# Spam and Ham


성균관대학교
최형기

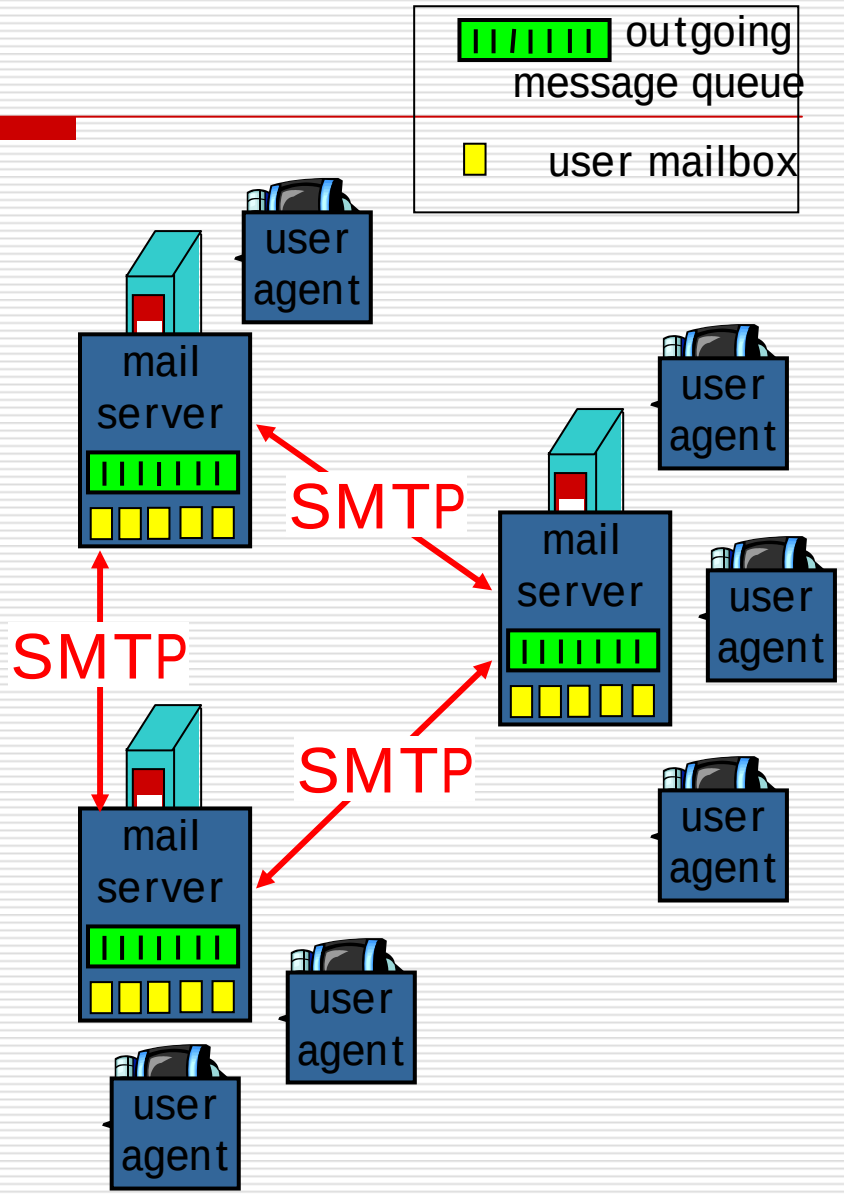
# Agenda

- Introduction to spam
- Techniques spammers use
- Kinds of spam
- Solutions to spam
- Conclusion

# Electronic Mail

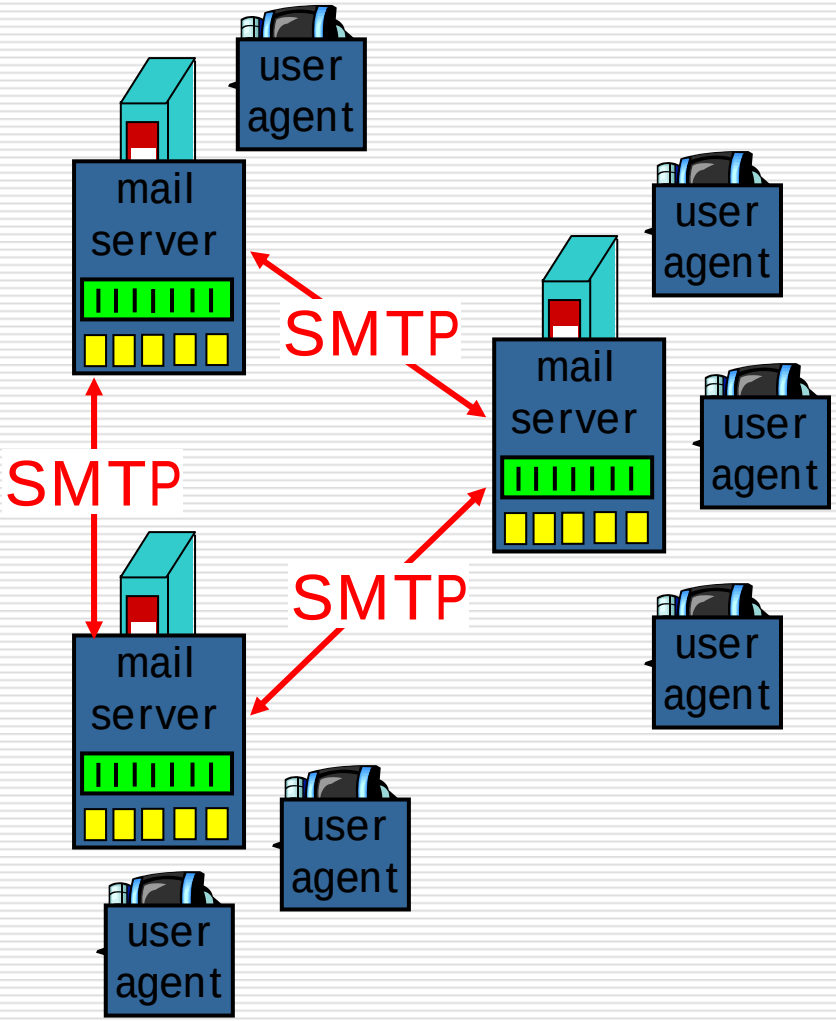
- Three major components:
  - user agents
  - mail servers
  - simple mail transfer protocol: SMTP

- User Agent
  - a.k.a. “mail reader”
  - composing, editing, reading mail messages
  - e.g., Eudora, Outlook, elm, Netscape Messenger
  - outgoing, incoming messages stored on server

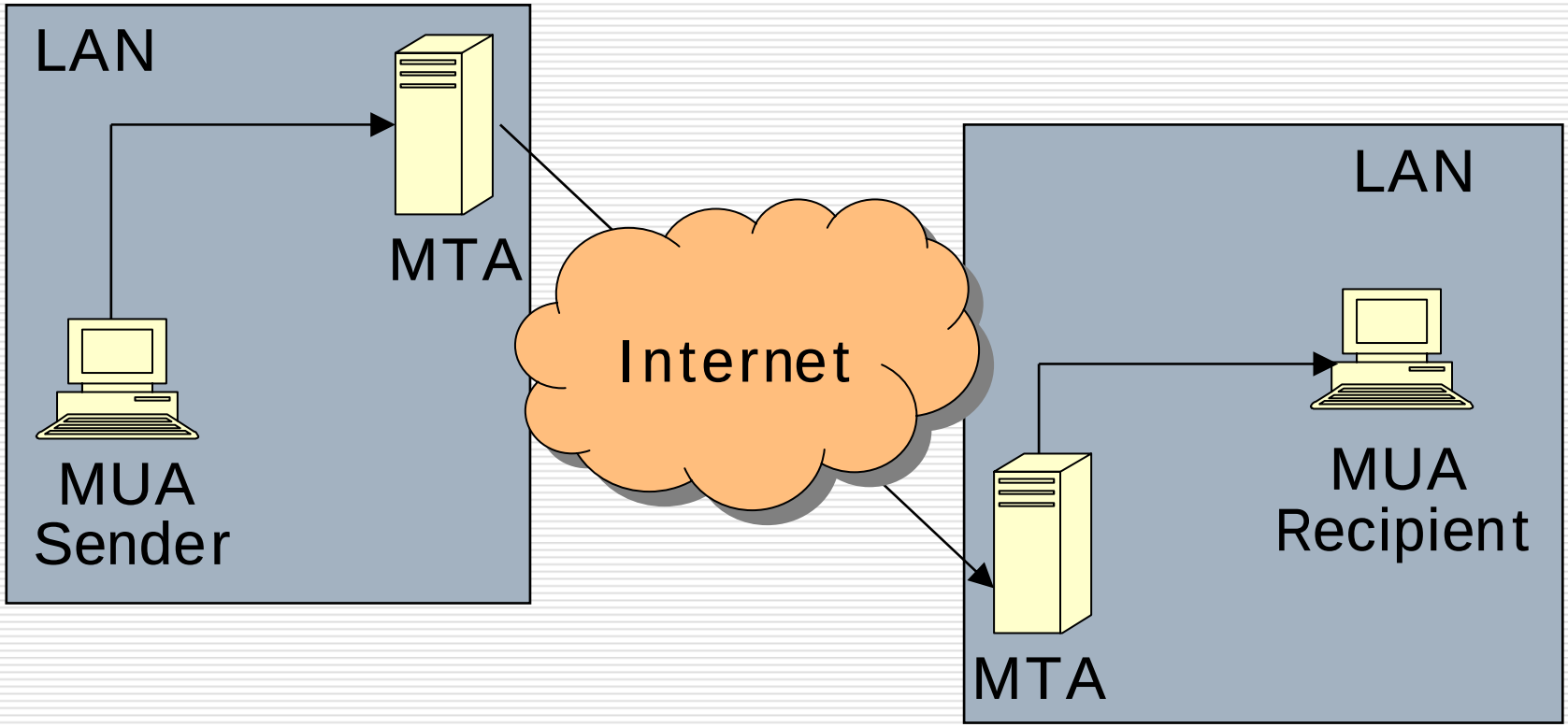


# Electronic Mail: mail servers

- mailbox contains incoming messages (yet to be read) for user
- message queue of outgoing (to be sent) mail messages
- SMTP protocol between mail servers to send email messages
  - client: sending mail server
  - server: receiving mail server



# How Are E-mails Transported?

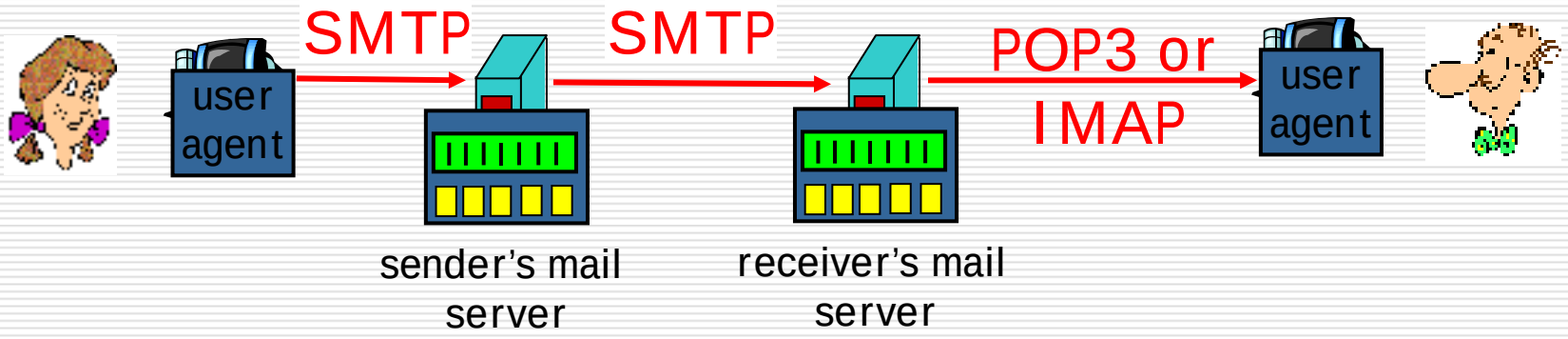



- MUA= Mail User Agent, *aka* Mail Client
- MTA=Mail Transport Agent, *aka* Mail Server

# Composition and Delivery

- Sender supplies `To:` and `Subject:` fields and message body.
- MUA translates into RFC 822 message and connects across LAN to MTA
- MUA instructs MTA using a protocol called SMTP (or a proprietary alternative; e.g. MS exchange) and sends RFC 822 message.
- Sender's MTA uses MX record in DNS to find IP address of recipient's MTA based on `To:` field.
- Sender's MTA opens connection to Recipient's MTA transfer RFC 822 message using SMTP.
- Recipient's MTA may deliver to Recipient's MUA or may store message locally for later retrieval.

# Mail access protocols




- SMTP: delivery/storage to receiver's server
- Mail access protocol: retrieval from server
  - POP: Post Office Protocol [RFC 1939]
    - authorization (agent <-->server) and download
  - IMAP: Internet Mail Access Protocol [RFC 1730]
    - more features (more complex)
    - manipulation of stored msgs on server
  - HTTP: Hotmail , Yahoo! Mail, etc.

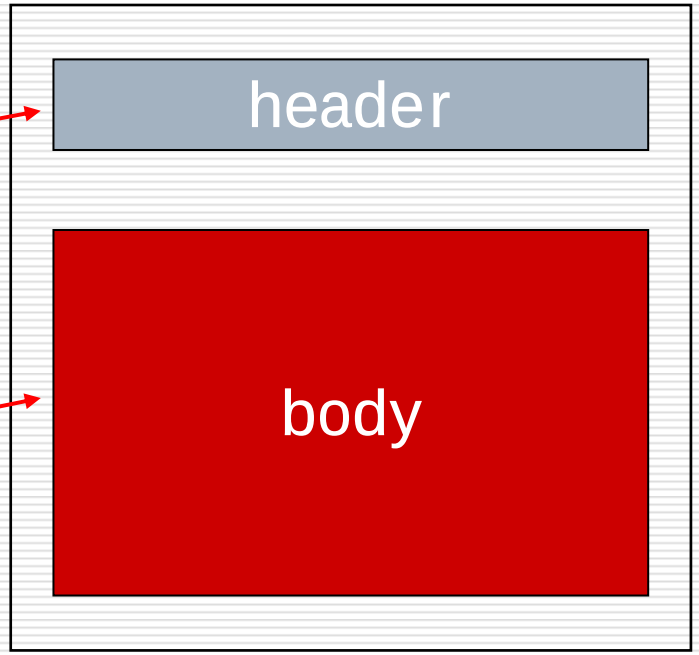
# Electronic Mail: SMTP [RFC 821]

- Uses TCP to reliably transfer email msg from client to server, port 25
- Direct transfer: sending server to receiving server
- Three phases of transfer
  - handshaking (greeting)
  - transfer of messages
  - closure
- Command/response interaction
  - commands: ASCII text
  - response: status code and phrase
- Messages must be in 7-bit ASCII

# Mail message format

- RFC 822: standard for text message format:
- header lines, e.g.,
  - To:
  - From:
  - Subject:

  *different from SMTP commands*!
- body
  - the "message", ASCII characters only



# Sample SMTP interaction

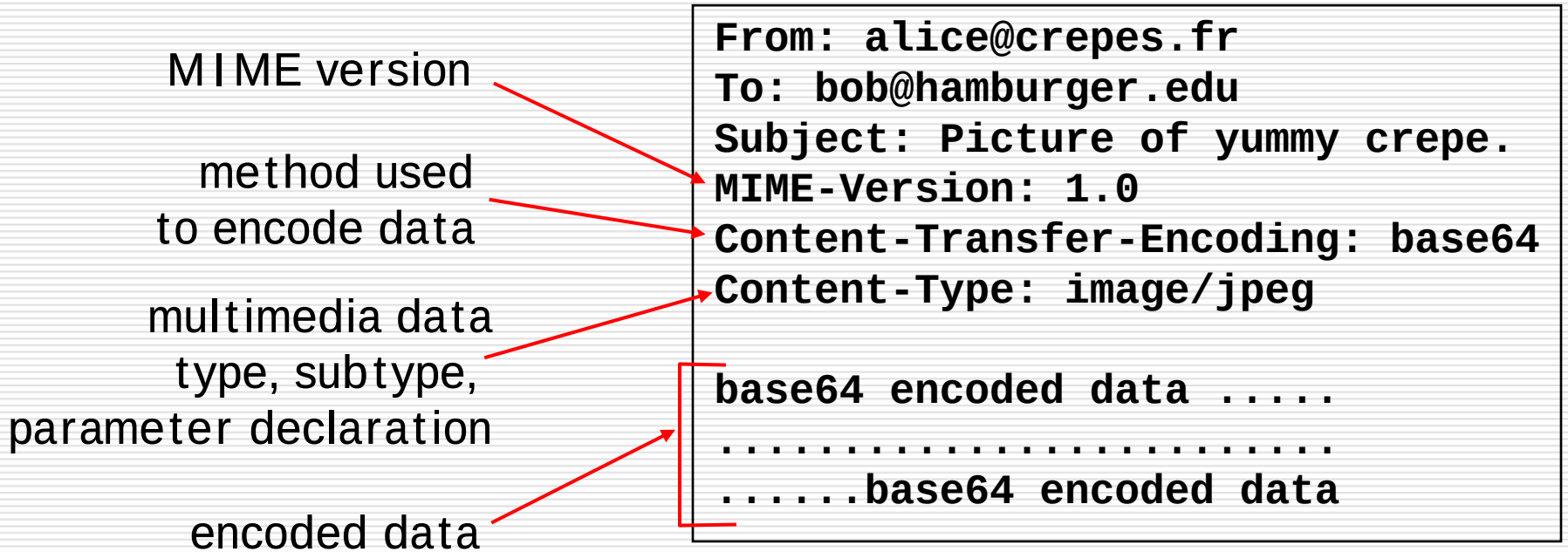
```
S: 220 hamburger.edu
C: HELO crepes.fr
S: 250  Hello crepes.fr, pleased to meet you
C: MAIL FROM: <alice@crepes.fr>
S: 250 alice@crepes.fr... Sender ok
C: RCPT TO: <bob@hamburger.edu>
S: 250 bob@hamburger.edu ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: Do you like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

# SMTP: final words

- SMTP uses persistent connections
- SMTP requires message (header & body) to be in 7-bit ASCII
- certain character strings not permitted in msg (e.g., `CRLF.CRLF`). Thus msg has to be encoded (usually into either base-64 or quoted printable)
- SMTP server uses `CRLF.CRLF` to determine end of message

# Multimedia Extensions

- MIME: multimedia mail extension, RFC 2045, 2056
- additional lines in msg header declare MIME content type



# Multipart Type

```
From: alice@crepes.fr
To: bob@hamburger.edu
Subject: Picture of yummy crepe.
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary=98766789

--98766789
Content-Transfer-Encoding: quoted-printable
Content-Type: text/plain

Dear Bob,
Please find a picture of a crepe.
--98766789
Content-Transfer-Encoding: base64
Content-Type: image/jpeg

base64 encoded data .....
.........................
......base64 encoded data
--98766789--
```

# MIME headers

- MIME specifies 5 new e-mail header fields:
  - **`MIME-Version`** (must be 1.0)
  - **`Content-Type`**
  - **`Content-Transfer-Encoding`**
  - **`Content-ID`** - optional
  - **`Content-Description`** - optional

# MIME Content-Type

- Seven major content types with 15 sub-types.
- `Multipart` content type has 4 subtypes.
- Most important is `Multipart/mixed`, indicating that the body contains multiple parts.
- Each part can be a separate MIME message – hence nesting of MIME messages to any level.
- Parts separated by a boundary string defined in `Content-Type` field.

# Content-Transfer-Encoding

- RFC 822 e-mails can contain only ASCII characters.
- MIME messages intended to transport arbitrary data.
- The **Content-Transfer-Encoding** field indicates how data was encoded from raw data to ASCII.
- **base64** is a common encoding:
  - 24 data bits (3 bytes) at a time encoded to 4 ASCII characters.

# What is Spam?

- Typical legal definition: unsolicited commercial email from someone without a pre-existing business relationship.

# Kinds of Spam

- Email spam (you already know about that)
- Newsgroup spam
- Chat rooms
- Popups
  - Web pages
  - Spyware
- Search engine spam
- Conclusion:
  If you can advertise for free, someone will

# Newsgroup Spam

- Just like email spam
- Cheaper for spammers – post once download many times
- Not big issue because … no one uses newsgroups
  - That's because they are all full of spam
  - Could the same thing happen to email?!

# Chat Room Spam

- Spambots come in and pretend to chat
  - But really just advertising porn sites
  - Some spambots trivial
    - Don't talk at all, but take up space
    - Link to porn spam in their profile
  - Some spambots very sophisticated
    - You can have a short conversation with them before they try to convince you to go to their website

# Popup Spam

- Web page popups
  - You go to a web page, and get a popup
  - May be a “pop under” that appears under all other windows, so you don’t even know where it came from
- Spyware (e.g. Gator)
  - Software installed on your computer either without your permission, or where permission is hidden deep in license agreement.
  - Creates popups all the time

# Search Engine Spam

- Link spam
  - Search engines use number of links to determine rankings
  - Spammers create millions of pages that link to their site
  - Fake pages may be realistic and may be returned as search results, too.
- Word spam
  - Spammers put misleading words on their page, e.g. celebrity names or technical terms
  - Page is actual porn
- Blog spam
  - Some web pages let anyone post comments
  - Spammers automate comment posting, add links to their pages

# Hence …

- Anywhere you can place free advertising, someone will
- Even after we solve email spam, there will be lots more to work on
- Some techniques from email spam transfer, some don't

# Techniques that Spammers Use

- Obscuring mail to avoid filters
  - Content in Images, Good word Chaff, Content Chaff, URL Spamming, Hidden Text, Character Encoding, etc…
- Getting email addresses
  - Dictionary attacks
  - Web Crawling
- Sending spam
  - Open proxies
  - Open relays
  - Zombies

# Obscuring mail

- Modify spam to avoid filters
  - Content in Images
  - Good word Chaff
  - Content Chaff
  - URL Spamming
  - Hidden Text
  - Character Encoding
  - Etc…

# Weather Report Guy

- Content in Image
- Good Word Chaff

Weather, Sunny, High 82, Low 81, Favorite…


hommy1 info
mojito ja kägu.

**Weather**

NA, NA - Sunny

High: 82 , Low: 81 degrees

**Favorites**

# Diploma Guy

- Word Obscuring

DpImoia Pragorm

Caer te a mroe prosoeprus

**Dlpmoia Pragorm**

Caerte a mroe prosoeprus frtuue for yorsuelf

Recieve a flul doplima from non areceditcd unvtersiiies baesd upon your rael lfie exneriepce

You will not be tseted, or irteeviewnd
Receive a Msetar's, Bachelor's or Dtctarooe

Clal 24 horus a day 7 days a week

1 - 2 7 0 - 8 1 7 - 8 2 4 7

# Trends in Spam Exploits

- Based on 1,200 spam messages sent to Hotmail

| Exploit | 2003 Spam | 2004 Spam | Delta (Absolute %) | Description |
|---|---|---|---|---|
| Word Obscuring | 4% | 20% | 16% | Misspelling words, putting words into images, etc. |
| URL Spamming | 0% | 10% | 10% | Adding URLs to non-spam sites (e.g. msn.com). |
| Domain Spoofing | 41% | 50% | 9% | Using an invalid or fake domain in the from line. |
| Token Breaking | 7% | 15% | 8% | Breaking words with punctuation, space, etc. |
| MIME Attacks | 5% | 11% | 6% | Putting non-spam content in one body part and spam content in another. |
| Text Chaff | 52% | 56% | 4% | Random strings of characters, random series or words, or unrelated sentences. |
| URL Obscuring | 22% | 17% | -5% | Encoding a URL in hexadecimal, hiding the true URL with an @ sign, etc. |
| Character Encoding | 5% | 0% | -5% | Pharmacy renders into Pharmacy. |

# Getting Email Addresses

- Dictionary Attacks:
  - Try millions or billions of possible email addresses
    - Put together first-name and last-name, or first-name + number, etc.
    - See if the mail "bounces": if not, you have a live address
    - Use "**Web beacons**" to check if mail is being read
- Web crawling:
  - Look for email addresses on web pages
- Send spam to these addresses or sell them to other spammers

# Sending Spam -- Open Relays

- In the old days, before the fully connected internet, it was “nice” to forward mail meant for someone else.
- A mail server is an “open relay” if it will forward on behalf of anyone.
- Spammers love open relays
  - A little harder to trace them
  - Shifts bandwidth and other burdens to someone else
- But open relays are added to **blackhole** lists, and quickly cannot send legitimate mail

# Sending Spam -- Open Proxies

- These are web-page proxy servers
  - Used for getting web-pages past firewalls
  - Should have nothing to do with email
- Spammers exploit holes
  - Exploit a hole that you can use some proxies to send email
  - Exploit another hole that anyone can access the proxy-server
- Spammers really love these
  - Almost impossible to trace spammer
  - Less incentive for owner to close the proxy than to close open mail relays: they don't care if their web proxy is on an email blackhole list

# Sending Spam -- Zombies

- Consumer computers taken over by viruses or trojans
  - Spammer tells them what to send
  - Very difficult to trace
  - Very cheap for spammer
  - Rent a zombie for about $3/month!
- As much as 40-60% of spam may originate from zombies now!

# Solutions to Spam

- Filtering
  - Matching/Fuzzy Hashing
  - Blackhole Lists (IP addresses)
- Postage
  - Turing Tests, Money, Computation
- Safe sender lists (Bonded Sender)
- Disposable Email Addresses
- Habeas Haiku
- Email authentication

# Filtering -- Machine Learning

- Learn spam versus good (ham)
- Problem: need source of training data
  - Get users to volunteer GOOD and SPAM
- But spammers are adapting to machine learning too
  - Images, different words, misspellings, etc.
- We use machine learning – details later

# Matching / Fuzzy Hashing

- Automatically get examples of known spam
  - Use “Honeypots” Use “Report Junk” button data
- Look for similar messages that arrive in real mailboxes
  - Exact match easily defeated
  - Use fuzzy hashes
    - How effective?
- The Chinese menu (madlibs) attack will defeat any exact match filters or fuzzy hashing

**Make thousands of dollars working at home !!!**

**Earn lots of money in the comfort of your own house .**

# Rule-based Filtering

- Keyword-based filtering
  - If **SUBJECT** contains
    "FREE", "TEENS", "Hello!!!"……
- How about "!!!!!!!!!hello!!!!hello!!!hello"?
- Miss-spelling: Porn → P0rn
- Strength
  - Easy to implement
  - Linear computational cost to # of rules
- Weakness
  - Low precision
  - Adding thousands of rules is a tedious work

# Challenge-Response Filtering

- Email from someone NOT on your whitelist, an automatic reply is sent telling what steps the sender should take to be considered for the whitelist
  - (e.g. send you a confirmation, make a donation, solve a puzzle, etc.)
- Very effective at stopping spam BUT has a number of drawbacks:
  - valid mail delayed
  - kind of harsh -- some may think of it as inconsiderate and never reply, extra work for senders etc.

# Content-based Filtering

- What is the signature in spam?
- People can find that 'join today', 'FREE' are good indicators for spam! But this doesn’t mean mail containing these words are definitely spam, but a high PROBABILITY they are!

# Naïve Bayes

- Basically, want $P(spam \mid words)$
- Use Bayes Rule:

$$P(spam \mid words) = \frac{P(words \mid spam) \times P(spam)}{P(words)}$$

$$P(words) = P(words \mid spam) \times P(spam) + P(words \mid good) \times P(good)$$

- Assume independence: probability of each word independent of others (wrong assumption)

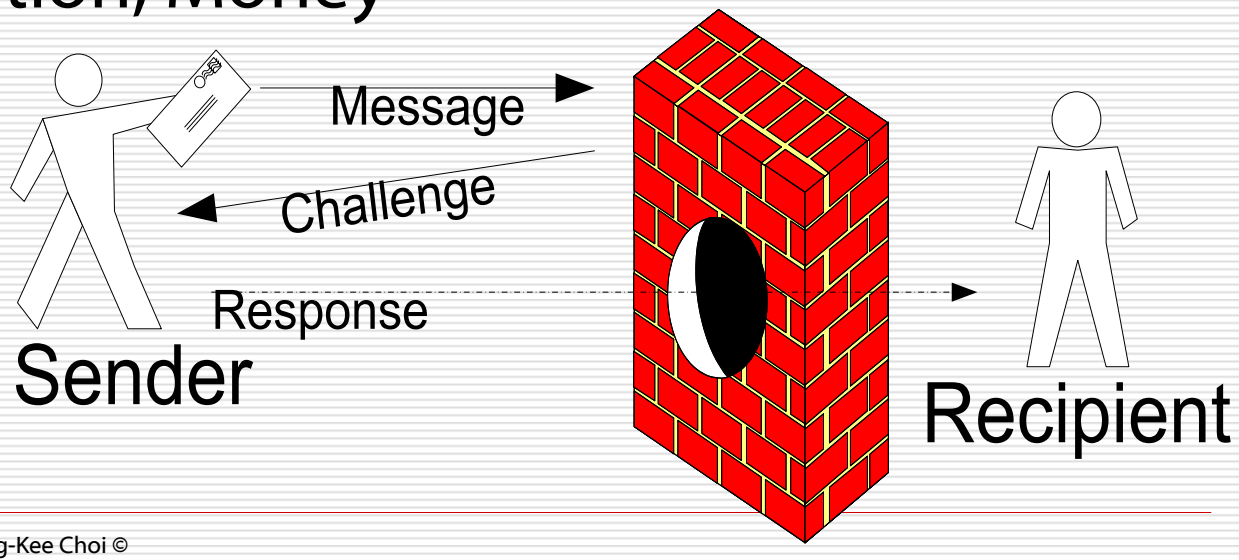
$$P(words \mid spam) \approx P(word1 \mid spam) \times P(word2 \mid spam) \times \ldots \times P(wordn \mid spam)$$

# Blackhole Lists

- Lists of IP addresses that send spam
  - Open relays, Open proxies, DSL/Cable lines, etc…
- Easy to make mistakes
  - Open relays, DSL, Cable send good and spam…
- Who makes the lists?
  - Some list-makers very aggressive
  - Some list-makers too slow

# Postage

- Basic problem with email is that it is free
  - Force everyone to pay (especially spammers) and spam goes away
  - Send payment pre-emptively, with each outbound message, or wait for challenge
- Multiple kinds of payment: Turing Test, Computation, Money



# Turing Tests (Naor '96)

- You send me mail; I don't know you
- I send you a challenge:
  `type these letters`


- Your response is sent to my computer
- Your message is moved to my inbox if correct.

# Money

- Pay actual money (1 cent?) to send a message
- My favorite variation: take money only when user hits “Report Spam” button
  - Otherwise, refund to sender
  - Free for non-spammers to send mail, but expensive for spammers
- Requires multiple monetary transactions for every message sent – expensive
- Who pays for infrastructure?

# Safe Sender Lists

- Block lists are hard
  - People get very angry when you call them a spammer
  - There are 2^32 IP addresses
  - Hard to know when to stop blocking
- Instead, make lists of known good senders
  - Amazon pays someone, e.g. Bonded Sender, to be added to list of known good senders
  - Users download or check Bonded Sender list
  - No one can get angry
  - Relatively small number
  - Sender can be given incentives to tell you when he changes IP addresses or use non-IP based methods

# Bonded Sender

- Program developed by IronPort systems, now working with TrustE
- List safe IP addresses
- Senders must “post a bond” – deposit money with Bonded Sender
- Portion of bond is confiscated based on complaints
  - However, some users make mistakes, so you get some complaints “for free.”
- Need to monitor volume rate (to compute complaint rate) and so that a spammer cannot post a small bond and then send billions of messages quickly
  - Record number of queries about each sender

# Habeas' Haiku

- Sender warranted email™

**The Habeas Warrant Mark**

```
X-Habeas-SWE-1: winter into spring
X-Habeas-SWE-2: brightly anticipated
X-Habeas-SWE-3: like Habeas SWE (tm)
X-Habeas-SWE-4: Copyright 2002 Habeas (tm)
X-Habeas-SWE-5: Sender Warranted Email (SWE) (tm). The sender of this
X-Habeas-SWE-6: email in exchange for a license for this Habeas
X-Habeas-SWE-7: warrant mark warrants that this is a Habeas Compliant
X-Habeas-SWE-8: Message (HCM) and not spam. Please report use of this
X-Habeas-SWE-9: mark in spam to <http://www.habeas.com/report/>.
```

- Vigilant to prevent spammers from misusing The Habeas Warrant Mark.
- Once a breach discovered, unlicensed IP addresses will be placed on Habeas blacklist.

# Disposable Email Addresses

- Also called Ephemeral Addresses
- You have one address for each sender
  - **JOSHUAGO1895422@microsoft.com**
  - All go to same mailbox
- If I give you my address, and you send me spam, I just delete the address
- How do new senders get an address?
- If I send mail to 3 people, which address is it From?
- Hard to remember!

# Miscellaneous

- Email masking
  - Hide email address from automatic email bot
- Limit issuing of email account
  - Must  involved HIP (Human Interactive Proof)
- Outbound control
  - Set max. send of mail per day
  - Limit interval between sending mails
- Mail server vulnerability check
  - Look for open relay, open proxy and zombie

# Email Authentication

- Traditional Email authentication
  - SMTP Auth and STARTTLS
  - Designed to restrict who can connect to a mail server
  - Closed system
- Traditional Email singing
  - S/MIME and PGP
  - Designed to cryptographically prove the individual author and content of a message
  - Secure but heavy
  - Requires enabled clients

# Domain Authentication

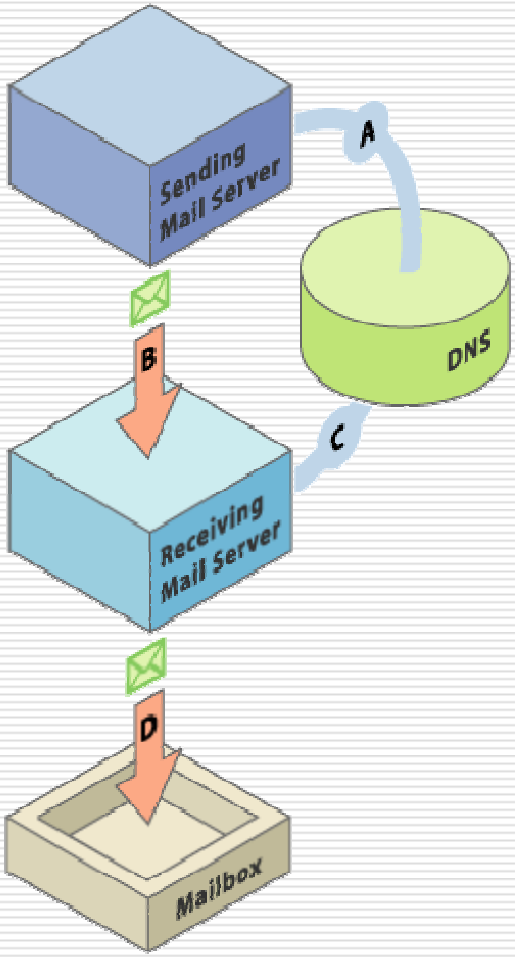
- Emerging standards designed to authenticate domains from which an email was sent
  - e.g. SenderID, DomainKeys
- Lighter weight than traditional email signing mechanism
  - Not required enabled clients
  - Only sending and receiving MTAs need to be aware
- Not break backward compatibility
- Not require a “flag day”

# DomainKeys by Yahoo

- Email spoofing
  - Forge another person's email to get users to trust
  - Without sender authentication, difficult to know for certain if message is legitimate or forged
- DomainKeys provides mechanism for verifying domain of email and message integrity
- Once domain can be verified, compare against domain in `From:` field in the header
- MTA such as qmail start to implement
- gmail started to implement DomainKeys
- Similar to Cisco's IIM (Identified Internet Email)

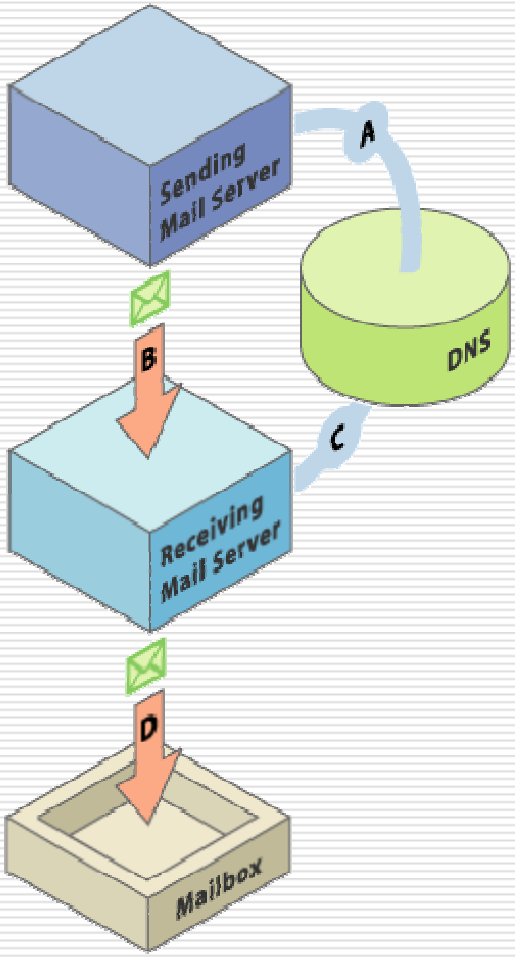
# DomainKeys – Sending servers

- Domain owner setup a public/private key pair.
  - This public key published in DNS (step A)
- Authorized end-users in domain send email email system generates digital signature using domain's private key
  - Signature is prepend to the email header as shown in step B



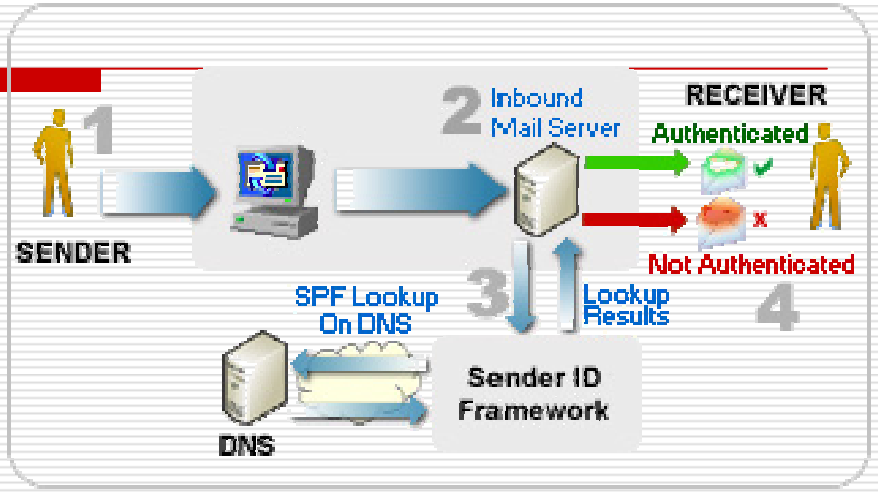
# DomainKeys – Receiving servers

- Extract signature and claimed `From:` field in the header and fetch public key from DNS
  - step C
- Verify signature
- This proves that
  - email was truly sent by claimed From: field
  - Not altered during tranfer
- Deliver
  - Step D



# SPF and senderID

1. Sender sends an e-mail message to Receiver.
2. Receiver's inbound mail server receives mail.




3. Receiver's server checks for SPF (Sender Policy Framework) record of the sending domain published in DNS.
   - ▷ Sending domain founded in `From:` field in the header
4. Inbound e-mail server determines if sending e-mail server's IP address matches the IP address that is published in DNS.

# Sender Authentication

- A way of "closing" your inbox to people you don't know
  - Similar to Instant Messaging
- Probably longer-term due to MUA changes, but paradigm shifts

# Email Authentication Summary

- Allows a Sender to prove they're who they claim to be
- Prevents Spammers/Phishers from pretending to be someone else
  - Anyone could have claimed to be `security@ebay.com`
- Enables more meaningful **Accreditation / Reputation**

# Conclusion

- Lots of different solutions
- Often, they work best in combination
  - Example: combine machine learning filters with postage or use machine learning filters and HIPs to help stop outbound spam
  - No single discipline can solve this
- Final solution will combine approaches
  - Safe sender lists and postage help ensure all good mail gets through
  - Filters, blackhole lists stop mail from unknown or known bad senders