

A Survey on an Integrated Approach for Web Pre-Fetching and Caching

Mrs. Kiran Jain¹ Er. Anuradha²

¹Asst. Professor ²P.G Student

^{1,2}Computer Science Engineering Department

^{1,2}DIET, Karnal, Haryana, India

Abstract--Web caching and web prefetching are the two major areas of research focused at reducing the user perceived latency. Both if used well can greatly help in reducing this latency. As web caching helps in exploiting temporal latency while web prefetching helps in exploiting spatial latency. However if prefetched pages are not visited by the users in their future accesses, they can increase the network traffic and overload the web server. This paper aims at surveying various research papers who have worked in this direction..

Keywords: Web caching, Web prefetching

I. INTRODUCTION

With the advancement in information technology, the World Wide Web (WWW) has become a huge information repository. Today, it covers almost every topic including text, image, audio, video, and metadata, in which a human user could be interested. The enormous spread of Internet and consequently, the growth of the World Wide Web have implied a dramatic increase in the number of connected users and therefore, a tremendous increase in the global traffic. Thus it can damage the quality of service (availability, reliability, and security) and specially the latency perceived by the users. People use the WWW to access information from remote sites. But they do not like to wait long for their results. The latency of retrieving a Web document depends on several factors.

- Speed of Servers
- Speed of clients
- Network Bandwidth and Propagation Delay

A. The Web Caching Approach:

Caching proved itself as an important technique to optimize the way the Web is used . In particular, many of the Web caching aspects are originated from the caching idea implemented in various computer and network systems whereas Web caching introduces new issues in Web objects management and retrieval across the network. Specifically, Web caching is implemented by proxy server applications developed to support many users. Proxy applications act as an intermediate between Web users and servers. Users make their connection to proxy applications running on their hosts. The proxy connects the server and relays data between the user and the server. At each request, the proxy server is contacted first to find whether it has a valid copy of the requested object. If the proxy has the requested object this is considered as a cache hit, otherwise a cache miss occurs and the proxy must forward the request on behalf of the user. Upon receiving a new object, the proxy services a copy to the end-user and keeps another copy to its local storage. From the above discussion follows that Web caching reduces bandwidth consumption, network congestion, and

network traffic because it stores the frequently requested content closer to users. Also, because it delivers cached objects from proxy servers, it reduces external latency (the time it takes to transfer objects from the origin server to proxy servers). Finally, caching improves reliability because users can obtain a cached copy even if the remote server is unavailable.

B. The Web Prefetching Approach:

Prefetching attempts to overcome these limitations by proactively fetching content before users actually request it .Web prefetching is the process of deducing user's future requests for Web objects by locating popular requested objects into the cache prior to an explicit request for them. Unlike Web data caching, which exploits the temporal locality, the Web prefetching schemes are based on the spatial locality of Web objects. In particular, the temporal locality refers to repeated users' accesses to the same object within short time periods, whereas, the spatial one refers to users' requests where accesses to some objects frequently entail accesses to certain other objects. Typically, the main benefit of employing prefetching is that it prevents bandwidth underutilization and reduces the latency. Therefore, bottlenecks and traffic jams on the Web are by passed and objects are transferred faster. Thus, the proxies may effectively serve more users' requests, reducing the workload from the origin servers. Consequently, the origin servers are protected from the flash crowd events as a significant part of the Web traffic is dispersed over the proxy servers. On the other hand, the main drawback of systems which have enhanced prefetching policies is that some prefetched objects may not be eventually requested by the users. In such a case, the prefetching scheme increases the network traffic as well as the Web servers' load. In order to overcome this limitation, high accuracy prediction models have been used .

II. PROBLEM FORMULATION

- (1) Caching is an important technique for improving the performance of web based applications with help of web caching techniques. Web caching provides great features like traffic reduction, less load on servers, user-end retrieval delays by replicating popular content on proxy caches that are strategically placed within the network. Web pre-fetching schemes have also been widely discussed where web pages and web objects are pre-fetched into the proxy server cache. In our research we will work on integration of web caching and web prefetching approaches to improve the performance of proxy server's cache.
- (2) In Domain Top approach for web prefetching, combination of knowledge of most popular domains and

most popular documents is done by proxy server. In this approach proxy is responsible for calculating the most popular domains and most popular documents in those domains, and then prepares a rank list for prefetching.

- (3) In Dynamic web pre-fetching technique, each user can keep a list of sites to access immediately called user's preference list. The preference list is stored in proxy server's database. Intelligent agents are used for parsing the web pages; monitoring the bandwidth usage and maintaining hash tables, preference list and cache consistency. It controls the web traffic by reducing pre-fetching at heavy traffic and increasing pre-fetching at low traffic.
- (4) In this research the concept of preference list from Dynamic technique into Domain Top approach is brought. Optimized Domain Top approach will consist of preference list along with the rank list.
- (5) Advantage will be that the pre fetching would have wider scope (will be more fast). Due to presence of rank list (which maintain top domain and top documents) and preference list (which consist of list of sites stored by user to have immediate access) model will work well. This model will cover the loopholes like overhead in proxy servers.

III. RELATED WORK

A. *A Survey of Web Caching and Prefetching*"(Waleed Ali , Siti Mariyam Shamsuddin, and Abdul Samad Ismail)(2011)

Web caching and prefetching are the most popular techniques that play a key role in improving the Web performance by keeping web objects that are likely to be visited in the near future closer to the client. Web caching can work independently or integrated with the web prefetching. The Web caching and prefetching can complement each other since the web caching exploits the temporal locality for predicting revisiting requested objects, while the web prefetching utilizes the spatial locality for predicting next related web objects of the requested Web objects. This paper reviews principles and some existing web caching and prefetching approaches. The conventional and intelligent web caching techniques are investigated and discussed. Moreover, Web prefetching techniques are summarized and classified with comparison limitations of these approaches. This paper also presents and discusses some studies that take into consideration impact of integrating both web caching and web prefetching together.

B. *A Survey On Web Pre-Fetching and Web Caching Techniques in a Mobile Environment*"(Greeshma G. Vijayan1 and Jayasudha J. S.)(2012)

As the Internet continues to grow in size and popularity, web traffic and network bottlenecks are major issues in the network world. The continued increase in demand for objects on the Internet causes severe overloading in many sites and network links. Many users have no patience in waiting more than few seconds for downloading a web page. Web traffic reduction techniques are necessary for accessing the web sites efficiently with the facility of existing network. Web pre-fetching techniques and web caching reduces the web latency that we face on the internet today. This paper describes about the various prefetching and

caching techniques, how they predict the web object to be pre-fetched and what are the issues challenges involved when these techniques are applied to a mobile environment

C. *Survey on Improving the Performance of Web by Evaluation of Web Prefetching and Caching Algorithms*" (Arun Pasrija) (2013)

Web caching and prefetching have been studied in the past separately. In this paper, we present an integrated architecture for Web object caching and prefetching. Our goal is to design a prefetching system that can work with an existing Web caching system in a seamless manner. In this integrated architecture, a certain amount of caching space is reserved for prefetching. To empower the prefetching engine, a Web-object prediction model is built by mining the frequent paths from past Web log data. We show that the integrated architecture improves the performance over Web caching alone, and present our analysis on the tradeoff between the reduced latency and the potential increase in network load.

D. *Survey of Recent Web Prefetching Techniques*" (Sonia Setia, Dr. Jyoti, Dr. Neelam Duhan) (2013)

Web caching and web prefetching are the two major areas of research focused at reducing the user perceived latency. Both if used well can greatly help in reducing this latency as web caching helps in exploiting temporal latency while web prefetching helps in exploiting spatial latency. However if prefetched pages are not visited by the users in their future accesses, they can increase the network traffic and overload the web server. This paper aims at surveying various research papers who have worked in this direction.

E. *Study of Web Pre-Fetching With Web Caching Based On Machine Learning Technique* " (K R Baskaran, Dr. C.Kalarasan, A Sasi Nachimuthu) (2013)

High bandwidth utilization, reduced load on the origin server, high access speed are possible by combining Web caching and pre-fetching techniques. Pre-fetching is the process of fetching few Web pages in advance which will be assumed to be needed by the user in near future and those pages are cached in the memory. Lots of work has been reported for caching and pre-fetching of Web pages in the literature. In this paper, pre-fetching using clustering technique is combined with SVM (Support Vector Machine) – LFU algorithm, a machine learning technique for Web proxy caching .By using real dataset it will be shown that the SVM technique will be better than clustering based pre-fetching technique using caching policy like LFU considering bandwidth utilization and access latency

IV. PROPOSED WORK

This research is focused on providing solution by enhancing web prefetching process. For experimentation we have used database with various web entries and have done cleaning process on the database This will include 3 algorithm. These are following:

- 1) Basic Algorithm
- 2) Apriori Algorithm
- 3) Advance Apriori Algorithm

I will do time comparison among these algorithms. Apriori Algorithm Half the Web Accessing Speed as

compare to Basic Algorithm. In the same way Advance Apriori Algorithm will half the the Web Accessing Speed as compare to Apriori Algorithm

V. CONCLUSION & FUTURE WORK

Lot of research is going on in web prefetching in various directions. In this paper various web prefetching techniques and other directions of web prefetching are analyzed and discussed. The web prefetching scheme focus on the property spatial locality of web objects. These techniques are applied to reduce the network traffic and improve the user satisfaction. Web prefetching and caching can also be integrated to get better performance. At present, web prefetching and caching in mobile environment opens a vast scope in the research area. The main focus of the research is to improve accuracy in the web mining process. Our research is started with information fetching of pre-fetching and caching techniques. In future anyone can develop advance algorithm to enhance the speed of Web prefetching and caching. To find loopholes and issues in new approach and to highlight the benefits for new approach.

REFERENCES

- [1] I. Dzitac, "Advanced AI techniques for web mining," Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems. Corfu Greece. 2008.
- [2] J. B. Patil and B. V. Pawar, "Improving Performance on WWW using Intelligent Predictive Caching for Web Proxy Servers," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011.
- [3] Alexandros Nanopoulos, Dimitrios Katsaros, and Yannis Manolopoulos, "Exploiting Web Log Mining for Web Cache Enhancement," LNAI 2356, Springer-Verlag Berlin Heidelberg, pp. 68–87, 2002.
- [4] Siddharth Jain, Ruchi Dave, Devendra Kumar Sharma, "An approach using Association Rule Mining Technique for frequently matched pattern of an Organization's web log data", International Journal of Engineering Sciences & Research Technology, Vol.1, No.5, pp 297-300, 2012.
- [5] Hendrik Blockeel, Raymond Kosala, "Web Mining Research: A Survey", ACM SIGKDD, Vol- 2, Issue- 1, pp 1, 2000.
- [6] P. Somrutai, "Improving the Performance of a Proxy Server using Web log mining," M.S. thesis, San Jose State University, 2011.
- [7] Rudeekorn Soonthornsutee, Pramote Luenam, " Web Log Mining for Improvement of Caching Performance ", Proceedings of the International Multi-conference of Engineers and Computer Scientists, Vol- 1, pp 14-16, March 2012.
- [8] E Alexandros Nanopoulos, Dimitrios Katsaros, and Yannis Manolopoulos, "Exploiting Web Log Mining for Web Cache Enhancement", WEBKDD 2001, LNAI 2356, pp. 68–87, 2002.
- [9] J. Srivastava, R. Cooley, M. Deshpande, P.-N. Tan, " Web usage mining: discovery and applications of usage patterns from web data", SIGKDD Explorations, Vol – 1, Issue- 2, pp 12–23, 2000.
- [10] Josep Dome` nech, Ana Pont, Julio Sahuquillo, Jose´ A. Gil, " A user-focused evaluation of web prefetching algorithms", Computer Communications, Science Direct, Vol- 30, pp 2213-2224, 2007.
- [11] Sarina Sulaiman, Siti Mariyam Shamsuddin & Ajith Abraham, (2010) "Intelligent Mobile Web Prefetching (IMWeP) Using XML Technology", 6th IEEE International Conference on Next Generation Web Services Practices, pp. 475-480.
- [12] Weifeng Zhang, Baowen Xu, William Song, Hongji Yang & Kecheng Liu, (2000) "Data Mining Algorithms for Web Pre-Fetching", Proceedings of IEEE International Conference on Web Information Systems Engineering