A Survey on Web Prefetching Techniques
Tanu Sihag
M. Tech Scholar
Department of Computer Science and Application
Kurukshetra University, Kurukshetra

Abstract—Web prefetching is an important aspect to find the possibility of finding which object would be requested in near future. Demand of internet and easy accessibility of information, communication and flexibility had put gigantic pressures on the principal infrastructure of WWW. Web prefetching is one of the techniques proposed to reduce user’s perceived latencies in the World Wide Web. The spatial locality shown by user’s accesses makes it possible to predict future accesses based on the previous ones. A prefetching engine uses these predictions to prefetch the web objects before the user demands them. Web prefetching is becoming important and demanding, even though the Web caching technique has been improved. Web prefetching is an effective tool for improving the access to the World Wide Web.

Key words: Internet measurement, World Wide Web, Traffic analysis, Web prefetching, Cache Latency

I. INTRODUCTION

The rapid growth in popularity of the WWW is leading to numbers of performance problems. The internet is becoming increasingly congested and popular. Web sites are suffering from overload conditions due to the large number of concurrent accesses. As a result, considerable latency is often experienced in retrieving web objects from the Internet. Web prefetching mechanisms are beneficial to web users by hiding the download latencies.

Web prefetching is a technique that made efforts to solve the problem of these access latencies. Specially, global caching methods that straddle across users work quite well. However, the increasing trend of generating dynamic pages in response to HTTP requests from users has rendered them quite ineffective. Prefetching is used as an attempt to place data close to the processor before it is required, eliminating as many cache misses as possible. Caching offers the following benefits: Latency reduction, Less Bandwidth consumption, Lessens Web Server load. Prefetching is the means to anticipate probable future requests and to fetch the most probable documents, before they are actually requested.

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 [8]. 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, it 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 prefetched web objects from its cache to the clients so there won’t be extra internet traffic. This paper describes about the various prefetching techniques, how they predict the web object to be pre-fetched and what are the issues involved in these techniques.

II. STUDY OF WEB PREFETCHING TECHNIQUES

A. Interactive Pre-fetching Scheme

In this scheme all the hyperlinks and inline images in linked pages are fetched. Since it retrieves all the hyperlinks, hit rate of 80% is possible. The disadvantage is it increases the load on to the host, and requires a lot of memory to store the pre-fetched web pages.

B. Link Pre-fetching

This mechanism, utilizes browser idle time to download documents that the user might visit in the future. A web page provides a set of pre-fetching hints to the browser and after the browser finishes loading the page, it starts pre-fetching specified documents and stores them in its cache. When the user visits one of the pre-fetched documents, it can be served up quickly out of the browser's cache. Fisher et al. proposed a server driven approach for ink pre-fetching. In this approach browser follows special directives from the web server or proxy server that instructs it to pre-fetch specific documents. This mechanism allows servers to control the contents to be pre-fetched by the browser. The browser looks for either HTML <link> tag or an HTTP Link: header tag to pre-fetch the subsequent links. The Link: header can also be specified within the HTML document itself by using a HTML <meta> tag. When the browser is idle, it observes these hints and queues up each unique request to be pre-fetched.

C. Top 10 Approach

Evangelos P. Markatos et al. proposes a top 10 approach to prefetching on the web, in which the server calculates the list of most popular documents. This approach is easy to implement in client server architecture. It considers frequency of access for prefetching, in which the server calculates the list of most popular documents.

D. Domain Top Approach

Seung Won Shin et al. proposed a domain top approach for web pre-fetching which combines the proxy’s active knowledge of most popular domains and documents [11,13]. In this approach proxy is responsible for calculating the most popular domains and most popular documents in those domains and for preparing a rank list for prefetching.

E. A Keyword based semantic prefetching approach in internet news services

This proposes a key word based semantic pre-fetching, in which prediction of future requests are based on semantic preferences of past retrieved web documents. This technique is applied to internet news service; it finds out semantic preferences by analysing keywords in URL anchor text of previously accessed documents in different news categories. The semantic representation is represented in an open self-learning capable model which collects the
knowledge about the client preferences. Client future request predictions are based on this knowledge. The pre-fetched documents are stored in internal cache. When client makes a request the web document is fetched from cache if it is available otherwise it fetches the documents from internet. This technique uses a neural network based semantics model that has the capability of self-learning.

**F. Dynamic Web Pre-fetching**

In dynamic web pre-fetching technique [12], 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 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. Thus it reduces the idle time of the existing network and makes the traffic almost constant. A hash table is maintained for storing the list of accessed URLs and their weight information [12, 13]. Depending upon the bandwidth usage and weights in the hash table, the prediction engine decides the number of URLs to be pre-fetched and gives the list to pre-fetch engine for pre-fetching the predicted web pages. After pre-fetching, the proxy server keeps the pre-fetched web pages in a separate area called pre-fetch area.

**G. Web Companion**

This approach employs a highly selective prefetching strategy based on estimated round-trip times for web resources.

Fig. 1: Semantic based pre-fetching scheme

![Semantic based pre-fetching scheme](image1)

**H. Markov Model for Predicting Web Access**

Markov model and hidden markov model are used for predicting the web pages to be pre-fetched based on the web access probability. Traditional markov model takes a sequence of web pages accessed by the user as an input and predict the next page to be accessed by the user. Low order markov models can’t accurately predict the subsequent requests of user. If higher order markov models are used the memory required increased rapidly with order.

**I. Prediction by Partial Match Web Prefetching**

The Prediction by Partial Match is a commonly used technique in Web prefetching, where prefetching decisions are made based on historical URLs in a dynamically maintained Markov prediction tree. Existing approaches either widely store the URL nodes by building the tree with a fixed height in each branch, or only store the branches with frequently accessed URLs. Building the popularity information into the Markov prediction tree, has proposed a new prefetching model, called popularity-based PPM. In this model, the tree is dynamically updated with a variable height in each set of branches where a popular URL can lead a set of long branches, and a less popular document leads a set of short one.
III. PERFORMANCE METRICS

There are several performance metrics exist to evaluate the performance of web prefetching techniques described below:

- **Hit Rate**: The hit rate (HR) is the ratio between the number of requests that hit in the cache and the total number of requests. Hit rate refers to the percentage of user access requests that are found in the prefetching cache. The byte-hit rate is an even more realistic measure of performance for web caching, it is the ratio between the number of bytes that hit in the proxy cache and the total number of bytes requested sent to its clients.

- **Byte Hit Rate**: Byte Hit Rate (BHR) is the percentage of the number of bytes that correspond to the requests served by the cache over the total number of bytes requested. A high HR indicates the user’s satisfaction and defines an increased user servicing. On the other hand, a high BHR improves the network performance and reduces the user-perceived latency (i.e., bandwidth savings, low congestion etc.).

- **Delay Saving Ratio**: It is a measure that takes into account the latency of fetching an object.

IV. CONCLUSION

This paper has given brief introduction about the different Web prefetching techniques. 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. With the help of above survey the cons and pros of different Web prefetching techniques are known. Techniques can be merged to overcome the drawbacks of one another.

Also to overcome the problem of long retrieval latency in the schemes a prefetching scheme using the association rules from the data mining technique is proposed. The data mining technique can provide some priority information such as the support, confidence, and association rules which can be utilized for prefetching. The prefetching policy can predict user behaviours and evaluate segments that may be accessed in near future. 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.

REFERENCES


[16] Jiawei Han, Kevin, Chen-Chuan Chang,"Data Mining for Web Intelligence", IEEE journal of computer society, Vol.34, Issue.9, pp.11-12, 2002.
