Infrastructure I

1. Web Caching Basics
   1. Idea:
      1. Storing of reusable responses
   2. Cache
      1. French roots 🡪 store
      2. Caching 🡪 Very common concept e.g. processors
      3. Locality of references
         1. Time and space
      4. (Cache) Hit 🡪 requested data is found
      5. Fresh/stale 🡪 ready to use? Need to validate?
      6. Locality of reference
         1. Time and space
      7. Good, necessary or just evil?
2. Why web caching
   1. Network
      1. Reduce load
   2. Client
      1. Faster access (cached content is fast)
      2. Handle intermittent connectivity
   3. Server
      1. Reduce load
3. Terminology
   1. Actors
      1. Server (aka origin server)
         1. Manages and provides access to resources
         2. Sends response to client initiated request
      2. Client (aka user agent)
         1. Initiate transaction by sending requests
      3. Proxy
         1. Intermediary in a web transaction
         2. Acts as client and server
   2. URI
      1. URL 🡪 most common type – location dependent
      2. URN 🡪 reference is location/server independent
   3. HTTP
      1. Application layer protocol
      2. Verbs, Headers, Response code
4. General pre-conditions for caching
   1. Locality of client requests
      1. Locality
   2. Costs of caching < costs of direct retrievals
      1. What costs?
5. Terminology II
   1. Latency
   2. Bandwidth
   3. Server Load
   4. Cache hit ratio
   5. HTTP defines 3 basic mechanisms
      1. Freshness
      2. Validation
      3. Invalidation
   6. Locations of cache
      1. Client (browser cache) / intermediary (proxy cache) / server (reverse cache)
6. HTTP verbs
   1. GET 🡪 Cacheable
   2. PUT 🡪 uncatchable
   3. POST 🡪 uncatchable by default 🡪 Cache-control can make it cacheable
   4. DELETE 🡪 uncatchable
   5. HEAD 🡪 helpful to update cached entry
   6. PATCH 🡪 uncatchable
   7. OPTIONS🡪 uncatchable
   8. TRACE 🡪 uncatchable
   9. CONNECT 🡪 uncatchable
7. Response Codes
   1. 200 OK
   2. 203 Non-authoritative information
   3. 206 partial content
   4. 301 moved permanently
   5. …
8. More caching
   1. Caching proxy
      1. Authentication
         1. Proxy-Authorization
      2. Request filtering
      3. Response filtering
      4. Prefetching
      5. Translation and transcoding
      6. Traffic shaping
      7. Web accelerator
         1. Cache
         2. Freshen responses
         3. Prefetch
         4. Compress and optimize
         5. Filter
         6. Persistent connections