International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 09 | Dec-2015 www.irjet.net p-ISSN: 2395-0072

PROXY SIDE WEB PREFETCHING SCHEME FOR EFFICIENT BANDWIDTH USAGE: DATA MINING APPROACH

Mr. Swapnil S. Chaudhari[1], Prof. Poonam Gupta[2]

[1] *PG Scholar, Computer Networks, G. H. Raisoni College of engg. & Mgmt, Maharashtra, India.*
[2] *Assistant Professor, Computer Networks, G. H. Raisoni College of engg. & Mgmt, Maharashtra, India.*

Abstract - ***As the number of World Wide Web (Web) users grows, Web traffic continues to increase at an exponential rate and has become one of the major components of Internet traffic. One of the solutions to reduce Web traffic and speed up Web access is Web caching and prefetching. Web prefetching is one of the methods to condense user's latencies in the World Wide Web professionally. User's accesses makes it possible to predict future accesses based on the previous objects and previous sites. A prefetching engine makes use of these predictions to prefetch the web objects and performed site before the user demands them on the behalf of user. Web prefetching is becoming important and demanding, even though the Web caching system has been recover because of bandwidth usage. Web prefetching is a helpful implement for upgrading the access to the World Wide Web and it also diminish the bandwidth usage at the time. We propose proxy-side web prefetching scheme for efficient bandwidth usage that improves cache hit rate with a small amount of additional storage space. We have simulated our scheme based on logs of a real Web proxy server and the results indicate that the proposed prefetching scheme can reduce bandwidth during peak-periods. Also, the scheme can be adaptively applied to any Web usage patterns by changing prefetching parameters.***

Key Words: *Web prefetching, Proxy side web prefetching, web prefetching objects, probabilistic method for web prefetching, Improvement of web caching followed by web prefetching.*

1. INTRODUCTION

The WWW can be considered as a large distributed information system where users can access to shared data objects. Its usage is inexpensive and accessing information is faster using WWW than using any other means [1]. The main problem of these may result to extreme congestion on the network and load on the servers, all resulting in unacceptable degradation of the Quality of Service (QOS) at the user end. The prediction model inside Proxy server is required to alleviate the situation. But the cache management has many challenges in balancing the process of meeting the demands of the users on the one hand and ensuring optimal utilization of system resources on the other hand [3]. The most software based solution is web caching and prefetching techniques .The caching is introduced at three level client level, proxy level and original server level. Prefetching technique is used according to prediction on web proxy .Successful proxy servers play the key roles between users and web sites, which could reduce the response time of user request and save network bandwidth. Therefore, an efficient caching approach should be built in a proxy server for achieving better response time. This prefetching scheme reflects Web access patterns of users. In fact, this scheme may increase total bandwidth usage slightly in comparison with standard Web caching. However, the proposed scheme can efficiently reduce Web traffic bandwidth usage during peak periods by consuming unused bandwidth during off-peak periods. As the number of World Wide Web (Web) users grows, Web traffic continues to increase at an exponential rate. Currently, Web traffic is one of the major components of Internet traffic. This will reduce Web access time and make more efficient use of Internet links. One of the solutions to reduce Web traffic and speed up Web access is through the use of Web caching. [1, 2, 5] and [7, 8, 10].

1.1 Prediction Based Proxy server

Web Logs: consists of a series of entries arranged in reverse chronological order, often updated on frequently with new information about particular topics. The information can be written by the site owner, gleaned from other Web sites or other sources, or contributed by users. A weblog often has the quality of being a kind of "log of our times" from a particular point-of-view. Generally, weblogs are devoted to one or several subjects or themes, usually of topical interest, and, in general, can be thought of as developing commentaries, individual or collective on their particular themes. A weblog consist of the recorded ideas of an individual (a sort of diary) or be a complex

collaboration open to anyone. Most of the latter are moderated discussions. [10]

Prediction Engine: The number of requests is predicted as per the number of request; this is more accurate, efficient and access requests faster than access request than from original server. It also reduces access latency using prediction .It improves hits over internet and become popular and user can enjoy good browsing over internet.

Performance: Performance of hits is the best practice used to avoid load from server as most of the accesses are handled at proxy server level. To reduce server load and Fever request goes over internet. A Prediction Based Proxy Server Using Apriori algorithm To Improve Hit Ratio queuing is not created traffic burden over network. Performance of hits is improved and reduces access latency and finally cost of internet is less than without using prediction.

1.2 Motivation and Challenges

We studied analytical models to compare the performance of both hierarchical and distributed caching. We derive models to calculate the latency experienced by the clients, the bandwidth usage, the disk space requirements, and the load generated by each cache cooperating scheme. Using analytical models, we can explore the different tradeoffs of the different cache cooperating schemes and simulate different scenarios. To make our model as realistic as possible, we take a set of reasonable assumptions and parameters from recent literature and try to validate our results with real data when possible. Denote N as the total number of documents in the WWW. Denote S as a size of a certain document. We assume that documents change periodically every update period. Requests for document i, in an institutional cache are Poisson distributed with average request rate. Therefore, the total request rate for document is i, which is also Poisson distributed. We consider that each document is requested independently from other documents, so we are neglecting any source of correlation between requests of different documents

2. METHODOLOGY

2.1 Problem Statement

The goal of dissertation is to design proxy side web prefetching scheme for efficient bandwidth usage using Apriori Algorithm to Improve hit Ratio which will help to improve performance of hits on internet. To Design and Implement our own Proxy server which have caching and Prefetching Mechanism for each User.

- To design and implement probabilistic cache prediction scheme using Apriori Algorithm to improve cache hit ratio in web proxy servers.
- To design Prefetch List Generator, Cache Request Generator and cache Refreshing Algorithm.
- To analyze bandwidth.

2.2 System Architecture

A proxy server is a computer that offers a computer network service to allow clients to make indirect network connections to other network services. A client connects to the proxy server, and then requests a connection, file, or other resource available on a different server. The proxy provides the resource either by connecting to the specified server or by serving it from a cache. In some cases, the proxy may alter the client's request or the server's response for various purposes. A common proxy application is a caching Web proxy. This provides a nearby cache of Web pages and files available on remote Web servers, allowing local network clients to access them more quickly or reliably.

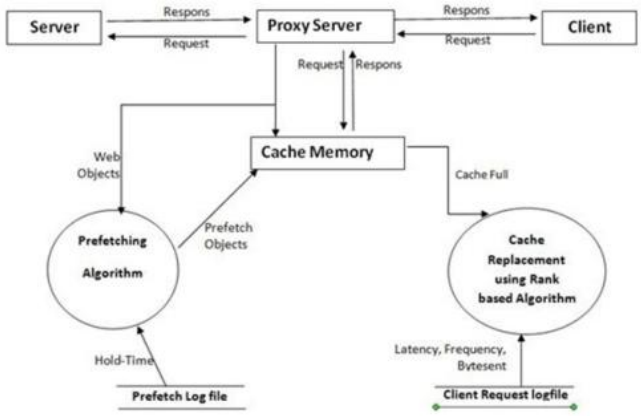

Fig -1: System Architecture

2.3 Prediction:

As per Proxy server generated log file which is useful for prediction it will take this log as input parameter and converted into prediction file which is predicted according to following.

- The client request transfer, the body length in the client request to Proxy server (in Bytes).
- The response length (in Bytes) from origin server to Proxy server.
- The length of Proxy server response to the client (in Bytes).
- The Time Proxy server spends processing the client request, the no. of seconds between the times at which the client establishes the

connection with Proxy server at the time at which Proxy server sends the last Byte of the Response Back to client.

3. PROPOSED ALGORITHMIC MODEL

The Apriori Algorithm is an influential algorithm for mining frequent item sets for association rules.

Frequent Item sets: The sets of item which has minimum support (denoted by Li for ith-Item set).

Apriori Property: Any subset of frequent item set must be frequent.

Join Operation: To find Lk, a set of candidate k-item sets is generated by joining Lk-1with itself.

Pseudo-code
Ck: Candidate item set of size k
Lk: frequent item set of size k
L1= frequent items;
for(k= 1; Lk!=0; k++) do begin
Ck+1= candidates generated from Lk;
for each transaction tin database do
increment the count of all candidates in Ck+1that
are contained in t
Lk+1= candidates in Ck+1with min support
end
return U kLk;

3.1 System Model

PAi(t):-is defined as the probability that the previous access to object i was t time units before the present time, and **PBi(t)** :-is the probability that no updates were done to object i since its last access t time units in the past. Then, the probability of a hit on a request to a demand cache is

$$Phitd = \sum_{i}^{n} PAi(t)PBi(t)dt$$

Here we provide a model H/B for the measurement of their balance, where the hit ratio and network resource consumption of demand cache serve as baseline for evaluation. This model defines the ratio of hit ratio upgrading over increased bandwidth cost.

$$\frac{H}{B} = \frac{Hitprefetching/Hitdemand}{BWprefetching/BWdemand}$$

Hit prefetching/Hit demand is the hit ratio step up of prefetching over demand caching.
BW prefetching=BW demand is network bandwidth boost over demand caching. [5]

4. IMPLEMENTATION DETAILS AND RESULT

Firstly we register the user in the proxy server for authentication and for that we use two parameters first one is users IP address and second one is Users Machine MAC address which is get by calling function GetMacAddress(). After this proxy server stores the Users Information that is IP address and Mac Address In XML file storage format.

```
<? Xml version="1.0" encoding="UTF-8" standalone="no"?>
  <Register>
    <User>
      <IP>192.168.100.143</IP>
      <MAC>00-22-68-4D-7C-EF</MAC>
    </User>
  </Register>
  <? Xml version="1.0" encoding="UTF-8" standalone="no"?>
    <Register>
      <User>
        <IP>192.168.1.197</IP>
        <MAC>4C-72-B9-25-EC-0E</MAC>
      </User>
    </Register>
```

Fig -2: Xml File

After storing Users information it will check for the any block IP address If it is in Block List Then it doesn't give authority to proceed. Secondly all request of browser are stored in proxy server as a log and frequently accessed Requests are store in Cache. Prediction engine takes this log file and predict frequent access pattern and future request using data mining (Apriori Algorithm) approach for each user and fetches this from origin server and stores in cache on behalf of user. Finally in this way we are calculating the hit ratio in the cache for user and bandwidth latency for system.

User experience is good over internet and performance of group is given to show the performance of internet according to cache hit and delay simulation. The result obtained will improve the performance of the proposed search method using "Proxy side web prefetching scheme using data mining approach technique" It will process the total number of users of web proxy and total number of pages visited by each user in one month at different time and according to Prediction The proposed system would be implemented under Microsoft Windows operating system. This web log contains all the HTTP requests collected from Web. We have showed that predictive system performance in term of hit ratio and it is observe that hit ratio that we calculi is growing in our experiment. Prediction engine which we used in our experiment have a simple data mining approach which is used to predict the users next request in web prefetching by extracting useful knowledge from historical user requests. The number of web pages on the

Internet has grown explosively in the past and such growth is expected to be more acute in the future .Fig shows the hits ratio and graph for System. We examine each set of requests on prediction based proxy server using Proxy side web prefetching system for efficient bandwidth usage to improve the hit ratio and to improve user's latency. After completion of prediction process system calculates hit rate. These calculations are based on predicted requests and when System predicts more requests then it increases hit rate. The Proxy side web prefetching scheme search will complete with performance and design parameter which is useful for prediction based web proxy server to improve hit ratio.

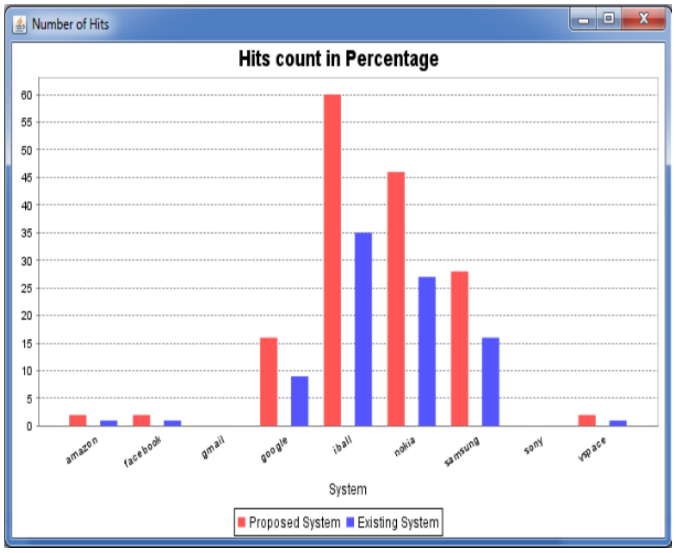

Chart -1: Hit Count in Percentage for both system (Proposed and Existing)

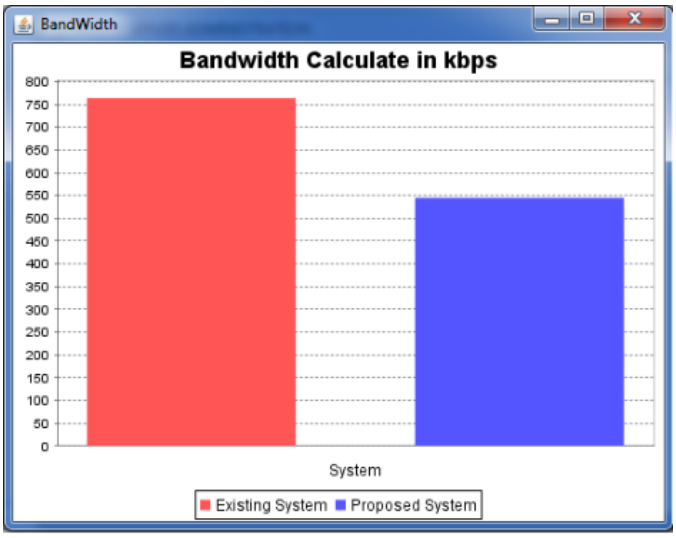

Chart -2: Bandwidth Calculation in Kbps (For both Systems)

5. CONCLUSIONS

Increasing attraction of World Wide Web over the earlier period of few years has imposed a significant load upon the internet and World Wide. Web is huge distributed information system and users can access shared data object. So results of internet service are slow down such as retrieve page from server and decrease the performance of system. Improvement of cache performance in web proxy servers is extremely important. This can be achieved through variety of ways. In our dissertation we aim to achieve this performance improvement using probabilistic method. To solve this problem, we have applied Proxy side web prefetching scheme for efficient bandwidth usage by Apriori method. It improve hit and byte rate. This project presents our research work on Apriori algorithm which is data mining approach. The main contributions of this project is search directions will process monthly request as per time to increase hit ratio, Prediction is periodical. Also every page is predicated whether user likes it or not. It prefetches only those pages from candidate links which users like.

REFERENCES

[1] Jae-young Kim "A Statistical Prefetching Web Caching Scheme for Efficient Internet Bandwidth Usage" Distributed processing and network Management Lab Dept of computer Engg. San 31,Hyoja,Nam-gu Pohang, Korea 790784,Jan,2009

[2] Abhay Singh Anil Kumar Singh "Web Pre-fetching at Proxy Server Using Sequential Data Mining" 2012 Third International Conference on Computer and Communication Technology 978-0-7695-4872-2/12 2012 Crown Copyright DOI 10.1109/ICCCT.2012

[3] N Nandini, H K Yogish G T Raju "Pre-fetching Techniques for Effective Web Latency Reduction - A Survey" 978-1-4673-5943-6/13 IEEE 2013

[4] C. Vignesh Manikandan, P. Manimozhi, B. Suganyadevi, K. Radhika, M. Ash "Efficient Load reduction and Congestion control in Internet Through Multilevel Border gateway Caching" 978-1-4244-5967-4/10/ 2010 IEEE

[5] Xiaorong, Cheng, Hong Liu, "Personalized Services Research Based on Web Data Mining Technology", Computational Intelligence and Design, 2009, ISCID 09, 177-180, 2009.

[6] Pallis G., Vakali A. Pokony J., "A Clustering-based prefetching scheme on a web cache environment", Journal Computers and Electrical Engineering, 34(4), 309-323, 2008.

[7] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, On the Implications of Zipfs law for web caching, in Proc. IEEE INFOCOM99, New York, Mar. 1999.
[8] W. Feng and H. Chen, A matrix algorithm on Web cache pre-fetching, to appear in the Proceedings of the 6th International Conference on Computerand Information Science (ICIS 2007), Melbourne, Australia July, (2007).
[9] OngChen, HuaiyZhu, Xian-He Sum, "An Adaptive Data Pre-fetcher for High Performance Processors," 2010, 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.
[10] Chi-Keung Luk, C. Todd, Mowry, "Compiler-based pre-fetching for recursive data structures." 7th Conference on Architectural Support of Programming Languages Operating systems. New York, NY, USA: ACM. pp. 222-233.
[11] Farhan, Intelligent Web Caching Architecture. Master thesis. Faculty of Computer Science and Information System, UTM University, Johor, Malaysia, (2007).
[12] George Pallis, Athena Vakali, and Jaroslav Pokorny. A clustering-based prefetching scheme on a Web cache environment. Computers and Electrical Engineering, 34(4):309323, 2008.

BIOGRAPHIES

Mr. Swapnil S. Chaudhari received the B.E. degree in Computer Engineering from JSPM's Imperial College of Engg, Pune. He is currently doing his M.E in Computer Networks in G. H. Raisoni College of engg. & Mgmt, Pune, (MH). His Research area is web prefetching, Web services, Proxy servers.