Apache Web Server Architecture

Chaitanya Kulkarni
chaitanya.kulkarni1@marist.edu

MSCS 521 23rd April 2008

What is “the web”?

- A collection of cross-linked “websites”
- A set of “web” technologies
- The consistent use of URIs to represent resources
- HTTP, HTML, and everything built around them
- Every person's cheap’n’easy client-server computing

What is a Server?

- A server is a computer or device on a network that manages network resources. Most servers are dedicated. This means that they perform only one task rather than multiple tasks. On multiprocessing operating systems, however, a single computer can execute several programs at once. A server in this case could refer to the program that is managing resources rather than the entire computer.

What is Web Server ?

- A computer program that is responsible for accepting HTTP requests from clients, which are known as web browsers, and serving them HTTP responses along with optional data contents, which usually are web pages such as HTML documents and linked objects (images, etc.).

A brief history of Apache (1)

- First viable alternative to Netscape Communications Corporation web server
- It is an open-source software with cross-platform functionality.
- First version was created by Robert McCool

A brief history of Apache (2)

- Named Apache because:-
  - Out of respect for the Native American Indian tribe of Apache
  - due to the project's roots as a set of patches to the codebase of NCSA HTTPd 1.3 - making it "a patchy" server.
- Chief competitors include Microsoft’s IIS and Sun Java Systems Web Server

A brief history of Apache (3)

- Apache is used by giants like:-
  - Google Web Servers
  - Wikimedia
- The Apache License allows for the distribution of both open and closed source derivations of the source code. The name *Apache* is a registered trademark.

Apache Server Architecture(1)

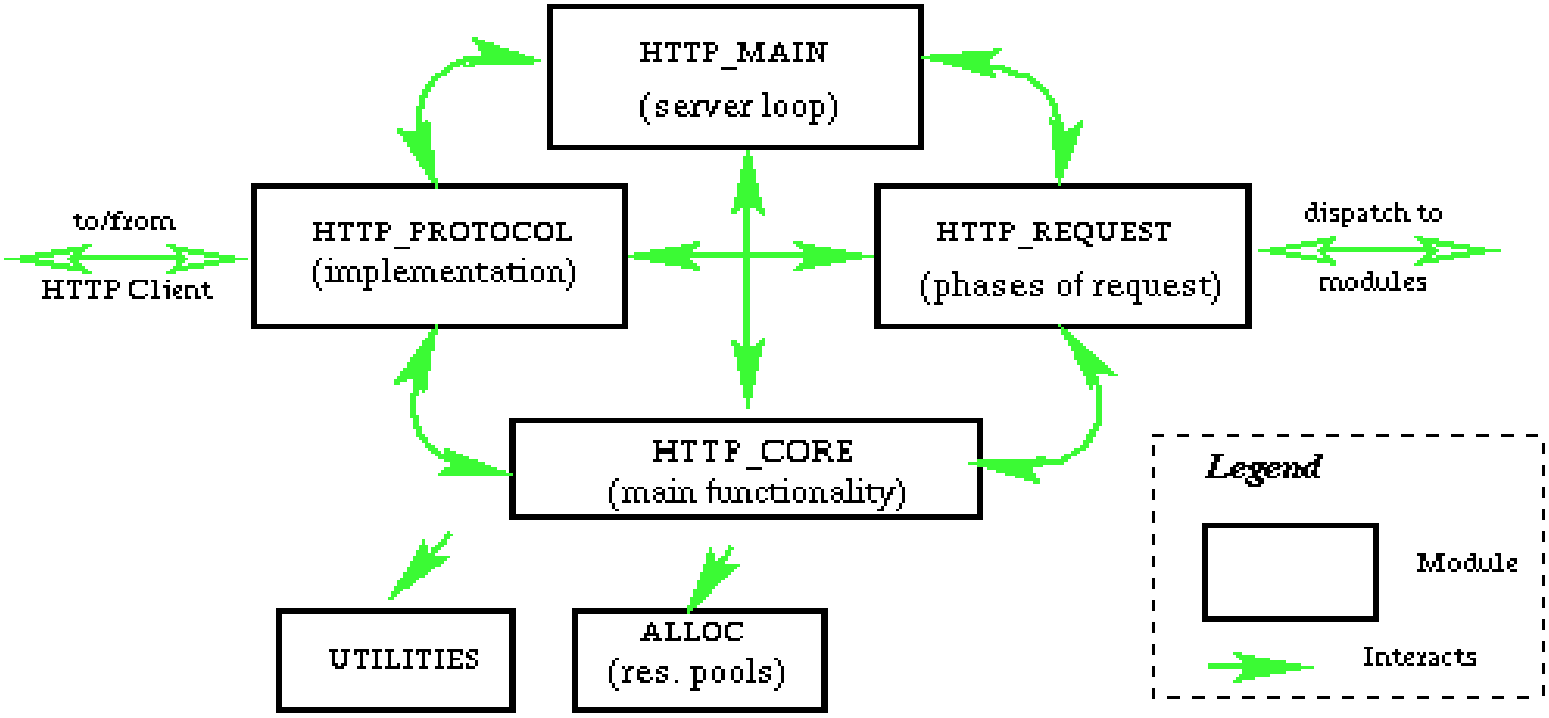
- Apache supports a variety of features, many implemented as compiled modules which extend the core functionality.
- These can range from server-side programming language support to authentication schemes.

Apache Server Architecture(2)

- Popular compression methods on Apache include the external extension module, mod_gzip, implemented to help with reduction of the size (weight) of web pages served over HTTP.
- The core of a Apache Server implements the basic functionality of the server. In addition it implements a number of utility functions

Core Component (1)

- Following are the components of the Apache core:
  - http_protocol.c
  - http_main.c
  - http_request.c
  - http_core.c



Picture Depicting the Core Component

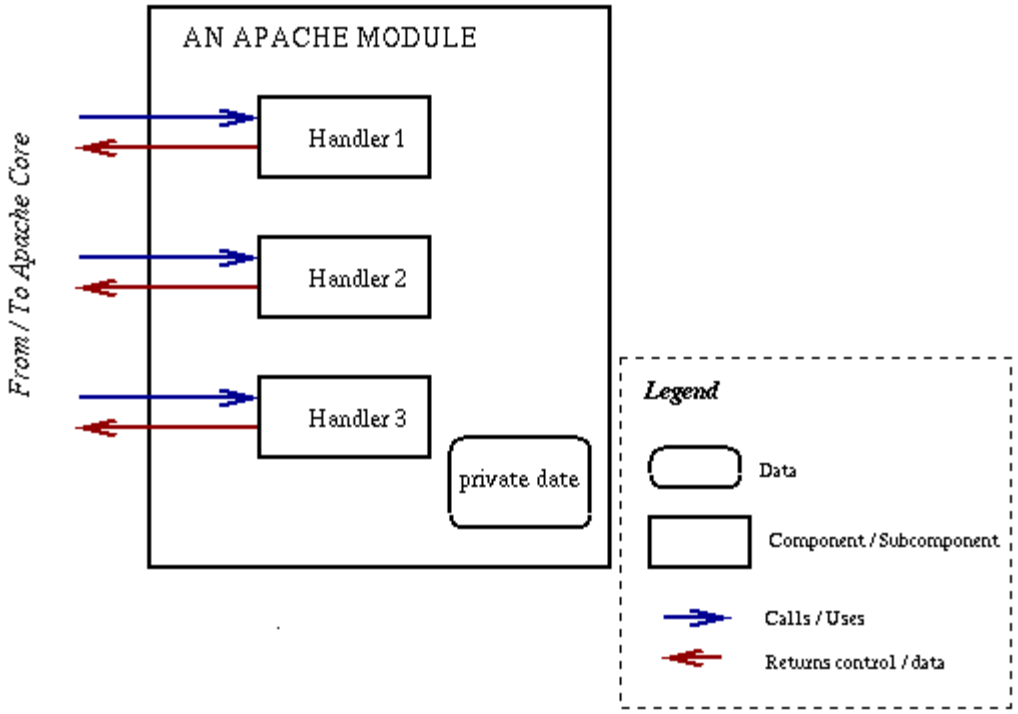
How Requests are Handled? (1)

- URI to filename translation.
- Check access based on host address, and other available information
- Get an user id from the HTTP request and validate it
- Authorize the user

How Requests are Handled?(2)

- Determine the MIME type of the requested object (the content type, the encoding and the language).
- Fix-ups (for example replace aliases by the actual path).
- Send the actual data back to the client.
- Log the request.

Architecture of an Apache Module



- Picture Depicting Apache Module Architecture.

The concept of a Handler

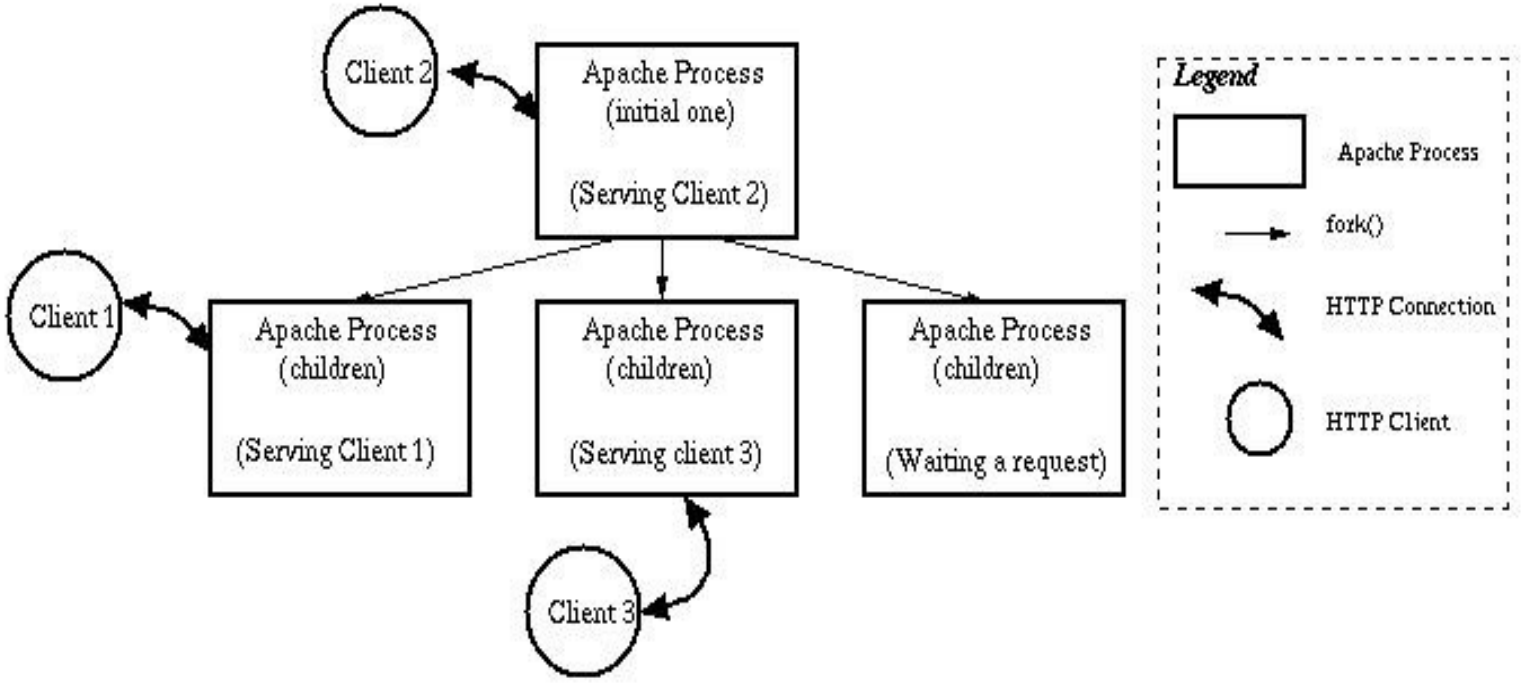
- A handler is for Apache the action that must be performed in some phase of servicing the request
- They are defined by modules. A module might specify handlers for one, many or none of the phases of the request

Persistent Server Processes

- 'Persistent Server Processes' is a concept that explains the implementation of Concurrency in Apache.
- To handle large incoming requests to website TCP/IP servers fork a new child to handle new incoming request from clients.
- However in the situation of a busy web site the overhead of forking a huge number of children will simply suffocate the machine.

Persistent Server Processes contd..

- To solve this problem Apache uses Persistent Server Process.
- It forks a fixed number of children, right from the beginning. The children service incoming requests independently (different address spaces).
- It is interesting that Apache server can dynamically control the number of children it forks (i.e. increasing or decreasing it), based on current load.



Picture depicting concurrency in Apache.

Configure Apache Web Server

- The configuration details are given at this site .
- We have to first configure the DNS. Then configure the following file /etc/httpd/conf/httpd.conf.
- The root directory of Web server is /etc/httpd, which is divided into three parts.

Configure Apache Web Server contd..

- The three parts are:
  - /etc/httpd/conf (where configuration files stays)
  - /etc/httpd/logs (where the logs of Web server and site accessing stay)
  - /etc/httpd/modules (where the module stays, which enables the server side programmer to do programming in the languages supported by Web server)

Apache Usage.

- Apache is primarily used to serve both static content and dynamic Web pages on the World Wide Web. Many web applications are designed expecting the environment and features that Apache provides.
- Apache is the web server component of the popular LAMP web server application stack, alongside MySQL, and the PHP/Perl/Python programming languages.

Apache License

- The Apache License is a free License authored by Apache Software Foundation or ASF.
- It allows use of the source code for the development of free and open-source software as well as free software.
- It is compatible only with the version 3 of GNU Public License or GPL.

Reference material

- Apache License

  http://en.wikipedia.org/wiki/Apache_License

- Apache Server Architecture

  http://www.cs.ucsb.edu/~tve/cs290i-sp01/papers/Concept_Apache_A

Concluding remarks

- Apache Web Server has a modular architecture. The core provides the basic functionality and separate set of supporting modules for handling HTTP requests.
- *Implicit Invocation* is the architectural style.
- Concurrency exists only between persistent identical processes that service incoming HTTP requests on the same port.
- Functionality is easily changed by writing new or editing existing modules.