

# Comp 112 Assignment 5: HTTP Proxy

Jim Mao and Max Cohen

April 8, 2015

## 1 Running the Proxy

Run the proxy with the following command:

```
python proxy.py [portnumber]
```

The default port number is 8080 if no port is supplied.

## 2 Content Filtering

One important functionality of proxies is filtering content that is potentially harmful. When implementing connections from client to server, the proxy will create a unique thread for each connection and its associated client address. We plan on implementing content filtering such that when a client requests a certain URL, the proxy will not send the GET request to the server if the URL is in the list of blocked URLs, specified by the proxy administrator. Upon receiving a GET request from the client, the proxy will check its list of blocked URLs. If the requested URL happens to be in this list, the proxy will never attempt to get data from the specified server. Instead it will end the communication with the client by making the associated connection's thread exit.

## 3 Caching Policy

Caching benefits a client-server relationship by reducing wait times experienced by the client and reducing traffic flow to the server, thereby speeding up server response times. Our caching policy will create a new entry in the cache each time a new URL is requested by some client. The proxy will retrieve the data from the server and store it in the cache. The next time this URL is requested, the proxy will first check the cache to see if the requested data is stored in the cache. If it is, the data will be retrieved from the cache and forwarded to the client. We will implement the cache by maintaining a cache directory in which we will store data from requested URLs in files with the exact name of the URL. In order to maintain the accuracy of the data, we will store a timestamp of when the data was retrieved from the server. In our implementation, we will store this timestamp as the first line of the file corresponding to the requested URL. When the proxy receives a request from a client, it will look through the cache directory and attempt to find a file with the exact name of the URL requested. If the URL matches an entry and the timestamp is not expired (we will parse the first line of the file in order to determine the timestamp), the proxy will not send a GET request to the server. It will instead retrieve the data from the file and send it along to the client. If the timestamp is expired, the proxy will get the data from the server and update the cache.