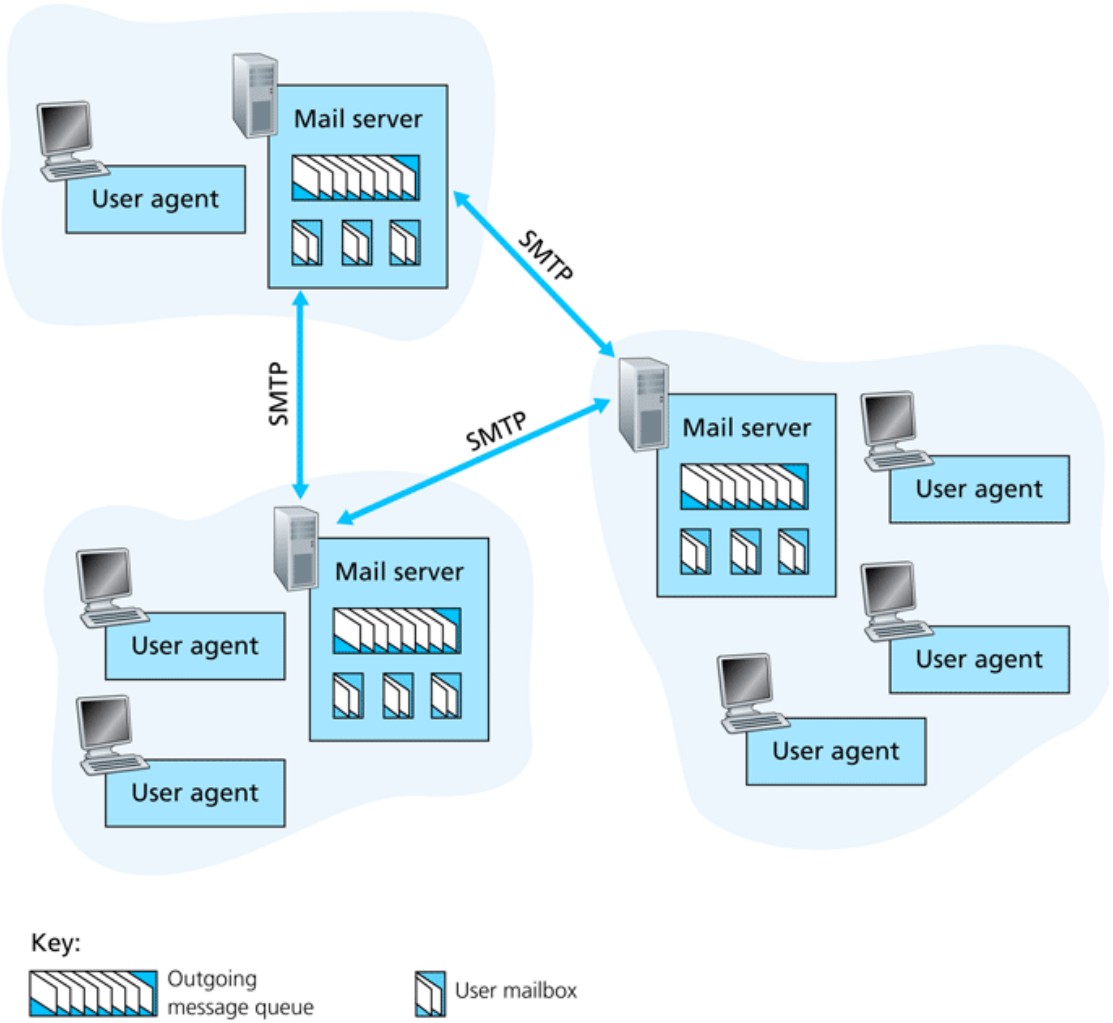
Email Protocols and Spam Prevention

CS 360 Internet Programming

Daniel Zappala

Brigham Young University
Computer Science Department

Mail Protocols

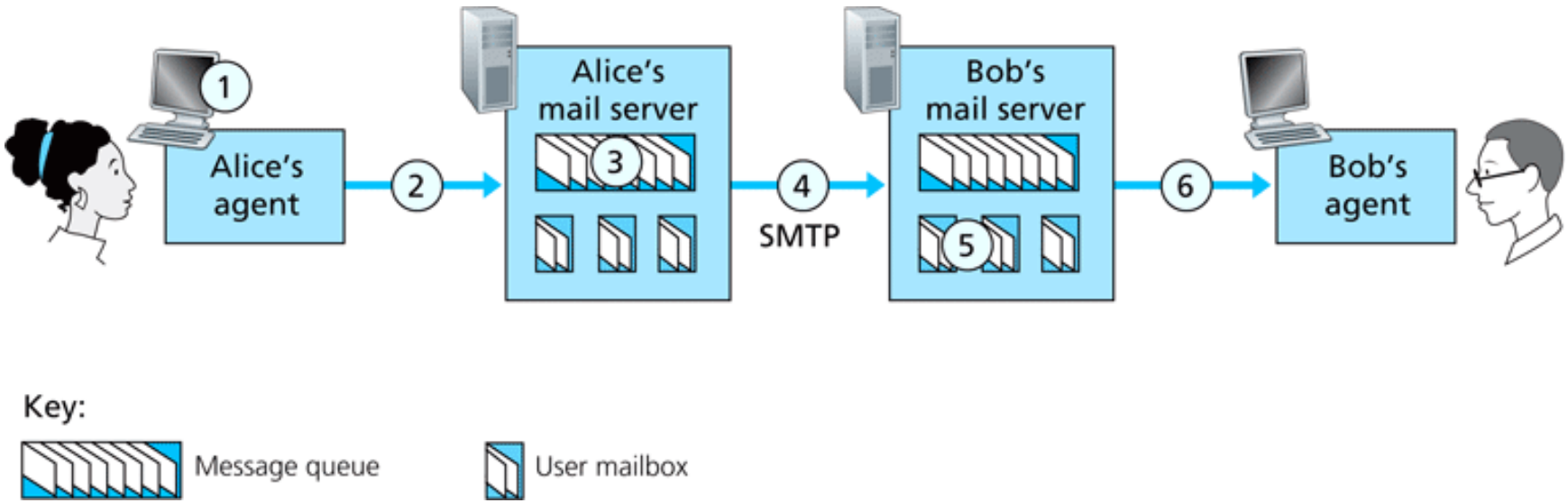


- user agents
  - POP, IMAP, or HTTP to exchange mail
- mail transfer agents (MTAs)
  - mailbox to hold incoming messages for users
  - message queue of outgoing mail messages
- Simple Mail Transfer Protocol (SMTP)
  - between MTAs

SMTP

- specified in RFC 2821
- uses TCP to transfer mail between servers, uses port 25
- three phases
  - handshaking
  - transfer of messages
  - closure
- command/response interaction
- commands: ASCII text
- response: status code and phrase
- messages must be in 7-bit ASCII

Mail Transfer

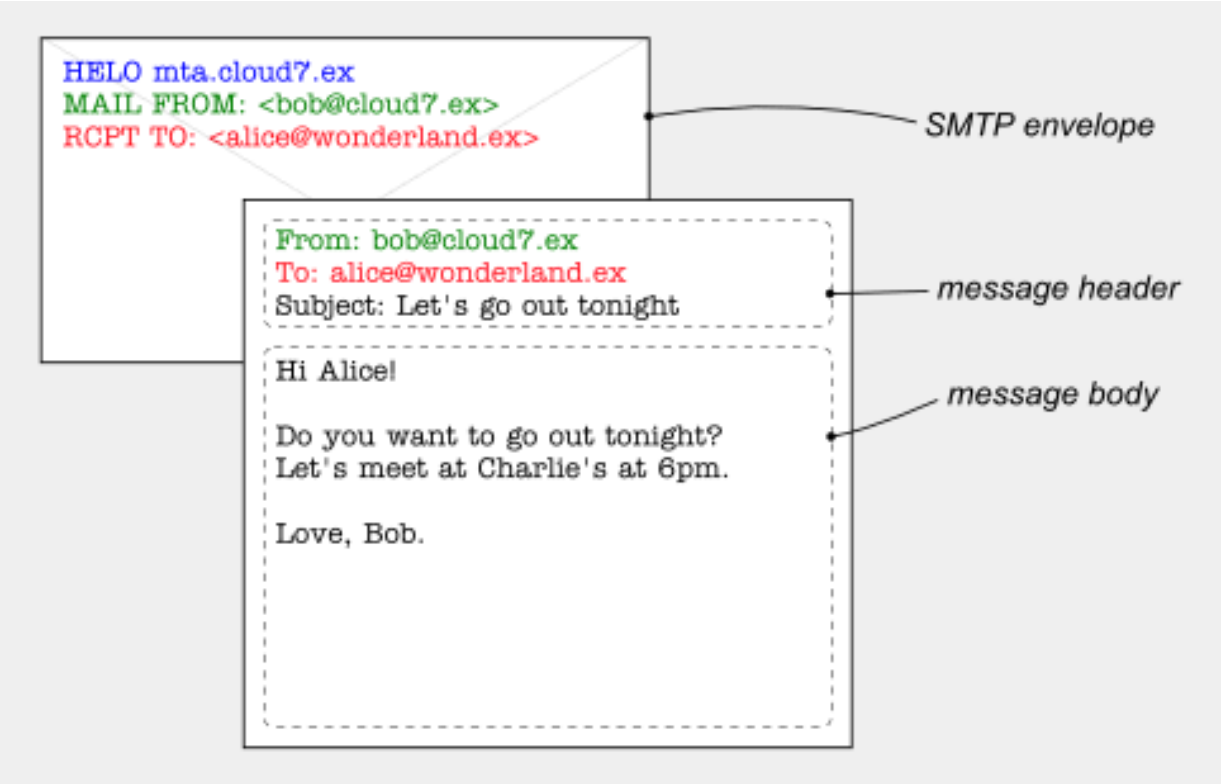


Example Session

```
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
```

SMTP Protocol

- creates the SMTP envelope for the transaction
- other RFCs define the message header and message body



Text Message Format

- RFC 822 standardizes text message format
- header
  - To:
  - From:
  - Subject:
  - etc.
- blank line
- body
  - the message in ASCII characters

MIME Format

- RFCs 2045, 2056 describe MultiMedia Mail Extension
- additional lines in message header declare MIME type

```
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
```

Economic Reasons

- low cost
  - text is low bandwidth
  - bandwidth is cheap
  - modern computers can send many messages per second
- huge audience
- need only a small fraction of users to respond in order to make money
- low risk – hard to track who sent the email

Tragedy of the (Unregulated) Commons

- free access and unrestricted demand for a finite resource will eventually lead to destruction of the resource through over-exploitation
  - motivation for individuals to maximize their use of the resource
  - individuals reap benefits of exploitation
  - costs of exploitation distributed to everyone
- examples
  - air, water and soil pollution
  - deforestation, overfishing, water shortages
  - traffic congestion, light pollution
  - open WiFi frequencies

Technical Reasons

- open relays: classically, in SMTP, anyone could send email through your server
  - from anywhere – because any IP address can connect to the mail server
  - from anyone – because the FROM address is not verified
  - to anyone – because the mail server would relay email to any user
- most ISPs now setup a closed relay
  - you can originate mail to anyone if you are a user of the ISP - verified by IP address or login
  - you can deliver mail to the server for a local user from anywhere
- if everyone ran a closed relay, then the originator of spam could be easily identified and punished
- many ISPs have blacklisted open relays: DSBL
  - detect open relay by trying to send email through it
  - add to a blacklist

Sending Email Remotely

- open relays used to be necessary when sending email while traveling
- better solutions
  - VPN: secure connection from remote network to local network
  - SMTP-AUTH: login to mail server before sending email
    - need a good password policy to prevent brute force attacks
    - vulnerable to any user that is compromised
  - POP before SMTP: must have successfully downloaded mail via POP before sending mail

Older Methods

- open relays
  - remedy: blacklists
- web mail
  - treat it as a throwaway account – send until you get shut down
  - remedy: account verification procedures, rate limiting
- disposable ISP accounts
  - sign up with a falsified or stolen CC numbers – use until shut down
  - remedy: better verification, paper trail

Botnets

- a large set of PCs taken over by a cracker and used to perpetrate various attacks on the Internet
  - spam
  - denial-of-service attacks
  - ad click fraud
- account for 85% of spam according to some studies
- greater access to bandwidth allows image spam – harder to filter and higher costs

Bayesian Spam Filtering

- use a Naive Bayes classifier to identify spam
  - user trains filter by marking messages as spam
  - may generate false negatives until trained properly
  - may generate false positives
- weaknesses
  - training time
  - false negatives and positives
  - spammers place random text in email to fool the filter

Micropayments

- require sender of email to pay a fraction of a cent per email
  - make the cost low enough that it is extremely cheap for the average user
  - spammers who send very large amounts of email will have to bear a significant cost
- weaknesses
  - people want absolutely free email
  - botnets can cause ordinary people to pay for the spam someone else sends
  - ultimately doesn't get rid of spam if spammers are willing to pay

SPF

- Sender Policy Framework: SPF
- validate who can send email from a domain
  - the domain owner publishes a DNS record listing which IP addresses can send email from that domain
  - the receiving MTA checks whether the email originates from one of these IP addresses
  - policy determines whether to reject or filter email
- advantages
  - hard to forge IP address for the duration of a TCP connection
  - can reject email connection before any email is transmitted
  - can use a reputation system to determine whether to accept email from a domain that passes the SPF test
- weaknesses
  - forwarding agents will get rejected unless whitelisted
  - forwarders should use Sender Rewriting Scheme (SRS), which essentially rewrites and remails the email from them

DKIM

- Domain Keys Identified Mail: DKIM
  - email authentication
  - sending MTA signs on behalf of the sender using a header added to the email message
  - receiving MTA authenticates using the public key of the domain
  - find the public key through DNS
- advantages
  - positively identifies sending domain: prevents phishing attacks
  - can discard forged domains easily
  - can use reputation systems, whitelists, blacklists to determine whether to accept mail from a particular domain
- weaknesses
  - doesn't encrypt the envelope of the message (return path, recipients, etc), just the message itself

SPF vs DKIM

- SPF
  - validate who can send email from a domain
  - based on IP addresses, which are hard to forge
  - can reject a message before it is received
  - lightweight computation
- DKIM
  - verify that the author is legitimate
  - authentication is better than authorization
  - works on the message body, not the SMTP envelope
  - requires the entire message to be received before it is verified
  - computationally expensive

Practical Considerations

- combining SPF and DKIM
  - use SPF to limit who can send mail
  - check DKIM to verify sender is legitimate
  - use reputation system to block spammers
  - to send spam from a domain you would need to get their key AND send from one of the valid IP addresses for that domain
- weaknesses of both schemes
  - requires a reputation authority to revoke the certification of a domain as "spam-free"
  - requires ISPs to terminate the accounts of spammers or suffer the consequences

Throwaway Domains

- the next challenge for spam
  - spammer signs up for a new domain
  - registers SPF record and DKIM keys
  - sends spam and his reputation is new/unknown
  - throw away domain when enough MTAs start blocking you
- remedies
  - rapid blacklisting
  - reputation system based on age of domain
  - follow legal records to see who paid for domain and prosecute where applicable