Email routing in the University

http://www-uxsup.csx.cam.ac.uk/~fanf2/

Tony Finch
⟨fanf2@cam.ac.uk⟩ ⟨postmaster@cam.ac.uk⟩

Wednesday 9 March 2011

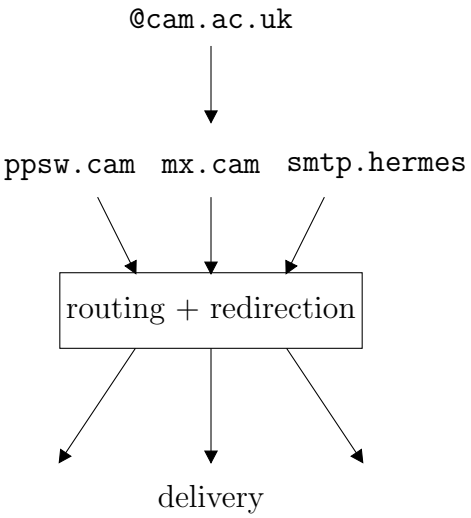
Abstract

This talk is about the functions of ppswitch, our central email relay, and how it works with other email systems in the University.

What special facilities can we provide? What does our peculiar terminology mean? What are the notable special cases that result from institutions doing their own thing? How do we help institutions with big changes in their email setups? What other weird things sometimes occur?

1 External faces of ppswitch

Overview



This talk is about ppswitch, in particular its SMTP delivery functions. PPswitch also runs the IMAP and POP proxies that make the many Hermes message store servers look like a single system to our users.

PPswitch has several service names which differ in the way that they decide which email to accept. After a message has been accepted, its routing and delivery is the same regardless of its origin.

What mail is accepted where

- `mx.cam.ac.uk`
  - From any non-spammer
  - To local addresses
  - Anti-spam content analysis
  - Anti-virus
- `smtp.hermes.cam.ac.uk`
  - From Hermes users
  - To anywhere
  - Rate limits + anti-virus
- `ppsw.cam.ac.uk`
  - From CUDN computers
  - To anywhere
  - Rate limits + anti-virus

`mx.cam.ac.uk` is our main outward face, where most of the anti-spam action happens. We will only accept mail to addresses at the 377 local mail domains that we support.

`smtp.hermes.cam.ac.uk` accepts mail from anywhere to anywhere, but you have to be an authenticated Hermes user.

`ppsw.cam.ac.uk` is for general outgoing email from computers on the University data network, mainly other mail servers within the University, or transactional mail from web servers, or documents from printer/scanner machines, etc. usw.

We have some anti-spam checks for outgoing email, mainly to protect against the consequences of a compromised web server or a user's email account credentials being stolen. Both are fortunately not too frequent.

2 DNS

Numbers of domains

4 *UCS test*

3 *independent*

8 secondary MX

76 special routes

9 long-form

2 obsolete

245 managed mail domains

15 https domains

22 internal redirects

384 *total*

44 *non-cam.ac.uk*

This is a break-down of the University domains that we know about.

Almost all Cambridge mail domains have an MX record pointing at `mx.cam.ac.uk`. The top two italic lines count University domains that ppswitch has nothing to do with.

I have classified the domains according to how we route and deliver mail for them, and the rest of the talk is basically going to follow this list and hopefully find some interesting things to say about it.

Unusual MX records

- `ccdc.cam.ac.uk`
- `gene.cimr.cam.ac.uk`
- `cup.cam.ac.uk`
- `cambridge.org`
- *`econ-1m.econ.cam.ac.uk`*

- eng.cam.ac.uk
- *ebulletin.foundation.cam*
- jesus.cam.ac.uk
- *his.path.cam.ac.uk*
- *mrc-bsu.cam.ac.uk*
- mrc-cbu.cam.ac.uk
- srl.cam.ac.uk
- *lib-cat.trin.cam.ac.uk*
- *cambridge-assessment.org*

The number of domains that aren't handled by `mx.cam.ac.uk` is small enough to list on one slide (apart from the UCS test domains!). Only the italic domains in the list above do not use us at all; the others just use us as a backup MX, for which we will accept email if their main servers are temporarily unavailable.

The `ebulletin` server is separate because we were concerned by the possible backlash from very large alumni mailshots.

The Trinity library catalogue server is run by external consultants.

Cambridge Assessment is the most notable department that doesn't even use us for secondary MX service. I didn't count their domains on the previous slide, except for `ucles.cam.ac.uk` for which we do handle email but which they don't use.

We route email to these domains using their MX records, just like non-University domains.

Local restrictions

- Port 25 blocked
- No mail to A or AAAA records under `cam.ac.uk`

When Cambridge joined the Internet the Computing Service took a very conservative attitude to some of the loose practices that were common then. In particular we kept strict control over the DNS and email, so only authorized servers would work. The idea was to ensure that the servers were run well, there weren't dangling NS and MX records in the DNS.

Back then this was often seen as unreasonable control-freakery but since 2000 it has saved us a fair amount of work since sensible anti-spam measures were already in place.

One thing we are still unusually strict about is that we will not deliver mail to arbitrary hosts within `cam.ac.uk`. A Cambridge mail domain must have an MX record. This has the advantage of stopping undeliverable junk from half-configured servers clogging up the queue on ppswitch. On the other hand we need a facility to relax this rule in a controlled manner, about which more later.

3 Whole domain configuration

Special Routes

`@admin.cam.ac.uk`	→	`mailgate.admin.cam.ac.uk`
`@cl.cam.ac.uk`	→	`mta.cl.cam.ac.uk`
`@tex.ac.uk`	→	`mta.cl.cam.ac.uk`
`@emma.cam.ac.uk`	→	`mail.emma.cam.ac.uk`
`@lists.cam.ac.uk`	→	`lists.cam.ac.uk`
`@medschl.cam.ac.uk`	→	`me-mail.medlan.cam.ac.uk`
`@srcf.ucam.org`	→	`mx.srcf.societies.cam.ac.uk`

First a fairly simple case.

There are 76 special routes of which 15 are non-`cam.ac.uk` domains, of which the above is a small selection.

A "special route" is a mail domain whose mail should all be delivered to a system separate from ppswitch. We have a table listing each domain and the name of the host the mail should be delivered to.

From ppswitch's point of view there is nothing special about `lists.cam.ac.uk`.

Long-form domains

- darwin.cam.ac.uk dar.cam.ac.uk
- cont-ed.cam.ac.uk ice.cam.ac.uk
- interfaith.cam.ac.uk cip.cam.ac.uk
- math.cam.ac.uk maths.cam.ac.uk
- murrayedwards.cam.ac.uk newhall.cam.ac.uk

- queens.cam.ac.uk quns.cam.ac.uk
- varsity.co.uk varsity.cam.ac.uk
- woolfinstitute.cam.ac.uk woolf.cam.ac.uk
- group.caret.cam.ac.uk camtools.cam.ac.uk

There are a couple of other setups we have for dealing with a domain as a whole.

Back in the old days of JANET there was Grey Book email instead of SMTP and there was the NRS instead of the DNS, and every name had a long form and a short form. We ditched the long form as quickly as we could, but some institutions were left with rather ugly domain names. So we have brought back a limited version of long form aliases.

This facility redirects any address at a long form domain to the same address at the corresponding short form domain. It's actually pretty general-purpose - in fact sometimes the short form is longer than the long form, as in the cases of maths/math and Varsity.

As well as permanent redirections, we also use this feature to help out with departmental renames, as is the case for cont-ed/ice. We set the new domain to run in parallel with the old domain for several months while the department gets everyone to move over.

Most of the target domains are managed mail domains, but we can also do long-form redirecttions before delivering via a special route, as is the case for Darwin.

In addition to this list, we also have about seven special routes for long-form aliases which are handled by the institution's mail server.

Obsolete domains

- cei.group.cam.ac.uk
- sps.cam.ac.uk

Finally, when a domain has been decommissioned we put it into "friendly bounce" mode for a few months, just in case there are still a few correspondents who haven't heard about the name change.

This table is usually pretty empty.

4 Managed mail domains

Managed mail domains

- Unix-style `aliases` file
- Edited using `ssh` to Hermes menu system
- Uses RCS for locking
- Converted to `cdb` for mail routing on ppswitch
- Live copy updated hourly
- Managers can get an audit of broken addresses

The vast majority, 245 of our mail domains are "managed mail domains", of which 22 are non-`cam.ac.uk`. The more common terminology for the same thing is a "virtual domain".

The user interface is now about 15 years old and rather shows its age.

We use RCS to stop more than one domain manager trying to edit a domain at the same time. The fact that it keeps a history of changes is very rarely useful.

If a domain manager kills their `ssh` session and leaves the aliases file locked, the same person can log back in and unlock it by just saying "yes" to the scary questions. If the person owning the lock isn't available then we have to unlock it for them.

The validity checks in the user interface are not especially strong, but domain managers can ask for a very thorough batch-mode check of their domain, and the results will be mailed to them after the next update.

We use `cdb`, which is a very simple read-only database, for fast lookups on the live systems.

Managed mail domains

```
f.a.n.finch:fanf2@ucs.cam.ac.uk  tony.finch:fanf2@ucs.cam.ac.uk  fanf2:fanf2@cam.ac.u
mail-support:msupport@hermes.cam.ac.uk msupport:fanf2@hermes,dpc22@hermes
```

An aliases file just lists the local parts (before the `@`) of the email addresses in the domain and the corresponding destinaton addresses.

Managed mail domains are designed to support friendly names for individuals, group role addresses, or any other kind of redirect.

There are no restrictions on the destination address, so they can accommodate external collaborators etc.

HTTPS Managed mail domains

- Automated management

- Our update job fetches aliases file from manager's web server instead of from Hermes menu system
- Live copy updated hourly
- Managers get monthly broken address audit

This feature is for those who don't like driving the menu system manually. Apart from the update procedure they are exactly the same as normal managed mail domains.

We currently have 15 https domains of which 6 are non-`cam.ac.uk`

Addresses at `cam.ac.uk`

- 40376 aliases
- 38918 go to Hermes
- 12 special cases

The main `@cam` domain is in fact just a slightly weird managed mail domain.

Users can control where their `@cam` address redirects to using a web form on `https://jackdaw.cam.ac.uk/cammail`.

Most of the aliases are fetched from Jackdaw hourly. There is an additional list of special case aliases, which is managed using the menu system like other managed mail domains.

Note that the Hermes account management job is run daily. When accounts are cancelled, the users' `@cam` email addresses stop working almost immediately, but their Hermes account continues to work until the next morning.

Addresses at `hermes.cam.ac.uk`

- 40355 accounts
- 33 special case aliases
- 8 live servers
- 8 hot spares
- 6TB storage each

Delivery to the Hermes message store servers is handled directly by ppswitch using special case code. There is a table listing which user is on which server.

Internal redirects

- For all the awkward cases
- Extend managed mail domain
- RT servers are common
- Also used for migrations

When I was discussin our local restrictions I mentioned that we have a facility to relax the restrictions in a controlled manner. That is the "internal redirect" table.

In the past we had an "escape route" table which did a similar job, but it allowed hostnames to appear in mail domains of messages between ppswitch and the destination mail server.

Internal redirects avoid that problem, which also makes them better for supporting gradual migrations from managed mail domains to institution mail servers.

Internal redirects

- `atj20:atj20@ucs-exchange.internal-redirect`
- `atj20@ucs.cam.ac.uk` → `atj20@ucs.cam.ac.uk`
- But delivered to an alternate destination

To use an internal redirect you put an alias in your managed mail domain that delivers to an address with a special domain name.

This doesn't change the address that the mail is delivered to, but instead changes the destination server.

This allows you to move addresses one-by-one from a managed mail domain to a server, so you can redirect just a few to support a ticketing system, or you can gradually migrate from a managed mail domain to your own server without any sudden cutovers or nasty surprises.

When all users have been migrated we do some final checks to make sure everything is ready before replacing the managed mail domain with a special route. Because the recipient email addresses on the mail from ppswitch to the server are the same, a managed mail domain full of internal redirects works just like a special route, so the cutover has no visible effect.

Current migrations in progress include Careers, the Hutchison MRC, Continuing Education, and Plant Sciences.

Migration

- "disabled" managed mail domains

When migrating in the other direction we again make it a gradual process. The department moves their users from their own server to Hermes leaving behind a redirect on their server. When all users have been migrated they should be left with no mail configuration other than an aliases file.

We create a managed mail domain in "disabled mode" so that the domain managers can edit it using the Hermes menu system wihtout having any effect on mail routing.

When they have populated their managed mail domain we can enable it so it takes over from the old mail server.

5 End

This talk has been terminated