Web Proxy Server

Team Members:
Royal Jain
Pratyaksh Sharma
Paramdeep Singh
Shyam JVS

What is a Proxy Server ?

- A computer system or application that acts as an intermediary for requests from clients seeking resources from other servers
- Web Proxy: A proxy server which intermediates HTTP requests

Why need Proxy Server ?

1. Monitoring & Filtering
   - Content control
   - Bypassing filters/censorship
   - Logging and eavesdropping
2. Improving Performance
   - Caching
   - HTTP Compression
3. Accessing Services Anonymously
   - Access Control
   - Geotargeted advertising
4. Security
   - NAT for intranet anonymity
   - Firewalls
5. And many more….

Features of our Proxy

1. HTTP request types supported:
   - GET
   - POST
   - CONNECT

Multithreading enabled and hence each request spawns it’s own thread

Features (…continued)

2. Support for both HTTP & HTTPS

- For HTTP, connection is over a normal TCP channel
- For HTTPS, connection is over a secure TCP channel
- Both are initially established using HTTP CONNECT exchange between client and proxy

Features (…continued)

3. Caching of Web pages
   - For improving performance and throughput of requests served
   - Web content hash-mapped with key as the URL
   - Replacement policy for pages based on LFU

Features (…continued)

4. Content-Filtering
   - Filtering adult content (webpage requests) based on Hypertext within the response payload
   - Naive Bayes Algorithm used for training the classifier
   - Caching URLs of filtered requests for saving redundant effort

Features (…continued)

5. Firewalling
   - Blocking requests to certain IP addresses or hosts
   - Incoming request’s destination host compared against list of blocked IPs
   - Based on IP , as against text content in content-filtering

Features (…continued)

6. Request Prioritization
   - Proxy server orders threads for various requests based on clients' priority
   - Done dynamically while reducing window of pending requests
   - Priority however has been assigned statically

Features (…continued)

7. HTTP Compression
   - Compressing the payload of HTTP responses where possible
   - Using gzip for this
   - Done only when content length is above a minimum threshold
   - Most browsers can handle gzipped data

Features (…continued)

8. Proxy Authentication
   - Enabled for access control of users within the proxy's client domain
   - Client requests are served only after valid credentials have been entered, else HTTP 407 is returned
   - Active IP entries maintained in a list along with timestamps, for avoiding redundant authorizations from same client

Features (…continued)

9. Ad-Blocker
   - Provides support for blocking adware
   - Saves upon the proxy’s bandwidth
   - Done based on matching URL requests with a list of regExps that cover most common advertisements

Features (…continued)

10. GUI Based Proxy Configuration
    - Provide a manual configuration page to the user
    - Can block specific keywords, URL
    - Can specify a whitelist

Features (…continued)

11. Logging

- All the requests are continuously logged along with timestamps
- Kept for statistics and to maintain records for possible data analysis later on

Demo

Contributions: Pratyaksh

- LRU/LFU Caching
- Adblock
- Blacklist/Whitelist filtering
- User Interface (Config pages)

Contributions: Shyam

- HTTP Compression (Gzip)
- Logging
- Multithreaded operation
- Handled HTTP GET

Contributions: Royal

- Handled HTTPS tunneling
- Content Filtering (naive bayes classifier)

Contributions: Paramdeep

- IP blocking
- User authentication
- Prioritization
- Queueing

References

- Simple proxy code to start, from O’Reily’s Java book
- Code for Naive Bayes classifier, from MLlib

THANK YOU !!