

Date: Wed, 2 Mar 2016

From: "Matt Arnold" <mpa29@cam.ac.uk>

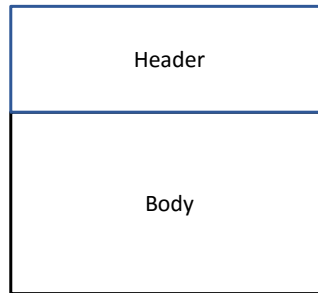
## Subject: Electronic Mail

# Electronic Mail

1. What does an email look like?
  - Some rules for ASCII email message format
  - Multipurpose Internet Mail Extensions (MIME)
2. How does it get sent?
  - Journey of an email
  - Simple Mail Transfer Protocol (SMTP)
  - Internet Message Access Protocol (IMAP)
3. Cambridge email
  - Development and current state of our system
  - Load balancing and protection
  - What about other unis?

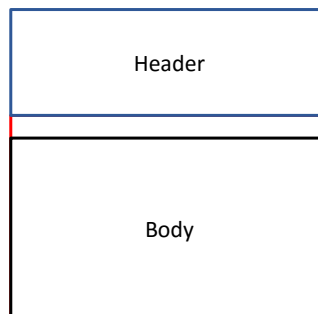
## What does an email look like?

- Messages must have two parts:
  1. Header
  2. Body
- ASCII-only



## What does an email look like?

- Messages must have two parts:
  1. Header
  2. Body
- ASCII-only



Email Composition      The Journey      Cambridge Email

# Header

- <CR><LF> terminated lines

```

sdj61VgDXzP8YyU7G6jyUvSvUmpzks+
Received: from [10.3.8.191] (heliovalvesoftware.com)
  by smtp08.steampowered.com with smtp (Exim 4.82)
  (envelope-from <noreply@steampowered.com>)
  id 1aWLSH-0003Nu-89
  for matux96@gmail.com; Thu, 21 Jan 2016 11:40:21 -0800
To: matux96@gmail.com
From: Steam Store <noreply@steampowered.com>
Reply-To: <noreply@steampowered.com>
Errors-To: <noreply@steampowered.com>
X-Steam-Message-Type: CEmailWishlistItemOnSale
X-DHE-Version: 1.0
Subject: An item on your Steam wishlist is on sale!
Content-Type: multipart/alternative;
  boundary="np56a134253699f"
Message-Id: <1aWLSH-0003Nu-89@smtp08.steampowered.com>
Date: Thu, 21 Jan 2016 11:40:21 -0800
  
```

Field: value

● 4/34

Email Composition      The Journey      Cambridge Email

# Header

- <CR><LF> terminated lines
- Examples of header fields:
  - To, From, Date, Subject, Cc, Bcc, Reply-To
  - Message-ID
  - In-Reply-To
  - Received

● 4/34

# Header

- <CR><LF> terminated lines
- Examples of header fields: To, From, Date, Subject, Cc, Bcc, Reply-To, Message-ID, In-Reply-To, Received
- Line length limits
- Folding

**Subject:** This is a tes

# Header

- <CR><LF> terminated lines
- Examples of header fields: To, From, Date, Subject, Cc, Bcc, Reply-To, Message-ID, In-Reply-To, Received
- Line length limits
- Folding

**Subject:** This  
is a test

## Header



- <CR><LF> terminated lines
- Examples of header fields: To, From, Date, Subject, Cc, Bcc, Reply-To, Message-ID, In-Reply-To, Received
- Line length limits
- Folding

**Subject:** This  
is a test

## MIME

- 1993, Multipurpose Internet Mail Extensions
- Allows more interesting content without breaking rules
- New header fields:
  - MIME-Version

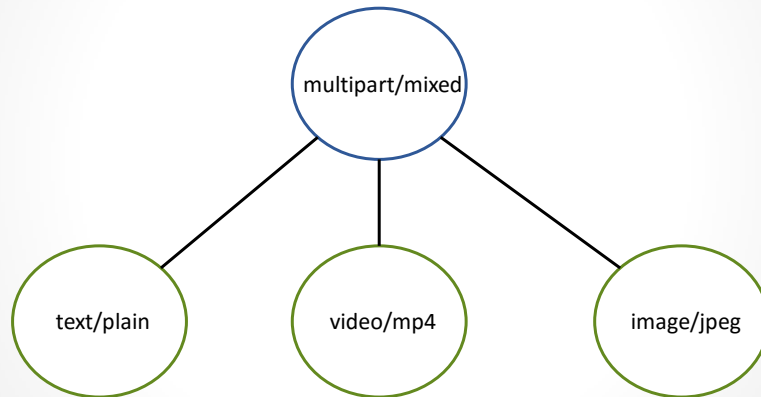
# MIME

- 1993, Multipurpose Internet Mail Extensions
- Allows more interesting content without breaking rules
- New header fields:
  - `MIME-Version: 1.0`
  - `Content-Type`

## MIME: Content-Type

- `text/plain`, `text/html`
- `image/jpeg`, `image/gif`, `audio/mp3`, `video/mp4`
- `application/pdf`
- `multipart/mixed`

## MIME: Content-Type



## MIME: Content-Type

- text/plain, text/html
- image/jpeg, image/gif, audio/mp3, video/mp4
- application/pdf
- multipart/mixed
- multipart/alternative

# MIME

- 1993, Multipurpose Internet Mail Extensions
- Allows more interesting content without breaking rules
- New header fields:
  - `MIME-Version: 1.0`
  - `Content-Type`
  - `Content-Transfer-Encoding`
    - 7bit
    - quoted-printable
    - base64

# base64

00100111 01001000 01100111 1000



Email Composition

The Journey

Cambridge Email

# base64

00100111 01001000 01100111

•

● 10/34

Email Composition

The Journey

Cambridge Email

# base64

00100111 01001000 01100111

•

● 10/34

# base64

001001 110100 100100 100111

0: A	8: I	16: Q	24: Y	32: g	40: o	48: w	56: 4
1: B	9: J	17: R	25: Z	33: h	41: p	49: x	57: 5
2: C	10: K	18: S	26: a	34: i	42: q	50: y	58: 6
3: D	11: L	19: T	27: b	35: j	43: r	51: z	59: 7
4: E	12: M	20: U	28: c	36: k	44: s	52: 0	60: 8
5: F	13: N	21: V	29: d	37: l	45: t	53: 1	61: 9
6: G	14: O	22: W	30: e	38: m	46: u	54: 2	62: +
7: H	15: P	23: X	31: f	39: n	47: v	55: 3	63: /

# MIME



- 1993, Multipurpose Internet Mail Extensions
- Allows more interesting content without breaking rules
- New header fields:
  - **MIME-Version:** 1.0
  - **Content-Type**
  - **Content-Transfer-Encoding**
    - 7bit
    - quoted-printable
    - base64
  - **Content-Disposition**
- Example message...

## Encoded Words

- Of the form:

**=?charset?encoding?encoded text?=?**

- Maximum of 75 characters
- Intended for header values
- For example:

To: "=?utf-8?B?TS5QLiBBcm5vbGQ=?=" <mpa29@cam.ac.uk>

↑  
base64

↙  
"M. P. Arnold"

## Journey of an email

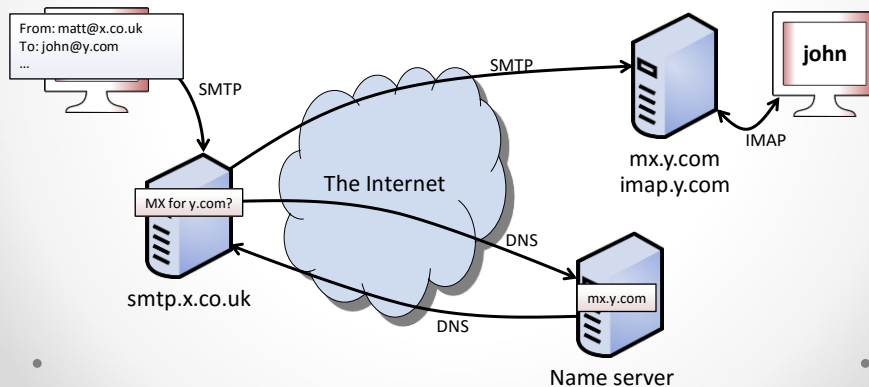
## Some TLAs

- MUA (Mail User Agent)
- MTA (Mail Transfer Agent)
- MSA (Mail Submission Agent)
- MDA (Mail Delivery Agent)
- SMTP – Simple Mail Transfer Protocol
- POP3 – Post Office Protocol
- IMAP4 – Internet Message Access Protocol

● 14/34

## Journey of an email

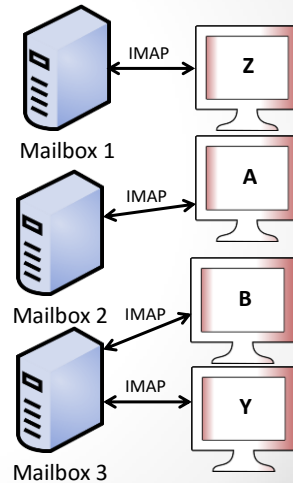
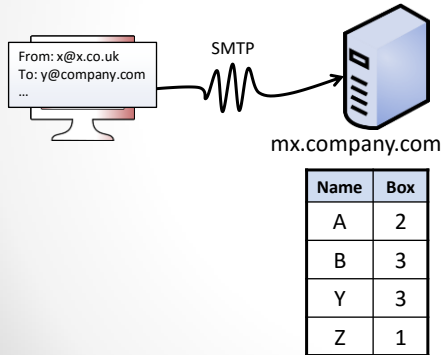
1. MUA uses SMTP on port 587 to submit email to MSA
2. Target host(s) is/are located using the DNS
3. MTA sends email to target MX server using SMTP on port 25
4. MX server gives the email to target's MDA
5. Authenticated MUAs pick up the email from the mailbox using (usually) either POP or IMAP



● 15/34

# Why multiple gateways?

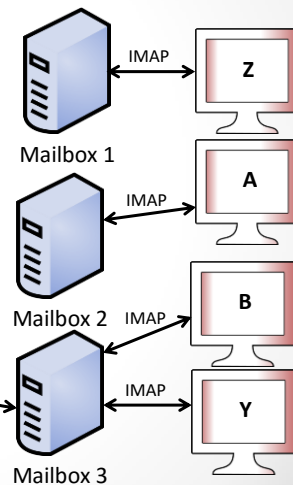
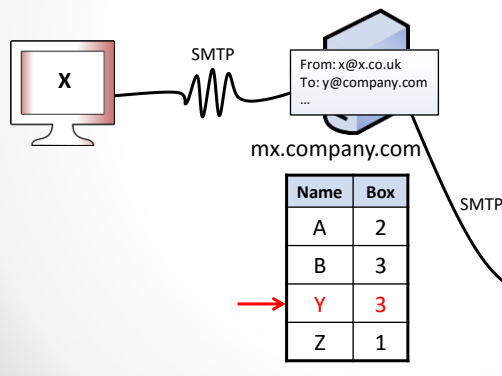
- Hiding recipient's machine
- Scale



● 16/34

# Why multiple gateways?

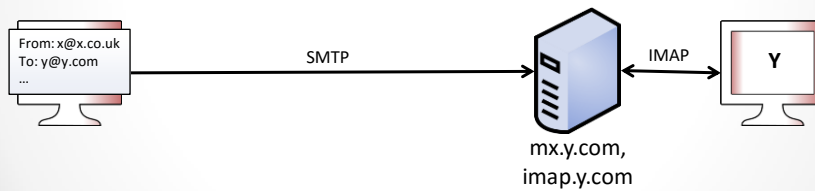
- Hiding recipient's machine
- Scale



● 16/34

# Why multiple gateways?

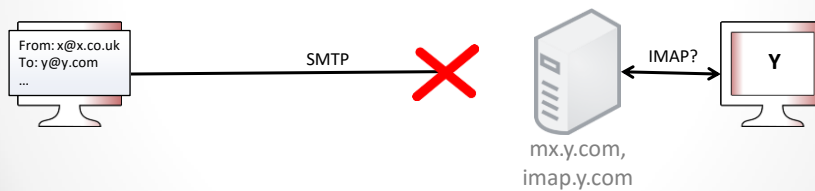
- Hiding recipient's machine
- Scale
- Target may not always be up or reachable



● 16/34

# Why multiple gateways?

- Hiding recipient's machine
- Scale
- Target may not always be up or reachable



● 16/34

## Why multiple gateways?

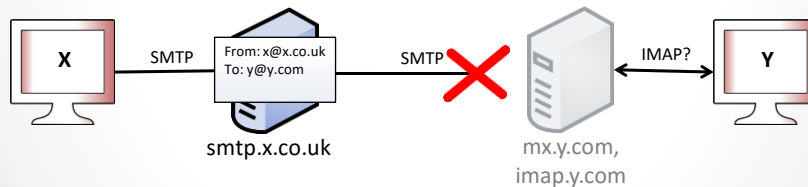
- Hiding recipient's machine
- Scale
- Target may not always be up or reachable



● 16/34

## Why multiple gateways?

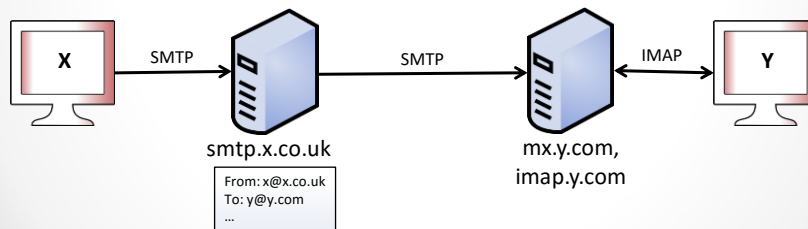
- Hiding recipient's machine
- Scale
- Target may not always be up or reachable



● 16/34

## Why multiple gateways?

- Hiding recipient's machine
- Scale
- Target may not always be up or reachable



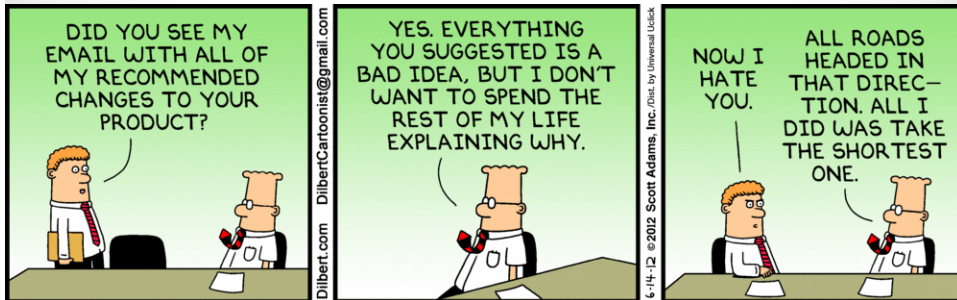
● 16/34

## Bounces

- Two categories of email bounces:
  - Soft – usually temporary, will retry
    - Temporarily unreachable
    - Mailbox full
  - Hard – not temporary, give up
    - Typo in email address
    - Sender address blocked

● 17/34





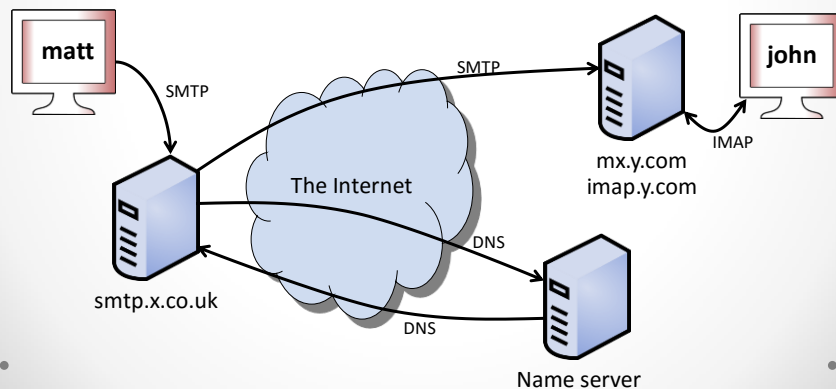
Email Composition

The Journey

Cambridge Email

## Journey of an email

1. MUA uses SMTP on port 587 to submit email to MSA
2. Target host(s) is/are located using the DNS
3. MTA sends email to target MX server using SMTP on port 25
4. MX server gives the email to target's MDA
5. Authenticated MUAs pick up the email from the mailbox using (usually) either POP or IMAP



● 18/34

Email Composition      The Journey      Cambridge Email

# SMTP

- HELO
- MAIL FROM
- RCPT TO
- DATA
- QUIT

```

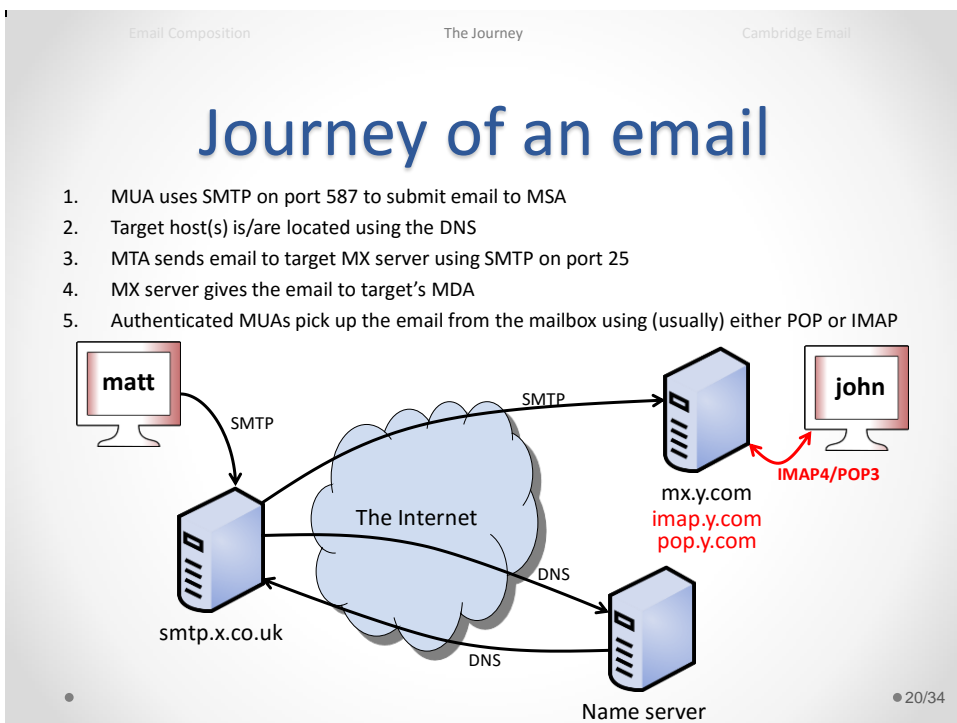
HELO smtp.example.co.uk
250 Hello smtp.example.co.uk, I am glad to meet you
MAIL FROM:<someone@example.co.uk>
250 Ok
RCPT TO:<someone.else@example.com>
250 Ok
RCPT TO:<other@example.com>
250 Ok
DATA
354 End data with <CR><LF>.<CR><LF>
From: "Some One" <someone@example.co.uk>
To: "Someone Else" <someone.else@example.com>
Cc: other@example.com
Date: Tue, 15 January 2008 16:02:43 -0500
Subject: Test message

Dear Mrs Else,



This is a test message.

Best wishes,
Some One
.
250 Ok: queued as 12345
QUIT
221 Bye
  
```

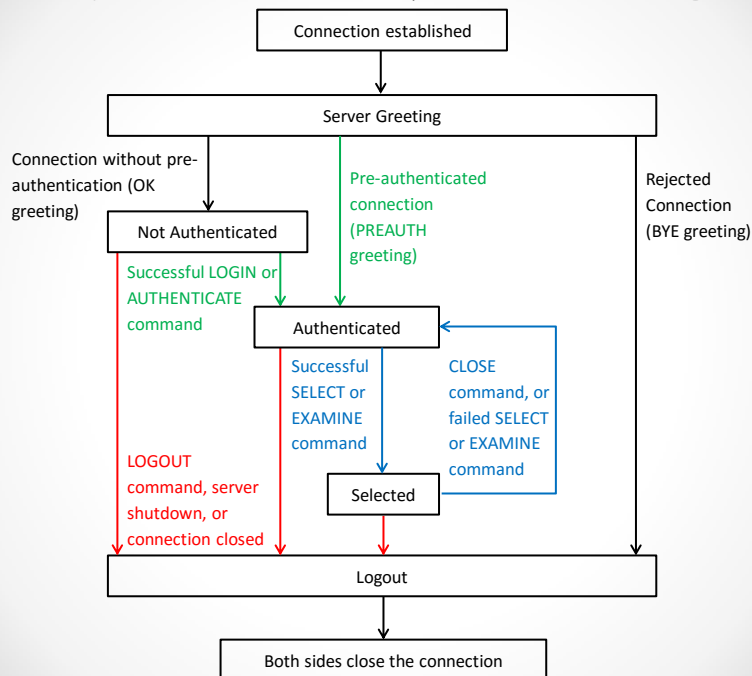
● 19/34



# POP3 vs IMAP4


POP3	IMAP4
<ul style="list-style-type: none"> <li>• Popular in early 2000's</li> <li>• Simple, <b>easy to implement</b></li> <li>• <b>Little storage required</b></li> <li>• Retrieve and delete</li> </ul> 	<ul style="list-style-type: none"> <li>• More <b>responsive</b></li> <li>• <b>Concurrent access</b></li> <li>• <b>Selective MIME fetch</b></li> <li>• Flags allow <b>synchronisation</b></li> <li>• <b>Folder structure</b></li> <li>• <b>Search mechanism</b></li> <li>• More <b>complicated</b></li> <li>• <b>Resource consumption</b></li> </ul> 

● 21/34




● 22/34

Email Composition      The Journey      Cambridge Email



# Cambridge email



● 23/34

Email Composition      The Journey      Cambridge Email

## Hermes and PPSW

- Two systems:
  - PPSW is the central mail relay system. It has three faces:
    1. [mx.cam.ac.uk](mailto:mx.cam.ac.uk) – Takes email from outside
    2. [ppsw.cam.ac.uk](mailto:ppsw.cam.ac.uk) – Outgoing relay
    3. [{imap, pop, smtp}.hermes.cam.ac.uk](mailto:{imap, pop, smtp}.hermes.cam.ac.uk) – Front-end to Hermes
  - Hermes is the central mail storage service for the University

● 24/34

# Development

- Early 90's – State of email was not great
- PPSW (1991) did protocol translation
- Hermes (1993)

60 cam members  
still use this!

```

----- Hermes Message Store Main Menu -----
Please enter the code for the action you require, and then
press RETURN.

1 PINE      - Run the Pine mail user agent to access your mail.
2 EDITOR    - Submenu for the Pico editor.
3 FILES     - Submenu for file management.
4 TERMINAL  - Change the terminal type setting for this session.
5 MISC      - Submenu for user info or printer and spell check setup.
6 DOMAINS   - Domain management (for authorised users only).
7 ALPINE    - Run the Alpine mail user agent to access your mail.

H HELP      - Help on Hermes
Q QUIT      - Terminate Hermes session.

Selection? █

```

# Development

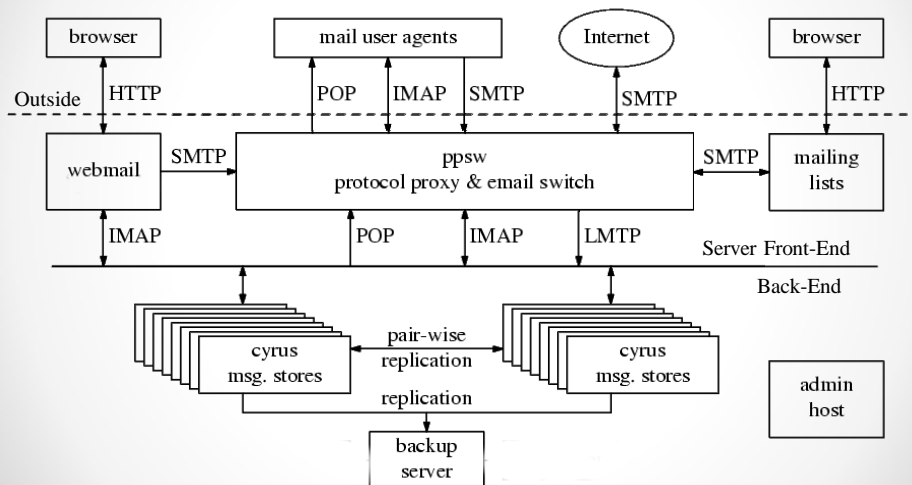
- Early 90's – State of email was not great
- PPSW (1991) did protocol translation
- Hermes (1993)
- Mid-90's – Email quotas increased to 10MB
- Late 90's – Exponential growth in number of users
  - Total of 280GB
  - 4 servers running UW IMAP

## Development

- Early 90's – State of email was not great
- PPSW (1991) did protocol translation
- Hermes (1993)
- Mid-90's – Email quotas increased to 10MB
- Late 90's – Exponential growth in number of users
- 1999 – Old Webmail until relatively recently
- 2003 – Major overhaul. Switched from UW IMAP to Cyrus  
Delayed Expunge  
Replication  
Performance Improvements

● 25/34

## Hermes and PPSW now



● <http://www-uxsup.csx.cam.ac.uk/~fanf2/hermes/doc/talks/2007-11-law-cio/notes.html>

● 26/34

# In the flesh (or metal!)

● 27/34



● 28/34

## Load balancing and protection


- 8 emails per second
- 4 PPSW servers needed, each with a different MX IP address
- Round-robin DNS for load balancing
- Reliable hardware and machine rooms
- 2 spare running PPSW server just in case

## Most Significant Difficulties

- Email virus/spam explosion around 2002
- Compromised accounts – a few per month
  - Rate limits: 5000 per day for on-site users, 200 for off-site
  - Fewer Cambridge students fall for phishing



## What about other Unis?

- PPSW/Hermes is a mostly custom solution
  - Exim, written by Philip Hazel 
  - Custom proxy software
  - Old webmail written locally
  - Original Cyrus had significant custom modifications
- This is quite unusual. Other Universities use:
  - Cloud mail providers (Google Apps, Office 365 email)
  - Microsoft Exchange Server

● 31/34

## The future of Hermes

- Departmental reorganisation
- Tidying up servers
- Hard to tell what the details will end up like for email servers
- Could be the death of Hermes!
- If so, would switch to Office365 mail services

● 32/34

## Summary

- Email format is limited to 7bit ASCII by legacy systems. We get around this by using MIME
- The journey of an email is based on a store-and-forward model
- SMTP and IMAP are the most important email transmission protocols right now, with independent sessions following a client-server format
- Cambridge University uses a custom system: Hermes and PPSW

## Questions?