TNE30019/TNE80014 – Unix for Telecommunications

Apache – Forwarding Web Sites & Proxy Servers

Dr. Jason But

Swinburne University

Outline

- Forwarding Web Sites from within Apache
  - Redirection
  - Transparent Forwarding
- Caching Proxy Servers
  - Why use them
  - Named vs. Transparent Proxies
  - Unix Proxy Servers

Motivation for Forwarding

Why would we do this?

- Use URL on our web site to map directly to other web site
- Internal renaming of URLs while keeping site available under old URLs
- Internal renaming or moving of content
- Present short URLs to user that are internally expanded

Forwarding Web Sites – Redirecting

Option 1: `Redirect` directive

- Redirects browser to different web site
- Specify path for web server and full URL of remote server

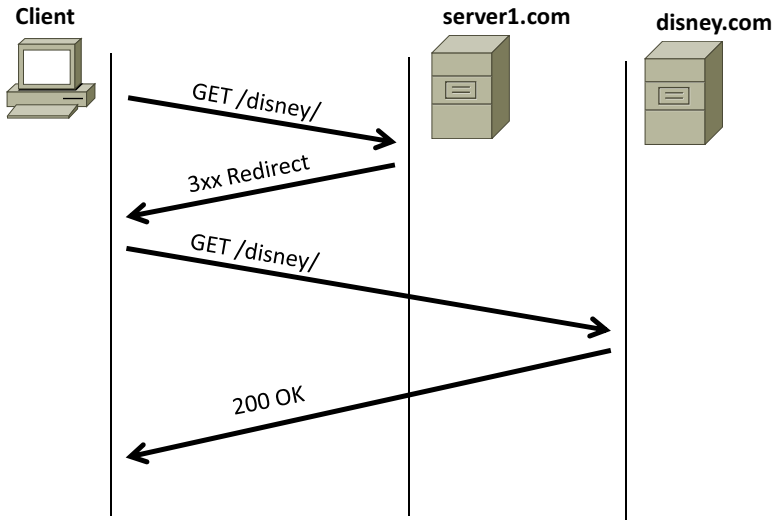
Example

- Configuration

```
<VirtualHost *:80>
    ServerName www.server1.com
    DocumentRoot /home/www/server1
    Redirect /disney/ http://www.disney.com
</VirtualHost>
```

- Apache server hosts **www.server1.com** website
- Files in `/home/www/server1` contain web site content
- Browser requests for **http://www.server1.com/disney/** will be redirected to **www.disney.com**

Forwarding Web Sites – Redirecting



Forwarding Web Sites – Proxying

Option 2: `ProxyPass` and `ProxyPassReverse` directives

- Instead of redirecting client Apache now acts as proxy
- Forwarded web site can be hidden behind a firewall
- Can do load balancing between multiple servers hosting forwarded site
- `ProxyPassReverse`: Apache replaces name of actual web server with its own name

Forwarding Web Sites – Proxying

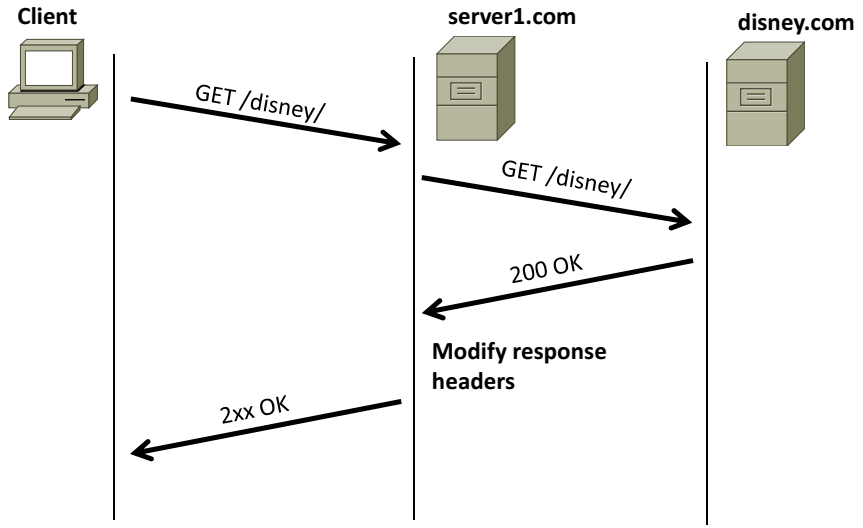
Example

- Configuration

```
<VirtualHost *>
    ServerName www.server1.com
    DocumentRoot /home/www/server1
    ProxyPass /disney http://www.disney.com
    ProxyPassReverse /disney http://www.disney.com
</VirtualHost>
```

- Apache server hosts **www.server1.com** website
- Files in `/home/www/server1` maps to web site content
- Browser requests for **http://www.server1.com/disney** will automatically show contents of **www.disney.com** even if this server is behind firewall
- User will see **http://www.server1.com/disney** in their browser window

Forwarding Web Sites – Proxying



Forwarding Web Sites

Option 3: `RewriteRule` and `RewriteCond` directives

- Very powerful framework to rewrite URLs on the fly
- `RewriteCond` specifies which URLs are to be rewritten and `RewriteRule` specifies the rewrite rule

Example: Moved document root

- Redirect URL / to /www/

```
RewriteEngine on
RewriteRule    ^/$  /www/  [R]
```

Example: Move home directories to other server

- Redirect /~user/anypath to http://server/~user/anypath

```
RewriteEngine on
RewriteRule    ^/~(.+)  http://server/~$1  [R,L]
```


Caching Web Proxies

Improve performance of Internet by caching content

- Retrieving content from local proxy at high speed
- Requesting content from server on other side of world is slow (TCP transfer ramps up slowly)
- Data transfer across congested/slow link

Provide access through firewall

- Many sites have firewall to protect network from outside
- Proxy is allowed to access Internet
- Proxy can be configured to protect against malicious attacks
- User computers are protected by virtue of proxy

Caching Web Proxies

Web access through caching proxy

- All HTTP requests go to **proxy** instead of actual server
- If requested content is cached, cached copy is returned
- Otherwise proxy downloads and forwards content
- Downloaded content cached by proxy for future requests

Caching

- Web server specifies content cache-ability and duration
- Proxy can run out of room to store content
- Content on remote servers can change
- Proxy must periodically flush content from cache
  - **HTTP HEAD** request indicates content changed
  - Cache period for content has expired
  - Remove Last Recently Used (LRU) if cache is full

Caching Web Proxies

Named Proxies

- Traditional implementation
- Browser must be configured with proxy information

Transparent Proxies

- Browser does not know about proxy
- Network configured to intercept and redirect HTTP requests via (transparent) proxy
- Proxy services HTTP requests (like named proxy)
- Proxy responds to browser as if it was chosen web server

Caching Web Proxies – Squid

Squid

- Popular free caching proxy server
- Available at `http://www.squid-cache.org`
- See online documentation at Squid website

FreeBSD install

```
cd /usr/ports/www/squid26
make
make install
```

Or

```
pkg_add -r squid26
```