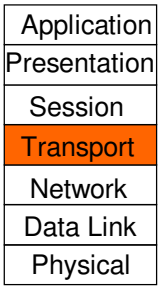
CSE/EE 461 – Lecture 18

HTTP and the Web

David Wetherall
djw@cs.washington.edu

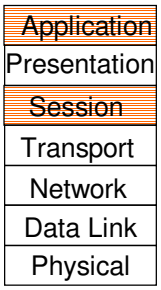
Last Time …

- The Transport Layer
- Focus
  - How does TCP share bandwidth?
- Topics
  - AIMD
  - Slow Start
  - Fast Retransmit / Fast Recovery

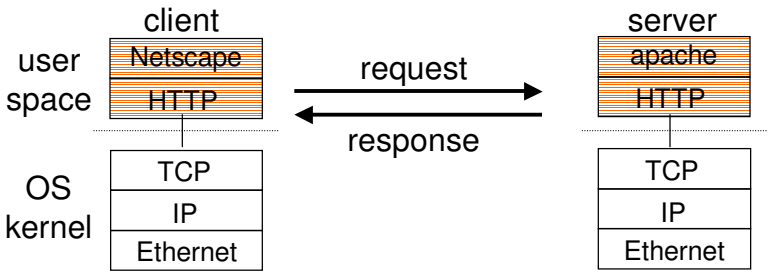


This Lecture

- HTTP and the Web (but not HTML)
- Focus
  - How do Web transfers work?
- Topics
  - HTTP, HTTP1.1
  - Get-If-Modified
  - Caching and Consistency

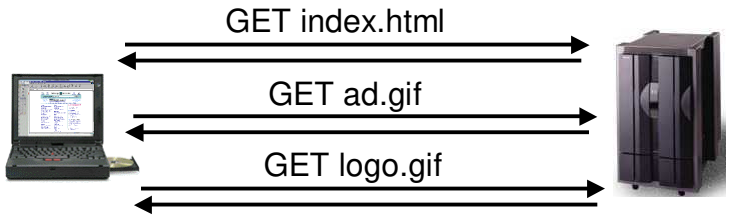


Web Protocol Stacks



- To view the URL http://server/page.html the client makes a TCP connection to port 80 of the server, by it's IP address, sends the HTTP request, receives the HTML for page.html as the response, repeats the process for inline images, and displays it.
- We cover the mapping from server names to IP later!

Simple HTTP

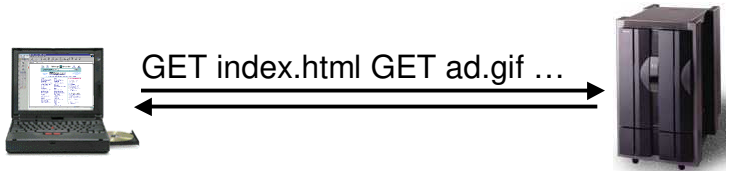


- HTTP is a tiny, text-based language
- The GET method requests an object
- There are HTTP headers, like "Content-Length:", etc.
- Try "telnet server 80" then "GET page.html"

Other HTTP Methods

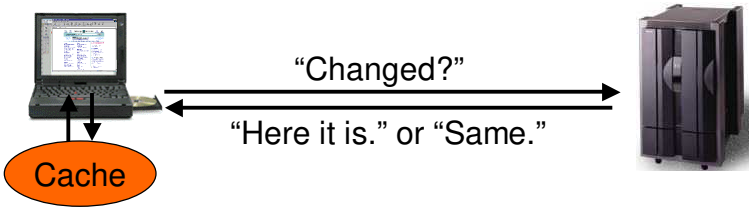
- We can do more than just retrieve pages …
- POST uploads data to the server
  - For when you hit Submit on a form
- HEAD gets the page headers but not content
  - Can check to see how new the page is, etc.

HTTP1.1: Persistent Connections



- Bright Idea: Use one TCP connection for multiple page downloads (or just HTTP methods)
- Q: What are the advantages?
- Q: What are the disadvantages?

Browser Caches

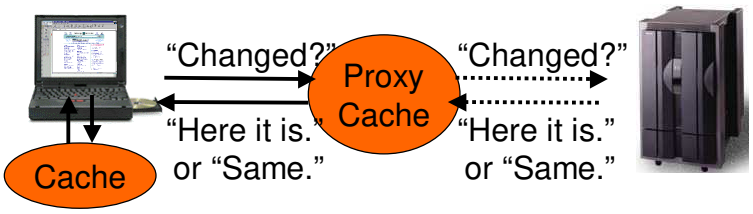


- Bigger win: avoid repeated transfers of the same page
- Check local browser cache to see if we have the page
- GET with If-Modified-Since makes sure it's up-to-date
- Q: What are the advantages and disadvantages?

Consistency and Caching Directives

- Key issue is knowing when cached data is fresh/stale
  - Otherwise many connections or the risk of staleness
- Browsers typically use heuristics
  - To reduce server connections and hence realize benefits
  - Check freshness once a "session" with GET If-Modified-Since and then assume it's fresh the rest of the time
  - Possible to have inconsistent data.
- Caching directives provide hints
  - Expires: header is basically a time-to-live
  - Also indicate whether page is cacheable or not

Proxy Caches



- Insert further levels of caching for greater gain
- Share proxy caches between many users (not shown)
  - If I haven't downloaded it recently, maybe you have
- Your browser has built-in support for this

Next Steps?

- Different types of content (streaming media, XML)
- Content Delivery Networks (caching alternative)
- Security (for all those purchases)

Key Concepts

- HTTP and the Web is just a shim on top of TCP
  - Sufficient and enabled rapid adoption
  - Many "scalability" and performance issues now important