Assignment 7
19.11.2014

HTTP Proxy

Purpose:

In this assignment, we will develop a simple HTTP proxy server. The proxy server will have an initial filter to block content based on a blacklist.

The blacklist will consist of a file with contains a list of wildcards to look for (eg: *ads.** should match *ads.doubleclick.com*). Note that the proxy should be designed in a way that you can extend it with additional filters later (for future assignments).

Layout:

Workstation (browser) ☐ Proxy Server (your program) ☐ HTTP server

Notes:

A very basic HTTP proxy server will act (almost) as a transparent gateway. Only the GET requests need to be parsed in order to identify the target HOST, e.g:

GET http://www.google.com/about.html HTTP/1.1 ☐

1. Connect to www.google.com (port 80, http)
2. Send *GET /about.html HTTP/1.1*

All received text after the GET should be forwarded to the actual host.

When a blocked URL is detected, the appropriate response can be a 403. This is a minimal reply that will be properly parsed by browsers:

HTTP/1.1 403 Forbidden[CRLF]
Content-Type: text/plain; charset=UTF-8[CRLF]
[CRLF]
Content blocked by proxy[CRLF]

Note that *[CRLF]* refers to the way new-lines must be send in HTTP commands, in most computer languages it is expressed as *\r\n (carriage return, line feed).*

Blocking based on URL should match only the hostname part of it (eg: *ads.** shouldn't block *www.google.com/ads.ring/about.html*).

It is recommended to use threads to proxy the replies from servers. You can use text streams to analyze the responses, but in doing so some data might be lost (eg: when transferring images or other non-textual data).

References:

HTTP made easy tutorial: http://www.jmarshall.com/easy/http/#proxies
RFC 2616 (HTTP): http://www.w3.org/Protocols/rfc2616/rfc2616.html
(Section 6, Responses):
http://www.w3.org/Protocols/rfc2616/rfc2616-sec6.html#sec6

Hand-In:

A demonstration of the software. The source files and project files must be submitted to the appropriate Ilias assignment.

Deadline:

One week (26.11.2014)