Improving the Performance of Squid Proxy Server by using SCSI HDD and Blocking the Media Streaming

Sanjay S. Kadam
Author
M.Tech. Student
BVDUCOE, Pune
Maharashtra, India

Y. C. Kulkarni
Co-Author
BVDUCOE, Pune
Maharashtra, India

ABSTRACT

This paper discussed the configuration of squid proxy server and how to improve the performance of the squid proxy server by using the various methods and techniques.

As internet users, N/W administrators know that the Squid[1] is a web caching proxy server for the supporting HTTPS, HTTP, File Transfer Protocols, and many more. It saves the bandwidth and advance the response times by caching and repeatedly using frequently-requested web sites. Squid proxy server has wide-ranging access controls and it delivers squid proxy server as great server accelerator. It runs on Linux and Windows platforms and is licensed under the GNU GPL[1] .

So the improving of the performance squid proxy server is essential. This paper introduced two ideas that if use the SCSI HDD(for improving the cache management system) and block the media streaming that is frequently used by various users because media streaming takes lot of bandwidth if administrator blocks unwanted media streaming that also improves the performance of the squid proxy and it accelerated the content delivery.

General Terms
squid proxy server, Cache Management, HDD, Bandwidth.

Keywords
Squid Proxy server, SCSI, Media Streaming, Blocking, ACL

1. INTRODUCTION
In today's Internet savvy world there is lot of people or many educational institutes, companies, organizations started to establish the internet facility in there offices' and working place. So the demand for the internet providers is very high they want to give the internet connection for everyone but the major problems they are facing are the 'Bandwidth allocation and restriction or control of internet accesses.

Squid proxy server is acquiescent to the High Cache Archi. and uses the Inter Cache Protocol (ICP) to transmit the data between two clients and /parent/child server Squid proxy can improve drastically and able to fast the traffic from the inside LAN to the world wide web or it can be employed to act as a front-end accelerator for a Web server, increasing access to the web sites on the server.

Following details give you the Squid can do:
Faster Internet Connections for Internal LAN
Secure the Internal LAN When browsing the Internet
Stores the log of User Activity on the WWW
Blocks the unwanted activity by clients on the www
Allow the internet access to the legitimate users only
Always watch and filter the sensitive data
Do the acceleration of web pages

Squid proxy server works both as a proxy, working in behalf of a client, a cache. When squid proxy server acts as a proxy and a client asks for a request for a web site, squid retrieves the web site and then provides it to the client. The client, in actually never goes to the www as the proxy server gets it and caches all the pages and sites the client asks requests for. So improving the performance of the squid proxy is always essential. This paper gives the idea to use various techniques, hardware's. In first part discussion is based on why SCSI HDD is important over IDE/ATA HDD and second Web Cache Manager.[2] and third part is blocking media streaming.

1.1 SCSI (Small Computer System Interface)

SCSI is normally useful in various servers, and most of in industrial area and educational institutes than home uses.
Table 4 compares the performance of IDE and SCSI disks on the various benchmarks[5], and computes the percentage differences in the last column. We use the UNIX fast file system [4] for the application benchmarks. In this table we use a tag setting of 32 in the driver.

Table 1: SCSI versus IDE Performance.

Benchmark		IDE/ATA	SCSI	%per. Diff
Read Micro benchmark	Sequential	31.6 mbps	15.8 mbps	-100%
	Interleave	23.5 mbps	17.4 mbps	-35%
	Random	1.5 mbps	2.5 mbps	40%
Write Micro benchmark	Sequential	27.1 mbps	19.2 mbps	-41%
	Interleave	19.2 mbps	16.8 mbps	-14%
	Random	1.4 mbps	2.1 mbps	33%
SDET	1 job	10.1 sec	12.3 sec	-18%
	5 jobs	78.2 sec	63.8 sec	23%
	10 jobs	178.5 sec	129.4 sec	38%
	20 jobs	377.1 sec	283.0 sec	33%
SSH		88.5	74.9	18%
PostMark		502	387	30%
NetNews		1100	809	36%

The micro benchmark results indicate that the IDE disk is about 33% and 40% slower on random writes and reads; respectively.[5] Also from the observation of the above table following advantages and disadvantages of SCSI is observed.
Advantages
Very Fast,   Supports large set of applications
Full scalability and flexibility in Arrays (RAID)
Backward compatible with old age SCSI devices
Better for storing, moving & copying large amounts of data

24/7 operations, More reliable.

Disadvantages

Expensive, Full Support is not available

Lot of different types of SCSI interfaces

SCSI drives have a higher RPM, very noisy and generated heat

1.2 Web Cache

What is Web Caching?

'A cache miss means that the request could not be filled from the cache but had to be filled with an actual connection to the web page'

Web Caching [6] is the most common method used by ISP's to reduced the bandwidth all over the world and to improve and FTP data temporary in the ISP server to all over the universe, to improve the performance of the bandwidth also save the bandwidth So we Web caching is nothing but to store the HTTP and FTP data temporarily.

Web Cache Manager:

The cache manager is a convenient way to manage the cache and view statistics without logging into the server. Following fig shows how cache manager works.
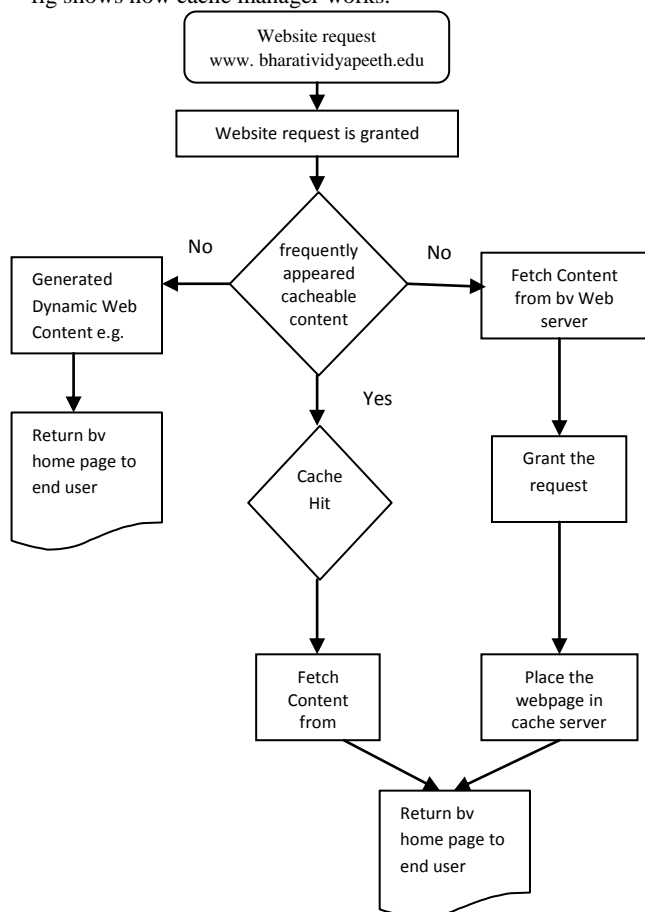

Fig 1: Cache Manager Flow chart

Cache has made the Bharati Vidyapeeth site much easier and faster to download because all the static graphics are already on hand and the only thing you need to complete the picture is the news, results content. Next page describe the how the cache manager is working in the web server and also it is possible to store a temporarily data of the webpage at the server side.

Cache validation refers to the testing of the data so that Squid provides information that is current and not stale information. Often before providing a web page Squid will verify the information and replace it if it is out of date. The way that Squid will verify the information is that each time it saves data to the cache a timestamp is placed on it. This use of a timestamp maintains the integrity of updated information.

After the cache manager flow chart next step is to store the all the cache data in proper place and access it easily and quickly so which will be best HDD for storing the cache data so far normally in inexperienced administrator use the IDE/ATA HDD which is cheap but the performance is not up-to-date but when tried to setup the squid proxy and space allocation for cache data the SCSI HDD the performance is total different.

In a small cache, the chance of a HIT is little, because the cache is easily filled so the less requested objects are replaced by newer objects. On the other hand, if, for example, 2 GB is available for the cache and the client only surf 20 MB a day, it would take more than more than hundred days to fill the cache data.

The easiest way to determine the needed cache size is to consider the maximum transfer rate of the connection. With a 1 Mbit/s connection, the maximum transfer rate is 125 KB/s. If all this traffic ends up in the cache, in one hour it would add up to 450 MB and, assuming that all this traffic is generated in only eight working hours, it would reach 3.6 GB in one day. Because the connection is normally not used to its upper volume limit, it can be assumed that the total data volume handled by the cache is approximately 2 GB. This is why 2 GB of disk space is required in the example for Squid to keep one day's worth of browsed data cached.

The cachemgr.cgi (cache manager) is a CGI utility for displaying statistics about the squid process as it runs.

By using above CGI utility. IDE/ATA HDD and approximately 230 users the following result of the cache manager displayed.

Summary		
Calamaris statistics		
lines parsed:	lines	1510566
invalid lines:	lines	6
skipped lines:	lines	0
parse time:	sec	5205
parse speed:	lines/sec	313
Proxy statistics		
Total amount:	requests	2121461
unique hosts/users:	hosts	237
Total Bandwidth:	Byte	12815M
Max. Bandwidth usage:	MBit/sec	0.03
Proxy efficiency (HIT [kB/sec] / DIRECT [kB/sec]):	factor	5.12
Average speed increase:	%	19.42
UDP response time of 0%% requests:	msec	76
TCP response time of 100%% requests:	msec	25146
Cache statistics		
Total amount cached:	requests	607147
Request hit rate:	%	30.07
Bandwidth savings:	Byte	2321M
Bandwidth savings in Percent (Byte hit rate):	%	20.02
Average cached object size:	Byte	3057
Average direct object size:	Byte	6171
Average object size:	Byte	4920

Fig. 2: Cache Manager Data report when using IDE/ATA and without blocking media streaming

Above picture of report shows the how useful is the cache manager in squid proxy server. This example is the case study of the internet access LAN of bharati vidyapeeth college of engineering, Navi Mumbai.

This statics shows that by using the proper setup of squid proxy server and web cache manager we can save almost 20%

current bandwidth. But various companies, educational institutes and organizations target is to improve more

performance of the squid proxy server. So we proposed new system.

2. PROPOSED SYSTEM ARCHITECTURE
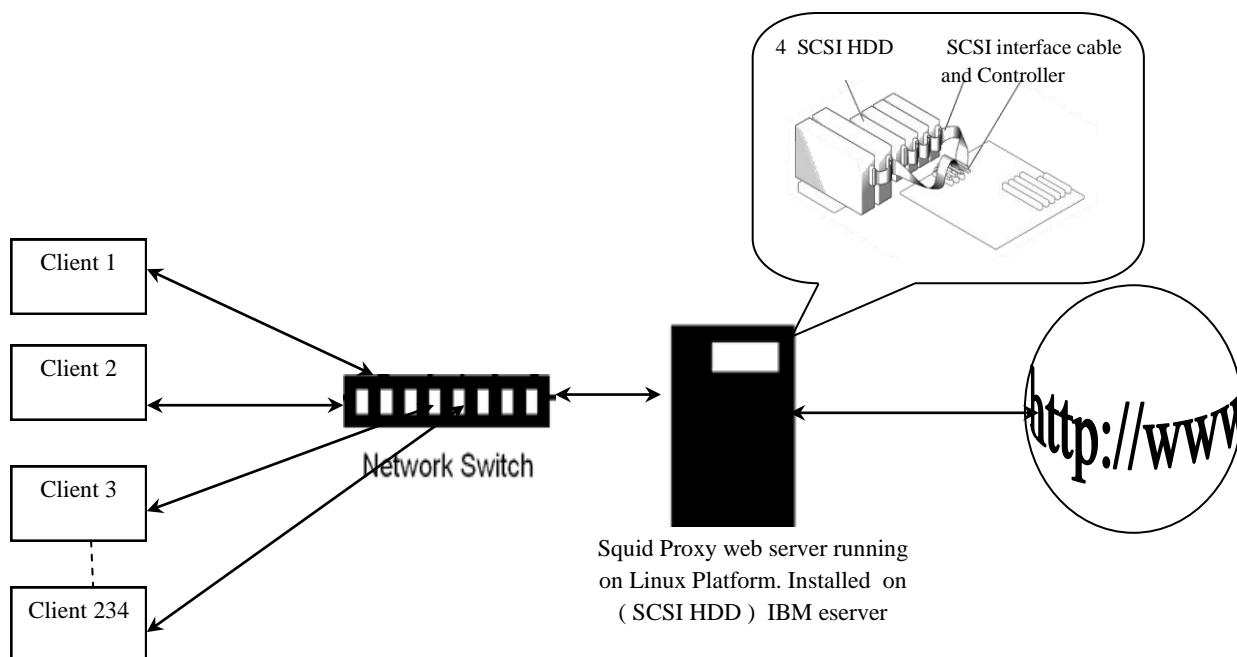## (Using SCSI HDD, Squid and Web Caching and Blocking Media Streaming)

Fig 3: Proposed System Architecture

Speed is an important role in the caching process, so this point deserves special concentration. For hard disks, this parameter is described as random seek time, measured in milliseconds. Because the data blocks that Squid reads from or writes to the hard disk will tend to be rather small, the seek time of the hard disk is more important than its data throughput. For the purposes of a proxy, hard disks with high rotation speeds are probably the better choice, because they allow the read-write head to be positioned in the required spot much quicker. Fast SCSI hard disks nowadays have a seek time of under 4 milliseconds.

One possibility to speed up the system is to use a number of disks concurrently or to employ striping RAID arrays.

This paper explain the all the procedure which adopted at Bharati Vidyapeeth College of Engineering(BVCOE), Navi Mumbai by using SCSI HDD and configuring it with various application and squid.conf.

This paper discusses how the performance of BVCOE, Navi Mumbai institutes internet access and squid proxy server is improved drastically with step by step procedure and advantages of SCSI which is discussed above.

As discussed earlier the first advantage of SCSI to other HDD's Very Fast , yes it is true, after installation of Linux OS and simultaneously squid package it works fine the execution of various packages and rpm packages it is very fast when administrator stop the various services and restart it. The response is very quick as compared with other IDE/ATA HDD's.

The second advantages which discussed are supports of large set of applications. After installation of large applications such as open office. Org, Pl sql, developer tools, and database tools the performance is very good and fast. For configuration of squid need to do lot of administration work for that it is useful web interface such as webmin, calamaris etc.

And also most important it is useful for caching the data.

Squid is proxy server which is open source and lot of administrator easily use it instead of high priced routers and hardware's so it need lot of support from the server hardware system which is essential for the running the squid properly and consistently and the experience shows that SCSI is fully scalable and flexibility in Arrays(RAID) configuration.

Now the most important part is storing the web caching data, access log and cache log which is in large amount, also need to be copy to DVD ROMs through the DVD writers which is also possible by SCSI controller DVD writers.

IN BVCOE for copying the entire access log, cache data used second SCSI hdd. This is better for storing, moving and copying large amounts of data.

Now most important part of the system or internet access management is that storing the cache data as all N/W administrator's know that squid proxy server is also cache server[8]. So that for storing lot of data on the normal hard disk it take lot of time to take backup of the data which is not possible to do quickly and reliably in other hard disk here the SCSI comes and SCSI is better for storing, moving and copying large amounts of data and also it can be work on 24/7 operations and More reliable.

Next topic is also vital for improving the performance of the squid proxy web server.

3. BLOCKING THE MEDIA STREAMING AND IMPROVING THE PERFORMANCE AND REDUCING THE BANDWIDTH.

Often multimedia streaming[8] is not good for the office environment. It will dry out the bandwidth of our network, thus slows it down. Administrator can block the multimedia

streaming using squid filter. Below is the example of how squid is modified:

After lot of trial and error following codes are included in the squid.conf[1] file and it works fine and following codes are the example of it.

```
acl fails rep_mime_type ^.*mms.*
acl fails rep_mime_type ^.*ms-hdr.*
acl fails rep_mime_type ^.*x-fcs.*
acl fails rep_mime_type ^.*x-ms-asf.*
acl fails2 urlpath_regex dvrplayer mediastream mms://
acl fails2 urlpath_regex \.asf$ \.afx$ \.flv$ \.swf$
acl deny_rep_mime_flashvideo rep_mime_type -i video/flv
acl deny_rep_mime_shockwave rep_mime_type -i
^application/x-shockwave-flash$
acl x-type req_mime_type -i ^application/octet-stream$
acl x-type req_mime_type -i ^video/x-ms-asf$
acl x-type req_mime_type -i application/octet-stream
acl x-type req_mime_type -i ^application/x-mplayer2$
acl x-type req_mime_type -i application/x-mplayer2
acl x-type req_mime_type -i application/x-oleobject
acl x-type req_mime_type -i ^application/x-oleobject$
acl x-type req_mime_type -i application/x-pncmd
acl x-type2 rep_mime_type -i ^application/octet-stream$
acl x-type2 rep_mime_type -i application/octet-stream
acl x-type2 rep_mime_type -i application/x-oleobject
acl x-type2 rep_mime_type -i ^application/x-mplayer2$
acl x-type2 rep_mime_type -i application/x-mplayer2
acl x-type2 rep_mime_type -i ^application/x-oleobject$
acl x-type2 rep_mime_type -i application/x-pncmd
acl x-type2 rep_mime_type -i ^video/x-ms-asf$
http_reply_access deny deny_rep_mime_flashvideo
http_reply_access deny deny_rep_mime_shockwave
#streaming files
http_access deny fails
http_reply_access deny fails
http_access deny fails2
http_reply_access deny fails2
http_access deny x-type
http_reply_access deny x-type
http_access deny x-type2
http_reply_access deny x-type2
```

And of course, you have to force your users to use your squid proxy server.

1. CONCLUSION

 All above architecture and modifying squid.conf file works fine and it shows improvement of the performance of squid proxy and lot of users are satisfied by the speed of the internet access system. And also the N/W administrator for using the various techniques and ideas for configuration of squid.conf file and storing access.log file in big size so the SCSI is useful. Also blocking the media streaming is also very useful for improving the performance of the same. Following report shows that the performance of the bandwidth saving and speed of the internet access is almost increased by the

30%. So the conclusion comes that for educational institutes, companies, organizations proposed squid proxy server architecture (using SCSI HDD) with blocking media streaming is very useful.

Summary		
Calamaris statistics		
lines parsed:	lines	2010566
invalid lines:	lines	6
skipped lines:	lines	0
parse time:	sec	4205
parse speed:	lines/sec	213
Proxy statistics		
Total amount:	requests	2119780
unique hosts/users:	hosts	247
Total Bandwidth:	Byte	16815M
Max. Bandwidth usage:	MBit/sec	0.025
Proxy efficiency (HIT [kB/sec] / DIRECT [kB/sec]):	factor	5.12
Average speed increase:	%	30.42
UDP response time of 0%% requests:	msec	56
TCP response time of 100%% requests:	msec	15146
Cache statistics		
Total amount cached:	requests	617147
Request hit rate:	%	32.07
Bandwidth savings:	Byte	3326M
Bandwidth savings in Percent (Byte hit rate):	%	30.02
Average cached object size:	Byte	2957
Average direct object size:	Byte	5671
Average object size:	Byte	4220

Fig. 4 : Cache Manager Data report when using SCSI and blocking media streaming

2. ACKNOWLEDGMENTS

3. REFERENCES
[1] Squid proxy server official website, http://www.squid-cache.org/Doc/

[2] Saini, "Squid Proxy Server Beginners Guide" Packt Publishing, Open Source Community, 2011

[3] F. Schmidt, The SCSI Bus & IDE Interface, Protocols, Applications & Programming. Addison- Wesley, 1998

[4] M. McKusick, W. Joy, S. Leffler, R. Fabry, "A Fast File System for Unix", ACM Trans. On Comp. Sys. 2(3), pp. 181-197, Aug. 1984.

[5] Brian White, Wee Teck Ng, Bruce K. Hillyer "Performance Comparison of IDE and SCSI Disks" http://citeseerx.ist.psu.edu.

[6] Jia Wang, "A Survey of Web Caching Schemes for the Internet," ACM Computer Communication Review, 29(5):36--46, October 1999.

[7] The Automatization of Information Security of Local Calculation Network with the Open Access to Internet on FreeBSD Base" IEEE Paper CADSM'2009, 24-28 February, 2009, by Writer "Andrian Piskozub, Dmytro Levytskyy, Igor Rudyk, Larisa Rakobovchuk".

[8] "Data Stream Splicing for Web Proxy Cache Optimization" IEEE 2008 Edition Paper, Proceedings of the Japan-China Joint Workshop on Frontier of Computer Science and Technology (FCST'06)by writer "Jingli Zhou, Jifeng Yu, Hongtao Xia".