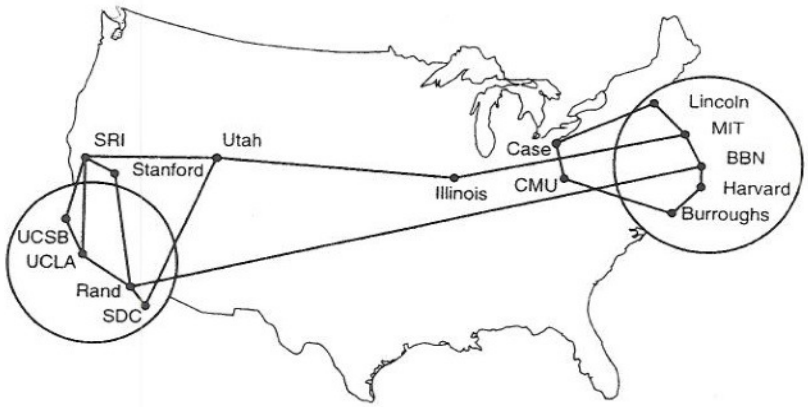


Applications & MAIL

Outline/Goals

- Approaches to Application Protocols
- EMAIL
- Chapter 9

Goals of Today's Lecture

- Application-layer protocols
  - Applications vs. application-layer protocols
  - Tailoring the protocol to the application
- Electronic mail
  - E-mail messages, and MIME
  - E-mail addresses, and role of DNS
  - E-mail servers and user agents
- Electronic mail protocols
  - Transferring e-mail messages between servers (SMTP)
  - Retrieving e-mail messages (POP, IMAP, and HTTP)

Application-Layer Protocols

- Network applications run on end systems
  - They depend on the network to provide a service
  - … but cannot run software on the network elements
- Network applications run on multiple machines
  - Different end systems communicate with each other
  - Software is often written by multiple parties
- Leading to a need to explicitly define a protocol
  - Types of messages (e.g., requests and responses)
  - Message syntax (e.g., fields, and how to delineate)
  - Semantics of the fields (i.e., meaning of the information)
  - Rules for when and how a process sends messages

Application vs. Application-Layer Protocols

- Application-layer protocol is just one piece
  - Defining how the end hosts communicate
- Example: World Wide Web
  - HyperText Transfer Protocol is the protocol
  - But the Web includes other components, such as document formats (HTML), Web browsers, servers, …
- Example: electronic mail
  - Simple Mail Transfer Protocol (SMTP) is the protocol
  - But e-mail includes other components, such as mail servers, user mailboxes, mail readers

Protocols Tailored to the Application

- Telnet: interacting with account on remote machine
  - Client simply relays user keystrokes to the server
  - … and server simply relays any output to the client
  - TCP connection persists for duration of the login session
  - Network Virtual Terminal format for transmitting ASCII data, and control information (e.g., End-of-Line delimiter)
- FTP: copying files between accounts
  - Client connects to remote machine, "logs in, and issues commands for transferring files to/from the account
  - … and server responds to commands and transfers files
  - Separate TCP connections for control and data
  - Control connection uses same NVT format as Telnet

Protocols Tailored to the Application

- SMTP: sending e-mail to a remote mail server
  - Sending mail server transmits e-mail message to a mail server running on a remote machine
  - Each server in the path adds its identifier to the message
  - Single TCP connection for control and data
  - SMTP replaced the earlier use of FTP for e-mail
- HTTP: satisfying requests based on a global URL
  - Client sends a request with method, URL, and meta-data
  - … and the server applies the request to the resource and returns the response, including meta-data
  - Single TCP connection for control and data

Comparing the Protocols

- Commands and replies
  - Telnet sends commands in binary, whereas the other protocols are text based
  - Many of the protocols have similar request methods and response codes
- Data types
  - Telnet, FTP, and SMTP transmit text data in standard U.S. 7-bit ASCII
  - FTP also supports transfer of data in binary form
  - SMTP uses MIME standard for sending non-text data
  - HTTP incorporates some key aspects of MIME (e.g., classification of data formats)

Comparing the Protocols (Cont.)

- Transport
  - Telnet, FTP, SMTP, and HTTP all depend on reliable transport protocol
  - Telnet, SMTP, and HTTP use a single TCP connection
  - … but FTP has separate control and data connections
- State
  - In Telnet, FTP, and SMTP, the server retains information about the session with the client
  - E.g., FTP server remembers client's current directory
  - In contrast, HTTP servers are stateless

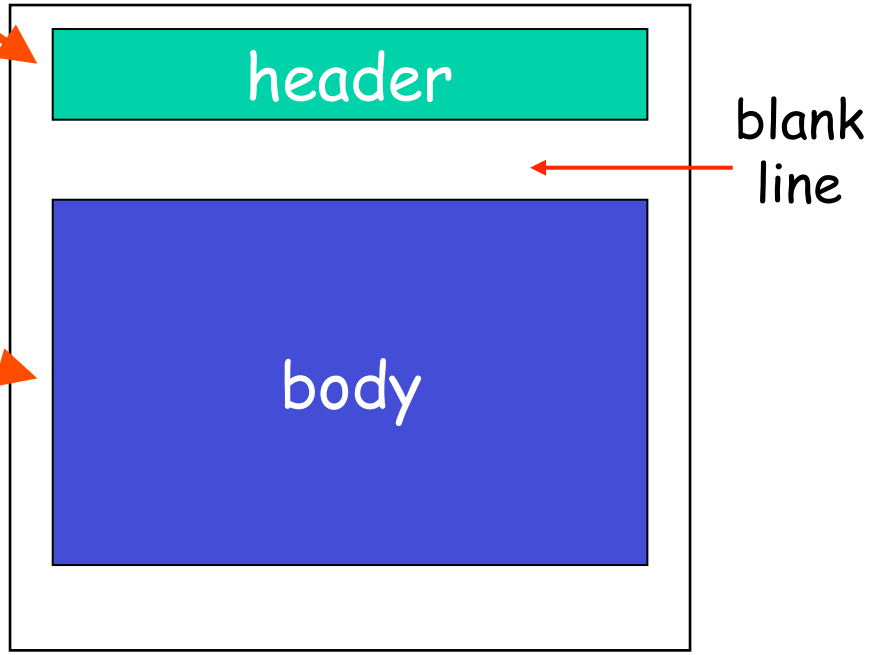
Reflecting on Application-Layer Protocols

- Protocols are tailored to the applications
  - Each protocol is customized to a specific need
- Protocols have many key similarities
  - Each new protocol was influenced by the previous ones
  - New protocols commonly borrow from the older ones
- Protocols depend on same underlying substrate
  - Ordered reliable stream of bytes (i.e., TCP)
  - Domain Name System (DNS)
- Relevance of the protocol standards process
  - Important for interoperability across implementations
  - Yet, not necessary if same party writes all of the software
  - …which is increasingly common (e.g., P2P software)

Electronic Mail

E-Mail Message

- E-mail messages have two parts
  - A header, in 7-bit U.S. ASCII text
  - A body, also represented in 7-bit U.S. ASCII text
- Header
  - Lines with "type: value"
  - "To: jrex@princeton.edu"
  - "Subject: Go Tigers!"
- Body
  - The text message
  - No particular structure or meaning



E-Mail Message Format (RFC 822)

- E-mail messages have two parts
  - A header, in 7-bit U.S. ASCII text
  - A body, also represented in 7-bit U.S. ASCII text
- Header
  - Series of lines ending in carriage return and line feed
  - Each line contains a type and value, separated by ":"
  - E.g., "To: jrex@princeton.edu" and "Subject: Go Tigers"
  - Additional blank line before the body begins
- Body
  - Series of text lines with no additional structure/meaning
  - Conventions arose over time (e.g., e-mail signatures)

Limitation: Sending Non-Text Data

- E-mail body is 7-bit U.S. ASCII
  - What about non-English text?
  - What about binary files (e.g., images and executables)?
- Solution: convert non-ASCII data to ASCII
  - Base64 encoding: map each group of three bytes into four printable U.S.-ASCII characters
  - Uuencode (Unix-to-Unix Encoding) was widely used

```
begin 644 cat.txt
#0V%T
`

end
```

Limitation: filename is the only cue to the data type

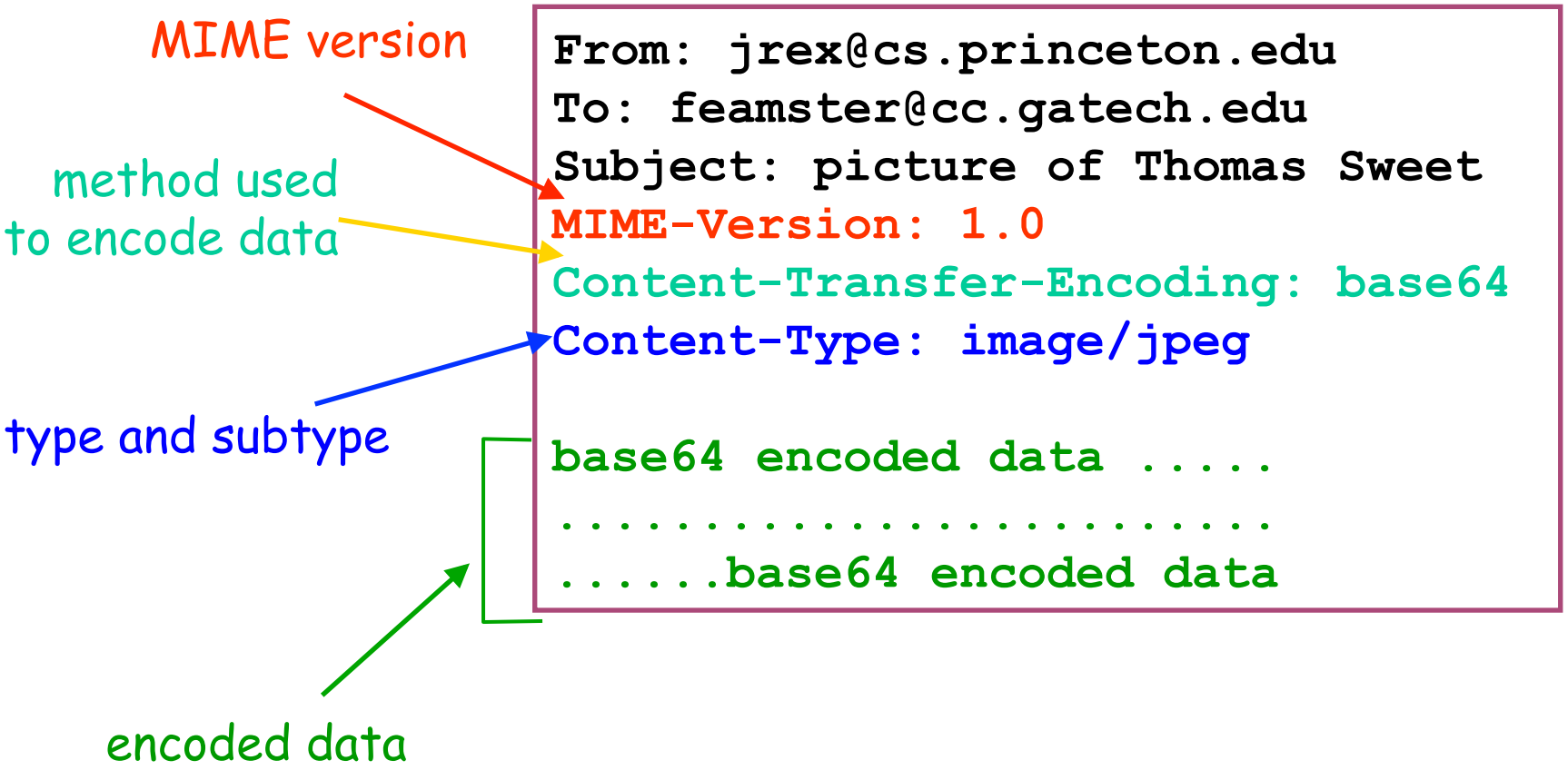
Limitation: Sending Multiple Items

- Users often want to send multiple pieces of data
  - Multiple images, powerpoint files, or e-mail messages
  - Yet, e-mail body is a single, uninterpreted data chunk
- Example: e-mail digests
  - Encapsulating several e-mail messages into one aggregate messages (i.e., a digest)
  - Commonly used on high-volume mailing lists
- Conventions arose for how to delimit the parts
  - E.g., well-known separator strings between the parts
  - Yet, having a standard way to handle this is better

Multipurpose Internet Mail Extensions

- Additional headers to describe the message body
  - MIME-Version: the version of MIME being used
  - Content-Type: the type of data contained in the message
  - Content-Transfer-Encoding: how the data are encoded
- Definitions for a set of content types and subtypes
  - E.g., image with subtypes gif and jpeg
  - E.g., text with subtypes plain, html, and richtext
  - E.g., application with subtypes postscript and msword
  - E.g., multipart for messages with multiple data types
- A way to encode the data in ASCII format
  - Base64 encoding, as in uuencode/uudecode

Example: E-Mail Message Using MIME



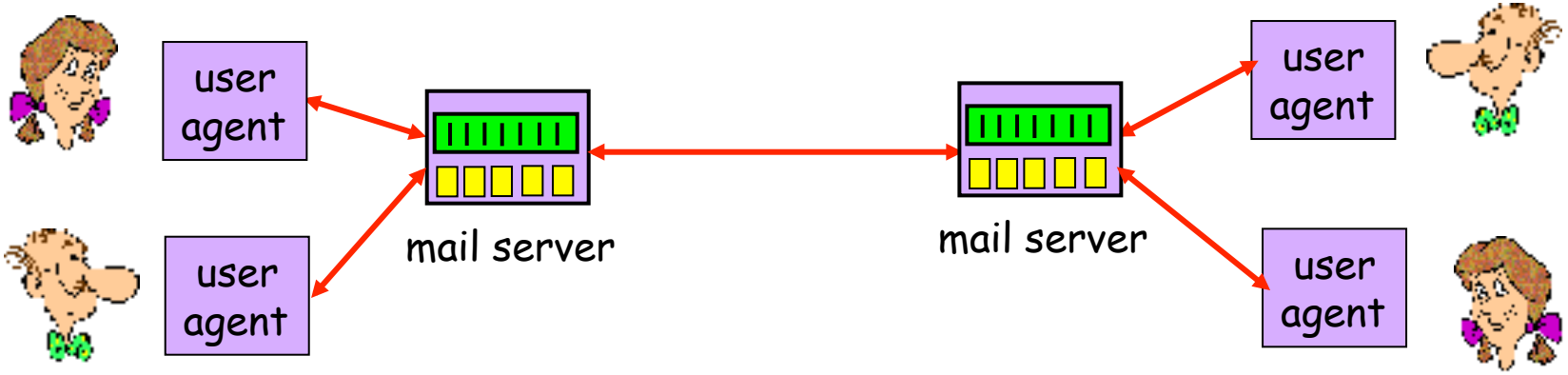
Distribution of Content Types

- Example content types in e-mail archive
  - Searched on "Content-Type", not case sensitive
  - Extracted the value field, and counted unique types
  - At UNIX command line: **`grep -i Content-Type * | cut -d" " -f2 | sort | uniq -c | sort –nr`**
- Out of 44343 matches
  - 25531: text/plain
  - 7470: multipart to send attachments
  - 4230: text/html
  - 759: application/pdf
  - 680: application/msword
  - 479: application/octet-stream
  - 292: image (mostly jpeg, and some gif, tiff, and bmp)

E-Mail Addresses

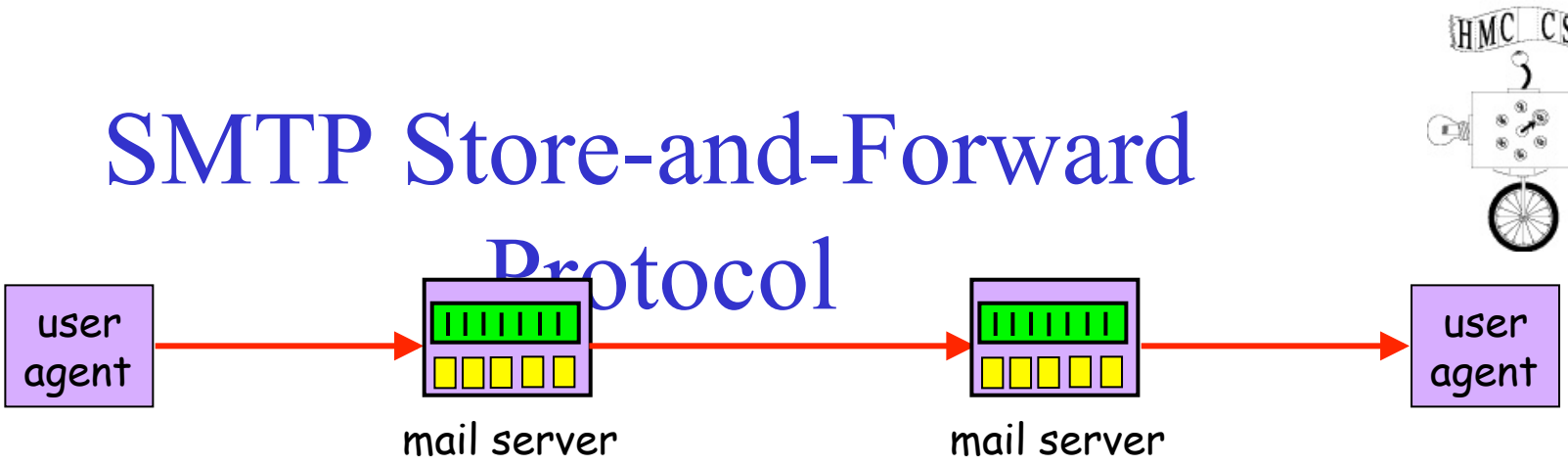
- Components of an e-mail address
  - Local mailbox (e.g., jrex or bob.flower)
  - Domain name (e.g., cs.princeton.edu)
- Domain name is not necessarily the mail server
  - Mail server may have longer/cryptic name
    - E.g., cs.princeton.edu vs. mail.cs.princeton.edu
  - Multiple servers may exist to tolerate failures
    - E.g., cnn.com vs. atlmail3.turner.com and nycmail2.turner.com
- Identifying the mail server for a domain
  - DNS query asking for MX records (Mail eXchange)
    - E.g., nslookup –q=mx cs.princeton.edu
  - Then, a regular DNS query to learn the IP address

Mail Servers and User Agents



- Mail servers
  - Always on and always accessible
  - Transferring e-mail to and from other servers
- User agents
  - Sometimes on and sometimes accessible
  - Intuitive interface for the user

SMTP Store-and-Forward Protocol



- Messages sent through a series of servers
  - A server stores incoming messages in a queue
  - … to await attempts to transmit them to the next hop
- If the next hop is not reachable
  - The server stores the message and tries again later
- Each hop adds its identity to the message
  - By adding a "Received" header with its identity
  - Helpful for diagnosing problems with e-mail

Example With Received Header


```
Return-Path: <casado@cs.stanford.edu>
Received: from ribavirin.CS.Princeton.EDU (ribavirin.CS.Princeton.EDU [128.112.136.44])
      by newark.CS.Princeton.EDU (8.12.11/8.12.11) with SMTP id k04M5R7Y023164
      for <jrex@newark.CS.Princeton.EDU>; Wed, 4 Jan 2006 17:05:37 -0500 (EST)
Received: from bluebox.CS.Princeton.EDU ([128.112.136.38])
      by ribavirin.CS.Princeton.EDU (SMSSMTP 4.1.0.19) with SMTP id M20060104170536079 46
      for <jrex@newark.CS.Princeton.EDU>; Wed, 04 Jan 2006 17:05:36 -0500
Received: from smtp-roam.Stanford.EDU (smtp-roam.Stanford.EDU [171.64.10.152])
      by bluebox.CS.Princeton.EDU (8.12.11/8.12.11) with ESMTP id k04M5XNQ005204
      for <jrex@cs.princeton.edu>; Wed, 4 Jan 2006 17:05:35 -0500 (EST)
Received: from [192.168.1.101] (adsl-69-107-78-147.dsl.pltn13.pacbell.net [69.107.78.147])
      (authenticated bits=0)
      by smtp-roam.Stanford.EDU (8.12.11/8.12.11) with ESMTP id k04M5W92018875
      (version=TLSv1/SSLv3 cipher=DHE-RSA-AES256-SHA bits=256 verify=NOT);
      Wed, 4 Jan 2006 14:05:32 -0800
Message-ID: <43BC46AF.3030306@cs.stanford.edu>
Date: Wed, 04 Jan 2006 14:05:35 -0800
From: Martin Casado <casado@cs.stanford.edu>
User-Agent: Mozilla Thunderbird 1.0 (Windows/20041206)
MIME-Version: 1.0
To: jrex@CS.Princeton.EDU
CC: Martin Casado <casado@cs.stanford.edu>
Subject: Using VNS in Class
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: 7bit
```

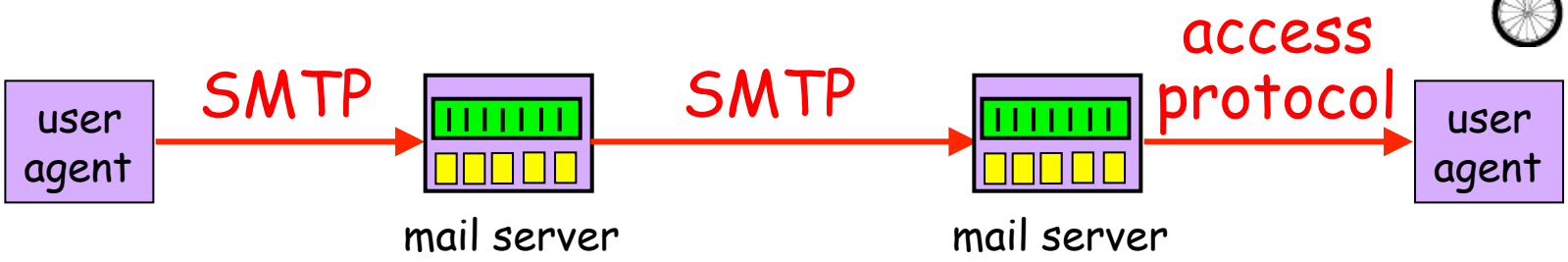
Multiple Server Hops

- Typically at least two mail servers
  - Sending and receiving sides
- May be more
  - Separate servers for key functions
    - Spam filtering
    - Virus scanning
  - Servers that redirect the message
    - From jrex@princeton.edu to jrex@cs.princeton.edu
    - Messages to princeton.edu go through extra hops
  - Electronic mailing lists
    - Mail delivered to the mailing list's server
    - … and then the list is expanded to each recipient

Electronic Mailing Lists

- Community of users reachable by one address
  - Allows groups of people to receive the messages
- Exploders
  - Explode a single e-mail message into multiple messages
  - One copy of the message per recipient
- Handling bounced messages
  - Mail may bounce for several reasons
  - E.g., recipient mailbox does not exist; resource limits
- E-mail digests
  - Sending a group of mailing-list messages at once
  - Messages delimited by boundary strings
  - … or transmitted using multiple/digest format

Simple Mail Transfer Protocol



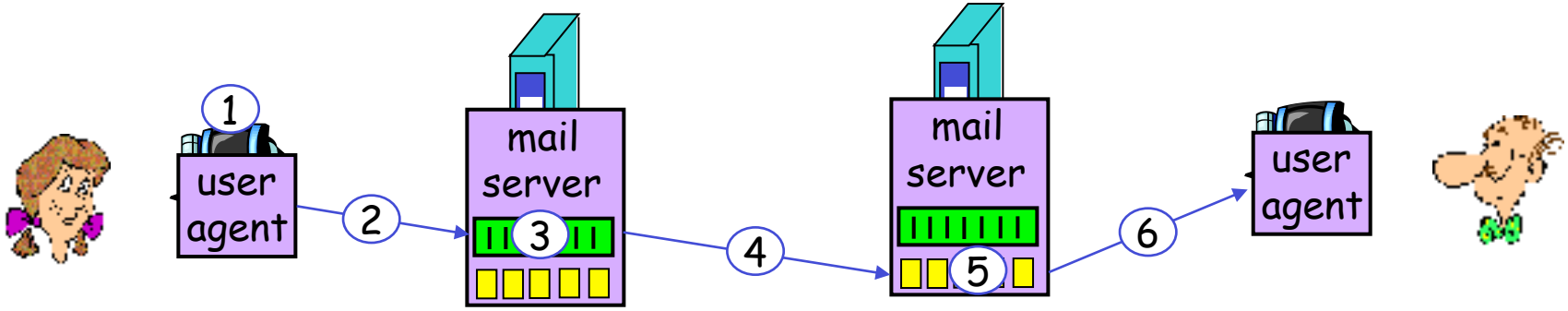
- Client-server protocol
  - Client is the sending mail server
  - Server is the receiving mail server
- Reliable data transfer
  - Built on top of TCP (on port 25)
- Push protocol
  - Sending server pushes the file to the receiving server
  - … rather than waiting for the receiver to request it

Simple Mail Transfer Protocol (Cont.)

- Command/response interaction
  - Commands: ASCII text
  - Response: three-digit status code and phrase
- Synchronous
  - Sender awaits response from a command
  - … before issuing the next command
  - Though pipelining of commands was added later
- Three phases of transfer
  - Handshaking (greeting)
  - Transfer of messages
  - Closure

Scenario: Alice Sends Message to Bob

1) Alice uses UA to compose message "to" `bob@someschool.edu`
2) Alice's UA sends message to her mail server; message placed in 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 mailbox
6) Bob invokes his user agent to read message



Sample SMTP interaction


```
S: 220 hamburger.edu
C: HELO crepes.fr
S: 250  Hello crepes.fr, pleased to meet you
C: MAIL FROM: <alice@crepes.fr>
S: 250 alice@crepes.fr... Sender ok
C: RCPT TO: <bob@hamburger.edu>
S: 250 bob@hamburger.edu ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: Do you like ketchup?
C: How about pickles?
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 hamburger.edu closing connection
```

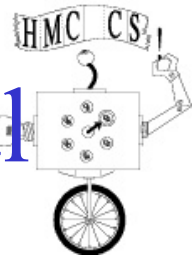
Try SMTP For Yourself

- Running SMTP
  - Run "telnet servername 25" at UNIX prompt
  - See 220 reply from server
  - Enter HELO, MAIL FROM, RCPT TO, DATA commands
- Thinking about spoofing?
  - Very easy
  - Just forge the argument of the "FROM" command
  - … leading to all sorts of problems with spam
- Spammers can be even more clever
  - E.g., using open SMTP servers to send e-mail
  - E.g., forging the "Received" header

Retrieving E-Mail From the Server

- Server stores incoming e-mail by mailbox
  - Based on the "From" field in the message
- Users need to retrieve e-mail
  - Asynchronous from when the message was sent
  - With a way to view the message and reply
  - With a way to organize and store the messages
- In the olden days…
  - User logged on to the machine where mail was delivered
  - Users received e-mail on their main work machine

Influence of PCs on E-Mail Retrieval



- Separate machine for personal use
  - Users did not want to log in to remote machines
- Resource limitations
  - Most PCs did not have enough resources to act as a full-fledged e-mail server
- Intermittent connectivity
  - PCs only sporadically connected to the network
  - … due to dial-up connections, and shutting down of PC
  - Too unwieldy to have sending server keep trying
- Led to the creation of Post Office Protocol (POP)

Post Office Protocol (POP)

- POP goals
  - Support users with intermittent network connectivity
  - Allow them to retrieve e-mail messages when connected
  - … and view/manipulate messages when disconnected
- Typical user-agent interaction with a POP server
  - Connect to the server
  - Retrieve all e-mail messages
  - Store messages on the user's PCs as new messages
  - Delete the messages from the server
  - Disconnect from the server
- User agent still uses SMTP to send messages

POP3 Protocol

Authorization phase

- Client commands:
  - **`user:`** declare username
  - **`pass:`** password
- Server responses
  - **`+OK`**
  - **`-ERR`**

Transaction phase, client:

- **`list:`** list message numbers
- **`retr:`** retrieve message by number
- **`dele:`** delete
- **`quit`**

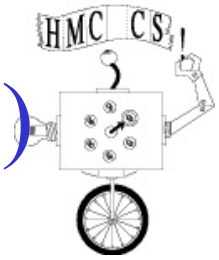
```
S: +OK POP3 server ready
C: user bob
S: +OK
C: pass hungry
S: +OK user successfully logged on
```

```
C: list
S: 1 498
S: 2 912
S: .
C: retr 1
S: <message 1 contents>
S: .
C: dele 1
C: retr 2
S: <message 1 contents>
S: .
C: dele 2
C: quit
S: +OK POP3 server signing off
```

Limitations of POP

- Does not handle multiple mailboxes easily
  - Designed to put user's incoming e-mail in one folder
- Not designed to keep messages on the server
  - Instead, designed to download messages to the client
- Poor handling of multiple-client access to mailbox
  - Increasingly important as users have home PC, work PC, laptop, cyber café computer, friend's machine, etc.
- High network bandwidth overhead
  - Transfers all of the e-mail messages, often well before they are read (and they might not be read at all!)

Interactive Mail Access Protocol (IMAP)



- Supports connected and disconnected operation
  - Users can download message contents on demand
- Multiple clients can connect to mailbox at once
  - Detects changes made to the mailbox by other clients
  - Server keeps state about message (e.g., read, replied to)
- Access to MIME parts of messages & partial fetch
  - Clients can retrieve individual parts separately
  - E.g., text of a message without downloading attachments
- Multiple mailboxes on the server
  - Client can create, rename, and delete mailboxes
  - Client can move messages from one folder to another
- Server-side searches
  - Search on server before downloading messages

Web-Based E-Mail

- User agent is an ordinary Web browser
  - User communicates with server via HTTP
  - E.g., Gmail, Yahoo mail, and Hotmail
- Reading e-mail
  - Web pages display the contents of folders
  - … and allow users to download and view messages
  - "GET" request to retrieve the various Web pages
- Sending e-mail
  - User types the text into a form and submits to the server
  - "POST" request to upload data to the server
  - Server uses SMTP to deliver message to other servers
- Easy to send anonymous e-mail (e.g., spam)

Conclusions

- Application-layer protocols
  - Rich and constantly evolving area
  - Tailoring communication to the application
- Electronic-mail protocols
  - SMTP to transfer e-mail messages
  - Several retrieval techniques (POP, IMAP, and Web)
- Evolution from text to a wide variety of formats
  - Text-based e-mail in RFC 822
  - MIME to represent a wide variety of data formats