제20회 한국정보처리학회 추계학술발표대회 논문집 제10권 제2호 (2003. 11)

분할 영역 기반의 웹 캐싱 알고리즘

고일석*, 나윤지**, 임춘성***
*,***연세대 컴퓨터산업시스템공학과
**한국COD공학
e-mail:isko0326@korea.com

A Web Caching Algorithm with Divided Cache Scope

Il-Suk Ko*, Yun-Ji Na**, Chun-Seong Leem***
*,***Dept of Computer and Industrial Eng., Yonsei Univ.
**Korea COD Company

Abstract

Efficient using of web cache is becoming important factor that decide system management efficiency in web-Based system. Cache performance depends heavily on replacement algorithms, which dynamically select a suitable subset of objects for caching in a finite space. However, replacement algorithm on web cache has many differences than traditional replacement algorithm. In this paper, a web-caching algorithm is proposed for efficient operation of web base system. The algorithm is designed based on a divided scope that considered size reference characteristic and heterogeneity on web object. The performance of the algorithm is analyzed with an experiment. With the experiment results, the algorithm is compared with previous replacement algorithms, and its performance is confirmed with an improvement of response speed.

Key words caching algorithm, web object, object-hit ratio, response speed

1. Introduction

Cache performance depends heavily on replacement algorithms, which dynamically select a suitable subset of objects for caching in a finite cache space[1,2]. A general goal of replacement algorithm in a uniform caching environment is to reduce cache misses. Researches on the replacement algorithm have been conducted actively using web object at storage scope of a cache[3,4,7].

A replacement algorithm for web cache must reflect characteristics of web object, and the web object which an web user requests has the following reference characteristics[5,6]. First, web object size referred to by a user is greatly variable, and the variable object that web user requests must be efficiently supported by web cache. Second, reference characteristics are variable by an object preference of web user, and a variation of the size is also very large. Third, user reference characteristics have the time locality and the space locality. These characteristics of web object are to have been based on heterogeneity of web object fundamentally. It is becoming issue to reflect heterogeneity to occur by a reference characteristic of web object on replacement algorithm effectively.

Reducing total cost incurred due to cache misses is more important in these caching environments. Therefore, a replacement algorithm is required that reflects these reference characteristics of web object. However, the traditional algorithm cannot reflect enough reference characteristics of the web object. This paper proposed a web cache replacement algorithm based on divided web cache scope. Emphasis is placed on improvement of the object-hit ratio and the response time. In experiment, we divided web cache storage scope into 6:4, 7:3 for evaluation of suggested algorithm. With the experiment results, the algorithm is compared with previous replacement algorithms, and its performance is confirmed with an improvement of response speed.

2. Related work

Web-caching algorithms to relieve network congestion and access latency have proliferated in recent years. The following descriptions indicate how each algorithm selects a victim to purge from the cache.

1) **LRU**(least recently used): Removes the least recently referenced object first. LRU is a algorithm to replace an unused object in storage scope so that a new object gets a storage space.
2) **LFU** (least frequently used): Removes the least frequently referenced object first[4]. LRU and LFU have applied the traditional replacement algorithm to web caching field to have the specialty that was an object of variable size among the traditional replacement algorithms.
3) **Size**: Removes the largest object first. SIZE is a algorithm to replace the largest object among the objects saved in storage scope so that a new object gets a storage space. The web cache is different from traditional cache as a hardware cache and a file system cache. In web cache the unit of an exchange is a web object, and the size of an object is very variable. Therefore, in a case where the size of many objects is small, they are removed from storage scope by one object whose size is large[3]. The algorithm of SIZE improved this issue by replacing the greatest object among objects of cache storage scope for new object.
4) **LRU-min**: Removes the least recently referenced object whose size is larger than desired free space of size s[4]. If enough space is not freed, LRU-min sets s to $s/2$ and repeats the procedure until enough space has been freed. LRU-min is a transformation of LRU.
5) **SLRU** (size-adjusted LRU): $Value(i) = (1/tik) \times (1/si)$, where si is the size of object i, and tik is the number of references since the last time i was referenced.

Thus, web object have various characteristics. Among these, its affect much performance of web caching that is big size deflection and many object requests and diversity of object change. If We cannot reflect this characteristic properly, so that an large-sized object that use frequency is very low is saved in storage scope of a cache, it is replaced a lot of small-sized object of that a use frequency is high. Consequently the result to decrease performance of a cache is occurred in. Therefore, a study on the web-caching algorithm that reflected various characteristics of web object is required.

3. Replacement algorithm

Idea of a proposal algorithm is as follows. First, It is possible to classify objects based on size characteristics, and to manage the divided storage of a cache efficiently. Second, Reference characteristics of an object are variable. Therefore, efficiency based on the divided size of cache storage is variable, too.

The number of division scope, the volume of scope to be allocated to divided scope, and the determination of size to classify an object have an important influence on the performance of web caching in this algorithm. Storage scope of the object that has an influence on web caching must be assigned in order to increase the hit ratio of cache.

The object storage scope of 10 k or above is divided into scope LARGE, and the object storage scope of 10k or less is divided into scope SMALL. Figure 1 shows the replacement algorithm that has divided cache scope.

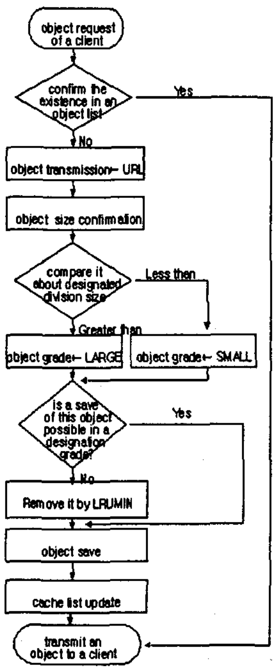

Figure 1 Cache replacement algorithm

When object requests of a client arrive, the cache administrator confirms the existence of an object in the corresponding division scope according to the size of an object. The service of the object that a client required would be provided if a cache-hit occurs. Then, the cache administrator updates the time record on when it was used so that high ranking is assigned to this object in an LRU-MIN replacement.

If a cache miss occurs, cache server requests the corresponding URL for the service of the object, and the object is transmitted to the cache server. Then, a transmitted object is classified into a corresponding grade according to size, and a cache administrator confirms whether there is a space for this object to be saved in a cache scope of a corresponding grade.

If there is a space to save in cache scope, this object is saved, and this object is saved by LRU-MIN replacement algorithm if it is not there. Then, web object saved in each scope is substituted for among the objects of the same grade. Also, a time record of the newly arrived object is saved, and high ranking is assigned to this object in an LRU-MIN replacement process.

As was mentioned in the previous reference characteristics analysis of an object, the reference characteristics and the heterogeneity of web objects would be affected by the characteristics of web service and users aging characteristics and users academic background, and a timing factor. The web service that includes many different kinds of multimedia data and the object reference of a comparatively young age user increase the object reference to large size. According to this, the object reference characteristic has an extreme variation. Therefore, the size of cache scope division must be variable.

4. Experiment

Response time is the time required to provide the requested web object to a client. Figure 2- Figure 4 show the results of the experiment on response time. And Figure 5- Figure 6 show the gain ratio on response time.

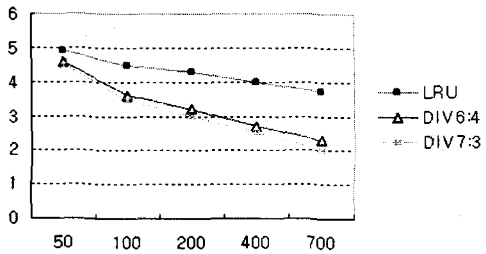

Figure 2 Response time(sec.): compare with LRU

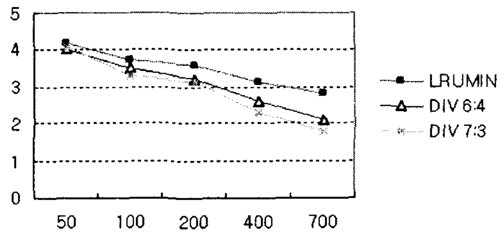

Figure 3 Response time(sec.): compare with LRU-MIN

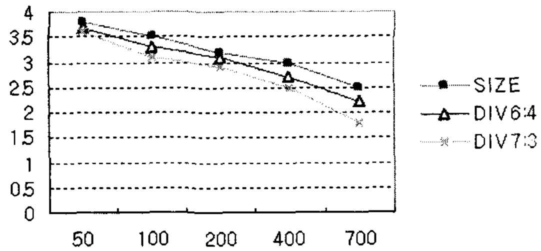

Figure 4 Response time(sec.): compare with SIZE

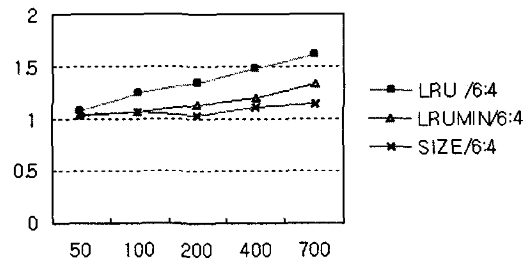

Figure 5 Gain ratio of response time on 6:4

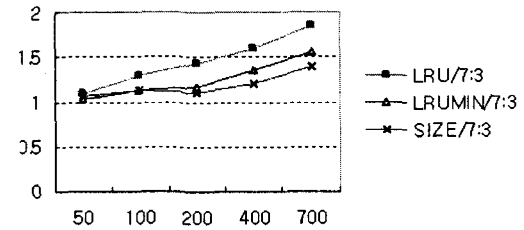

Figure 6 Gain ratio of response time on 7:3

The experiments were also performed as an object-hit ratio in experiment one. Also, experiments were conducted on response time performance of this algorithm and LRU, LRU-MIN, SIZE. We reached the following conclusion by the experiments results of a response time.

1) As the cache scope grows larger, the proposed algorithm and LRU, LRU-MIN, SIZE all improved

their response time.

2) The performance of traditional algorithms and the proposed algorithm were almost similar on a small-sized cache.
3) As the capacity of cache grew larger, the response time performance of the proposed algorithm is more efficient than traditional replacement algorithms.
4) As for the gain ratio of response time, we can get 30% or above performance enhancement than LRU. Also, we can get 15% or above performance enhancement than LRUMIN and SIZE. The reason that performance enhancement of gain ratio is higher than performance enhancement of object hit ratio originated in size of the object which user refer to. There was comparatively a lot of reference on large-sized object in the experiment. According to this, response time was affected delay time of network greatly.

5. Conclusion

This study represented and experimented on the reference characteristics analysis of web object references, the elevation of an object-hit ratio and the improvement of response time on supposed algorithm mostly. Through an experiment, we can got the following analysis conclusions about the influencing factors on the performance of web caching and the performance of the proposed algorithm.

1) Performance of LRU, LRU-MIN, SIZE and the proposed algorithm was almost similar about small-sized cache. But proposed algorithm performed more excellently than traditional algorithms in a large-sized cache.
2) Most of the current web cache server systems are produced with the large volume. Therefore, we can confirm the superior quality of the proposed algorithm by the experiment result.

The experiment results were variable depending on the diverse object reference characteristics and various traffic conditions of the network. Further researches are needed on the division-ratio of storage scope and the operation method of cache that considers this diversity dynamically.

References

[1] G. Barish, K. Obraczka, World Wide Web Caching: Trends and Algorithms. 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, 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. Almcida, 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.

[6] J. C. Bolot and P. Hoschka, "Performance engineering of the World-Wide Web: Application to dimensioning and cache design," Proc. of the 5th Int'l WWW Conf., 1996.

[7] M. Abrams, C. Standridge, G. Abdulla, S. Williams and E. Fox, "Caching Proxies: Limitations and Potentials," Proc. 4th Int'l World Wide Conf., 1995.