Techlinks email update

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

Tony Finch
⟨fanf2@cam.ac.uk⟩ ⟨dot@dotat.at⟩

Wednesday 29 October 2008

Abstract

This talk covers recent changes to the Computing Service's email services, including quota changes, webmail changes, and especially anti-spam changes.

1 Hermes

Hermes update

1.1 Quotas

Standard Hermes quota

~~1 GB~~

In my previous talk 6 months ago, I announced that we had increased the standard Hermes quota to 1GB.

Standard Hermes quota

500 MB

Unfortunately we had to reduce it again. The increase was based on the assumption that we would replace some old hardware over the summer, and that the new hardware would provide more space.

Budget problems

Still waiting for confirmation of this year's hardware budget... Last year's money disappeared before we could spend it. We hope that we will get some money this year, otherwise further austerity measures will be imposed.

Quota increases

- Quota upgrades still available
- `http://www.cam.ac.uk/cs/request/quota.html`
- Good reason required!

The previous regime was fairly laissez faire about quota increases. We're now being rather stricter about the reasonableness of quota requests, especially for students - though less so for staff.

1.2 Redstone

Redstone

- Hermes now split half and half
- Computing Service (city centre)
- Redstone (north Cambridge)
- Continuous replication
  from live half to backup half
- Manual fail-over

This means our disaster recovery plan does not inevitably require a week to restore stale data from tape. We don't yet know if this will improve availability. A lot of the network still depends on the Computing Service machine room, especially our link to the outside world.

The nuclear weapon illustrating the slide is a "Redstone" medium-range ballistic missile.

1.3 Webmail

Webmail

A few comments about Hermes webmail.

Old webmail style

New webmail style

We have reskinned webmail according to the new University house style.

New webmail front page

The webmail front page provides quick links to UCS news items.

Webmail survey

We received a lot of feedback in the first week after introducing the new look webmail. We're now running a survey to encourage more feedback. It'll continue until the end of term.

Log in via Raven

Starting Monday 17th November

Webmail will support the University's web single sign-on service for login. This was announced in September.

I used a cartoon featuring Kerberos (Cerberus) to illustrate this even though Raven isn't based on the Kerberos single sign-on protocol.

2 Anti-spam

Anti-spam upgrade

Now I'm going to talk a bit about the upgrade to the spam and virus filters that we did over the summer, five years after the last big upgrade.

2.1 Overview

Constraints on late scanning

- We must not emit autoreplies
  - spam and viruses have bogus sender addresses.
- So our options are:
  - deliver the message (perhaps after modification)

- quarantine it
- throw it away

The old scanner scanned email after we accepted responsibility for delivering it, which put some constraints on what the software could do when it found a dodgy message. The most important one is that we mustn't cause backscatter by emitting messages in response to spam or viruses.

Annoying things

- Stripped attachments are awkward
- Risky phish handling
- Complaints about forwarded spam
- Quantity of stored spam

This led to some annoying consequences, which were the motivation for the upgrade.

I'll discuss these in more detail in a moment.

Scanning early

- Reject messages as soon as possible
- doesn't cause backscatter
- legitimate email doesn't disappear
- performance is more important

There's a point in the SMTP protocol after the message data has been transmitted, and before the receiving MTA confirms to the sending MTA that it has received the message OK. The receiver can scan the message at this point and reject it if it looks bad.

This usually causes spam and viruses to disappear, because spamming software does nothing when its messages are rejected. Legitimate senders, on the other hand, will inform the message's author of the problem using a bounce message. In this case the sending MTA usually knows the author is correct so the bounce will not be backscatter.

The disadvantage is that there's relatively little time in which to do the scan. If load gets too high the recipient can't delay scanning until capacity is available.

2.2 Problems and improvements

Stripped attachments

If a legitimate message had an attachment with a "dangerous" filename, then it was stripped and delivered to the recipient. This required some back-and-forth to resolve because the sender was not directly informed of the problem.

When we scan early, the sender gets immediate nofication of a problem, and can correct it without wasting the recipient's time.

Phishing risks

We decided, because of the volume of phishing email, to treat it like mass-mailing viruses and just discard it. This was less safe than is ideal, because heuristic phishing detection is not as accurate as anti-virus signature matching.

A recent example was an email from Barclays that contained a link allegedly to `www.stockbrokers.barclays.co.uk` but which actually went to `www.barclayswealth.com`, just like a phishing attack.

Because early rejection generally does the right thing, false positives caused by this kind of foolishness are not such a disaster, so we can use more aggressive scanners which have much better coverage but a slightly higher false positive rate.

Other virus risks

If the old scanner detected a non-mass-mailing virus it stripped the attachment rather than discarding the message, in case it was something like a macro virus attached to a legitimate message.

We can now treat all viruses the same.

Forwarded spam

About 90% of email is spam, and about 90% of spam is blocked by DNS blacklists. This means the email that gets past this first hurdle is about half-and-half spam and legitimate.

We dealt with spam by tagging and delivering it, expecting the recipient to filter it. This is not good for our email servers' reputation if the recipient forwards their email offsite, and the destination site observes a large proportion of spam in email from the University.

We now reject spam that scores more than 10. This eliminates about 2/3 of spam, which substantially reduces this problem.

Size of spam folders

An advantage of delivering spam to a different folder is that if a legitimate message is mis-classified, the recipient can retrieve it without the sender having to jump through awkward hoops. However people are very inaccurate at finding legitimate email in a pile of spam, and usually don't bother looking unless an expected message hasn't arrived in their inbox.

We also had to implement automatic pruning of old messages in users' spam folders, to prevent them from using all the available disk space.

The new setup's blocking threshold of 10 is quite conservative, so the advantage of being able to recover mis-classified messages is preserved. The smaller quantity of delivered spam should make this less of a chore.

2.3 Other changes

Policy variations

The filtering policy we apply to email differs for internal and external email.

Internal email

- Anti-virus only

External email

- Anti-virus
- Anti-spam

Minor changes

- More informative spam details header
- Faster delivery, especially for internal email

Faster delivery because the scanner is no longer batch-based, and because internal email doesn't have to wait for the spam scanner to process external email.

2.4 Last few words

Future

- Greylisting
- Adaptive scoring

At the moment I'm working on a greylisting implementation. The aim is to delay suspicious email to allow time for 3rd party spam and virus filters to update. If it is successful, then new threats will be old enough that they are detected by the time we see them.

After that, a system for users to provide feedback to the filters, e.g. marking undetected messages as spam so that the filters can detect them in the future.

```
221 closing connection
```