

Integration of Caching and Prefetching in Proxy Server

Anish Kumar Saha

Assistant Professor

Dept of Comp. Science & Engg
NIT Arunachal Pradesh, India

Hrituparna Paul

Assistant Professor

Dept. Comp. Science & Engg
NIT Agartala, India

Partha Pratim Deb

Lecturer

Dept. of Comp. Science & Engg
Techno India College of
Technology, West Bengal, India

Abstract--Web caching is the technique to caching of web documents into cache memory. The web objects are generally cache in the proxy servers, which are closer to the clients for quick retrieval of data. Only web caching in proxy server is not sufficient for good caching management in any proxy server. Web Prefetching is the technique in which to predict the future web objects from the current used web objects in the cache memory with the high probability of chances of cache hit in the proxy server. Here web caching and web Prefetching are integrate into the proxy server, and provides higher chances of cache hit in the proxy server to deliver web objects to the clients. The rank based cache replacement policy is used here. The rank is depends on frequency, Latency and bytesent of the web objects. Here the History based and cluster based technique for Web Prefetching is proposed. The proposed Prefetching is find out the list of Prefetching web objects for the requested web objects and placed into the cache memory in proxy server.

Keywords--Web Caching, Web Prefetching, Latency, Frequency, Cluster

I INTRODUCTION

Caching and Prefetching in proxy server provides quick retrieval of information from cache memory and deliver the requested web objects to the clients without communicate to the main web server. As a result, the bandwidth, latency time reduces, since the proxy server is closer to the clients. Web caching is also reducing the server load. Proxy server also provides load balancing for web servers and delivers the request to different servers on their loads basis. Already different types of caching algorithm are there. It is good to place the web objects to the cache and deliver to the clients based on their requests. The main problems of web caching are replacement policies and cache consistency. Authors [3] provides information of different web cache replacement policies like First In First Out (FIFO), Least Recently Used (LRU), Least Frequently Used (LFU) and so on. A K Saha et al [1] proposed different parameter based cache replacement policies of the web objects.

Consistency is the problem in which the cached objects in the cache memory should be valid with respect to the original web objects contents in the original server. There are different techniques to overcome the problem. Some examples among them are TTL (Time-To-Live), PCV (Piggyback Client Validation), PSI (Piggyback Server

Invalidation), Polling-every-time, Invalidation information from server, Lease in Client/server Interaction etc [4]. Only caching technique in the proxy server may not be deliver optimal solution. Because, Web Caching is dully depend on the current usages. Another solution is Web Prefetching. In web prefetching, the prefetch engine, which is an algorithm, finds the prefetch web objects from the current used web objects in the cache memory. That means the perfecting technique predicting the future web objects from the current used web objects with higher probability chances of cache hits. Some examples are Prefetching hyperlinks, Association relation, History based, Page rank based [5].

Other than, Web mining based Prefetching algorithm are also proposed [6] [7] [8] [9]. The Neuro fuzzy System in Client side caching is discussed [11].

II. WEB CACHING

Web caching is the technique in which to keep the clients used web objects into the cache memory of proxy server. By placing the web objects into proxy server, it is reduces response time or latency time, bandwidth uses and improves traffic on Internet. Proxy server is of different types, says Forward Proxy Server, Reverse Proxy Server, Load balancing Proxy server. Here Rank based Web caching & cache replacement policy is use as discussed in the paper [1]. Briefly discuss, the cache replacement algorithm [1] usages Frequency, Latency and Bytesent as input parameters to calculate the rank of any web objects in the cache memory. This rank gives, which web objects would replaced by the coming web objects for placing to the cache memory. The low rank web objects are removed to provide space for coming web objects. The frequency means the number of cache hits for a period, Latency means the response time to get the web object from the original web server, and Byte Sent is the size of the web object. These values passes to Suegno type Fuzzy Inference System to calculate the rank of the web objects. The parameters values can be collect from Proxy server log file. The samples of rank values for web objects are given in Fig 1.

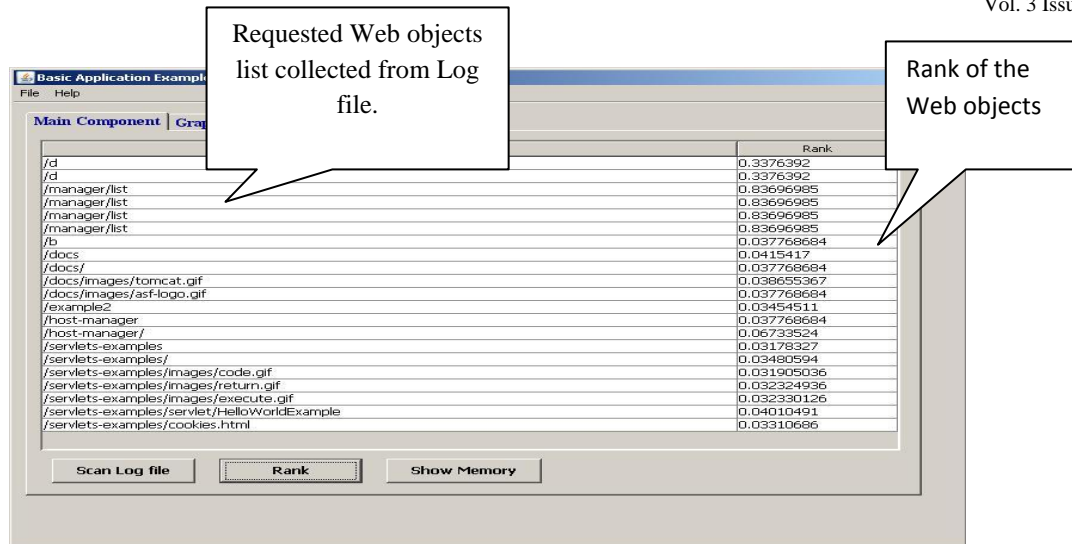


Fig 1: Rank vale samples.

In this approach [1], adding extra prefetch search engine whose task is to place prefetch web objects in the cache memory.

Whenever the low rank web object is replace, another Log-file named as Prefetch-Log file maintain the data of web object name and its last Hold-Time in the Cache memory of proxy server. The Hold-Time calculation can be calculated from simple subtraction of (Current time – Arrived time). The Arrived time of the Web object can get from Log-file contents. Higher value of Hold-Time means good to Prefetch, when its parent web object surf by clients and higher chance to surf the hyperlinked web object from parent web object. The Proxy Server maintains & updates the Prefetch-Log file and the data in Prefetch-Log file use in Web Prefetching.

III WEB PREFETCHING

Different types of Prefetching technique already discovered [10]. Here both Hyperlinks based Prefetching and History based Prefetching used. In Hyperlink based prefetching, all the hyperlink web objects are called and place into the cache memory. Whenever clients surf websites, there are higher tendency to uses any hyperlink web objects from the current web objects. Generally, any web document contains many hyperlinks. Therefore, it would not be feasible to load all the hyperlinked objects to the cache memory of the proxy server due to limited cache memory size as well as cache hit ratio would reduced. In case of History based prefetching, the client's uses history are used. Here both the techniques are integrated to get optimal result of Prefetching. Here Prefetching Engine algorithm finds the best N (Any value) number of hyperlinked web objects from the current requested web object. The value of N is varying depending on the cache management system for different proxy server. The value of N can be 1 or 2 or 3 or 4 or 5 or more. The Prefetch engine select N web objects based on the association

between one web object to another web objects by means of hyperlinks.

All the hyperlinks are cluster into one group for the particular web objects and the best of the N web objects are select based on last Hold-time of each web object. So, the system is based on both History based as well as hyperlink based.

IV THE PROPOSED ARCHITECTURE

The proposed architecture is explain

1. The clients are requesting web objects. The proxy server will check the availability of the web objects. If available, the web objects delivered to the clients, otherwise client request forwarded to the main server. When delivering the web objects from server to the clients, the proxy server keeps a copy of the web objects into cache memory as cached object.
2. Whenever any web objects insert into the cache memory of the proxy server, then the Prefetch Engine will find the Prefetching web objects for the client and the Prefetch web objects will called by the proxy server to main server in backgroundly and place the prefetched web objects into the cache memory of proxy server.
3. The Prefetch engine first finds the hyperlinks of the web document. This is acting as association between parent web document to the hyperlinked web document and all the hyperlinked web objects are classified as clusters for the parent web document. Among the hyperlinks, only N number of hyperlinks would be select as Prefetch Web objects. The selection of N numbers of hyperlinks Web objects are depends on its last Hold-Time and this value already saved in Prefetch Log file. The hyperlink Web objects are search in Prefetch-Log file and from the list of found hyperlink web objects, select the best N number of Web objects.

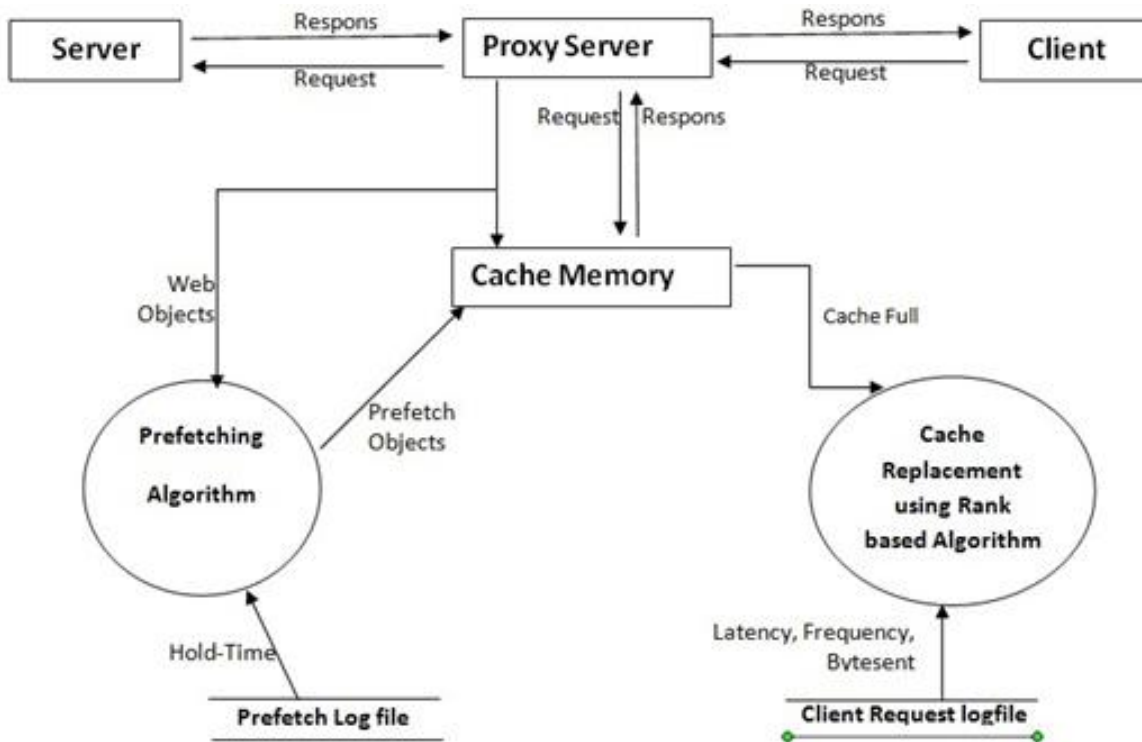


Fig 2: Data Flow Diagram of Proposed Architecture

The higher value of Hold-Time would give higher priority to prefetch.

4. The Prefetch log-file maintains and updates the last Hold-Time of all the web objects during removal of web objects from the Cache memory in Proxy Server. Hold-Time means the cached time of the web objects in the cache memory.
5. The cache replacement policy is based on rank basis. The rank is calculated on the basis of frequency, Latency & bytesent and passes to the FIS. The Fuzzy Inference System (FIS) gives the output acting as rank. Lower the value of the rank means lower priority to cache. This replacement policy is discussed in the paper [1].

V CONCLUSIONS AND FUTURE WORK

The main disadvantage of the proposed method is the complexity to implement and high computation needed for the proxy server. Only web caching provides gives higher cache hit rather than caching & Prefetching. The whole architecture is only for static web objects. Recent web sites generally contain dynamic web documents. But, Dynamic web documents are tough to cache due to its low degree of reusability and strong dependent on backend data. Our plan is to caching the Dynamic Web documents. The dynamic web documents may be cache by caching the data rather than the web documents.

REFERENCES

- [1]. Anish Kumar Saha et al "An Optimization Technique of Web Caching using Fuzzy Inference System" on International Journal of Computer Applications (IJCA) Volume 43- No.17, April 2012, ISBN: 973-93-80867-65-4
- [2]. Ahmad, N.; Malik, O.; ul Hassan, M.; Qureshi, M.S.; Munir, A., "Reducing user latency in web prefetching using integrated techniques," Computer Networks and Information Technology (ICCNIT), 2011 International Conference on , vol., no., pp.175,178, 11-13 July 2011
- [3]. Jaeeun Jeon, Gunhoon Lee, Ki Dong Lee and Byoungchul Ahn, "An Adaptive Prefetching Method for Web Caches" Yeungnam University, School of Electrical Engineering and Computer Science.
- [4]. L. Y. CAO and M. T. OZSU "Evaluation of Strong Consistency Web Caching Techniques" University of Waterloo, School of Computer Science, Waterloo, ON, Canada N2L 3G1.
- [5]. Wei-Guang Teng, Cheng-Yue Chang, and Ming-Syan Chen "Integrating Web Caching and Web Prefetching in Client-Side Proxies" Fellow, IEEE.
- [6]. Qiang Yang, Henry Haining Zhang, Ian T.Y. Li, and Ye Lu "Mining Web Logs to Improve Web Caching and Prefetching" School of Computing Science Simon Fraser University Burnaby, BC, Canada V5A 1S6 (qyang, hzhangb, tie, yel)@cs.sfu.ca
- [7]. Areerat Songwattana "Mining Web logs for Prediction in Prefetching and Caching" School of Engineering and Technology, Asian Institute of Technology Rangsit, Pathumthani, 12000, Thailand areerat@rangsit.rsu.ac.th, sareerat@gmail.com
- [8]. Qiang Yang and Haining Henry Zhang "Web-Log Mining for Predictive Web Caching".

- [9]. Francesco Bonchi¹, Fosca Giannotti², Giuseppe Manco³, Chiara Renso⁴, Mirco Nanni⁵, Dino Pedreschi⁶, Salvatore Ruggieri⁷
“Data Mining for Intelligent Web Caching”.^{1,2,3,4} CNUCE-CN -
Institute of Italian National Research Council^{5,6,7} Department Of
Computer Science, University of Pisa
- [10]. Wei-Guang Teng, Cheng-Yue Chang, and Ming-Syan Chen
“Integrating Web Caching and Web Prefetching in Client-Side Proxies”
Fellow, IEEE.
- [11]. Waleed Ali & Siti Mariyam Shamsuddin “Neuro-Fuzzy System in Web
Client-side Caching” Faculty of Computer Science & Information
System, University Technology of Malaysia

IJERT