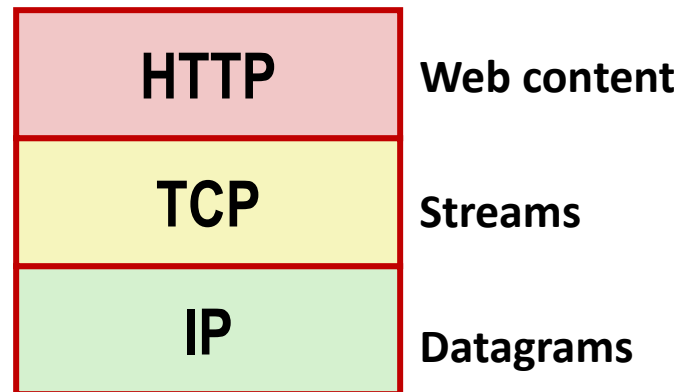
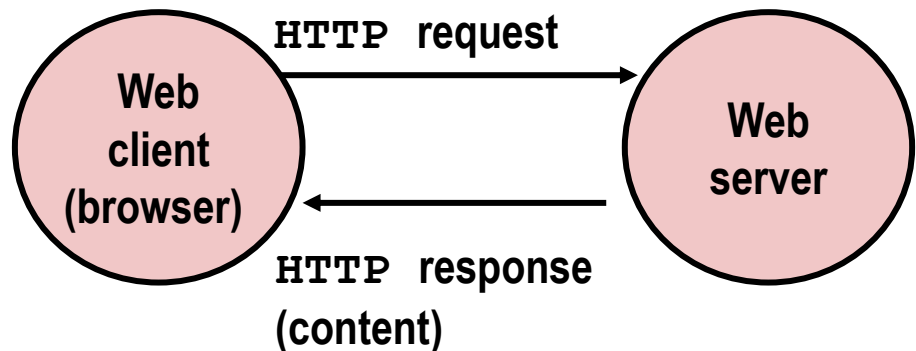


Network Programming:

Web servers

Web Server Basics

- **Clients and servers communicate using the HyperText Transfer Protocol (HTTP)**
 - Client and server establish TCP connection
 - Client requests content
 - Server responds with requested content
 - Client and server close connection (eventually)
- **Current version is HTTP/2**
 - RFC 7540, May 2015
 - Backward compatible with HTTP/1.1



<https://tools.ietf.org/html/rfc7540>

Web Content

■ Web servers return *content* to clients

- *content*: a sequence of bytes with an associated MIME (Multipurpose Internet Mail Extensions) type

■ Example MIME types

- | | |
|---------------------------|-------------------------------------|
| ■ <code>text/html</code> | HTML document |
| ■ <code>text/plain</code> | Unformatted text |
| ■ <code>image/gif</code> | Binary image encoded in GIF format |
| ■ <code>image/png</code> | Binary image encoded in PNG format |
| ■ <code>image/jpeg</code> | Binary image encoded in JPEG format |

■ Files are sent as-is after a set of *headers*

You can find the complete list of MIME types at:

<http://www.iana.org/assignments/media-types/media-types.xhtml>

Static and Dynamic Content

- The content returned in HTTP responses can be either *static* or *dynamic*
 - *Static content*: content stored in files and retrieved in response to an HTTP request
 - Examples: HTML files, images, audio clips
 - Request identifies which content file
 - *Dynamic content*: content produced on-the-fly in response to an HTTP request
 - Example: content produced by a program executed by the server on behalf of the client (e.g., PHP)
 - Request identifies file containing executable code
- Bottom line: ***Web content is associated with a file that is managed by the server***

URLs and how clients and servers use them

- Unique name for a file: URL (Universal Resource Locator)
- Example: `http://www.depaul.edu:80/index.html`
- Clients use *prefix* (`http://www.depaul.edu:80`) for:
 - What kind (protocol) of server to contact (HTTP)
 - Where the server is (`www.depaul.edu`)
 - What port it is listening on (80)
- Servers use *suffix* (`/index.html`) to:
 - Determine if request is for static or dynamic content.
 - No hard and fast rules for this
 - One convention: executables reside in `cgi-bin` directory
 - Find file on file system
 - Initial “/” in suffix denotes home directory for requested content.
 - Minimal suffix is “/”, which server expands to configured default filename (usually, `index.html`)

HTTP Requests

- HTTP request is a *request line*, followed by zero or more *request headers*
- Request line: `<method> <uri> <version>`
 - `<method>` is one of GET, POST (less commonly OPTIONS, HEAD, PUT, DELETE, or TRACE)
 - `<uri>` is typically URL for proxies, URL suffix for servers
 - A URL is a type of URI (Uniform Resource Identifier)
 - See <http://www.ietf.org/rfc/rfc2396.txt>
 - `<version>` is HTTP version of request (HTTP/1.1 or HTTP/2)
- Request headers: `<header name>: <header data>`
 - Provide additional information to the server
 - Importantly: the host name used by the client

HTTP Responses

- HTTP response is a *response line* followed by zero or more *response headers*, possibly followed by *content*, with blank line (“\r\n”) separating headers from content.
- Response line:
`<version> <status code> <status msg>`
 - <version> is HTTP version of the response
 - <status code> is numeric status
 - <status msg> is corresponding English text
 - 200 OK Request was handled without error
 - 301 Moved Provide alternate URL
 - 404 Not found Server couldn't find the file
- Response headers: `<header name>: <header data>`
 - Provide additional information about response
 - Content-Type: MIME type of content in response body
 - Content-Length: Length of content in response body

Example HTTP Transaction

```
$ telnet www.perdu.com 80
Trying 208.97.177.124...
Connected to www.perdu.com.
Escape character is '^]'.
```

```
GET / HTTP/1.1
Host: www.perdu.com
```

**Client
Request**

```
HTTP/1.1 200 OK
Date: Sun, 23 Aug 2020 18:18:07 GMT
Server: Apache
Last-Modified: Thu, 02 Jun 2016 06:01:08 GMT
Accept-Ranges: bytes
Content-Length: 204
Cache-Control: max-age=600
Expires: Sun, 23 Aug 2020 18:28:07 GMT
Vary: Accept-Encoding,User-Agent
Content-Type: text/html
```

**Server
answer**

```
<html><head><title>Vous Etes Perdu ?</title>[...]</html>
Connection closed by foreign host.
```

- HTTP standard requires that each text line end with “\r\n”
- Blank line (“\r\n”) terminates request and response headers

GET Request to Apache Server From Chromium Browser

URI is just the suffix, not the entire URL

```
GET /~user/test.html HTTP/1.1
Host: www.depaul.edu
User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/84.0.4147.135 Safari/537.36
Accept: text/html,application/xhtml+xml;q=0.9,*/*;q=0.8
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 115
Connection: keep-alive
CRLF (\r\n)
```

GET Response From Apache Server

```
HTTP/1.1 200 OK
Date: Fri, 29 Oct 2010 19:48:32 GMT
Server: Apache/2.2.14 (Unix) mod_ssl/2.2.14 OpenSSL/0.9.7m
mod_pubcookie/3.3.2b PHP/5.3.1
Accept-Ranges: bytes
Content-Length: 479
Keep-Alive: timeout=15, max=100
Connection: Keep-Alive
Content-Type: text/html
```

```
<html>
<head><title>Some Tests</title></head>

<body>
<h1>Some Tests</h1>
. . .
</body>
</html>
```