The Application Layer: HTTP, SMTP

Lecture 6
http://www.cs.rutgers.edu/~sn624/352-S22
Srinivas Narayana
Quick recap of concepts

HyperText Transfer Protocol (HTTP)

HTTP is a client/server application

Methods: GET/POST/…

Headers
User-agent/server/…

Response codes: 200, 404, etc.

Persistence

Connection initiation

RTT

HTTP req

HTTP resp

Cookies

Set-cookie: XXX

Cookie: XXX

Cookie file

Caching
e.g., proxy server
Web caches

Web caches: Machines that remember web responses for a network

Why cache web responses?

• Reduce response time for client requests
• Reduce traffic on an institution’s access link

Last lecture: Caches implemented in the form of a proxy server
Content Distribution Networks (CDNs)

A global network of web caches
- Provisioned by ISPs and network operators
- Or content providers, like Netflix, Google, etc.

Uses (overlaps with uses of web caching in general)
- Reduce traffic on a network’s Internet connection, e.g., Rutgers
- Improve response time for users: CDN nodes are closer to most users than origin servers
- Reduce bandwidth requirements on content provider
- Reduce $$ to maintain origin servers
Without CDN

Clients distributed all over the world

- Problems:
  - Huge bandwidth requirements for Rutgers
  - Large propagation delays to reach users

<table>
<thead>
<tr>
<th>DOMAIN NAME</th>
<th>IP ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cs.rutgers.edu</td>
<td>128.6.4.2</td>
</tr>
<tr>
<td><a href="http://www.google.com">www.google.com</a></td>
<td>74.125.225.243</td>
</tr>
<tr>
<td><a href="http://www.princeton.edu">www.princeton.edu</a></td>
<td>128.112.132.86</td>
</tr>
</tbody>
</table>

Cluster of Rutgers CS origin servers (located in NJ, USA)
Where the CDN comes in

• Distribute content of the origin server over geographically distributed CDN servers

• But how will users get to these CDN servers?

• Use DNS!
  • DNS provides an additional layer of indirection
  • Instead of returning IP address, return another DNS server (NS record)
  • The second DNS server (run by the CDN) returns IP address to client

• The CDN runs its own DNS servers (CDN name servers)
  • Custom logic to send users to the “closest” CDN web server
With CDN

NS record delegates the choice of IP address to the CDN name server.

CDN Name Server (124.8.9.8)

Popular CDNs:
- CloudFlare
- Akamai
- Level3
...

Custom logic to map ONE domain name to one of many IP addresses!

Most requests go to CDN servers (caches). CDN servers may request object from origin.
Few client requests go directly to origin server.
Summary of HTTP

- Request/response protocol
- ASCII-based human-readable message structures
- Improve performance using connection persistence, caching, and CDN
- Enhanced stateful functionality using cookies
- Simple, highly-customizable protocol
  - Just add headers
- Protocol that forms the basis of the web we enjoy today!
Simple Mail Transfer Protocol
We’re all familiar with email. How does it work?
Electronic Mail

Three major components:

1. **User agents**
   - a.k.a. “mail reader”
   - e.g., Applemail, Outlook
   - Web-based user agents (ex: gmail)
Electronic Mail: Mail servers

2. Mail Servers
- Mailbox contains incoming messages for user
- Message queue of outgoing (to be sent) mail messages
- Sender’s mail server makes connection to Receiver’s mail server
  - IP address, port 25

3. SMTP protocol: client/server protocol
- Used to send messages
- Client: sending user agent or sending mail server
- server: receiving mail server
Scenario: Alice sends message to Bob

1) Alice (alice@rutgers.edu) uses UA to compose message to bob@nyu.edu
2) Alice’s UA sends message to her mail server; message placed in outgoing message queue
3) Client side of SMTP opens TCP connection with Bob’s mail server
4) SMTP client sends Alice’s message over the TCP connection
5) Bob’s mail server places the message in Bob’s incoming mailbox
6) Sometime later, Bob invokes his user agent to read message

A set of durable files on the machine. Persisted on disk.
Observations on these exchanges

• Mail servers are the “infrastructure” for email functionality
  • Receiving the email on behalf of Bob, should Bob’s machine be turned off
  • Retrying the delivery of the email to Bob on behalf of Alice, should Bob’s mail server be unavailable in the first attempt

• The same machine can act as client or server based on context
  • Rutgers’s mail server is the server when Alice sends the mail
  • It is the client when it sends mail to Bob’s mail server

• SMTP is push-based: info is pushed from client to server
  • Contrast to HTTP or DNS where info is pulled from the server
Sample SMTP interaction

• telnet <mail-server> 25
• HELO <sender-domain>
• MAIL FROM: <name>@<sender-domain>
• RCPT TO: <user>@<mail-server-domain>
• DATA
• Put data in, then [enter].[enter] Don’t forget the “.”
• You can add mail headers (later) to make your email look good
### MAIL command response codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>System status or help reply</td>
</tr>
<tr>
<td>214</td>
<td>Help message</td>
</tr>
<tr>
<td>220</td>
<td>Service ready</td>
</tr>
<tr>
<td>221</td>
<td>Service closing transmission channel</td>
</tr>
<tr>
<td>250</td>
<td>Request command completed</td>
</tr>
<tr>
<td>251</td>
<td>User not local; the message will be forwarded</td>
</tr>
<tr>
<td>354</td>
<td>Start mail input</td>
</tr>
<tr>
<td>421</td>
<td>Service not available</td>
</tr>
<tr>
<td>450</td>
<td>Mailbox not available</td>
</tr>
<tr>
<td>451</td>
<td>Command aborted: local error</td>
</tr>
<tr>
<td>452</td>
<td>Command aborted: insufficient storage</td>
</tr>
<tr>
<td>500</td>
<td>Syntax error; unrecognized command</td>
</tr>
<tr>
<td>501</td>
<td>Syntax error in parameters or arguments</td>
</tr>
<tr>
<td>502</td>
<td>Command not implemented</td>
</tr>
<tr>
<td>503</td>
<td>Bad sequence of commands</td>
</tr>
<tr>
<td>504</td>
<td>Command temporarily not implemented</td>
</tr>
<tr>
<td>550</td>
<td>Command is not executed; mailbox unavailable</td>
</tr>
<tr>
<td>551</td>
<td>User not local</td>
</tr>
<tr>
<td>552</td>
<td>Requested action aborted; exceeded storage location</td>
</tr>
<tr>
<td>553</td>
<td>Requested action not taken; mailbox name not allowed</td>
</tr>
<tr>
<td>554</td>
<td>Transaction failed</td>
</tr>
</tbody>
</table>

220: Service ready  
250: Request command completed  
354: Start mail input  
421: Service not available  
500: Unrecognized command
Mail message format (stored on server)

SMTP text message exchange standardized in RFC 822

- **Header lines**, e.g.,
  - To:
  - From:
  - Subject:
    *These are different from SMTP commands!*
    (these would still be under “DATA”)

- **body**
  - the “message”.
  - ASCII characters only
Message format: multimedia extensions

- MIME: multimedia mail extension, RFC 2045, 2056
- additional headers in DATA header declare MIME content type
- A message can have many parts

```
From: alice@crepes.fr
To: bob@hamburger.edu
Subject: Picture of yummy crepe.
MIME-Version: 1.0
Content-Transfer-Encoding: base64
Content-Type: image/jpeg

base64 encoded data ..... 
..........................
......base64 encoded data
```
Dine, Shop, Play & Stay... It All Happens Here! VALENTINE'S WEEKEND =EF=BB=BF
live music & theater entertainment, comedy shows, a variety of activities,
chocolate shops, and high-end restaurants make New Brunswick an excellent
choice for couples on Valentine's Day! It's easy to see why this city is
romantic with so many fun date night options, luxurious spas, and deliciou
s dining selections. Book Your Reservations Now Jersey's Choice Restaurant =
Poll Presented by New Brunswick's Performing Arts Center, it's your time to-
show your support for your favorite spots in New Brunswick as NJ Monthly h=
osts their 39th Annual Restaurants of NJ Poll! New Brunswick is represented
strongly with many of your favorites including, Roosterspin, Stage Left St=
Valentine’s Day in New Brunswick - Weekend Specials &amp; Events

Valentine’s Weekend

Live music & theater entertainment, comedy shows, a variety of activities, chocolate shops, and high-end restaurants make New Brunswick an excellent choice for couples on Valentine’s Day!

It’s easy to see why this city is romantic with so many fun date night options, luxurious spas, and delicious dining selections.

Book Your Reservations Now

Jersey’s Choice Restaurant Poll

Presented by New Brunswick’s Performing Arts Center, it’s your time to show your support for your favorite spots in New Brunswick as NJ Monthly hosts their 39th Annual Restaurants of NJ Poll!

New Brunswick is represented strongly with many of your favorites including, Roosterspin, Stage Left Steak, Harvest Moon Brewery, The Frog & The Peach, Delta’s, Cambo Box,
Mail Access Protocols
Mail access protocols

• SMTP: delivery/storage to receiver’s server. Focused on push

• Mail access protocol: retrieval from server
  • POP: Post Office Protocol [RFC 1939]
    • Client connects to POP3 server on TCP port 110
  • IMAP: Internet Mail Access Protocol [RFC 1730]
    • Client connects to TCP port 143
  • HTTP: gmail, outlook, etc.
POP vs IMAP

- POP3
  - Stateless server
  - UA-heavy processing
  - UA retrieves email from server, then typically deleted from server
  - Latest changes are at the UA
  - Simple protocol (list, retr, del within a POP session)

- IMAP4
  - Stateful server
  - UA and server processing
  - Server sees folders, etc. which are visible to UAs
  - Latest changes are at the server
  - Complex protocol
  - Heavily used: email sync across devices, reliable, …
What about web-based email?

• Connect to mail servers via web browser
  • Ex: gmail, scarletmail, etc.

• Browsers speak HTTP
• Email servers speak SMTP
• Need to bridge these two
Web based email

Application process on the web server machine uses SMTP to push mail

HTTP server
scarletmail.rutgers.edu

HTTP server
outlook.com

SMTP server
aspmx4.googlemail.com

SMTP server
outlook-com.olc.protection.outlook.com

Internet

Alice
alice@scarletmail.rutgers.edu

Bob
bob@outlook.com

Alice’s mail provider’s server(s)

Bob’s mail provider’s server(s)

May run on the same or different machines (owned by your webmail provider)

IMAP

Application process on the web server uses access protocol to pull email

HTTP
Comparing SMTP with HTTP

- HTTP: pull
- SMTP: push

- both have ASCII command/response interaction, status codes
- HTTP: each object encapsulated in its own response msg
- SMTP: multiple objects sent in multipart msg

- HTTP: can put non-ASCII data directly in response (dedicated entity body for binary data)
- SMTP: need ASCII-based encoding (base64)
More themes from app-layer protocols

• Keep it simple until you really need complexity
  • Start with ASCII-based design; stateless servers. Then introduce:
  • Cookies for HTTP state
  • Stateful mail (IMAP, folders, etc.) for email organization
  • Security extensions (e.g., TLS)
  • Performance optimizations: persistence, caching, indirection, …
  • Use headers as much as possible to non-intrusively evolve functionality

• Partition functions based on what’s done best at the user (app) and protocol. Examples:
  • Content rendering for users (browser, UA) separate from protocol operations (mail server)
  • mail UA doesn’t need to be “always on” to send or receive email reliably. That’s the mail server’s job