

A Survey on Web Cache Techniques used for Web Performance Tuning

Kushal Mehta¹ Jasmine Jha²

¹M.E. Student ²Assistant Professor

^{1, 2}Department of Computer Engineering

^{1, 2} L. J. Institute of Engineering and Technology Ahmedabad, Gujarat, India

Abstract— Internet is growing exponentially in today's world. Data on web servers is getting high and high in terms of size. Many servers reach their limit of serving web pages to requests. That is why the concept of web cache arises. So the server does not have to serve each request actually and requester gets requested data also. This paper represents the survey of different web caching techniques and their performance ratios. Web caching is one of the best techniques to improve web performance. It reduces server load and user wait time. In this paper we will describe caching elements and its characteristics. Then we will do a survey of different existing cache techniques.

Key words: Server, Client, Cache Replacement

I. INTRODUCTION

The internet is growing very fast day by day and lots of new websites are creating every day. As increasing data over internet also increasing space and load on servers. A server getting too much requests may hang or crash if not configured well to handle load. Lots of techniques are available to handle load on server without making it crash.

Caching is a technique which helps a server to focus on new web page request and serve already requested pages from cache repository. Caching is done two ways.

- 1) Client side caching
- 2) Server side caching

A. Client side caching

All latest browsers support client side caching. When a user request a web page using URL hit, that request goes to server. Server executes that request and if a request is valid server response client in form of HTML. This HTML is saved in browser cache. [1]

Command to enable cache at browser side:

```
<meta http-equiv="Cache-control" content="public">
```

Generally a HTML tag <meta> is used to enable cache at browser side. When we clear browser cache all cached pages at browser side are flushed. Client side caching is a light weight caching system to get the pages faster from local system.

B. Server side caching

Server is a processing element which processes all requests from client. It is a good idea to cache (store) already requested pages in a repository. When request comes for same page again there is a no need to execute that page on server and we can direct the request to cache repository. [1]

Cache repository is placed between server's processing unit and response engine. It compares with its history and makes a decision whether it has to be served

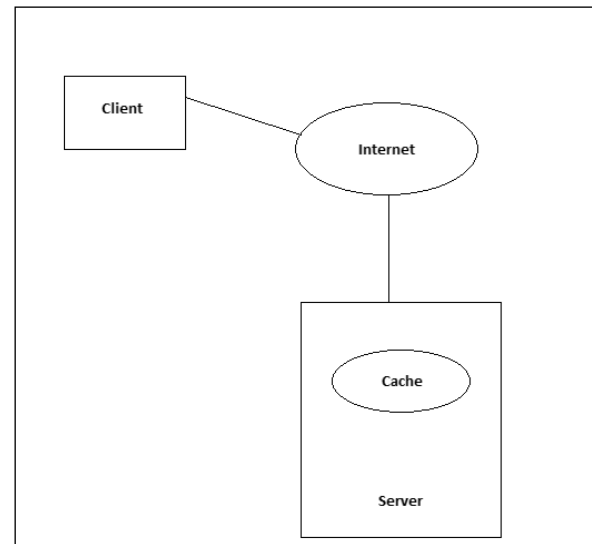


Fig. 1: Server side cache mechanism

from cache or divert request to server for execution. Sometimes rather to keep cache repository on server a proxy server is introduced between server and network. It saves server space for caching. A proxy server is a mediator between client and server.

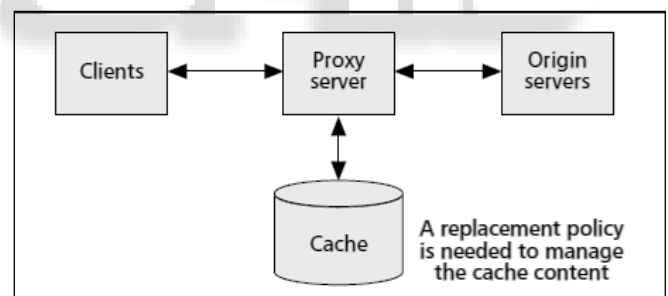


Fig. 2: Proxy server used for caching [2]

Client sends a request to server and proxy server compares requested URL to existing cached pages. If page is available then it is server using cached page. Advantages of using proxy server are it increases speed because already executed pages are available with it. If main server crashes still client can get cached pages from proxy server.

II. CACHE REPLACEMENT TECHNIQUES

Cache replacement policy is required because a server stores cache pages into local storage. Server storage has some limit in terms of memory. If storage for cache is full and some page needed to be written into storage then a cache replacement policy should be there who decides which page will be discarded from cache storage. Generally all cache replacement techniques work basic attributes like hit rate and byte rate.

A. Cache replacement policies categorization

- 1) Recently used
- 2) Frequently used
- 3) Size based
- 4) Function based
- 5) Random policy

Most frequently used cache replacement policies are Least recently Used (LRU), Least frequently Used (LFU), Lowest Relative Value (LRV), Greedy Dual Size (GDS) and smart cache. [3] A summary of all cache replacement algorithms is shown in Table 1.

Algorithm	Details	Advantage	Disadvantage
LRU	Least recently used cached document will be replaced first	Easy to implement	It only use time as decision factor
LFU	Least frequently used document will be replaced first	Simple to use	It only use time as decision factor
Function based	Works on some function	Best fit if function is proper	Cannot be used universally
Size based	Large disk space occupied documents will be removed first.	High hit rates	Lower size files will not be removed
GDS	Use of H factor to remove files	Overcome the drawback of size based	It will not consider frequency and time

Table 1: Comparison of Cache techniques

III. PROPERTIES OF WEB CACHING

Properties of web caching system include quick access, robust, efficiency, heterogeneity, load balancing and simplicity. We will look each property one by one. [1]

- 1) Quick Access: Web cache implemented systems have quick access of website than not implemented system. At user's side it decrease download time of a web page.
- 2) Robust: for user a robust system is that which is available whenever it is needed. Cache system should not crash at single point. And if it crashes it should have an easy way to restore the system.
- 3) Efficiency: A cache system should be efficient that it does not vary on speed time or download time very often.
- 4) Heterogeneity: Now a day result from web is delivered through nearest server geo-location. A cache system or architecture supports this kind of heterogeneity.
- 5) Load balancing: This concept works on multiple caching or proxy servers. All proxy servers communicate with each other to balance load of requests.

- 6) Simplicity: Simple things are always acceptable in technological world rapidly. A simple cache system which is easy to deploy is accepted by world fast.

IV. MEASUREMENT OF WEB PERFORMANCE

Web caching performance measurement is needed to scale web performance techniques. Some milestones are required which proves that some caching technique is a best fit for some situation. Such milestones will help to choose a perfect cache solution for a site, and also it show some specific structure is efficient for specific kind of application.

Similarly, some firms may choose for proxy based caching system. They may try to overcome the problem of configuration Web browsers by forcing the use of browsers that provide auto configuration. In the case of massive organizations, they will may use of network components such as routers and switches. Another alternative, they can employ transparent caching. Some firms may prefer highly scalable solutions anticipating future needs. Besides, organizations which Web sites contain highly dynamic content might occupy Active Cache or possibly will utilize Web server accelerators. Obviously, the subject of measurement of performance is controlled not just to find the competence of a given Web caching solution but also to cover evaluation of performance of cache consistency protocols, cache replacement algorithms.

A. Parameters for measuring Web performance

Several parameters are used when scaling Web caching policies. [3] These include the following:

- 1) Hit rate: The hit rate is generally a percentage ratio of documents obtained through using the caching mechanism versus the total documents requested. In addition, if measurement focuses on byte transfer efficiency, weighted hit rate is a better performance measurement.
- 2) Bandwidth utilization: A reduction in the amount of bandwidth consumed shows the cache is better.
- 3) Response time/access time: The response time is the time takes for a user to get a document.

V. CONCLUSION

As increment of web traffic an effective solution to minimize delay to user is needed. Web caching is a very good solution for this purpose. In this paper we analyze different web caching techniques and basics of web caching mechanism. Web caching can be easily be implemented in current websites and future web sites also.

REFERENCES

- [1] Wang, Jia. "A survey of web caching schemes for the internet." ACM SIGCOMM Computer Communication Review 29.5 (1999): 36-46.
- [2] Wong, Kin-Yeung. "Web cache replacement policies: a pragmatic approach." Network, IEEE 20.1 (2006): 28-34.
- [3] Sarina Sulaiman,Siti Mariyam Shamsuddin, Ajith Abraham, Shahida Sulaiman, "Web Caching and Prefetching: What, Why, and How?" Network, IEEE 20.1 (2008): 28-34

- [4] Shi, Lei, and Yan Zhang. "Optimal model of Web caching." *Natural Computation, 2008. ICNC'08. Fourth International Conference on*. Vol. 7. IEEE, 2008.
- [5] Benevenuto, Fabricio, et al. "Web cache replacement policies: properties, limitations and implications." *Web Congress, 2005. LA-WEB 2005. Third Latin American*. IEEE, 2005.
- [6] Cardenas, L. G., et al. "Emulating web cache replacement algorithms versus a real system." *Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on*. IEEE, 2005.

