CONTENT SQUID PROXY SERVER PRODUCTIVITY RESEARCH

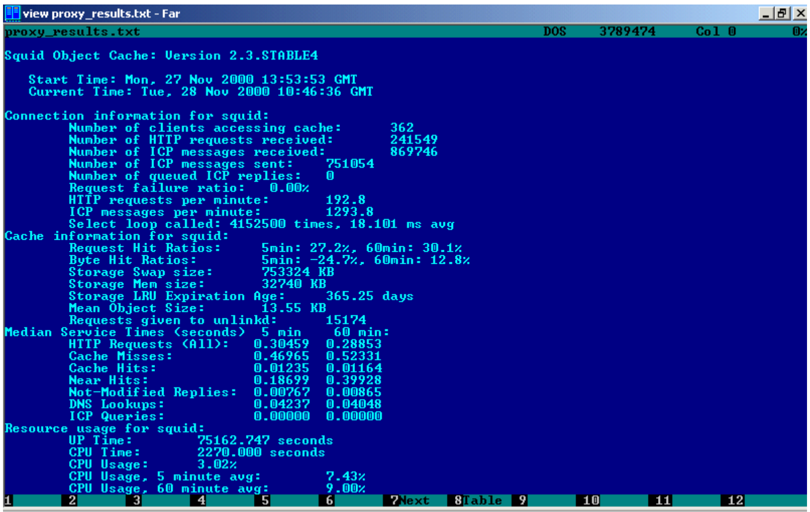
Alexander Berezhnoy, Pavel Danilov

Transport and Telecommunication Institute
JUNIK Internet Service Provider, Riga, Latvia
Lomonosova 1, Riga, LV 1019, Latvia
Ph.: (+371) - 7100504, fax: (+371) – 7100510. E-mail: gpb@junik.lv. avb@junik.lv

The dynamics of web caching technology implementation for the conditions of Internet service provider activity is observed. Basically, two main rates characterizing the variation of resource usage and its effectiveness – request hits ratio and bytes hits ratios are proceeded. Another aspect deals with defining and introducing of formal dependencies and the further analysis of its functional connection with influencing factors. The scope is performed on the basis of real statistics for the case of explicit caching model. Alongside the change of users number and their activity is presented tracing the possible growth trend for the near time period.

Preface

Proxy caching still remains actual for local Internet providers by the reasons of comparatively low amount of information that could be easily applied for their environment. On the other hand both sides (both providers and users) are acting in terms of well-known resource restrictions, meaning not only limited throughout traffic capability but also the time, wasted on obtaining the necessary information.



```
view proxy_results.txt - Far
proxy_results.txt                                   DOS     3789474     Col 0     0%
Squid Object Cache: Version 2.3.STABLE4

   Start Time: Mon, 27 Nov 2000 13:53:53 GMT
   Current Time: Tue, 28 Nov 2000 10:46:36 GMT

Connection information for squid:
        Number of clients accessing cache:      362
        Number of HTTP requests received:       241549
        Number of ICP messages received:        869746
        Number of ICP messages sent:    751054
        Number of queued ICP replies:   0
        Request failure ratio:   0.00%
        HTTP requests per minute:       192.8
        ICP messages per minute:        1293.8
        Select loop called: 4152500 times, 18.101 ms avg
Cache information for squid:
        Request Hit Ratios:     5min: 27.2%, 60min: 30.1%
        Byte Hit Ratios:        5min: -24.7%, 60min: 12.8%
        Storage Swap size:      753324 KB
        Storage Mem size:       32740 KB
        Storage LRU Expiration Age:     365.25 days
        Mean Object Size:       13.55 KB
        Requests given to unlinkd:      15174
Median Service Times (seconds)  5 min    60 min:
        HTTP Requests (All):   0.30459  0.28853
        Cache Misses:          0.46965  0.52331
        Cache Hits:            0.01235  0.01164
        Near Hits:             0.18699  0.39928
        Not-Modified Replies:  0.00767  0.00865
        DNS Lookups:           0.04237  0.04048
        ICP Queries:           0.00000  0.00000
Resource usage for squid:
        UP Time:        75162.747 seconds
        CPU Time:       2270.000 seconds
        CPU Usage:      3.02%
        CPU Usage, 5 minute avg:        7.43%
        CPU Usage, 60 minute avg:       9.00%
1  2  3  4  5  6  7Next  8Table  9  10  11  12
```

Fig. 1. *Squid proxy log file*

One of the exciting research dimensions proceeds the question concerning the choice of an optimal cache size for one's company web proxy caching needs. Obtained ultimate solution may lead to a substantial decrease of direct and indirect expenses connected with traffic supplies. The present research is performed on the basis of the squid cache proxy technology.

One of the sharpest edges is the balance between the desire of the proxy cache operator to maximize system cache hit ratio and the demand of most web-page owners to count the number of work request for their resource. Generally web site owner may not be interested in gathering such kind of information, but in some cases getting the precise amount of web-page visitors works as method of request load distribution that allows to minimize costs of local server productivity. That is a reason for web publishers to use directives denying cached pages usage more than once. In order to avoid this type of defense and increase the cache-server hit coefficient is to ignore the field "Expires" at the header of web-page.

General characteristics for cache measurements

Program logger contains a parameter set that could be logically divided into the following groups: connection information for squid, cache information for squid, median service times, resource usage for squid, memory usage for squid and file descriptor usage, as well as internal data structures (Fig. 1).

Cache information for squid contains the most significant evaluation data allowing to find out possible dependencies. Median service times and connection information groups provide additional data, suitable for this case.

Measurements of described characteristics have been handled for the period of more than two months, starting from Mon, 7 Nov 2000 till Thu, 18 Jan 2001. The first part of gathered information has been cut off, because of initially small size of used swap space. This goes from the well-known fact, stating the lower limit of initial cache size not less than 1Gb. Due to this reason all the information taken into account begins from Mon, 27 Nov 2000, the crucial point when the obtained data may be considered as more reliable and capable to demonstrate the explicit trend of results distribution.

For these purposes a comparatively low swap size of nearly 750Mb has been chosen, allowing to trace the prescribed tendency. Value gaining time period has been set up equal to one hour.

The number of clients accessing cache changed in a random way. The registered minimum of clients one-time presence was near ten people, while the maximum limit of access reached the number of 956. Most part (about 70%) of users were using the dial-up access technology. Selected case also covered the segment of hardusers connected directly to the LAN of Internet service provider and some groups working through the radio links.

The most significant parameters for the further analysis are considered to be Request Hit Ratios (RHR) and Byte Hit Ratios (BHR), that are counted for its average values on a 5 min and 60 min time intervals.

Some more useful information could be gained from the Cache Hit, Cache Misses rates and Storage Memory size.

Dependency of Request Hits and Byte Hit Ratios on a Swap Size

As it has been already mentioned, the dependency of hit ratios on a size of the cache size is a primary object for this research. Figures in the table 1 show the fragments of data measurements within the consequent growth of a proxy cache.

SwapSize	Request Hits Ratio	Byte Hits Ratio
753324	30,1	12,8
766922	29,8	9,9
821078	33,9	12,3
876580	32,6	6,8
933525	28,7	9,7
990783	35	2,9
1035698	37,9	26,4
…	…	
29076751	70	99,3
29079872	1,6	2,9
29081536	29,7	-46,3
29084649	33,4	22,2
29084614	59,6	26,8
29084663	44,2	25,5

Table 1. Fragments of data measurements for the instantly growing swap size

The negative figures in a tab of a byte hit ratios are explained in the following way. Squid counts the number of bytes read from the network on the server-side, and the number of bytes written to the client-side. The resulted byte hit ratio is calculated as

$$\frac{(client\_bytes - server\_bytes)}{client\_bytes}.$$

If the meaning of server_bytes is greater than client_bytes, the result ends up with a negative value. The server_bytes variable may be greater than client_bytes for a number of reasons, including:

- Cache Digests and other internally generated requests. Cache Digest messages are really quite large. They are counted in the server_bytes, but since they are consumed internally, they do not count in client_bytes.

- User-aborted requests. If quick_abort setting allows it, Squid sometimes continues to fetch aborted requests from the server-side, without sending any data to the client-side.

- Some range requests, in combination with Squid bugs, can consume more bandwidth on the server-side than on the client-side. In a range request, the client is asking for only some part of the object. Squid may decide to retrieve the whole object anyway, so that it can be used later on. This means downloading more from the server than sending to the client. This behavior can be affected with the range_offset_limit option.

Results

As a result of consequent logarithmic approximation the slowly growing trendline was obtained both for the Request and Byte hits ratio value series (Fig. 2 & 3).

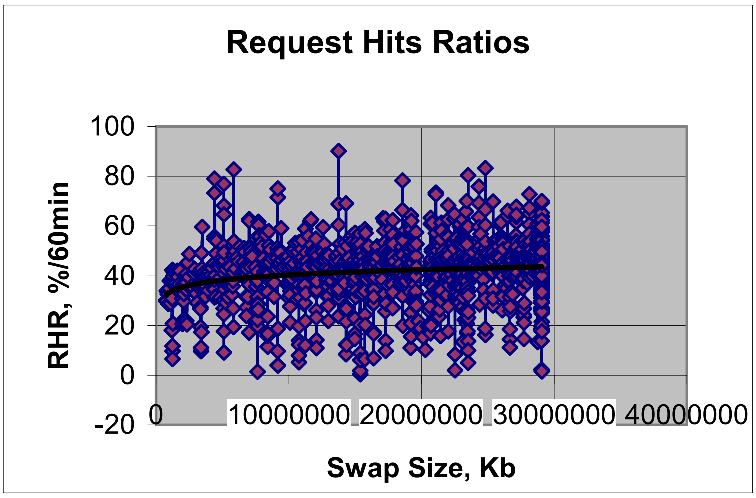


Fig. 2. *Dependency of request hit ratio on a swap size*

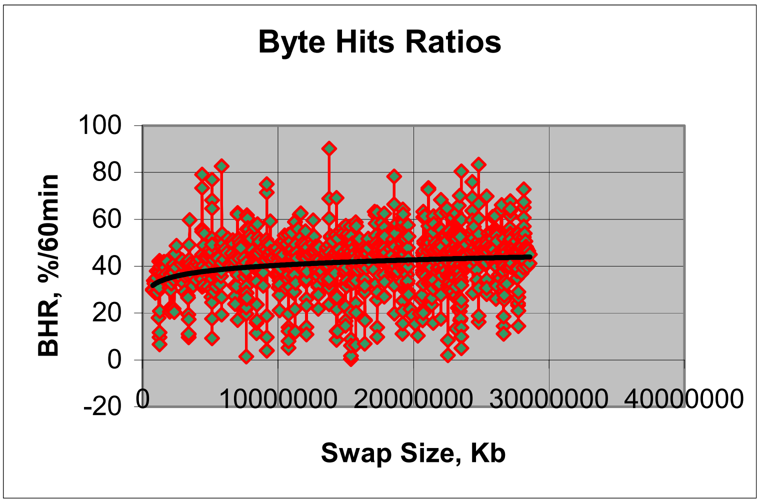


Fig. 3. *Dependency of byte hit ratio on a swap size*

The observed trend dynamics proves the hypothesis of a comparatively low binding level between the initially chosen size of the proxy cache and a possibility feature of its hits function. One of the practical issues that could be marked out in this connection is equally effective meaning of taking the benefit choosing different disk size over one gigabyte, targeted for the proxy cache purposes and resulting in relatively similar amount of its proxy cache effectiveness.

References

1. Geoff Huston, Telstra, "Web Caching", The Internet Protocol Journal (IPJ), Volume 2, Number 3, September 1999.
2. Melve, I. "Inter Cache Communications Protocols", Internet Draft, draft-melve-intercache-comproto-00.txt, November 1998.
3. "Squid Internet Cache" – Resource Web page. http://squid.nlanr.net.