

Email Traffic: A Quantitative Snapshot

Richard Clayton

CEAS, Mountain View

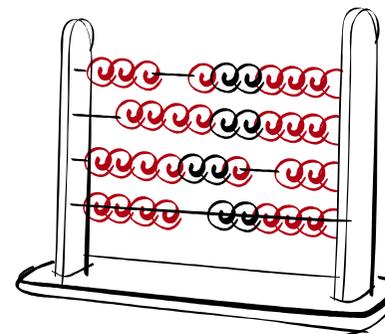
3rd August 2007



**UNIVERSITY OF
CAMBRIDGE**

Computer Laboratory

Demon



Dataset

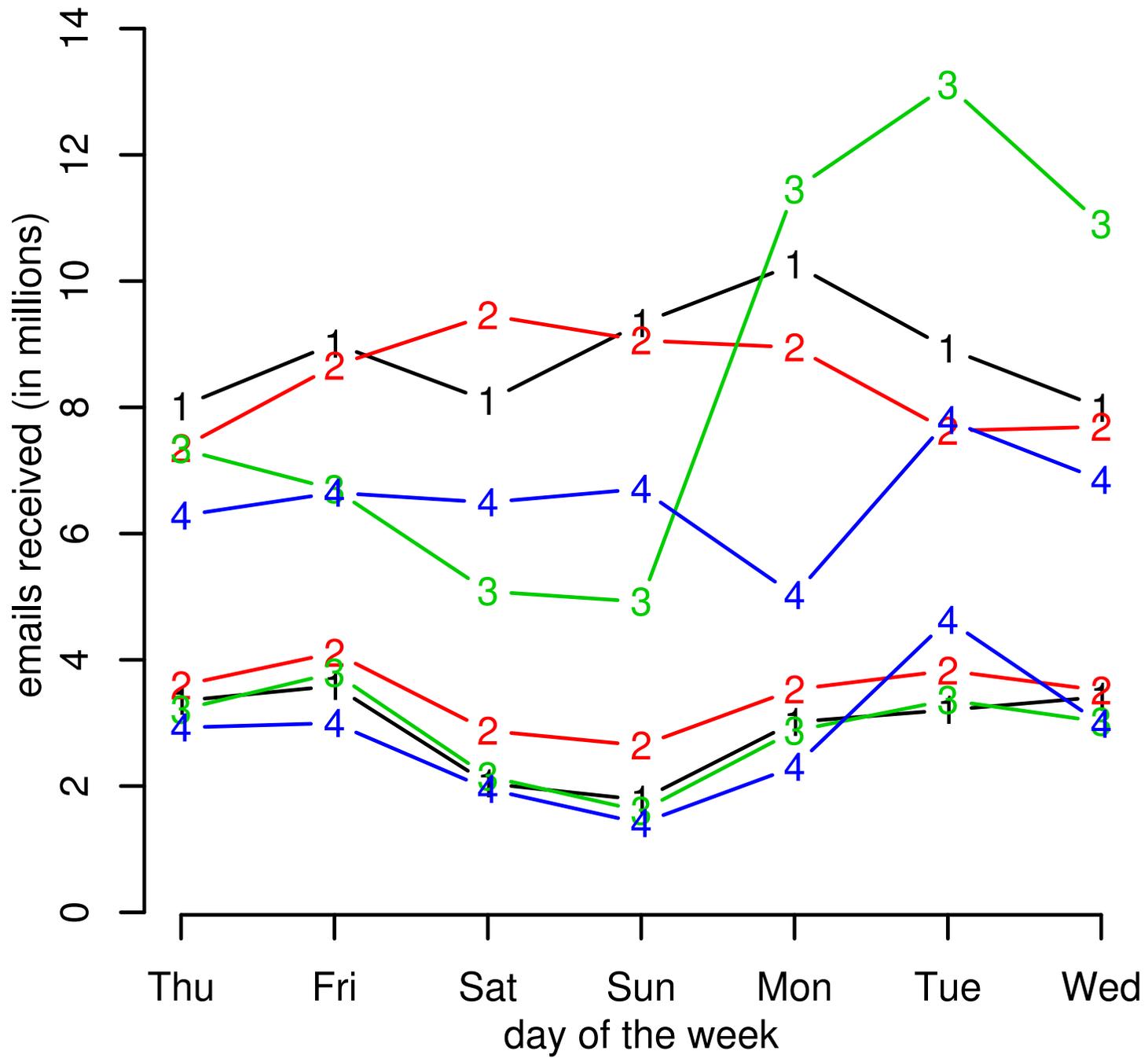
- Incoming email to Demon Internet
 - medium sized, long established UK ISP
 - c 150,000 customers, mainly ADSL, some dialup
 - mix of consumers, small & medium business
- Four week dataset (1-28 March 2007)
 - no public holidays
 - trying to measure “normal” activity

Raw numbers

- 41 565 269 connections didn't deliver
 - 1.2 million/day failing to complete SMTP protocol
- 331 858 366 emails, 355 559 372 destinations
 - 12 million/day
 - 1.15 addresses per email
- 22 400 218 of these had null < > sender
 - mainly backscatter (800K/day)

Spam

- Demon is running a Cloudmark product
- 73% of the email was “spam” & not accepted
 - 1.10 destinations per email
- Only 83 720 106 not spam
 - 2.99 million per day
 - 1.18 destinations per email
- Viz: spammers no longer multiply address
 - Gomes (2004) found 1.7 for spam (1.4 overall)



Forwarded email

- Remote sites sending to one email address more than 10 items per day on 14 or more days
- This accounts for 11 709 190 emails to 7 943 email addresses (which belonged to 5 427 different customers).
 - ie 3.5% of email is auto-forwarded and 66% is spam
- But huge differences – some no spam, some little spam, some all spam (see paper)

Mailing lists

- Heuristic approach based on **-bounce-**, **listmanager@**, or **lists.example.com** host etc
 - failed to find workable *n items/n destinations* test
 - excludes ad hoc lists (**brenda@**, **sales@**)
- Accounts for 7% of email (with 10% of it being spam, and half entirely spam free)
- Way down on previously reported Demon 2004 figure (on same basis) of about 40%

This matters!

- Spam is changing (fewer 100 destination emails) so classic heuristics are dated
- Spam is very bursty, suggesting “200 spam gangs” is rather out of date
- Proposals such as SPF/SenderID have trouble with forwarding – but it seems to be rarer than some people suggest
- Mailing lists are dying! (helps proof of work)

OR maybe it's just like this at Demon?

We need LOTS MORE PUBLISHED DATA to permit academic analysis, to allow proper assessment of anti-spam proposals and to inform public policy debates

It would be nice to copy my methodology for comparison purposes – but almost any more detailed data beyond “73% spam ” or “x billion spams a day” would be useful!

Email Traffic: A Quantitative Snapshot

<http://www.lightbluetouchpaper.org/>



UNIVERSITY OF
CAMBRIDGE
Computer Laboratory

Demon