Apache as a Reverse Proxy

Daniel López Ridruejo
daniel@bitrock.com
About me

- Open source: Author of mod_mono, Comanche, several Linux Howtos and the Teach Yourself Apache 2 book
- Company: founder of BitRock, multiplatform installers and management software
About this presentation

- Introduction to proxies
- Apache mod_proxy architecture and configuration
- Usage scenarios
  - Performance
  - Availability
  - Management
  - Security
- Common Issues
- Resources
What is a Proxy?

- HTTP Proxy server: Issues HTTP requests on behalf of clients
- Forward Proxy: Security, Monitoring, Performance
What is a Reverse Proxy?

Proxy is in front of the servers, transparent to the clients.
Apache Proxy Architecture (1)

- Implemented as a module
- Multiple backends: HTTP, FTP, CONNECT...
- Interacts with other Apache modules: SSL, URL mapping, Compression, Caching
- Apache 2.0: Filtering architecture
- Apache 2.1: Improved caching, new features (load balancing, AJP protocol)
Apache Proxy Architecture (2)
Related modules

- mod_cache
- mod_ssl
- mod_deflate
- mod_headers
- mod_charset_lite
- mod_injection
- mod_proxy_add_forward
- mod_ext_filter
- mod_rewrite
- mod_proxy_html
- mod_replace
- mod_ip_forward (1.3)
- mod_proxy_add_uri
- many others...
Compiling mod_proxy

- Use APXS or enable support at compile-time:
  --enable-proxy
  --enable-proxy-connect
  --enable-proxy-ftp
  --enable-proxy-http
  --enable-proxy-balancer (2.1)
  --enable-proxy-ajp (2.1)
Configuration

● Load appropriate modules
  LoadModule proxy_module modules/mod_proxy.so
  LoadModule proxy_http_module modules/proxy_http.so
  etc...

● Does *not* need ProxyRequests On
  ● Security/Spam implications
Configuration : Basic Scenario

ProxyPass /downloads http://downloads.example.com
(or inside <Location /downloads>)
   GET /download/file.exe HTTP/1.1
   Host: www.example.com

turns into
   GET  /file.exe HTTP/1.1
   Host: downloads.example.com

Problems : redirections, self-referential URLs, absolute paths. Need to rewrite URLs in HTTP responses

Use “!” not to proxy certain URLs
Hiding backend server

- ProxyPassReverse
  ProxyPassReverse /downloads http://downloads.backend.com
- ProxyErrorOverride
- mod_proxy_html, mod-replace, rewrite embedded URLs on the fly
- ProxyPassReverseCookiePath
- ProxyPassReverseCookieDomain
Client Information Loss

- Client Information: IP, hostname, port, SSL information...
- Server information

This is added/can be added to headers and passed to the backend server. Apache 1.3 requires third party modules to do this X-Forwarded-*, ProxyPreserveHost, etc.
New functionality in 2.1

- Load Balancer
  - BalancerMember
  - ProxySet
  - ProxyStatus
- Support for AJP protocol
  - Communicate with Tomcat or any other engine that supports the protocol
mod_rewrite

- Advanced URL manipulation, allows [P] option for reverse proxying
- Simple load balancing example:

  ```
  RewriteMap SERVERS rnd:/Library/Services/Apache/conf/servers.conf
  <Location /webapp>
  RewriteEngine On
  RewriteCond %{HTTP_COOKIE} "(^|;\s*)jsessionid=\w+.(\w+)(\$|;)"
  RewriteRule "(.*)" "http://${SERVERS:%2}%{REQUEST_URI}" [P,L]
  RewriteRule "^.*;jsessionid=\w+.(\w+)(\$|;)" "http://${SERVERS:$1}%{REQUEST_URI}" [P,L]
  RewriteRule "(.*)" "http://${SERVERS:ALL}%{REQUEST_URI}" [P,L]
  </Location>
  ```

(Full details at http://wiki.apache.org/cocoon/LoadBalancingWithModProxy)
Scenarios

- **Performance**
  - Caching
  - SSL
  - Slow clients

- **Management**
  - Unified URL space
  - Logging
  - Troubleshooting

- **Security**
  - Interception
  - Single sign-on
  - Isolation

- **Availability**
  - Load balancing

- **Fancy stuff**
Performance / Scalability

- SSL processing
- Dealing with slow clients, lingering_close
- Can be more easily optimized/tuned
- Caching
- Serve static content locally, proxy rest
Management

- Unified URL Space
- Unified Logging
- Unified User Tracking
- Standardized front end
Security

- Single sign-on
- Isolation
- Interception: HTTP-level firewall
  - Protect unpatched backend servers
Availability

- Apache 2.1 includes mod_proxy_balancer
- Smart load-balancers thru custom modules, mod_rewrite, etc.
Fancy stuff

- Reverse Proxy + filtering architecture, lots of interesting possibilities
- Check/transform content on the fly
  - Server Side Includes
  - XSLT transformations
  - Fixing broken apps/clients
  - Antivirus
Online resources

- Apache Docs
  - http://httpd.apache.org

- mod_proxy_html
  - http://apache.webthing.com
  - http://www.apacheweek.com/features/reverseproxies

- mod_rewrite guide (dated but relevant)
  - http://httpd.apache.org/docs/misc/rewriteguide.html

- mod_replace
  - http://sourceforge.net/projects/mod-replace/

- http://modules.apache.org

Related sessions

- TU10: mod_proxy multi protocol framework for httpd-2.0

Also useful

- WE16: HTTP Caching and Cache-busting for Content Publishers
- TU18, TU22 : mod_rewrite
Q&A

Thanks for your time!

More info:
http://www.apacheworld.org/apachecon/

You can reach me at:
daniel@bitrock.com