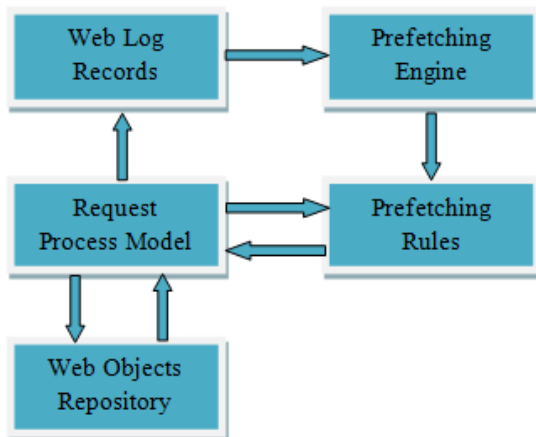


Figure1:- Prefetching Scheme

II. LITERATURE REVIEW

Pallis *et. al* [4] addressed the short-term prefetching problem on a web cache environment using an algorithm for clustering inter-site web pages. The proposed scheme efficiently integrates web caching and prefetching. According to this scheme, each time a user requests an object, the proxy fetches all the objects which are in the same cluster with the requested object.

Sharma and Dubey proposed a framework for web traffic reduction [5]. The approach first extracts data from proxy server web log, and then the extracted data is pre-processed. The pre-processed data is then mined using clustering, and sequence analysis is being done to know the patterns to be pre-fetched.

Zhaolei Duan in at al [6] proposed technique improve the web proxy cache cluster's performance, a novel web proxy cache cluster-WPCC is obtainable. WPCC divide all back ends into two groups: group A and group B. Group A takes charge for hit requests and group B takes charge for miss requests. WPCC uses different load balancing strategy in different back end group. Back end can be migrated from one back end group to another when load imbalance between back end groups happens. Simulations results show WPCC can achieve outstanding performance than obtainable web proxy cache clusters.

Umagandhi, R., Kumar, A.V.S [7]. The process of mining the query log file improves the performance of the search engine. The proposed algorithm mines the query log file to discover the similarity between the query keywords and URLs in its first phase. In the second phase, the query cluster and the URL cluster is created by using the combined similarity measure generated from the first phase. The clusters recommend query to the user to frame their future queries based on their previous search histories and click through data. This combined similarity based approach also recommends the user about the URL selection for the future queries.

Nanhay Singh in at al[8] proposed work, prefetching is done on the basic of proxy logs as it is the key requirement to make available user with best recommendations. Web log data is preprocessing for pattern discovery. It integrate the following two techniques together i.e. clustering and association using frequency support pruning, it achieves complete logs, better accuracy, less state space complexity and less number of rules. The predicted pages are pre-fetched and keep it in proxy server cache which reduce the accessing time of that page and increases the web proxy server performance.

Kasthuri *et. al*[9] shows that deduction of future references on the basis of predictive Prefetching, can be implemented by based on past references. The prediction engine can be residing either in the client/server side. But in our context, prediction engine resides at client side. It uses the set of past references to find correlation and initiates Prefetching that is driving user's future requests for web documents based on previous requests.

Sharma and Dubey proposed a framework for web traffic reduction [10]. The paper presented a framework for the prefetching and prediction in web. According to the framework, previous web requests of the user will be extracted from the proxy web log. From this web log, strong rules will be generated using FP Growth algorithm. These rules will be used to prefetch the upcoming requests of the current user.

III. THE PROPOSED METHODOLOGY

Web prefetching and web caching scheme is used to improve the performance of web even when more number of user are connected and demanding access to the content over the web. These two techniques can go together each other since the Web caching exploit the temporal locality, whereas Web pre-fetching utilizes the spatial locality of Web substance. It is complicated to search and display information applicable to user. Our objective is to enlarge relevancy in obtainable web search by gathering functional information from the web, based on user's interest. There are some problems arise like while doing caching and prefetching:-

- Extra Overhead
- Size of Cache

- Cache consistency difficulty
- Scalability

Our proposed Prefetching techniques are useful to reduce the latency, control the network traffic and server load and get better the user approval. Markov model and prediction by incomplete match based approach exploit memory efficiently. In Distributed Web Caching with geographical area base Clustering, the proxy servers those are geographically mutually are grouped into the similar cluster. This scheme is base on the clustering of proxy servers in dispersed situation. This provides simple supervision of metadata as in preceding strategy. It also provides low latency profit by a factor of preceding approach. Furthermore load balancing is managed in this by limiting the number of client's needs per cluster. If there is no limit on numeral of customers per proxy server, a number of the proxy servers will be congested and start dropping the desires or will add unlimited delays. This can be handling by limiting the number of customers for each proxy server and per cluster as well. This is complete by maintain a cluster queue length data construction. The proposed plan includes clients; each cluster has an additional node than proxy servers that is Metadata Server. This Metadata server is a category of data base server whose task is to preserve metadata of that cluster and it also keep track of neighboring cluster's metadata. All additional proxy servers of a cluster are linked with their Metadata server. Each proxy server only maintains metadata of its own and does not concerning about the metadata of other proxy servers of similar cluster or of neighboring clusters. In above strategy each proxy server itself maintains metadata of it's possessing cluster as well as of their neighboring clusters. So this strategy will reduce efforts and time of proxy servers.

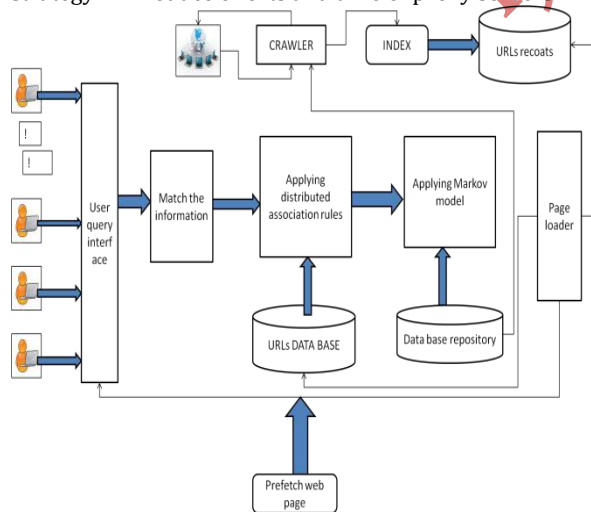


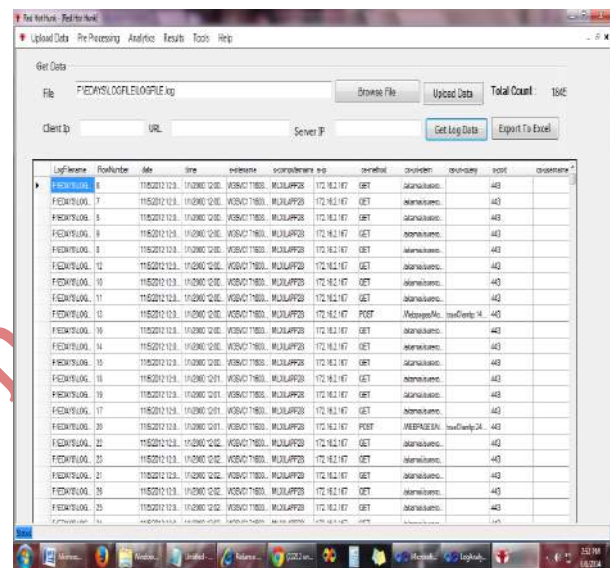
Figure 2: proposed system framework

IV. RESULTS ANALYSIS

In these project we have used the Visual Studio .NET and Log parser for development of the software. Log parser is a powerful, versatile tool that provides universal query access to text-based data such as log files, XML files and CSV files, as well as key data sources on the Windows® operating system such as the Event

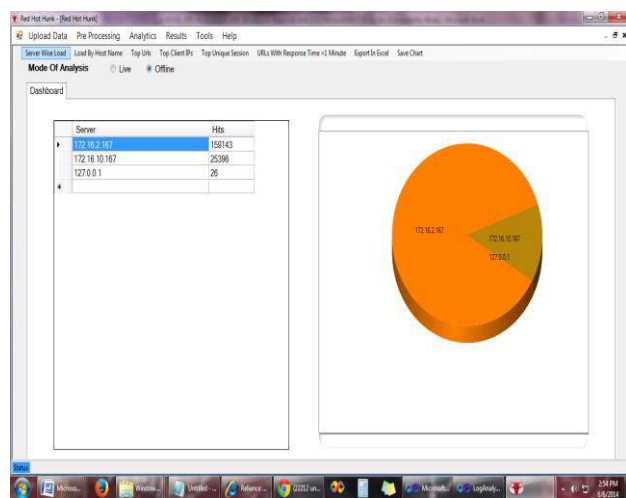
Log, the Registry, the file system, and Active Directory®. We have applied our technique to different classical benchmark problems including Log File of an insurance company.

4.1 Loading Log Files

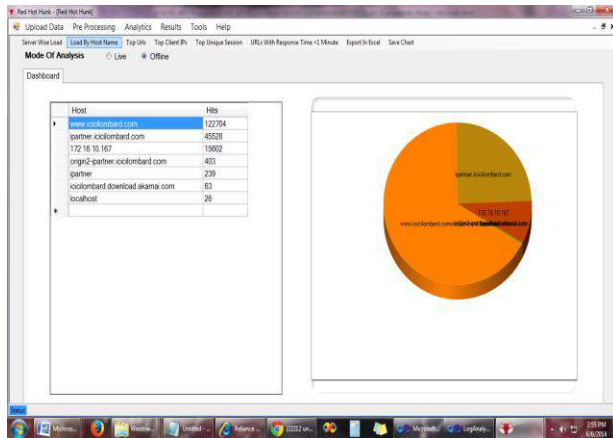


4.2 Performance Analysis

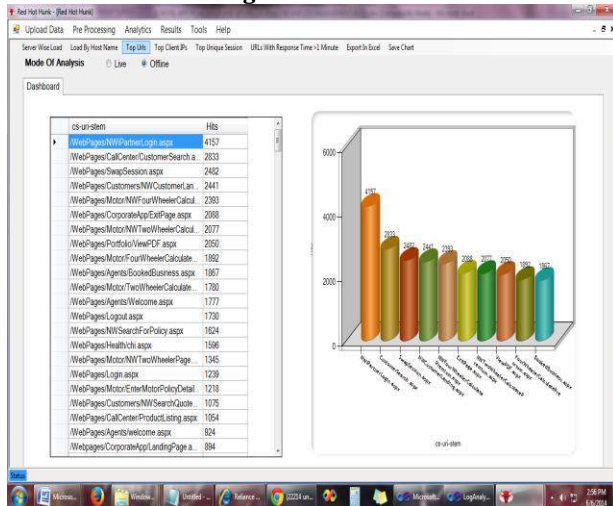
4.2.1 Server Load



4.2.2 Loads by Domain Name



4.2.3 Total Hits On Pages



V. FUTURE SCOPE

System can be enhanced to live log analysis as currently this analysis is of off line analysis. Also it can be further enhanced to greater performance if we use parallel tasking or multi threading concept in programming.

VI. CONCLUSION

For test and result we take log sample of web application of an insurance company. The data is taken for a period of one week. We analyze separately web log server from IIS and the application log generated through developed code in application. We got different links patterns that user navigates on web site. Also from this links we got number of error links that are coming repeatedly on application. While analyzing application log we got page and line number for cause of error.

From this application log we can trace top errors occurred during a day or week for application maintenance. Thus above system collecting log to a central log server is use full in application maintenance. This tool has been built and effectively used in performance analysis on different occasions.

VII. REFERENCES

- [1] A. Bestavros and C. Cunha, Server-initiated document dissemination for the WWW, IEEE Data Engineering Bulletin, Sept. 1996.
- [2] M. R. Korupolu and M. Dahlin, Coordinated placement and replacement for large-scale distributed caches, Proceedings of the IEEE Workshop on Internet Applications, July 1999 (Technical Report TR-98-30, Department of Computer Science, University of Texas at Austin, December 1998).
- [3] Seung Won Shin, Byeong Hag Seong & Daeyeon Park, (2000) "Improving World-Wide- Web Performance Using Domain-Top Approach to Prefetching", Fourth International Conference on High- Performance Computing in the Asia-Pacific Region vol. 2, pp. 738-746.
- [4]. Pallis G., A. Vakali and J. Pokorny, (2008) "A clustering-based prefetching scheme on a Web cache environment", Computers and Electrical Engineering 34, Elsevier, pg 309-323.
- [5]. N. Sharma and S.K. Dubey, (2013) "Fuzzy C-means clustering based prefetching to reduce web traffic", International Journal of Advances in Engineering & Technology, vol. 6, lissue 1, pp. 426-435, ISSN: 2231-1963,2013.
- [6] Zhaolei Duan, Zhimin Gu Xiaoguang Ding," WPCC: A novel web proxy cache cluster" Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on (Volume:03) Date of Conference: 15-18 Feb. IEEE-2009.
- [7] Umagandhi, R., Kumar, A.V.S.," Time independent query recommendations from search engine query logs" Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012), International Conference on- Date of Conference: IEEE-19-21 Dec. 2012.
- [8] Nanhay Singh, Arvind Panwar, and Ram Shringar Raw, "Enhancing the Performance of Web Proxy Server through Cluster Based Prefetching Techniques", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.
- [9]. I. Kasthuri, M.A. Ranjit Kumar , K. SudheerBabu, and Dr. S. S. S. Reddy, (2012) " An Advance Testimony for Weblog Prefetching Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4.
- [10]. N. Sharma and S.K. Dubey, (2013) "FP tree use in prefetching", Proc. of Int. Conf. on Advances in Computer Science, AETACS, Elsevier, 2nd December, 2013, pp 555-561.