A Web Cache Replacement Technique of the Divided Scope Base that Considered a Size Reference Characteristics of Web Object

Ko Il Seok
Chungbuk Provincial University of Science and technology
E-mail : isko@ctech.ac.kr

Abstract

We proposed a Web cache replacement technique of a divided scope base that considered a size reference characteristics of a Web object for efficient operation of a Web base system and, in this study, analyzed performance of the replacement technique that proposed it through an experiment. We analyzed a reference characteristics of size to occur by a user reference characteristics through log analysis of a Web base system in an experiment. And we divide storage scope of a cache server as its analysis result and tested this replacement technique based on divided scope. The proposed technique has a flexibility about a change of a reference characteristics of a user. Also, experiment result, we compared it with LRU and the LRUMIN which were an existing replacement technique and confirmed an elevation of an object hit ratio.

I . Introduction

These days Web traffic is taking up a lot of burdens among total network traffics. According to this, an efficient process of a HTTP request to process in Web becomes an important factor to decide on efficiency of a network management. As a result, the use of Web cache server exclusively responsible for HTTP requests increases.

If an object request of a client occurs in a web cache, a cache administrator determines whether a required object exists in storage scope of web cache server and read the object in storage scope of cache and transmit this to a client if existing. If an object required of storage scope of Web cache server is not existing, it is transmitted an object from ordered URL and provides it to a client.

Because an Internet object used often is saved in cache scope as using a web caching, we improve a computing performance of an Internet and decrease a load of a Web server and traffic of total network and can provide a fast response of an object saved in to cache to an Internet user finally[1-4]. A Web caching improves performance of an Internet as an object of an Internet used often being saved in cache scope[1,2]. Performance of a Web caching technique depends on effective management of a limited storage space. Because storage scope of a web cache is limited, we can increase performance of cache if we let the object that a use frequency is high save in storage scope of a cache.

An efficient replacement technique is an important factor to improve performance of a caching. Up to now, FIFO(First In First Out), LRU(Least Recently Used), LFU(Least Frequently Used), LUV, LRUMIN, a lot of replacement techniques have been researched[1-4]. As for the replacement technique of a Web caching, there is a difference with a traditional technique as a file system and a memory system. An Internet caching must efficiently support the object which a user requires. But, as for the object which an Internet user refers to, size is very variable in dozens of MB from several bytes[5]. And a preference degree of each object is variable, too, and a size deflection of an object is very large. Also, users have a characteristics of object reference according to time and an

area, age and a preference degree of the Internet use.

Therefore, a study about the replacement technique that considered an object reference characteristics of a user is required. Among existing replacement techniques, LRU, LFU is to have applied a traditional replacement technique to an Internet caching field to have the specialty that was an object of variable size. LRU is a way to delete from a storage space from an unused object most recently in order to get a replacement space of an object. And LRUMIN is modification of LRU.

But these techniques cannot reflect enough an object reference characteristics of a user using Web. Particularly, as for the size of the object that a preference degree is high to occur about a specific Web system, it is had a lot of influences on performance of a caching.

In this study, we proposed a Web cache replacement technique of the divided scope base that considered a size reference characteristics of a Web object for efficient operation of a Web base system. Also, through an experiment, we analyzed an object reference characteristics to occur because of the reference characteristics about object size of a user by log analysis.

And we applied this analysis result and divided a storage of a cache to two scope, and measurement did an object hit rate of a proposal technique here. We compared it with the experiment result, LRU and LRUMIN, and a proposal technique confirmed that an object hit rate was improved.

II. Log analysis of a reference characteristics

characteristics analysis of object reference is required for efficient operation of a cache storage space. We analyzed a characteristics of a Web object by log analysis in this study. It is to have shown the log analysis result with Figure 1.

Like seeing it in figure 1, there were a lot of object requests between 100k bytes from 100 bytes. Particularly, there were the most a lot of object requests between 10 kbytes from 1 kbyte. But the result of log analysis will differ if reference of the object which included a multimedia data is increased.

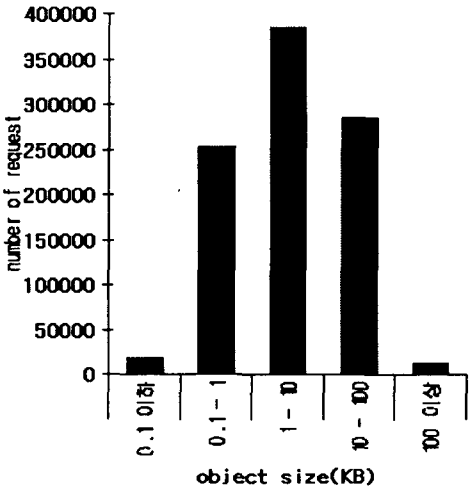

Figure 1 Size and the request frequency of an object

Table 1 Total transmission quantity about each object size

Object size(KB)	request frequency	request frequency rate	total transmission quantity
0.1 or below	17,823	1.86	891.15
0.1 - 1	253,654	26.53	139,509.7
1 - 10	385,428	40.31	2,119,854
10 - 100	285,673	29.88	15,712,015
100 or above	13,567	14.19	2,102,885
TOTAL	956,145	-	20,075,154.85

In a Table 1, we calculated a total transmission quantity and rate about the object request frequency and each object size. We processed the request frequency and object size of each object statistically and, in a table 1, calculated a total transmission quantity and rate. A total transmission quantity is the value that multiplied the request frequency by in an average of required object size, and a request frequency rate is to have shown a frequency rate about a request of a total object with a percentage. And a total transmission quantity rate is to have shown a rate about a

total transmission quantity with a percentage.

There are the most a lot of requests between 10 kbytes from 1 kbyte in the request frequency side as we can see it in a table 1. But, in the total transmission quantity side, the highest value is a value between 100 kbytes from 10 kbytes. We can get two following conclusions through this result. First of all there are more a lot of the request frequencies of a 10kb or above object than the request frequency of a 10kb or below object. With the second, a request of a few frequencies about a large object increases traffic of network at once.

Also, we must consider the following point in order to apply these points to a replacement technique of a cache. The first, a frequent object replacement request of a small object increase a cache miss and become a factor to decrease the efficiency of a replacement technique. With the second, a replacement becomes a factor to decrease the efficiency of networks traffic at once by a cache miss of a large object.

Therefore, what is required in order to increase the efficiency of a Web cache is the variable object replacement technique that considered size of an object.

III. The replacement technique that was based on size of an object

Divided number, size of scope to be allocated to each cache scope, size of an object of cache scope are with a factor to have an important influence on performance of a proposed technique. We can increase a hit rate of cache replacement as what cache size of scope for an object used often to use is allocated to bigger than cache size unused often. The technique that proposed it divided the large object to LARGE scope than 10 kbytes according to size, and the small object divided it to SMALL scope than 10 kbytes.

In Figure 2, it is to have shown the replacement technique how we used divided cache storage scope for. If a request of an object occurs in a client, a cache administrator classifies a required object according to size. And a cache administrator confirms whether an object exists in the cache scope. If an object required in a client exists in a cache, at the time of this, it is provided the object ordered to a client by a cache. Also, a cache administrator saves a use time record of this object, and this object is allotted high ranking for in a LRU replacement through this. If a required object does not exist in a cache, as for the cache administrator, it is transmitted an object from the Internet server which this object exists in.

And a transmitted object is classifieding according to size, and it is allotted a storage space of a cache, cache administrator confirms whether a space for this object to be saved in is at allotted cache scope. If there is the storage where this object can be saved in, this object is just saved, and if a storage is not, apply LRU and make a storage, and an object is saved here. Like having described at the above, a proposal technique is a Web cache replacement technique of the division scope base that considered a size reference characteristics of a Web object.

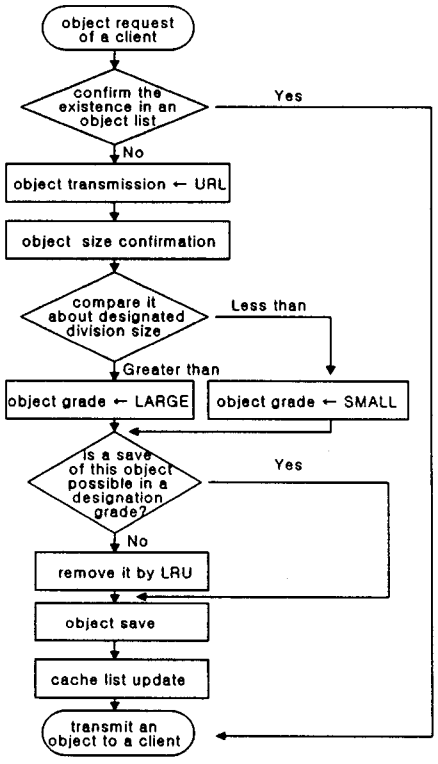

Figure 2 cache replacement technique

Ⅳ. An experiment and analysis

Generally, as for the measure of a replacement technique performance evaluation, cache hit ratio, byte hit ratio, delay saving ratio, cost saving ratio, and response time are used. Size of an object to use in Web is very variable. Therefore, we must use the measure that considered object size and a hit ratio all in order to evaluate performance of a replacement technique. We measure an average object hit ratio and response time and evaluate performance of a proposal technique through an experiment in this paragraph. An average object hit ratio is the evaluation measure that considered a hit ratio and size of an object all. An average object hit ratio can show like equation (1) below.

$$\text{Average object hit ratio} = \frac{\sum_{i=1}^{n} s\_o_i \cdot n\_hit_i}{\sum_{i=1}^{n} s\_o_i \cdot n\_req_i} \times 100 \quad - (1)$$

s_oi : Size of object i
n_hiti : Number of hits for object i
n_reqi : Number of request for object i

We evaluate performance through measurement of an average object hit ratio like equation (1) in this experiment. An average object hit ratio is calculated in a size rate of the object which a hit became about size of the total object which ordered in a client. If an average object hit ratio rises, a fast response can be provided about an object request of a client and can reduce traffic load of network. Also, it can efficiently decrease a load of a system.

Allot LARGE 60%, SMALL 40% to storage scope of a cache and measured an average object hit ratio and response time, and LARGE 70%, SMALL 30% was allocated too and, in an experiment, measured an average object hit ratio and response time. If a cache space increased, as for the LRU or the LRUMIN, the efficiency rose like Figure 3, Figure 4.

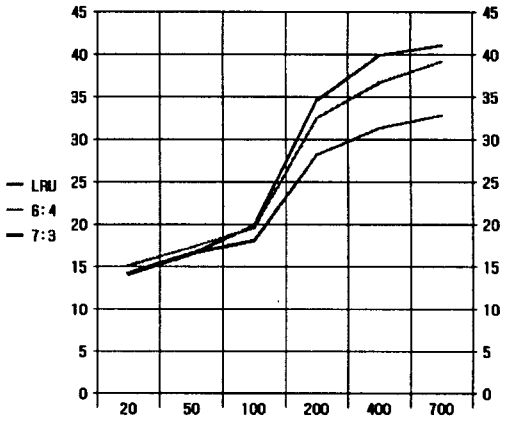

Figure 3 mean value of object hit ratio(%): compare with LRU

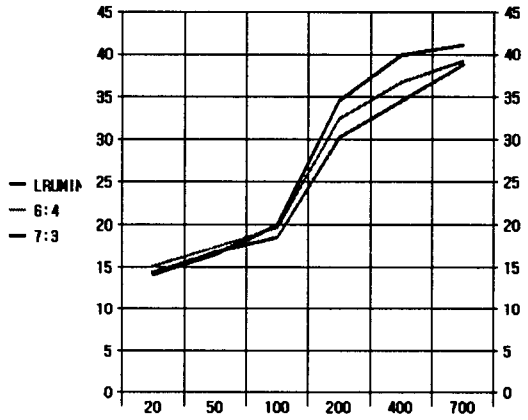

Figure 4 mean value of object hit ratio(%): compare with LRUMIN

The hit ratio of a LRU, LRUMIN technique and the hit ratio of 6:4, 7:3 technique were almost similar in a small-sized cache. But size of a cache grows larger, a hit ratio of a proposal technique rose than a hit ratio of an existing technique.

According to an influence of an environment, the result of this experiment is variable. Particularly, a large influence will occur in the experiment result according to a user preference degree related to an specific object like a frequent request of a multimedia object. A replacement of a large object is an important factor to let delay response time than a replacement of a small object. If it included a

large object very much, in this viewpoint, it will increase the efficiency of a replacement technique to divide division of a cache into 7:3 than 6:4. But a continuous researcher is required about the influence that size of division has on performance of a cache.

The result of log analysis will differ according to the network traffic that is an experiment place and a physical environment and experiment time. In this way a change of an experiment environment will change the measurement result of a hit ratio and response time. Therefore, a continuous study is required about this situation.

V. Conclusion

In this study, we proposed a Web cache replacement technique for efficient operation of a Web base system. The technique that proposed it was a Web cache replacement technique of the divided scope base that considered a reference characteristics about size of a Web object, and we analyzed performance of the technique that proposed it through an experiment too.

The proposed technique has a flexibility about a change of a reference characteristics of a user. Experiment result, we confirmed that a proposed technique improved an object hit ratio than LRU and the LRUMIN which were an existing replacement technique.

References

[1] Greg Barish and Katia Obraczka, World Wide Web Caching: Trends and Techniques. IEEE Communications, Internet Technology Series, May 2000.

[2] H. Bahn, S. Noh, S. L. Min, and K. Koh, "Efficient Replacement of Nonuniform Objects in Web Caches," IEEE Computer, Vol.35, No.6, pp.65-73, June 2002.

[3] L. Rizzo and L. Vicisano, "Replacement Polices for a Proxy Cache," IEEE/ACM Trans. Networking, vol.8, no.2, pp.158-170, 2000.

[4] S. Williams, M. Abrams, C. R. Standridge, G. Abhulla, and E. A. Fox, "Removal Policies in Network Caches for World Wide Web Objects," Proc. 1996 ACM Sigcomm, pp.293-304, 1996.

[5] V. Almeida, A. Bestavros, M. Crovella, and A. Oliveira, "Characterizing Reference Locality in the WWW," In Proc. of the 4th Int'l Conf. on Parallel and Distributed Information Systems, 1996.