Web Technologies: The W3C Technology Stack

Unit 2 - Tutorial: The HTTP Protocol

http://www.unibw.de/ebusiness/

Which are the four layers of the Internet Protocol Suite?

Four layers

- Link Layer
- Internet Layer
- Transport Layer
- Application Layer

Source: RFC 1122

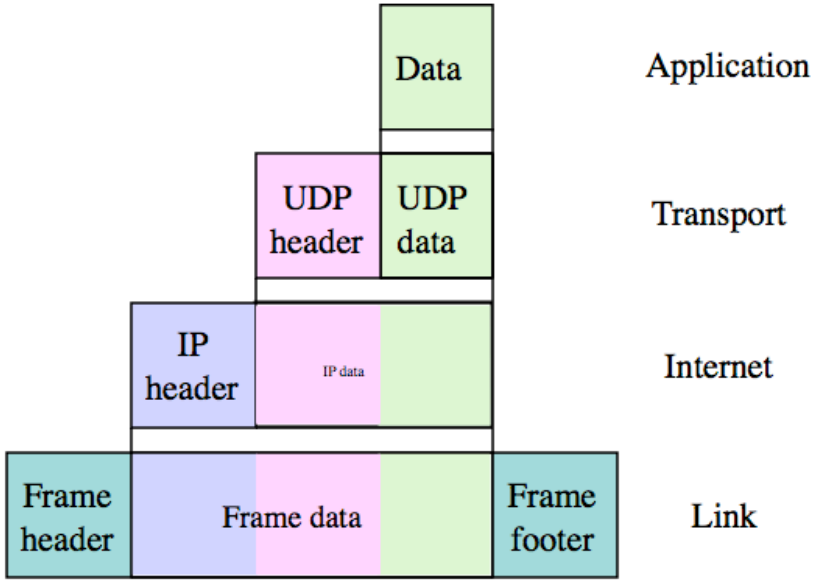

Illustration by Cburnett, available under GFDL from http://en.wikipedia.org/wiki/File:UDP_encapsulation.svg

Which of those four layers is addressed by the HTTP protocol?

- Application Layer
  - DNS
  - FTP
  - Gopher
  - **HTTP**
  - SMTP
  - Telnet

Explain the relationship between URIs and HTTP

- URIs: Unique and reliable identifiers
- HTTP: Application-layer protocol

```
http_URL =
"http:" "//" host [ ":" port ] [ abs_path
  [ "?" query ]]
```

- Relation
  - http://weather.example.com/
    - HTTP names a URI scheme
      - Scheme: specification that explains the scheme-specific details of how scheme identifiers are allocated and become associated with a resource

Explain the relationship between HTTP and HTML/XHTML.

- HTTP is the protocol to transfer web pages and HTML/XHTML is the markup language that defines the rendering of web content.

Can HTTP be used to transfer representations in formats other than HTML

- Yes, it can be used to retrieved images, video, audio → Anything specified as a Media type

What is the relationship between a URI, a resource, and a representation in HTTP terminology?

- **URIs**: Unique and reliable identifiers
- **Resource:** A network data object or service that can be identified by a URI.
- **Representation**: byte-stream manifestation of the resource at a particular URI
- See: http://www.w3.org/TR/webarch/#representation-reuse

Which are the two types of HTTP messages?

- **Request:** An HTTP request message
- **Response:** An HTTP response message

HTTP Messages: Request or Response

```
generic-message = start-line
                  *(message-header CRLF)
                  CRLF
                  [ message-body ]
start-line      = Request-Line | Status-Line
```

Explain the individual parts of the HTTP message on slide 20.

Request Header	Value
(Request-Line)	GET / HTTP/1.1
Host	feedregistrar.appspot.com
User-Agent	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.5; en-GB; rv:1.9.1.7)...
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language	en-gb,en;q=0.5
Accept-Encoding	gzip,deflate
Accept-Charset	ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive	300
Connection	keep-alive

Request line: Method and protocol version

Request header: Host and port
-user who originated the request
- Accept: Media type
-Language: Natural language
-Encoding: Content transformation
-Charset: ISO, Unicode (standard)
-Keep alive: Used to allow reusing
-Connection: Same connection for other request

http://www.w3.org/Protocols/rfc2616/rfc2616.html

Explain the individual parts of the HTTP message on slide 21.

- 1xx: Informational - Request received, continuing process - 2xx: Success - The action was successfully received, understood, and accepted - 3xx: Redirection - Further action must be taken in order to complete the request - 4xx: Client Error - The request contains bad syntax or cannot be fulfilled - 5xx: Server Error - The server failed to fulfill an apparently valid request

Response Header	Value
(Status-Line)	HTTP/1.1 200 OK
Content-Type	text/html; charset=utf-8
Cache-Control	no-cache
Expires	Fri, 01 Jan 1990 00:00:00 GMT
Content-Encoding	gzip
Date	Tue, 19 Jan 2010 13:03:28 GMT
Server	Google Frontend
Content-Length	955
X-XSS-Protection	0

- Status line: Protocol version and status code (success, fail)
- Content: Media type and charset
- Cache: Unidirectional directive
- Expires: Expiration date
-Encoding: Content transformation
- Date: When message was originated
- Server: software used by the origin server
-Length: Body size in octets
-Protection: 0 → do not check XSS attack (injected Javascript)

How will an HTTP/1.1-compliant server react if you use a „host“ parameter in combination with an absolute URI

- With an absolute URI, the host will be ignored
- Relative URI is use only when the client fetch subsequence pages (images in the same page or sections in the same page)
- See also: http://www.w3.org/Protocols/rfc2616/rfc2616-sec5.html#sec5.2

Explain the interplay between the HTTP protocol and the DNS when a client fetches a Web page

- HTTP: When a message involves a non numerical host, DNS gets the corresponding IP

If a client tries to dereference a URI Reference (i.e. one including a hash fragment), will the server be aware of the hash fragment?

- HTTP clients dereference hash URIs by stripping off the part after the hash and dereferencing the resulting URI
  → Server is not aware about the part after the hash

Explain the GET, HEAD, and POST methods of HTTP/1.1.

- **HEAD**: Similar to get but without the response body. This is useful for retrieving meta-information written in response headers, without having to transport the entire content.
- **GET**: Requests a representation of the specified resource. Note that GET should not be used for operations that cause side-effects, such as using it for taking actions in web applications.
- **POST** Submits data to be processed (e.g., from an HTML form) to the identified resource. The data is included in the body of the request. This may result in the creation of a new resource or the updates of existing resources or both.

See: http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol

http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html

Explain the role of status codes in HTTP

- Attempt to understand and satisfy the request
  - 1xx: Informational - Request received, continuing process
  - 2xx: Success - The action was successfully received, understood, and accepted
  - 3xx: Redirection - Further action must be taken in order to complete the request
  - 4xx: Client Error - The request contains bad syntax or cannot be fulfilled
  - 5xx: Server Error - The server failed to fulfill an apparently valid request

Explain the role of media types in the context of HTTP content negotiation

- Providing open and extensible data typing and type negotiation
  - With the media type it is possible to specify the representation you want
  - Indicate preferences
  - See: Quality Values

Use cURL to display the response header of the URI http://www.heppnetz.de/

- curl -I http://www.heppnetz.de/
  - HTTP/1.1 200 OK
  - Date: Mon, 25 Jan 2010 13:14:53 GMT
  - Server: Apache
  - Last-Modified: Fri, 11 Dec 2009 15:15:23 GMT
  - ETag: "edd0768-3ec1-4b22620b"
  - Accept-Ranges: bytes
  - Content-Length: 16065
  - Content-Type: text/html

cURL: Representation only if it was modified after Jan 1, 2010.

- curl -I -z "Fri, 01 Jan 2010 00:00:00 GMT" http://www.heppnetz.de
  - HTTP/1.1 304 Not Modified
  - Date: Mon, 25 Jan 2010 13:45:48 GMT
  - Server: Apache
  - ETag: "edd0768-3ec1-4b22620b"

cURL: fetch a representation in application/rdf+xml for the URI http://www.heppnetz.de/ontologies/goodrelations/v1

- curl -I "Accept: application/rdf+xml" http://www.heppnetz.de/ontologies/goodrelations/v1
  - curl: (6) Couldn't resolve host 'Accept: application'
  - HTTP/1.1 303 See Other
  - Date: Mon, 25 Jan 2010 14:10:45 GMT
  - Server: Apache
  - Location: http://www.heppnetz.de/ontologies/goodrelations/v1.owl
  - Content-Type: text/html; charset=iso-8859-1

- curl -I -H "Accept: application/rdf+xml" http://www.heppnetz.de/ontologies/goodrelations/v1
  - HTTP/1.1 303 See Other
  - Date: Mon, 25 Jan 2010 14:12:04 GMT
  - Server: Apache
  - Location: http://www.heppnetz.de/ontologies/goodrelations/v1.owl
  - Content-Type: text/html; charset=iso-8859-1

cURL: fetch a representation in application/rdf+xml for the URI http://www.heppnetz.de/ontologies/goodrelations/v1

- curl -L -I -H "Accept: application/rdf+xml" http://www.heppnetz.de/ontologies/goodrelations/v1
  - HTTP/1.1 303 See Other
  - Date: Tue, 26 Jan 2010 08:34:19 GMT
  - Server: Apache
  - Location: http://www.heppnetz.de/ontologies/goodrelations/v1.owl
  - Content-Type: text/html; charset=iso-8859-1

  - HTTP/1.1 200 OK
  - Date: Tue, 26 Jan 2010 08:34:19 GMT
  - Server: Apache
  - Last-Modified: Wed, 25 Nov 2009 14:41:38 GMT
  - ETag: "ee7c006-18ceb-4b0d4222"
  - Accept-Ranges: bytes
  - Content-Length: 101611
  - Content-Type: application/rdf+xml

Thank you.

These slides and additional materials are at

http://www.ebusiness-unibw.org/wiki/WWW-WT2010