객체 참조 특성 기반의 분할된 영역 웹 캐시 대체 기법

고 일 석[†] · 임 춘 성[††] · 나 윤 지[†††] · 조 동 욱[††††]

요 약

일반적으로 웹 기반 시스템의 성능을 높이기 위하여 웹 캐시를 사용하며, 대체 기법은 웹 캐시의 성능에 큰 영향을 미친다. 웹 캐시를 위한 대체 기법은 메모리 영역의 대체 기법과는 다르게, 대체되는 단위가 웹 객체이다. 또한 웹 객체는 사용자 참조 특성의 편차가 매우 크다. 따라서 웹 캐시를 위한 대체 기법은 이러한 웹 객체의 특성을 충분히 반영할 수 있어야한다. 하지만 기존의 기법에서는 이러한 특성을 충분히 반영하고 있지 못하다. 본 연구의 주된 관점은 참조특성 분석, 객체적중률의 향상, 응답시간의 개선이다. 이를 위해 로그분석을 이용하여 웹 객체의 참조특성을 분석하고, 이 참조특성의 분석 결과를 기반으로 분할된 웹 캐시 대체 기법을 제안하고 실험하였다. 실험 결과에서는, 제안 기법이 기존의 기법에 대해 객체 적중률과 응답속도의 성능이 개선되었음을 확인할 수 있다.

A Divided Scope Web Cache Replacement Technique Based on Object Reference Characteristics

Il-Seok Ko[†] · Chun-Seong Leem[††] · Yun-Ji Na[†††] · Dong-Wook Cho[††††]

ABSTRACT

Generally we use web cache in order to increase performance of web base system, and a replacement technique has a great influence on performance of web cache. A web cache replacement technique is different from a replacement technique of memory scope, and a unit substituted for is web object. Also, as for the web object, a variation of user reference characteristics is very great. Therefore, a web cache replacement technique can reflect enough characteristics of this web object. But the existing web caching techniques were not able to reflect enough these object reference characteristics. A principal viewpoint of this study is reference characteristic analysis, an elevation of an object hit rate, an improvement of response time. First of all we analyzed a reference characteristics of an web object by log analysis. And we divide web cache storage scope using the result of reference characteristics analysis. In the experiment result, we can confirm that performance of an object-hit ratio and a response speed was improved than a conventional technique about a proposal technique.

키워드 : Reference Characteristics, Hit Ratio, Web Caching, Object Replacement, Web Object, Response Speed

1. Introductions

Because web traffic takes up a lot of sections in total networks traffic, an efficient process of a HTTP request is becoming an important factor to evaluate the efficiency of a network management. Web cache can efficienta HTTP request, and improve performance of an Internet [1, 2]. Performance of web cache depends on effective management of limited web cache storage scope. Researches about a replacement technique to keep on is used often, web object at storage scope of a cache are performed actively for this [3-5]. A web cache replacement technique must reflect characteristics of web object, and the web object which an Internet user requests have following reference characteristics [6, 7].

① Size of web object referred to by a user is greatly variable, and the variable object that an Internet user requests must be supported web cache efficiently.
② Reference characteristics are variable by an object preference of Internet users, and the size variation is also greatly large.
③ Users have a variable object reference characteristics according to time and an area, age and a proficiency degree of the Internet use.

Therefore, the replacement technique that reflected enough that reference characteristics of web object is required. But the traditional technique cannot reflect enough reference

† 정 회 원 : 충북과학대학 전자상거래과 교수
†† 정 회 원 : 연세대학교 정보산업전공 교수
††† 정 회 원 : 충북대학교 대학원 컴퓨터공학
†††† 정 회 원 : 충북과학대학 정보통신과학과 교수
논문접수 : 2003년 4월 1일, 심사완료 : 2003년 8월 18일

characteristics of web object. In this study, we propose web cache replacement technique of the division scope base that considered size reference characteristics of web object. The important viewpoints are reference characteristics analysis, an elevation of an object-hit ratio, an improvement of response time in this study and an experiment about this. At first, we analyze reference characteristics on size to occur by user reference characteristics through log analysis about web base system for this. And using this analysis result we divide the storage scope of the cache server into 6 : 4, 7 : 3, and experiment on the replacement technique which used divided scope based on an object size. The supposed technique compare to conventional techniques. The experiment result, we can know that an object-hit rate and response time is improved.

2. Related work

Until now many web servers serviced mainly a text and an image object. But web is predicted it develops into service of a multimedia environment gradually, and multimedia service taking up 50% or above of web service since 2003 [10]. An action of web caching is as follows. If an object request of a client occurs in web cache, cache administrator determines whether a required object exists in storage scope of web cache server. If the object that a client required exists in storage scope of web cache, at the time of this, cache hit occurs. And cache miss occurs if not existing, and this becomes a factor to decrease performance of a cache.

In cache hit case, a corresponding object is transmitted from storage scope of a cache to a client. And in cache miss case, a corresponding object is transmitted from requested URL and transmit it to a client. In case of this, we must provide storage of a cache for a new object to use often. A replacement technique is required in order to replace an object to be new to use often with the unused object that is already saved. Web caching improves a computing performance of an Internet as an object used often being saved in a cache, and provide a fast response of an object to a client as what decrease a load of web server and traffic of total networks finally.

There are a client caching, a server caching, a proxy caching, a hierarchical caching, a cooperation server caching in a type of web caching, and the common subject of this web caching method is efficient operation of limited storage scope [1, 8, 9]. We can increase performance of web caching as the object that a use frequency is high being saved in storage scope of cache. Therefore, an efficient object replacement technique is an important factor to improve performance of a caching [8, 9].

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, 8]. LRU and LFU are to have applied a traditional replacement technique to an Internet caching field to have the specialty that was an object of variable size among the conventional replacement techniques.

① LRU is a technique to replace an unused object in storage scope most recently so that a new object gets a storage space to become.

② SIZE is a technique to replace the largest object among objects saved in storage scope so that a new object gets a storage space to become. As for the Internet cache, there is a difference with a hardware cache, a file system cache, and size of an object is variable, and a unit of an exchange is web object. Therefore, the case that many objects that size is small are removed in storage scope by one object that size is large occurs [3]. The case of SIZE improved this issue as replacing the greatest object among objects of cache storage scope.

③ LRUMIN is transformation of LRU, and an action is as follows. At first this technique sets up size of the object that newly comes as S. Then remove it with LRU among large objects of S/2 or above if there is not large space than S. If there is not a large object of S/2 or above, at the time of this, remove it with LRU among large objects of S/4 or above. LRUMIN repeats this process till storage of an object is possible.

Web object has various characteristics. Among these, a size variation of an object and a diversity of an object type have many influences on performance of web caching. Therefore, a study about an operation technique of the web cache that reflected characteristics of this web object is required.

3. Log analysis

We analyze the characteristics of an object using log analysis. The result is affected by characteristics of web service, a user's age, user's scholarship background, reference timing and same many different kinds of factors. Frequency characteristics of a user have a size locality. Web is affected by a state of networks and the same physical factor. Therefore, object reference distribution characteristics of specific time cannot reflect total reference characte-

ristics. So we need a total transmission quantity analysis.

<Table 1> Total transmission quantity and rate about each object size

Object size(KB)	request frequency	request frequency rate	transmission quantity	transmission quantity rate
0.1↓	17,823	1.86	891.15	-
0.1~1	253,654	26.53	139,509.7	0.7
1~10	385,428	40.31	2,119,854	10.6
10~100	285,673	29.88	15,712,015	78.3
100↑	13,567	14.19	2,102,885	10.5
TOTAL	956,145	-	20,075,154.85	-

The <Table 1> calculated a total transmission quantity and the rate about size of each object through the request frequency, and obtained an arithmetic value about a size rate of the object that a user requested. A total transmission quantity is the value that multiplied the request frequency by in an average of requested object size, and, in this Table, a request frequency rate is to have obtained a frequency rate about a request of a total object with percentage. Also, a total transmission quantity rate is to have shown a rate about a total transmission quantity with percentage. The request frequency of a 1~10KB is the highest in the request frequency side, but the 10~100KB ratio is the highest in the total transmission quantity side.

Through analysis of <Table 1>, we can get the following conclusion about the influence that reference characteristics of an object and heterogeneity of an object have on efficiency of web cache.

① The request frequency of a small-sized object (10k bytes or below) is frequent. According to this, we can know a small-sized object generates more object replacements than a large-sized object (10kbytes or above).
② A request of a few frequencies of a large-sized object (10kbytes or above) generates a sudden increase of networks traffic.

Therefore, we can get the following results in order to increase the efficiency of web caching.

① A frequent cache replacement occurring by cache miss of a small-sized object is a factor to lower the efficiency of a cache. Therefore, we must reduce a frequent replacement of a small-sized object.
② A cache replacement occurring by cache miss of a large-sized object becomes factors increase networks traffic at once, and to decrease the efficiency. Therefore, we can reduce a replacement of an large-sized object.

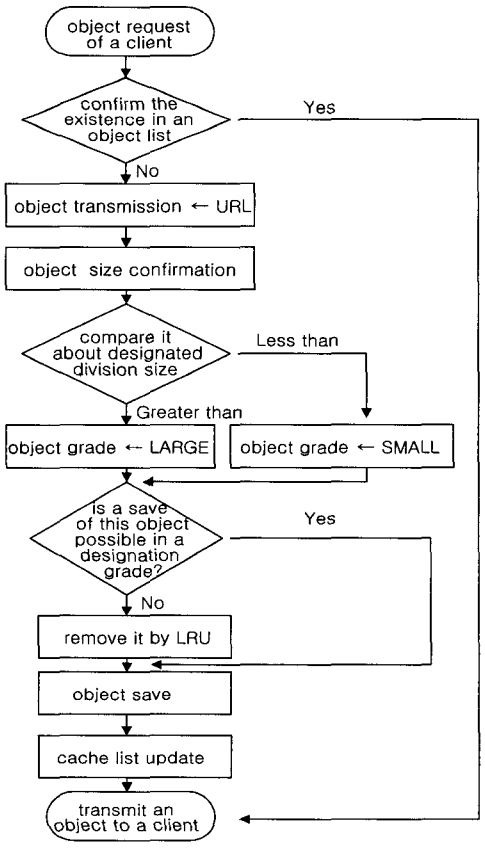
4. Replacement technique

The viewpoints of proposed technique approached is followings.

① We can classify an object as size characteristics. And we divide cache scope like the object size class.
② Reference characteristics of an object are variable. Therefore, efficiency to have been based on division size of cache storage is variable, too.

Numbers of division and scope size have important influences on web caching performance. In order to increase a hit ratio, the storage scope of the object class requested highly is assigned largely.

In this study, we divided cache storage scope of cache into two class. The object storage scope of 10kbytes or above divided it into LARGE class, and the 10kbytes or below object storage scope divided it into SMALL class. (Figure 1) shows the supposed replacement technique.



(Figure 1) supposed replacement technique

If the client object request comes, a cache administrator confirms existence of the object according to size class. If the requested object is exist in the cache, the object is offered to the client. At this time, a cache administrator updates the object using time record so that high ranking is assigned this object in a LRU replacement.

If the cache miss occurs, the cache requires objects to the corresponding URL server. Transmitted objects are classified into the class as size, and the cache administrator confirms whether there are a space for these object to be saved.

If there are a space to save in cache, these objects are saved and replaced by LRU replacement technique if there are not it. At the time of this, objects are replaced at the same class. Also, the timing record is recorded in the cache administrator and high ranking is assigned to these objects in a LRU replacement process.

Like what we mentioned in the previous analysis of web object reference characteristics, the reference characteristics and the heterogeneity of web object are affected according to the environment that are a character of web service, user's age characteristic and scholarship background, timing factor. The web service that include many multimedia data increase large-sized object reference. So, it is caused the very extreme variation of object reference characteristics. In this study, we experiment the performance of supposed technique by 6 : 4 and 7 : 3 cache division. But the efficiency of web caching may be increased more if division scope of variable size is used according to dynamic reference characteristics of web object than division scope to have been fixed.

5. Experiment and analysis

Generally, as for the performance evaluation of replacement techniques, a cache hit ratio, a byte hit ratio, a delay decrease ratio, a cost saving ratio and response time are used. In this experiment, we evaluate performance of the proposed technique by a evaluation of an average object-hit ratio and response time.

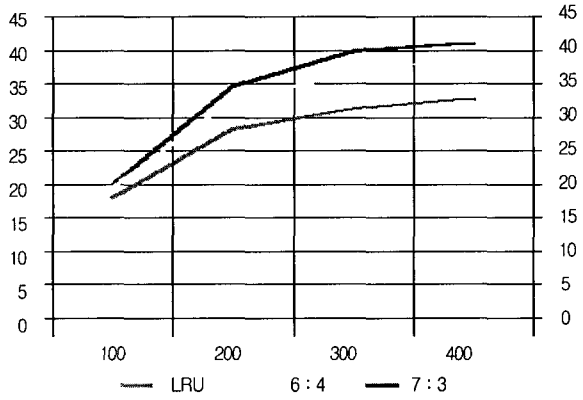
A cache hit ratio can indicate with an object-hit ratio in web caching. At the first, an average object-hit ratio can calculate an average value of an object-hit ratio on a requested object like a Formula (1).

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

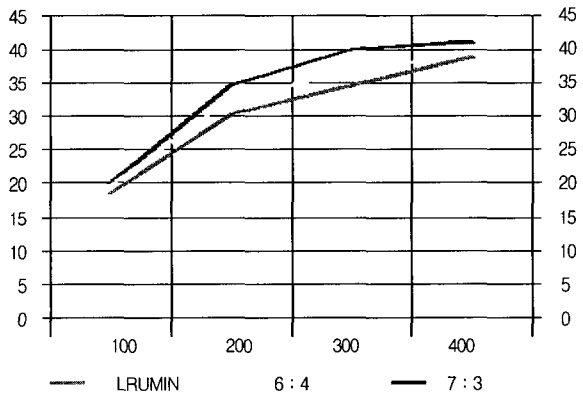
s_o_i : size of object i
n_hit_i : number of hits for object i
n_req_i : number of request for object i

An average object-hit ratio is calculated by a ratio of hit object size in size of the total object that requested in a client. The meaning that an high average object-hit ratio is a faster response to a client, and it is able to efficiently reduce a load of a system. Also, it can decrease traffic of a main server system and local server system.

Experiments method on the proposed technique are as follows. The first, 60% were assigned to a LARGE grade, and 40% were assigned to a SMALL grade. And experimented on performance of this technique and LRU, LRUMIN. In the second, 70% were assigned to a LARGE grade, and 30% were assigned to a SMALL grade. Also, experimented on performance of this technique and LRU, LRUMIN.



(Figure 2) Mean value of object hit ratio (%) : compare with LRU

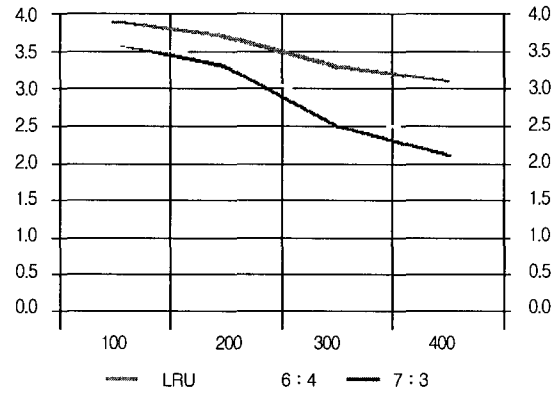


(Figure 3) Mean value of object hit ratio (%) : compare with LRUMIN

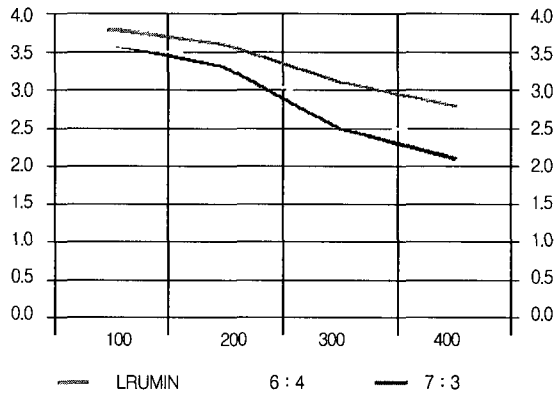
These experiments result show with (Figure 2) and (Figure 3). We can get the following conclusions by experiment results.

① In cache of a few capacities, a hit ratio of LRU and LRUMIN, and the proposed technique are almost similar.
② The cache capacity grows larger, the performance of LRU and LRUMIN is improved. In these cases, the performance of LRUMIN showed up in this experiment higher than performance of LRU.
③ If capacity of cache grows larger, the performance of proposed technique is better than LRU and LRUMIN, and there are an average of 7~10% performance enhancement.

The response time is the time when it takes a web object service request that occurred in a client to be provided to a client. (Figure 4) and (Figure 5) are the results that experimented on an object response time.



(Figure 4) Mean value of response time (sec.) : compare with LRU



(Figure 5) Mean value of response time (sec.) : compare with LRUMIN

A response speed has tight relationship on an object-hit ratio of experiment 1. The experiment was equally performed as an object-hit ratio, and the experiment methods are as follows. In the first experiment, we assign 60% : 40% to the LARGE : SMALL. And the experiment was performed on the response time performance of proposed technique and LRU, LRUMIN. In the second experiment, we assign 70% : 30% to the LARGE : SMALL. Also, the experiment was performed as the first experiment.

① As the web cache scope grew larger the proposed technique and LRU and LRUMIN all, it was improved response time.
② LRU or LRUMIN and performance of a proposed technique were almost similar about small-sized cache.
③ But proposal technique had excellent performance than LRU, LRUMIN in a large volume cache.

In this study, important viewpoints are the reference characteristics analysis and the elevation of an object-hit ratio, the improvement of response time and the experiment about these. Through the experiments, we can get the following analysis conclusions about factors to have an influence on performance of web caching.

① Physical characteristics as system performance and networks traffic have a large influence on performance of web caching. Because the proposed technique improved an object-hit ratio, it can decrease the influence of physical characteristics.
② User's reference characteristics are influenced by a user's preference degree and age, and it have large influence on performance of web caching. Because a proposed technique is based on reference characteristics of a user, it can improve performance of web caching.
③ Performance of LRU, LRUMIN and the proposed technique are almost similar on the small-sized cache. But proposed technique has more excellent performance than LRU, LRUMIN on the large-sized cache. The current most web cache server systems are produced with the large volume. Therefore, we can know superiority of a proposed technique.

6. Conclusion

Generally, we use web cache in order to increase performance of web base system, and a replacement technique has a great influence on performance of web cache. A unit substituted for in web cache is web object, and web object is affected by user reference characteristics greatly. Therefore, a web cache replacement technique must reflect enough

characteristics of web object.

In this study, we propose the divided scope web cache replacement technique based on reference characteristics, and we analyze performance of supposed technique by the experiment. The proposed technique reflect user's reference characteristics and we can confirm that improvement of the response speed and the elevation of the object-hit ratio than conventional techniques.

References

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

[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] 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.

[8] J. Wang, "A Survey of Web Caching Schemes for the Internet," ACM Computer Communication Review, 29, pp.36-46, October, 1999.

[9] C. Aggarwal, J. Wolf and P. Yu, "Caching on the World Wide Web," IEEE Trans. Knowledge and Data Engineering, Vol.11, No.1, pp.94-107, 1999.

[10] S. Sahu, P. Shenoy, D. Towsley, "Design Considerations for Integrated Proxy Servers," Proc. of IEEE NOSSDAV '99, pp.247-250, June, 1999.

고 일 석

e-mail : isko@ctech.ac.kr

1989년 경북대학교 컴퓨터공학(공학사)

2000년 미국 USIU 경영학과(MBA)

1996년 경북대학교 컴퓨터공학(공학석사)

연세대학교 컴퓨터산업시스템공학(박사수료)

현재 충북과학대학 전자상거래과 교수

관심분야 : 전자상거래 시스템, 웹기반개발, CBD

임 춘 성

e-mail : leem@yonsei.ac.kr

1987년 서울대학교 산업공학(공학사)

서울대학교 산업공학(공학석사)

1992년 UC Berkeley 산업공학과(박사)

1993년~1995년 미국 Rutgers University 산업공학과 조교수

1995년~현재 연세대학교 정보산업전공 부교수

관심분야 : 전자상거래, 기업정보화방법론

나 윤 지

e-mail : yjna2967@kebi.com

1994년 경북대학교 생명공학(이학사)

2000년 충북대학교 컴퓨터공학(공학석사)

미국 NYIT Communication ART 전공

충북대학교 컴퓨터공학(박사수료)

전)대전보건대학 컴퓨터정보처리과 (초빙교수)

관심분야 : 멀티미디어 콘텐츠, CBD, 웹기반 응용

조 동 욱

e-mail : ducho@ctech.ac.kr

1983년 한양대 공대 전자공학과(공학사)

1985년 한양대 대학원 전자통신공학과 (공학석사)

1989년 한양대 대학원 전자통신공학과 (공학박사)

1991년~2000년 서원대학교 정보통신공학과 교수

2000년~현재 충북과학대학 정보통신과학과 교수

관심분야 : 영상처리 및 인식, 생체측정, ITS, 정보보호, 퍼지이론