Algoritmer og Datastrukturer 1

Gerth Stølting Brodal

Greylisting

AARHUS UNIVERSITET

Greylisting

Greylisting er en teknologi anvendt på mail-serverne på cs.au til at begrænse mængden af spam brugerne modtager

Teknisk Forklaring

Greylisting keeps a database of where you receive mail from. The records in the database are pentuples listing the IP network, sender domain, recipient address, a counter and a timestamp.

When a new pentuple is seen, the mail transaction is aborted with a temporary failure. After some time, the remote mailserver will retry the transaction. As the pentuple is in the database by now, the mail will get processed normally and reach you.

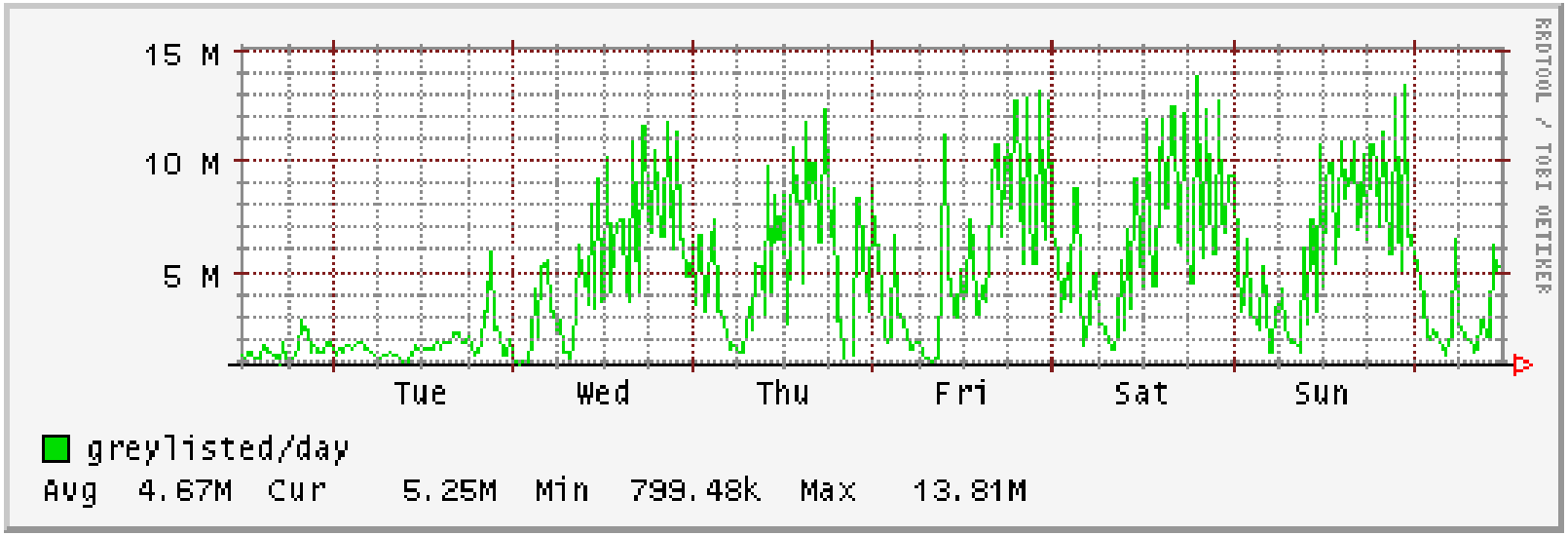
Mail servers (MTA's) has according to the standards to be able to handle temporary errors. However, the software used by the spammers - eg on home PCs acting as spam zombies does typically not implement this functionality.

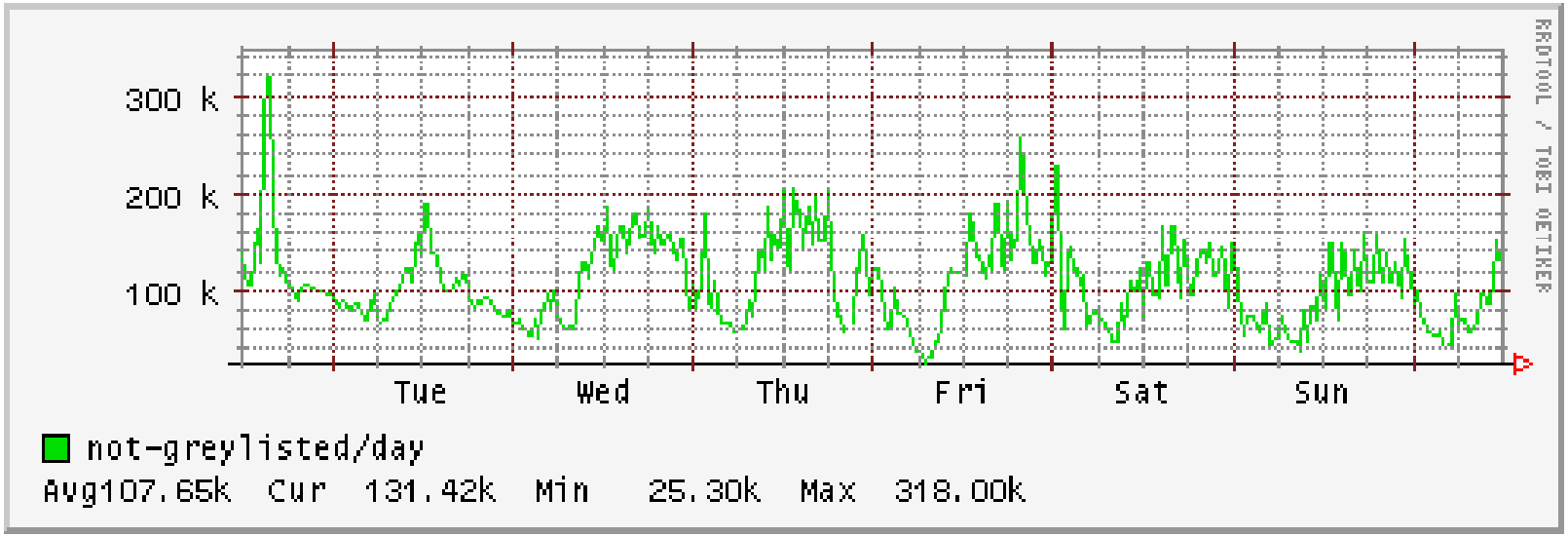
Pentuples with a usage count of 1 are removed after 24 hours as they most likely represent spam.

www.daimi.au.dk/local/system/mail/grey-listing

Greylisting cs.au.dk

25. februar 2008

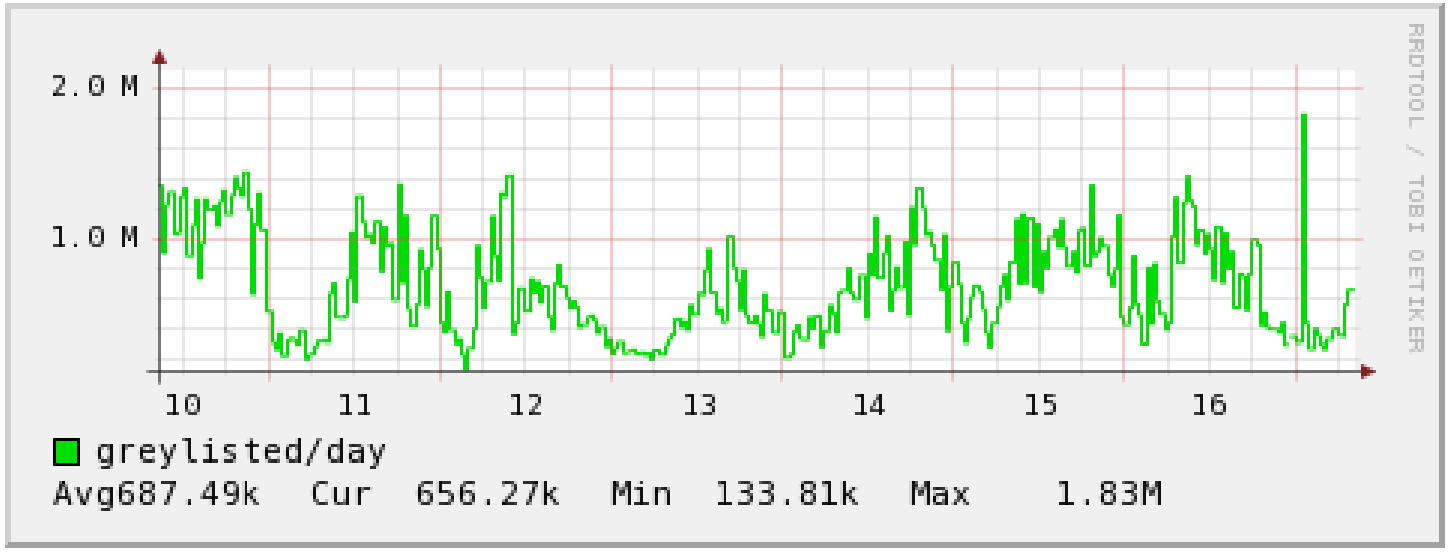


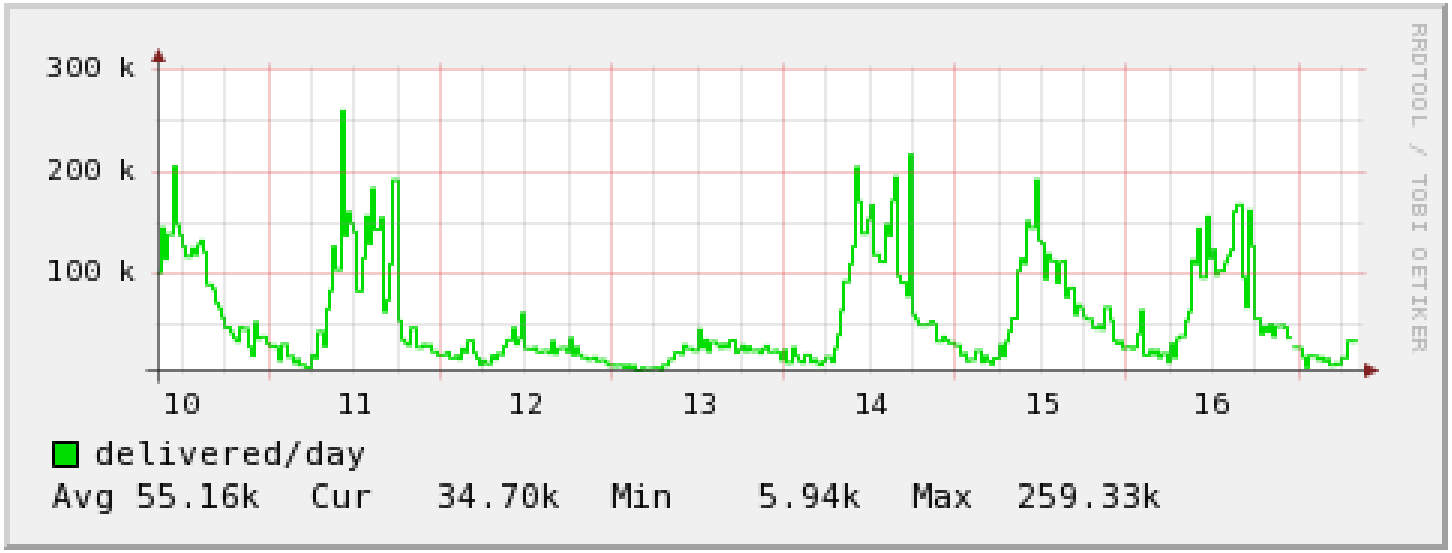


mailfilter.nfit.au.dk/greyops.html

Greylisting cs.au.dk

17. februar 2011





mailfilter.nfit.au.dk/greyops.html

Flaskehals

- 10 x 10^6 mails per dag
- Gennemsnitlig 8.6 ms mellem hver email
  - kan ikke nå at skrive til disk for hver email
- Pentuple = 128 bytes
- Pentupler per dag = 1.3 Gb
  - løber tør for hukommelse

Løsning

- Istedet for at gemme 128 bytes pentupler p, gem en 64 bit **hashværdi** $h(p)$
- Brug en ordbog implementeret ved **linear probing** til at gemme $h(p)$ startende søgningen på position $h(p) \bmod m$
- $10x10^6$ emails fylder 2 x 8 x $10x10^6$ = **160 MB**

- **Pris:** Enkelte spam mails hasher til samme værdi og slipper fejlagtigt igennem greylisting