Prelims

Hypertext Transfer Protocol (HTTP)

- Communication protocol used to access hypermedia and/or hypertext on the WWW

- Version history:

  - HTTP 0.9

    - using only the GET method

    - can only establish one connection, which means that there is no persistent connection

    - server responds HTML only

  - HTTP 1.0

    - RFC 1945

    - Methods added:

      - POST

      - HEAD

    - can already accommodate extensions (e.g., characters, etc.)

  - HTTP 1.1

    - RFC 2068, RFC 2616, and RFC 7230-7235

    - Protocol currently being used

    - 8 methods all in all

    - Features:

      - can reuse connections - can already accommodate persistent connections

      - data compression - body can already be compressed to increase speed

    - SPDY - Google's project to enhance data scheme

  - HTTP 2

    - RFC 7540

- Similar to HTTP 1, but with the following improvements:

  • data transmission

  • multiplexing and pipelining

    - pipelining - concurrent

  • accommodate header compression

  • accommodate server push - server anticipates possible requests so it can already send the response ahead of time

- HTTP Fundamentals

  • HTTP is a stateless protocol which means that the server does not keep information about the client

  • Runs on top of TCP/IP

    - TCP Port 80 - default

  • https

    - to secure information

    - uses TCP port 443

    - HTTP over SSL/TLS

  • Based on Client Server Architecture (CSA)

    - server

      • server - origin server processes requests

      • proxy server, gateways, tunnels

    - clients (user agents)

      • gives requests to server

      • web crawlers/spiders, etc.

  • Uses "request response" protocol

  • Provides support for the following functionalities:

    - Cache control

- Reload usually accessed Web pages faster the next time a client requests for it
- Content media type specification (MIME)
- Language and character set specification
- Content or transfer codings
  - Ways in transmitting the data
    - chunked
    - compressed
- Content negotiation
  - Client negotiates with the server to specify a specific format for the response
- Persistent connection
- Client-server protocol negotiations
- Request pipelining/multiplexing
- Authentication/authorization
- HTTP Resource Addressing
  - HTTP resources are identified using URI (Uniform Resource Indicator) RFC 3986
  - Parts of URL:
    - scheme (http or https)
    - authority
      - user information/authentication credentials (deprecated)
      - host - associated with the domain name (resolved to an IP address using DNS)
      - port number (optional)
    - path
      - refers to a resource
      - resource types:

- dynamic - creates new information on the fly

- static - does not create new information on the fly