REAL BENEFITS OF ACTIVE CACHING IN THE WWW

ANDREY NAUMENKO
ICA-EPFL, Swiss Federal Institute of Technology, Lausanne CH-1015,
e-mail: andrey.naumenko@epfl.ch

Abstract: We present our research on active caching in the WWW. We drew some key conclusions on the service properties and its significance, according to the results of our analysis of the WWW content. These conclusions include the percentage values of the potentially outdated Web content in the overall Web content after an object's lifetime has passed. This paper provides reasons for using active caching; it may be of interest for researchers of the WWW and for Internet Service Providers.

Keywords: WWW content, active caching, prefetching.

1. INTRODUCTION

WWW caching has been considered by many researchers as a robust mechanism, which improves WWW performance for users by significantly reducing the response time for user's requests. But it is well known that traditional proxy caching technology has limited performance results [1,2]. For example [2] reports that local proxy caching could reduce latency by at best 26%. Further improvements were found when the idea of prefetching or *active caching* of the Web content was developed. In this paper we will consider the advantages and drawbacks of the active caching and provide reasons for its use.

2. MECHANISMS OF CACHING

To deliver a good service to the Web users WWW caching systems support more and more sophisticated functions. Presently in large organizations it is not common to use only one caching proxy server, which stands between the Web users and WWW site, and keeps user requested documents locally, on the storage accessible to the users through a high bandwidth link. Caching performance has been improved, and as a consequence, a modern WWW cache supports many complementary functions. Hierarchical distribution of caches, caching arrays and active caching are among them.

In order to explore active caching functionality of Web proxies, we will need to review several issues related with WWW content. For better understanding it makes sense to consider the separation of WWW content in two following groups [3]:

- *Static content* consists of published on the WWW files such as HTML files, graphics files, video, or sound files.
- *Dynamic content* is dynamically generated WWW content, which was generated in response to an HTTP request that launched a program including CGI scripts, NSAPI scripts, ISAPI scripts, or PERL scripts. These scripts are often used to query databases.

The static content may be cached as long as it was not intentionally flagged as noncacheable. The dynamic content usually should not be cached, although often there are WWW sites with databases that generate dynamic content, which however consists of static objects; these objects can be cached as any static content.

3. ACTIVE CACHING

One of the problems in traditional proxy caching was the problem of stale content. Indeed, if a user was served with the copy of a document from cache and the cached document was not recently verified with the original WWW server, then the user could not be sure that he/she got a fresh document. From the moment the copy was stored in the cache, the original document may have changed and the user would then get outdated information from cache. After active caching was introduced, it became possible to keep in cache only valid content, so that users can be sure that they will always be served with a fresh copy of a document from local cache.

The active caching functionality is usually implemented on the WWW caching proxy server. Normally it should be determined by the webmaster on the original WWW server: which particular WWW object is to be cached and how long it can be kept in cache before it will become outdated. However this option is not used very often. Overall WWW content may be separated into three categories:

- WWW server flags some Web content as non-cacheable. This is often the case with the data which is updated several times during an hour. This kind of content will not be put in proxy cache and active caching is not applicable here.
- Web content is not flagged as non-cacheable and WWW server provides explicit expiration dates and times for it. This content may be cached and active caching may be applied here.

- Web content is not flagged as non-cacheable and WWW server does not provide time to live (TTL) value for it. This content may be cached and active caching may be applied here.

To support active caching, either a caching proxy takes original expiration time from the WWW server or if it was not provided, a caching proxy sets this parameter (TTL) for its cache according to the object's lifetime taken from the date of last modification of the object. Having values for the TTL parameter for the objects in its cache, a proxy with active caching checks when the TTL of an object in cache has expired, then makes a request to the WWW server, downloads a fresh copy of the document and puts it in the cache. All these events happen without user initiation and before the user, in reality, requests the copy of the document. That is why some sources refer to active caching as to prefetching.

There are some algorithms that predict the content, which might be of potential interest to users, and then prefetch it and put it into local cache. To avoid confusion, we do not consider this case as active caching in this paper. Here active caching operates only with the content that was once requested by users and cached locally.

With active caching the proxy will be able to keep up with the status of all documents in cache. To understand the importance of active caching we performed set of experiments.

3.1. DYNAMIC NATURE OF THE WWW

In our experiments we used Microsoft Proxy Server v. 2.0 as a caching proxy with enabled active caching. The proxy served for WWW users in our institute (approximately 40 people). The parameters of the proxy server were setup in a way that it was aware of the status of all cached objects. In the case when the TTL value was not assigned to a content by its originating WWW server, this value was assigned by the caching proxy and was set equal to the object's lifetime (time which passed after the date of object's last modification).

The caching proxy server checked the state of the original copy of the object every time the TTL of cached copy expired (see the description of active caching process in the previous section). A sample of the distribution and percentage for the status of WWW objects that were checked in this way is presented on the Figures 1 and 2.

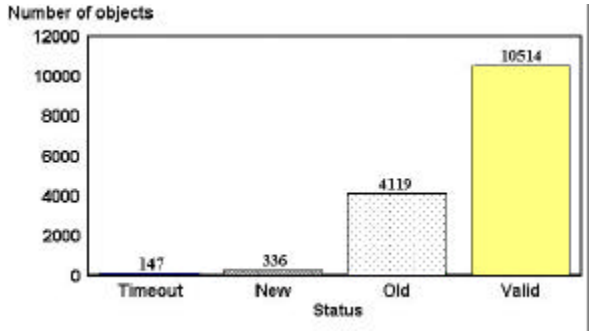

Figure 1: Sample distribution of WWW objects after expiry of object's TTL in cache.

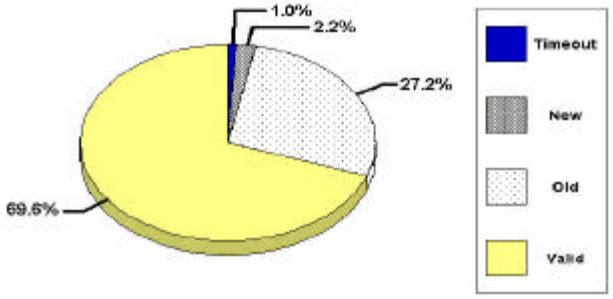

Figure 2: Percentage for the status of WWW objects after expiry of object's TTL in cache.

From these data we can see that that WWW content is very dynamic in its nature: after the TTL of an object in cache has expired, in approximately 30% of all cases this object is no longer valid. Either it does not exist any more on the WWW server where it was originally taken from (27.2% of all checked objects), or it was updated and there exists a fresh copy on its originating WWW server (2.2% of all checked objects). With active caching the former objects would be removed from the cache, the latter would be replaced with their corresponding fresh copies downloaded from the corresponding WWW server.

The dynamic nature of WWW content demonstrates the importance of active caching. Indeed, any cache system without the active caching functionality would not be able to check the status of cached objects and update them if necessary, so cache users would then be served with information, 30% of which would be wrong from statistics.

We can conclude that in addition to the well known benefit of prefetching in advance updated content from a WWW server (that is the significant reduction of the response time for WWW user), active caching has one more very important reason to be used. Active caching should not be ignored, because without it cache users will be served with wrong content.

3.2. INCREASE OF EXTERNAL TRAFFIC

One of the major advantages of traditional caching was the reduction of external traffic for Web caching servers. This point was very important, particularly for Internet Service Provider (ISP) organizations, since they sell traffic, which is internal for them, but buy themselves external traffic to the Internet. So, by having a caching proxy server for their WWW users, ISPs would significantly reduce expenses for external Internet traffic. After the introduction of active caching this point became not as strong, because the active caching functionality of a proxy generates external traffic by itself, without users' initiation, and in this way increases the expenses of the ISPs or any organization, which pays for its external traffic. Thinking about this problem, it would be interesting to have an estimation of this increase in external traffic, in order to understand how much the

expenses would be raised after the introduction of active caching.

Using the same caching proxy software as in the previous case and the same conditions where every cached object has been checked with active caching, we were able to bind the increase in the external traffic resulted from active caching. Of course, the obtained data are valid first of all for the concrete caching proxy software (that is Microsoft Proxy Server v.2.0). But these data should also be valid for other proxy software, since the difference for other caching proxy packages should not be significant because the principles of the work of active caching are similar in many products.

The ratios for the external traffic generated by caching proxy users and for the external traffic that was the result of active caching activities are presented on Figure 3 and Figure 4.

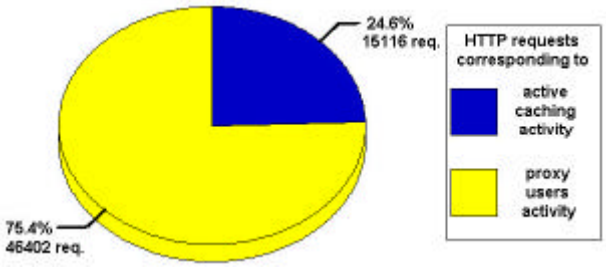

Figure 3: Ratio of external HTTP requests corresponding to active caching activity and caching proxy users activity.

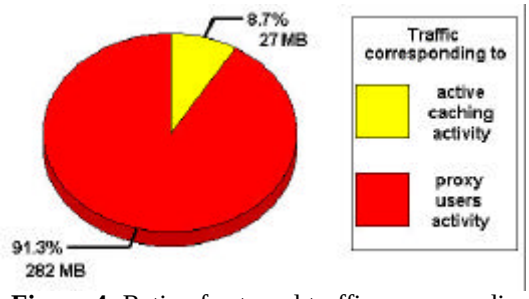

Figure 4: Ratio of external traffic corresponding to active caching activity and caching proxy users activity.

The difference in the percentage of HTTP requests and of traffic appears because active caching causes a download of an updated copy of an object only in 2.2% of all cases (see Figure 2). From the presented experimental data we can conclude that external traffic generated by the active caching functionality of the caching proxy has an upper bound. Since in the measured case the status of every cached object was checked, but with different possible scenarios, active caching may apply to different groups of the cached objects, but not necessarily to all of them. The upper bound may be considered approximately equal to 32.6% of users' HTTP requests (see 1) or 9.5% of the amount of users' traffic (see 2).

$$\frac{24.6\%}{75.4\%} \times 100\% \approx 32.6\% \qquad (1)$$

$$\frac{8.7\%}{91.3\%} \times 100\% \approx 9.5\% \qquad (2)$$

From our long-term statistics, the cache-hit ratio in the caching proxy without active caching was equal to 1/3, on average. This means that 1/3 (33.3%) of the overall initial users' traffic is served from cache by proxy and is not forwarded to the Internet. Let's imagine that it was decided to add the active caching functionality to the caching proxy. Then active caching would generate, at the maximum, 9.5% of traffic from the remaining 2/3 (66.7%) of the overall initial users traffic, which means that to the Internet there would be forwarded, at the maximum, 73.0% of the overall initial users' traffic (see 3).

$$\frac{2}{3} \times 100\% + \frac{2}{3} \times 100\% \times \frac{9.5\%}{100\%} \approx 73.0\% \qquad (3)$$

Comparing the 66.7%, that we may have without active caching and the 73.0% (maximum), that we may have with active caching, we can conclude that the increase of the external traffic caused by active caching functionality is about 6.3% of overall initial users' traffic. This value may be considered as relatively small.

4. CONCLUSIONS

We recommend using the active caching functionality in every WWW caching proxy. Ignoring active caching in the conditions of modern dynamic WWW will result in unsatisfactory service from cache, since invalid content will be given to cache users. According to our experimental data, up to 29.4% of the cached objects will be invalid if the caching proxy does not use active caching. Among them, 27.2% of the objects will not exist any more and 2.2% of the objects will have updated copy on the corresponding WWW servers.

Active caching does not significantly increase external traffic for caching proxy. From our results, proxy caching without the active caching saves only 6.3% more from original users' traffic than proxy caching with the active caching functionality.

5. REFERENCES

[1]. F. Douglis, A. Feldmann, B. Krishnamurthy, J. Mogul, Rate of Change and other Metrics: a Live Study of the World Wide Web, *Proceedings of USITS 97*, Monterey: USENIX association, December 1997.

[2]. T. Kroeger, D. Long, J. Mogul, Exploring the Bounds of Web Latency Reduction from Caching and Prefetching. *Proceedings of USITS 97*, Monterey: USENIX association, USA, December 1997, 13-22.

[3]. R. Lee. *A Quick Guide to Web Server Acceleration*, Novell Research, Advanced Development Group. Novell Inc, October 1997.

[4]. A. Naumenko. Some recommendations on building proxy caching service, *Proceedings of AACE, WebNet 98*, Orlando, USA, November 1998, 682-687.