

Name: Rabin Ranabhat
Program 4
CSCI 5273

README:

RESULTS:

pages_from_server - PASS
pages_from_proxy - PASS
pages_from_server_after_cache - PASS
prefetching - Didn't do
multithreading - PASS
connection keepalive – Didn't do

Video link: <https://youtu.be/iuq6F-7FCi0>

WebProxy:

For the webproxy, the proxy server will receive requests from the client. The requests are analyzed to determine the following:

- The request type is GET.
- The requested URL is valid.
- The HTTP message is properly formatted and valid versions are being sent.

If it is not a GET request, the server responds with the 'ERROR: 501 Not Implement' response. If the URL is not valid or the HTTP version is not valid, then the server responds with the 'ERROR: 400 Bad Request' response.

After all these checks pass, the proxy server will then try to determine if the request is in the cache. If it is in the cache, the other thing that is determined is, if the existing cache element is within the timeout limit. If the cache has not timed out, that data is sent to the client. However, if it is determined that the element has timed out, the cached data is replaced with fresh data that is obtained by making request to the server.

While making the request to the server, if all the above tests have passed, then the client request received is forward to the server without modification. While webproxy is making request to the server, the connection to the client is maintained. On receiving the request from the server, the response is sent to the client. At the end of receiving the request from the server, the webproxy closes it connection with the server and also the client.

Workings of the cache:

Based on the url request that is sent by the client, a md5sum is calculated. This md5sum is used as a key to determine whether or not the element being searched is in the cache. Once md5sum hash string has be calculated and determined that the cache element is not present in the cache, while receiving the data from the server, the web proxy writes to the file named the same as the md5sum hash string. Therefore, the response from the server is saved as a file in the same directory as the executable. The cache files have a preceding "." as the first character for the file name. Therefore, the cached files will be hidden when opening the folder.

Cache functions:

Name: Rabin Ranabhat
Program 4
CSCI 5273

1. Determine if the element is in the cache:

Based on the requested URL, md5sum hash code string is calculated. This string is used to determine if the file is present. On doing the file open function, if it is determined that the file is not present, determination is made that the file is not present.

2. Time out in cache:

The file modified time for the cache file is checked with the current time to determine if the cache file has timed out. If the difference between current time and the last modified time of the file is greater than the cache time out value given during the start of the program, the cache element is ignored and a request to the server is made.

Concurrency:

Webproxy server forks a child to handle concurrent requests that are made to the server. Concurrent requests are handled by the server as shown in the video submitted.