ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume No. 2, Issue No. 6, June 2013

Survey on Improving the Performance of Web by Evaluation of Web Prefetching and Caching Algorithms

Arun Pasrija

M-Tech Student, Department of Computer Engineering, Yadawindra College of Engineering, Talwandi Sabo, India

ABSTRACT:

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.

Keywords: Web Perfecting; Performance Evaluation; User Perceived Latency.

I. INTRODUCTION

In the Internet, proxy servers play the key roles between users and web sites, which could reduce the response time of user requests and save network bandwidth. Basically, an efficient buffer manager should be built in a proxy server to cache frequently accessed documents in the buffer, thereby achieving better response time. In the paper, we developed an access sequence miner to mine popular surfing 2-Sequences with their conditional probabilities from the proxy log, and stored them in the rule table. Then, according to buffer contents and the rule table, a prediction-based buffer manager also developed here will make appropriate actions such as document

Caching, document prefetching, and even cache/prefetch buffer size adjusting to achieve better buffer utilization. Through the simulation, we found That our approach has much better performance than the other ones, in the quantitative measures such as hit ratios and byte hit ratios of accessed documents. 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.

1) 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.
2) 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. 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.
3) 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. Clustering based pre-fetching methods make decisions using the information about the clusters

Containing pages that have been fetched previously, assumes that pages that are close to those previously fetched pages are more likely to be requested in the near future.

II. WEB LOG MINING

Web usage mining is the third category in web mining. This type of web mining allows for the collection of Web access information for Web pages. This usage data provides the paths leading to accessed Web pages. This information is often gathered automatically into access logs via the Web server. CGI scripts offer other useful information such as referrer logs, user subscription information and survey logs. This category is important to the overall use of data mining for companies and their internet/ intranet based applications and information access mining in this fashion.

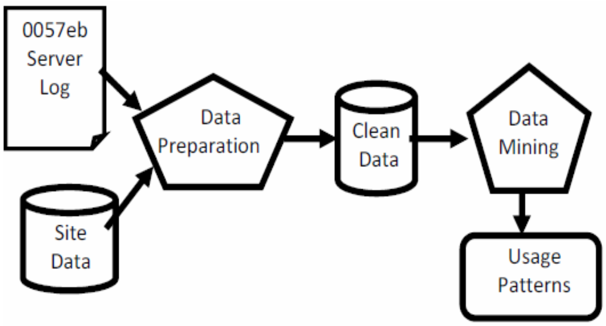

Fig1. Web Log Mining

A) The first is usage processing, used to complete pattern discovery. This first use is also the most difficult because only bits of information like IP addresses, user information, and site clicks are available. With this minimal amount of information available, it is harder to track the user through a site, being that it does not follow the user throughout the pages of the site.

B) The second use is content processing, consisting of the conversion of Web information like text, images, scripts and others into useful forms. This helps with the clustering and categorization of Web page information based on the titles, specific content and images available.

C) Third use is structure processing. This consists of analysis of the structure of each page contained in a Web site. This structure process can prove to be difficult if resulting in a new structure having to be performed for each page. Analysis of this usage data will provide the companies with the information needed to provide an effective presence to their customers. This collection of information may include user registration, access logs and information leading to better Web site structure, proving to be most valuable to company online marketing. These present some of the benefits for external marketing of the company's products, services and overall management.

III. LITERATURE REVIEW

Yin –fu huang in 2006[1] people often get various kinds of information and entertainments from the Internet. There, proxy servers play the key roles to transmit these data quickly between users and web sites. If a proxy server has an inapplicable buffer management, the network might suffer from a traffic jam when huge amounts of data are being transmitted At the same time. Hence, it is necessary and important to develop an efficient buffer management. However, some challenges or issues exist in designing a proxy buffer manager; e.g., there are many web sites, documents, and objects in the Internet. How can we judge what documents and objects are significant or frequently accessed for users in the Internet? Also, the users usually have their own browsing behaviors. Could we predict the next patterns they will browse, and prefetch them into the proxy buffer beforehand?

Josep Dome`nech in (2007) [2] the knowledge and comprehension of the behavior of a web user are important keys in a wide range of fields related to the web architecture, design, and engineering. The information that can be extracted from web user's behavior permits to infer and predict future accesses. This information can be used, for instance, for improving Web usability developing on-line marketing techniques or reducing user-perceived latency, which is the main goal of prefetching techniques.

These techniques use access predictors to process a user request before the user actually makes it. Several ways of prefetching user's requests have been proposed in the literature: the preprocessing of a request by the server, the transference of the object requested in advance, and the pre-establishment of connections that are predicted to be made. Despite the large amount of research works focusing on this topic, comparative and evaluation studies from the user's point of view are rare. This fact leads to the inability to quantify in a real working environment which proposal is better for the user. On the one hand, the underlying baseline system where prefetching is applied differs widely among the studies. On the other hand, different performance key metrics were used to evaluate their benefits. In addition, the used workloads are in most cases rather old, which significantly affect the prediction performance, making the conclusions not valid for current workloads. Research works usually compare the proposed prefetching system with a non-prefetching one. In these different workload and user characteristics have been used, so making it impossible to compare the goodness and benefits among proposals. Some papers have been published comparing the performance of prefetching algorithms**.**

Johann Marque et.al. In 2008[3] an intelligent technique for controlling web prefetching costs at the server side. Prefetching is an interesting technique for improving. Web performance by reducing the user-perceived latency when surfing the web... In this paper they propose an intelligent prefetching mechanism that dynamically adjusts the aggressiveness of the prefetching algorithm at the server side. To this end, they also propose a traffic estimation model that Permits to accurately calculate, in the server side, the extra load and traffic generated by the prefetching with the aim of reducing these negative effects we have developed an adaptive prefetching mechanism at the server side to control the traffic increase and its impact on the system. Applying the proposed adaptive mechanism, the prefetching technique can be also improved since the prediction algorithm requires neither a long period to reach a stable state nor extra resources. Since our proposal proves that the negative effects of web prefetching can be controlled even at the server side, the use of prefetching can be safely spread among users and system administrators.

V. Murali Bhaskaran in 2011[4] Web caching is an important technique for improving the performance of web based applications. Web caching is used to reduce network traffic, server load, and user-perceived 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. This paper presents an approach that integrates web caching and web pre-fetching approach to improve the performance of proxy server's cache.

By integrating Web caching and Web pre-fetching, these two techniques can complement each other since the Web caching technique exploits the temporal locality, whereas Web pre-fetching technique utilizes the spatial locality of Web objects

- By grouping the users and analyzing their previous access patterns, the system is able to predict pages that might be of interest to the users.
- By caching or pre-fetching the pages, the access speed of these pages can be considerably increased.
- This system also helps us in efficiently using the cache, since the pages are cached based on the access history of a group of users and not based on individual users.

P. Somrutai et al (2011) [5] explained that Proxy servers have been used widely to reduce the network traffic by caching frequently requested web pages by using web caching. Proxy server acts as an intermediary between the web server and the web user requesting the web page. The proxy servers try to serve as many requests at the proxy server level. Proxy servers first fetch the requested web pages from the origin web servers and store the web pages in the proxy server's cache. If a user makes a request to a web page already stored in the cache, the proxy server accesses the local copy of the web page stored in the cache and serves it to the user who requested the web page. The proxy server's cache has limited capacity in terms of size of web pages that can be stored in the cache at any given time. Once the cache capacity is reached, the temporally stale web pages in the cache are discarded and replaced by newly requested web pages. The web pages stored in the proxy server cache are managed by the cache replacement algorithms .This approach of caching is called as web caching. Web caching has been used to reduce the network traffic by caching web pages at the proxy server level. The work presented in this paper seeks to explore an analysis based pre-fetching scheme to improve the performance of the proxy server. When the user requests a web page that is part of such a cluster, other related web pages in the same cluster can be pre-fetched into the proxy server's cache in the expectation that the next set of web pages requested by the web user would be from the pre-fetched web pages. The approach presented in this paper integrates the pre-fetching approach with the web caching scheme with the objective of improving performance of the proxy server. The integrated scheme would boost the performance of the proxy server in terms of the Hit Ratio and the Byte Hit Ratio as opposed to a plain web caching approach.

Pramote lueman in (2012) [6] the concept of web log mining for improvement of caching performance. Authors explained that the objective of study is to build a model of cache replacement policy for improvement of web caching performance. The integration approach of cluster analysis and classification are used to create a classifier for predicting the cache life time. The data set was collected from the cache of the National Institute Development Administration's proxy servers. There are four main tasks in this study. First, the access log from proxy servers were collected and preprocessing tasks were performed. Second, the access log data were partitioned into clusters based on users' access patterns. Third, classifier models of the cache replacement policy were built and their accuracies were compared. Finally, the efficiency of the selected classifier was compared with other cache replacement algorithms. Results show that overall classification accuracy of the model is satisfactory and the model is efficient and very good in performance.

IV. CONCLUSION

In these papers, evaluation of web prefetching algorithms has been studied. Each algorithm has its own advantages and disadvantages and each algorithm has its own application area. Applying the proposed adaptive mechanism, the prefetching technique can be also improved since the prediction algorithm requires neither a long period to reach a stable state nor extra resources. Since our proposal proves that the negative effects of web prefetching can be controlled even at the server side, the use of prefetching can be safely spread among users and system administration. In Dynamic web pre-fetching technique, each user can keep a list of sites to access immediately called users preference list. The preference list is stored in proxy server's database. Intelligent agents are used for parsing the web page, monitoring the bandwidth usage and maintaining hash table, preference list and cache consistency. It controls the web traffic by reducing pre-fetching at heavy traffic and increasing pre-fetching at light traffic. In our future work we bring concept of preference list from Dynamic technique into Domain Top approach. Optimized top domain approach will consist of preference list along with the rank list.

V REFERENCES

[1] Yin-Fu Huang, Jhao-Min Hsu," Mining web logs to improve hit ratios of prefetching and caching", Knowledge-Based Systems, Science Direct, Vol- 21, pp 62-69, 2006.

[2] 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

[3] Johann Marquez, Josep Domenech, Ana Pont, Julio, Jose A.Gil, "An intelligent Technique for controlling web prefetching costs at the server side" 2008 IEEE

[4] v.sathiyamoorthi ,v.murali Bhaskaran "improving the performance of web pages retrieval through pre-fetching and caching using Web log mining" vol 66 no.2 pp 207-218@2011 EJSR

[5] P. Somrutai, "Improving the Performance of a Proxy Server using Web log mining," M.S. thesis, San Jose State University, 2011.

[6] 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.

Arun Pasrija received his B.Tech degree in Computer Science & Engineering from Guru Gobind Singh College Of Engineering & Technology, Talwandi Sabo, Punjab in 2011 and pursuing M Tech. (Regular) degree in computer engineering from Yadawindra College of Engineering Punjabi University Guru Kashi Campus Talwandi Sabo (Bathinda), batch 2011-2013. His research interests include improving the Performance of Web by Evaluation of Web Prefetching & Caching Algorithms

.