

Don Porter

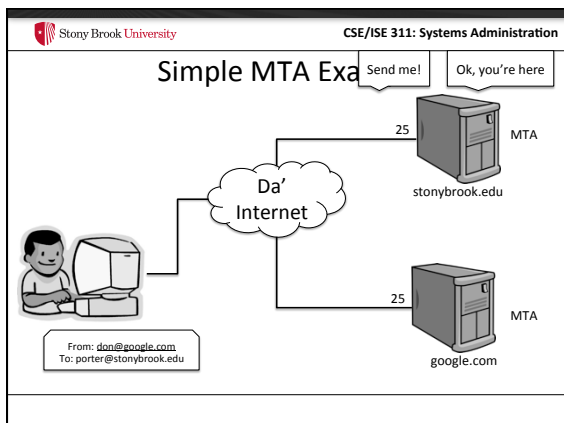
- Email systems started with a pretty simple design
  - Everyone trusted each other, it was nice
- Then spam came along
  - Lots of complexity and distrust to try to reduce spam

- A simple message format:
  - Envelope (server-internal routing info, not user-visible)
  - Headers (basically the history of the message handling)
    - These are viewable in your email app
  - Body (the text you see in your email app)

[illegible]

- Accepts emails from the Internet
  - Delivers to local users
  - Or sends outgoing messages
- Every site that accepts email runs one
  - Identified with an MX record in DNS
- Listens on Port 25

The diagram shows a user at a computer connected to a cloud labeled "Da' Internet". Two servers, labeled "stonybrook.edu" and "google.com", are also connected to the cloud. The user is sending an email to "porter@stonybrook.edu". The servers are labeled "MTA" (Mail Transfer Agent). The user's email client shows "From: don@google.com" and "To: porter@stonybrook.edu". The servers have buttons labeled "Send me!" and "Ok".



Stony Brook University CSE/ISE 311: Systems Administration

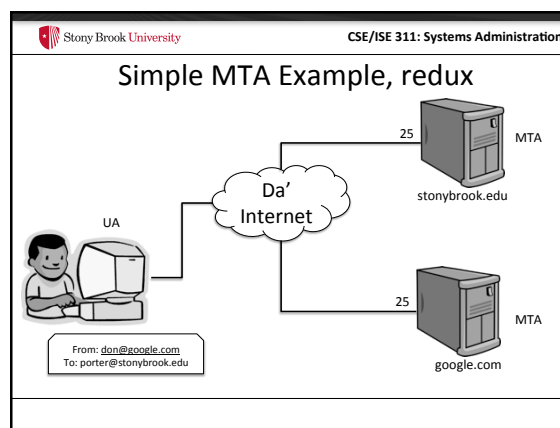
### SMTP

- Simple Mail Transport Protocol
- It really is simple.
- Main operations:
  - Send a message
  - Check if an address is valid
  - Expand an address (for lists and forwarding)
- Email basically works by lots of MTA servers passing messages to each other

Stony Brook University CSE/ISE 311: Systems Administration

### So what is Apple Mail, or Outlook?

- A User Agent (UA)
- Usually provides a program to type emails
- Ultimately packaged and sent to an MTA using SMTP



Stony Brook University CSE/ISE 311: Systems Administration

### So how does email get to my inbox?

- Once an email reaches its destination MTA,
  - Handed off to a mail delivery agent (MDA)
  - MDA can be the same program or a different daemon as MTA
- MDA finds inbox for the recipient and stores the email

Stony Brook University CSE/ISE 311: Systems Administration

### Where is my inbox?

- On Unix/Linux, it is either a file or directory
- 2 Popular formats:
  - mbox – single file
  - maildir – directory with one file per-message
- Literally, a file like: /home/porter/mail/mbox
  - Each message has a special delimiter between it
- Maybe shared with other machines over NFS
  - Users can get their mail without logging into MDA machine

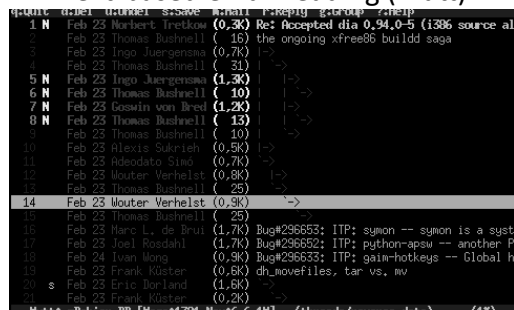
Stony Brook University CSE/ISE 311: Systems Administration

## Old-School Mail Reading

- Programs like mutt, pine, elm would read these mail files directly on a server
  - You could even open them in vi or emacs!

Stony Brook University CSE/ISE 311: Systems Administration

## Text-based email reading (mutt)



From: <http://greek0.net/mutt.html>

Stony Brook University CSE/ISE 311: Systems Administration

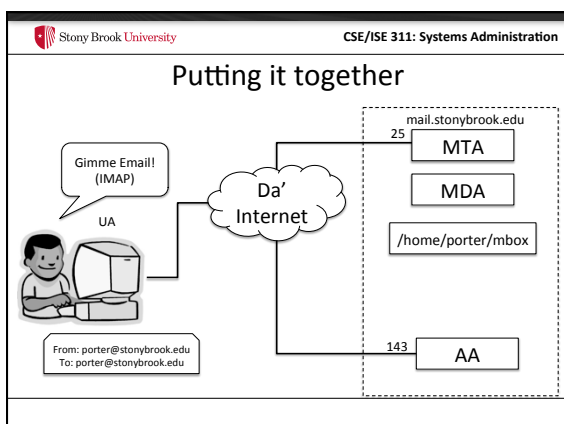
## Pointy-Clicky Mail Reading

- Not everyone runs a command line on a Linux server
- Mail for the rest of us?
  - 2 popular protocols: POP and IMAP
    - Make a nice GUI app, like Thunderbird or Apple Mail
    - Download inbox using POP or IMAP
    - Still send using SMTP
- Access Agent (AA) – serves POP or IMAP

Stony Brook University CSE/ISE 311: Systems Administration

## POP vs. IMAP

- Post Office Protocol (POP)
  - Download everything and (usually) delete it from inbox on server
  - Designed when people had one PC
- Internet Mail Access Protocol (IMAP)
  - Sync a local copy of mailbox with server
    - Work offline and sync later
  - Multiple clients
    - Laptop, desktop, iphone, ipad
  - Better choice for most modern users



Stony Brook University CSE/ISE 311: Systems Administration

## Recap so far...

- User Agent: Program for human user
- Sending mail?
  - Protocol: SMTP
  - Emails exchanged by MTAs
  - Delivered to inboxes by MDAs
  - Inboxes are just simple files (mostly)
- Receiving mail?
  - Read inbox through an AA
  - Protocols: POP or IMAP

Stony Brook University CSE/ISE 311: Systems Administration

## SMTP Review

- Main operations:
  - Send a message
  - Check if an address is valid
  - Expand an address (for lists and forwarding)
- **Anything missing?**

Stony Brook University CSE/ISE 311: Systems Administration

## Are my emails private?

- No
- Sent in plain text over internet, inbox in plain text
- If you want privacy, use encryption in your UA
  - PGP/GPG are good choices


Stony Brook University CSE/ISE 311: Systems Administration

## How do I know email came from sender?

- You don't
- Anyone can put any name/email address in the 'from' field
  - Of course, replies may not go to you...
- This is why people who care use PGP/GPG to *sign* messages

Stony Brook University CSE/ISE 311: Systems Administration

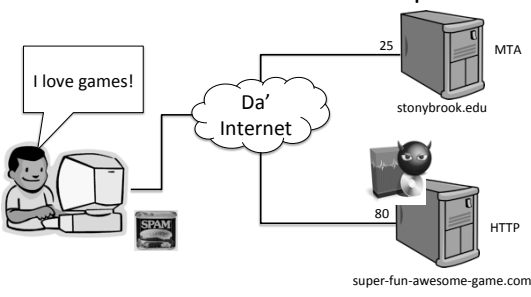
## Spam



- Junk email
  - Unwanted ads, harassing emails, etc.
  - Often selling illegal products
- Why is it called spam?
- A massive nuisance
  - But also a massive business
- A lot of modern system organization designed to limit/prevent spam
  - We will refine previous model of email to cope with spam

Stony Brook University CSE/ISE 311: Systems Administration

## An all-too-common example



Stony Brook University CSE/ISE 311: Systems Administration

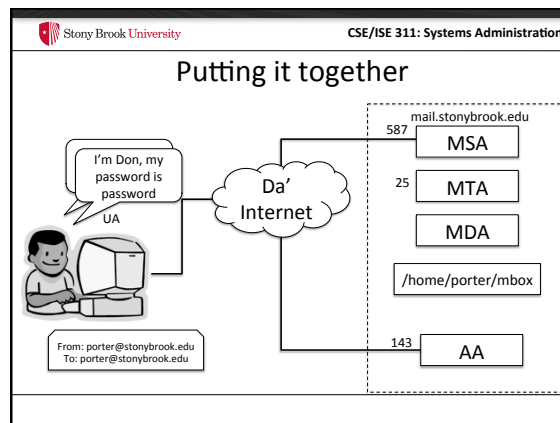
## Example Recap

- User downloads software
  - Software includes "trojan horse" malware
  - Malware connects to email servers and sends spam from user's computer over SMTP
  - Other SMTP servers then relay the email throughout the internet
- Ideas how to fix this?

Stony Brook University CSE/ISE 311: Systems Administration

### Refinement 1

- SMTP servers (as presented) are dumb
  - Accept and relay mail from anyone!
- Idea: Only send mail from authenticated users
- Mail Submission Agent (MSA)
  - Basically, provides a log-in system, and then forwards mail to MTA
  - MTA configured to only relay mail from MSA
  - MSA often listens on port 587



Stony Brook University CSE/ISE 311: Systems Administration

### Refinement, recap

- In the “good old days”, any email server would send your mail for you
  - Made email (and spam) easy
- Now, only servers that know you will relay email for you
  - Spam program at least has to steal your account info now

Stony Brook University CSE/ISE 311: Systems Administration

### Problem 2

- The first example was about *relaying*
  - Using stonybrook.edu to relay email to lots of other addresses (gmail, hotmail, etc)
  - Fixed in part by:
    - Only accepting new email for legitimate users in stonybrook.edu
    - Only relaying email from authenticated stonybrook.edu users
- What about spam for users of stonybrook.edu?

Stony Brook University CSE/ISE 311: Systems Administration


### Refinement 2


- Be choosier about who you accept email from
- Whitelists: known-good email servers (accept)
- Blacklists: known-bad email servers (reject)
  - Services that provide both for administrators
- Unknown servers? (Gray lists)
  - Reject: May lose email
  - Delay: Spammers often impatient


Stony Brook University CSE/ISE 311: Systems Administration

### Refinement 3: Spam Detection

- Before the MDA delivers spam, run it through a filter
- If it passes, deliver to inbox
- If it fails, put in Junk folder
- How do spam filters work?

 Stony Brook University	CSE/ISE 311: Systems Administration
<h2>2 Strategies for Spam Filters</h2> <ul style="list-style-type: none"> <li>• Fuzzy matching against known spam <ul style="list-style-type: none"> <li>– Verbatim matches foiled by including time of day</li> <li>– Services track these things for admins</li> </ul> </li> <li>• Bayesian learning filters <ul style="list-style-type: none"> <li>– Users mark things as 'spam' or 'ham'</li> <li>– Basically training a fancy-dancy machine learning classifier</li> <li>– Classifier applied to new mail</li> <li>– Learns user preferences over time</li> </ul> </li> </ul>	

 Stony Brook University	CSE/ISE 311: Systems Administration
<h2>2 Phase Spam Prevention</h2> <ul style="list-style-type: none"> <li>• My advice (and the book's): <ul style="list-style-type: none"> <li>– Do quick checks at the MTA (white list, black list)</li> <li>– Then do detailed checks (spam filtering) in MDA</li> </ul> </li> <li>• Why? <ul style="list-style-type: none"> <li>– Quickly drop obvious garbage</li> <li>– Shift load for heavier-weight scanning to MDA</li> </ul> </li> </ul>	

 Stony Brook University	CSE/ISE 311: Systems Administration
<h2>Summary</h2> <ul style="list-style-type: none"> <li>• MTAs exchange mail all over internet <ul style="list-style-type: none"> <li>– Only relay outgoing mail from MSA (to prevent spam)</li> <li>– Only accept incoming mail from white/gray list (to prevent spam)</li> </ul> </li> <li>• MDA delivers mail <ul style="list-style-type: none"> <li>– After it passes a filter (to prevent spam)</li> <li>– Stores it in a simple inbox file</li> </ul> </li> <li>• UA gets mail using IMAP/POP <ul style="list-style-type: none"> <li>– Sends it via MSA using SMTP</li> </ul> </li> </ul>	