Design and Analysis of Enhanced HTTP Proxy Cashing Server

Vimal Bibhu, Narendra Kumar, Mohammad Islam, Shashank Bhardwaj

Department of Computer Science & Engineering,
Galgotia College of Engineering & Technology, Plot -1, Knowledge Park – II, Greater Noida, Uttar Pardesh, India,
vimalbibhu@gmail.com
nkteotia2004@gmail.com
islam_cs1@yahoo.com

Abstract—— **A proxy acts as an intermediary agent between its clients and the servers which they want to access, performing functions directed towards a variety of purposes, like security, caching, etc, in its capacity as an intermediary. A server is anything that has some resource that can be shared. A server process is said to "listen" to a port until a client connects to it. A "port" is a numbered socket on a particular machine. Proxy servers are commonly used to create an access point to the Internet that can be shared by all users. In this paper we present a new method of caching HTTP Proxy servers which takes lower bandwidth by maintaining a cache of Internet objects like html files, image files, etc. which are obtained via HTTP in local machine. The size of cache is dynamic and automatically determined time to time. When the local system memory is not enough to store the internet objects then paging system is implemented automatically without interference of users. Once an internet objects comes into local machine then no external objects is needed.**

Keywords— **Hypertext Transfer Protocol, Proxy Server, Redundant Array of Inexpensive Disk, Transmission Control Protocol, Internet Service Provider, Domain Name System, File Transfer Protocol, Uniform Resource Locator.**

I. INTRODUCTION

In computer networks, a **proxy server** is a server (a computer system or an application program) which services the requests of its clients by forwarding requests to other servers. A client connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource, available from a different server [1]. The proxy server provides the resource by connecting to the specified server and requesting the service on behalf of the client. A proxy server may optionally alter the client's request or the server's response, and sometimes it may serve the request without contacting the specified server. In this case, it would 'cache' the first request to the remote server, so it could save the information for later, and make everything as fast as possible.

A proxy server that passes all requests and replies unmodified is usually called a gateway or sometimes tunneling proxy. A proxy server can be placed in the user's local computer or at various points between the user and the destination servers or the Internet [2].

A proxy server speaks the client side of a protocol to another server. It is a system or device that operates between computer application, such as a Web browser and a server. When users wish to read information from the Internet rather than requesting data directly from the object, they communicate with the proxy server that fills the request either from its cache or from the object itself. No direct communication is established between the system requesting the data and the Internet. The great thing that proxy servers provide, when configured correctly, is complete security [3].

HTTP (Hyper Text Transfer Protocol) is the protocol that web browsers and servers use to transfer hypertext pages and images. Here's how it works: When a client requests a file from an HTTP server, it simply prints the name of the file in a special format to a predefined port and reads back the contents of the file. The server also responds with a status code number to tell the client whether the request can be fulfilled and why.

A caching proxy server accelerates service requests by retrieving content saved from a previous request made by the same client or even other clients. Caching proxies keep local copies of frequently requested resources, allowing large organizations to significantly reduce their upstream bandwidth usage and cost, while significantly increasing performance. Most ISPs and large businesses have a caching proxy. These machines are built to deliver superb file system performance (often with RAID and journaling) and also contain hot-rodded versions of TCP [4]. Caching proxies were the first kind of proxy server. The HTTP 1.0 and later protocols contain many types of headers for declaring static (cacheable) content and verifying content freshness with an original server, e.g. ETAG (validation tags), If-Modified-Since (date-based validation), Expiry (timeout-based invalidation), etc. Other protocols such as DNS support expiry only and contain no support for validation. Some poorly-implemented caching proxies have had downsides (e.g., an inability to use user authentication). Some problems are described in RFC 3143 (Known HTTP Proxy/Caching Problems) [5].

II. PROBLEM DEFINITION

The proxy server performance is measured by the factors like waiting time, filtering, handling the http request and caching. The main problem is response time, when load increases in local intranet the proxy server waiting time becomes higher. By the way some connections with http references are refused [6]. Filtering the request is another problem. The filtering of request first matches the its internal database for http references. Hence, in this way time is taken to establish a connection via proxy and local system to internet sites.

The above mentioned problems are very authentic problems and it

is required that to develop such a caching proxy server that will reduce the problem existing in the current proxy servers. The only performance constraint that lies within the software is the congestion in the network [7]. This is a very common problem, which can arise in the network at any time. If the network is congested, and more users login to the network simultaneously, it will further to the problem.

III. PROPOSED PROXY SERVER APPROACH

The proposed new approach of caching http proxy server based upon two major factors which reduce the performance of the current proxies. The architectural diagram of basic working proxy server is shown in figure 1.
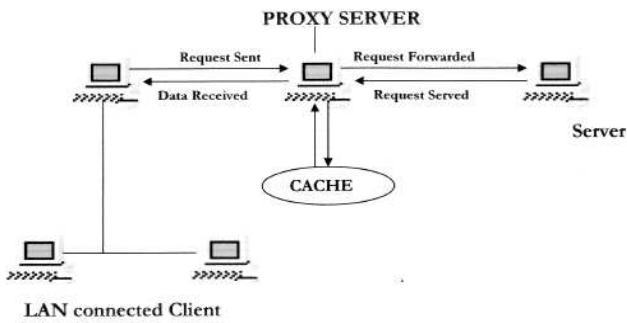

Fig. 1 Architectural Diagram of working HTTP Proxy Server.

Above diagram shows the basic functionality of current proxy server. The above proxy does not scan the http requests into the cache, only the current page is taken under the cache. Due to this lack of design the current proxy is not so efficient to function according to need [8]. The proposed architectural diagram of proxy is given in figure 2.
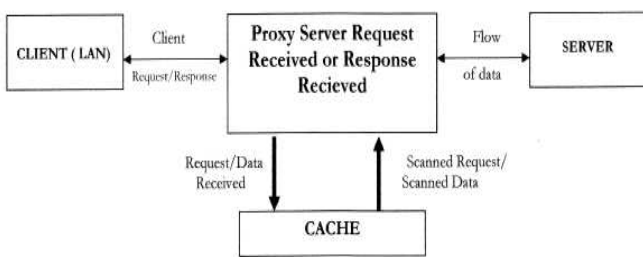

Fig. 2 Architectural Diagram of proposed HTTP Proxy Server.

The working and internal configuration of the new efficient proxy server is designed with help of given below in figure 3.
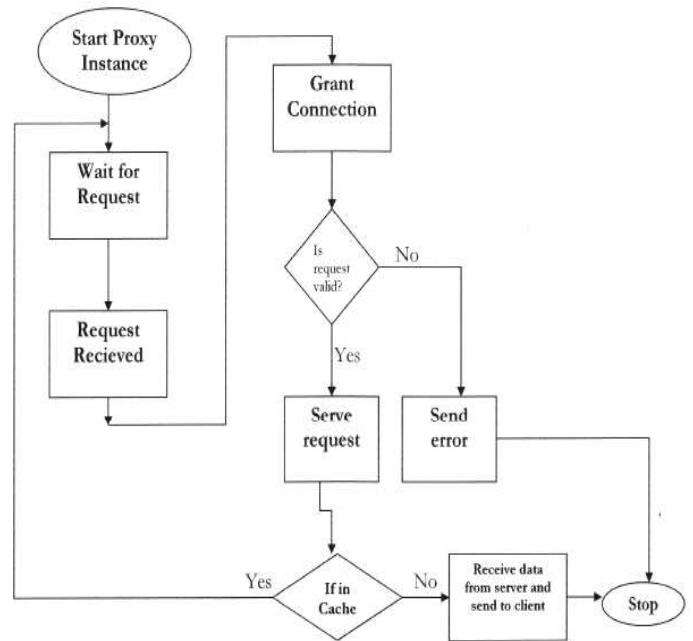

Fig. 3 Working and internal configuration of new efficient proxy server.

The above flow mechanism of working version of proposed proxy server based upon the request and the grant. This means that is request by client is being granted then caching the information with local system is being started automatically [9]. The internal caching mechanism design in figure 4.
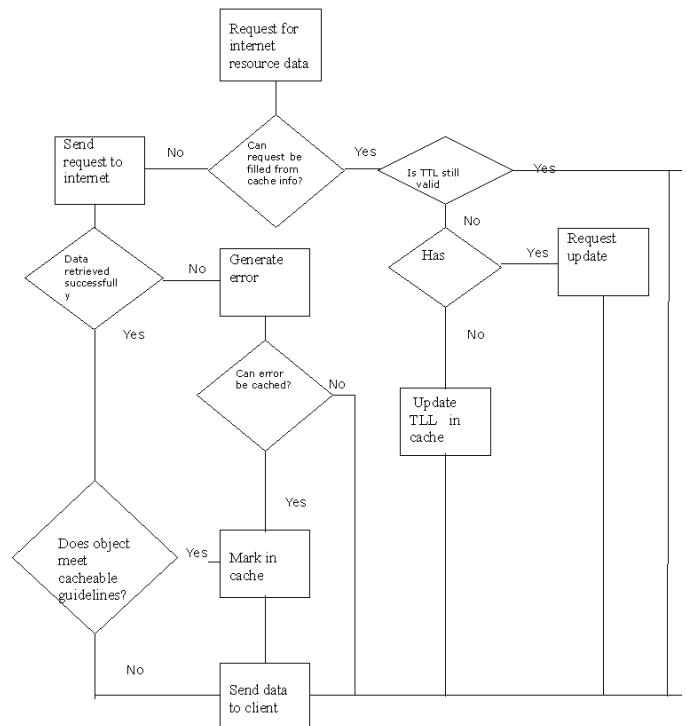

Fig. 4 Internal cashing mechanism design flow.

111

Without the user authentication unauthorised use restriction is not possible [10]. In this proxy server the new trend of authentication which very fast in nature. The flow mechanism of this fast authentication design is presented in figure 5.
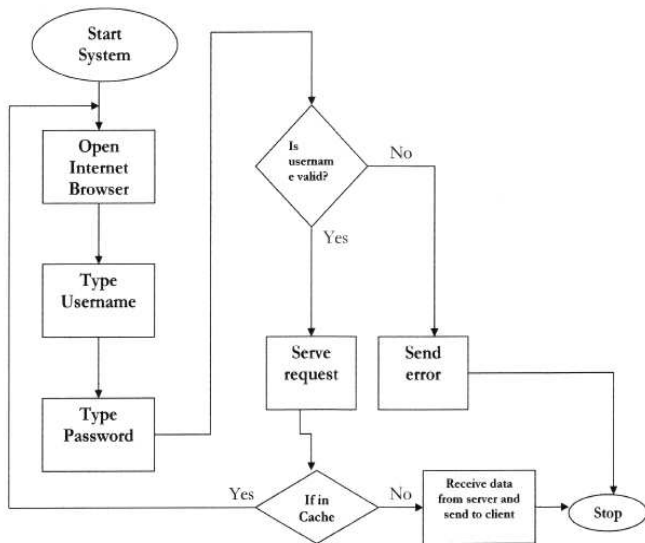

Fig. 4 Mechanism of fast authentication.

The fast authentication is taken due to search in cache. The above all designs are the part of the our proposed proxy server.

IV. ADVANTAGE OF THE PROPOSED VERSION OF PROXY SERVER

The new proposed version of proxy has the following advantage than the previous existing proxy servers.

1) Handling HTTP requests: The proxy server handles multiple HTTP requests from the clients. Concurrency issues are also handled in this process.

2) Cashing: it is one of the few mechanisms that are preventing the Internet from overloading. By caching frequently accessed sites, Information, and errors, proxy server significantly reduces total required bandwidth, which gives the appearance of a faster response time and save employees time and connectivity Expenses.

   Caching is not merely a copy of everything requested from the Internet. In order for an Internet resource to be cached, it must abide by the following Guidelines:

   Access to the Internet resource must be established via FTP or HTTP. Access to the Intemet resource must be via the get request.
   The URL line cannot contain any "? Keywords" as in Internet searches.

.

The expires HTTP header field must contain a later date than the date in the Date header field. (It would be ineffective to cache old information)

The HTTP result code must be 200 (success), 403 (Forbidden request), or 404 (URL not found).

3) Cookies: A cookie is a commonly used method for either delivering Information from a custom web page or authorizing or tracking a connection in a way that is insecure.

V. SECURITY RELATED FEATURES

The security is very important aspect in shared enviournment of working. In this proxy server the security related aspect is considered as follows.

1) Access Authority: Proxy server allows controlling access of inbound and outbound connections. Access authority may be used to limit a user's ability to access certain Internet sites.

   Outbound connections control a user's ability to use certain functions on the Internet. For instance, if a User wants to run FTP, the proxy server must grant them access to the protocol. If no access is granted, the user will not be able to transfer files with FTP.

   Inbound connections are limited based on the configuration of the proxy server. For instance, if a Company does not offer any Web-based services or Pages, there is no inbound traffic.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] Kroeger T.., Long D., Mogul J.: "Exploiting the bounds of Web latency reduction from caching and prefetching", in Proceedings od USENIX Symposium on Internet Technologies and Systems, pp. 13-22, 1997.

[2] David A. Malts and Pravin Bhagwat. Improving HTTP caching proxy performance with TCP tap. Technical report, IBM, March 1998.

[3] Ramsay J., Barbesi J., Preece J.: "Psychological Investigation of Long Retrieval Times an world wide web" Interacting with computers, 10, 1, (1998).

[4] A. Cohen, S. Rangaraan, and H. Slye. On the Performance of TCP Splicing for URL- Aware Redirection. In Proceedings of the 2nd Usenix Symposium on Internet Technologies and Systems, Boulder, CO, Oct. 1999.

[5] Feldman A,, Caceres R., Douglis F., Glass G., Rabinovich M.: "Performance of Web proxy caching in heterogeneous bandwidth environments", in Proceedings of INFOCOM, pp 107-116, 1999

[6] Jiantao Kong, Daniela Rosu and Marcel Rosu. Towards Enabling Web Proxy Control of TCP Splice Transfer Rates. 3rd New York Metro Area Networking Workshop, 2002.

[7] Sieminski A.: "The Potentials of Client Oriented Prefetching",, in Intelligent Technologies for Inconsistent Knowledge Processing, Advanced Knowledge International, 2004 Dikaiakos M.: "Intermediary Infrastructures for the WWW", url: "citeseer.ist.psu.edu/dikaiakos02intermediary.html"

[8] Gourley D., Totty B.: "HTTP: the Definite Guide", O'Reilly & Associates, 2002.

[9] Jung J., Krishnamurthy B., Rabiniwich M.: "Flash crowds and denial of service attacks Characterization and implications for CDN's and web sites", Proceedings of the International World Wide Web Conference, pp. 252--262, IEEE May 2002.

[10] Sieminski A.: "Buforowalnosc stron WWW", Multimedialne i sieciowe systemy informacyjne, MISSI 2004.