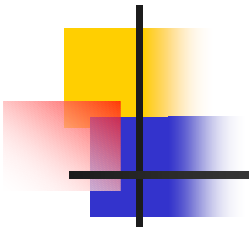
"E-mail Forensics: eliminating spam, scams and phishing"

Les Hatton

CISM, Kingston University
L.Hatton@kingston.ac.uk

Version 1.1: 08/Feb/2012

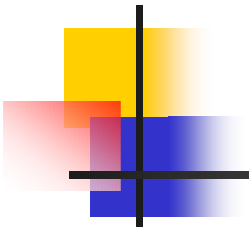
Types of email forensics



- Investigative forensics
  - Generally poring over immense logs trying to determine their provenance
- Preventative forensics
  - Attempting to prevent future attacks by analysing their pathology

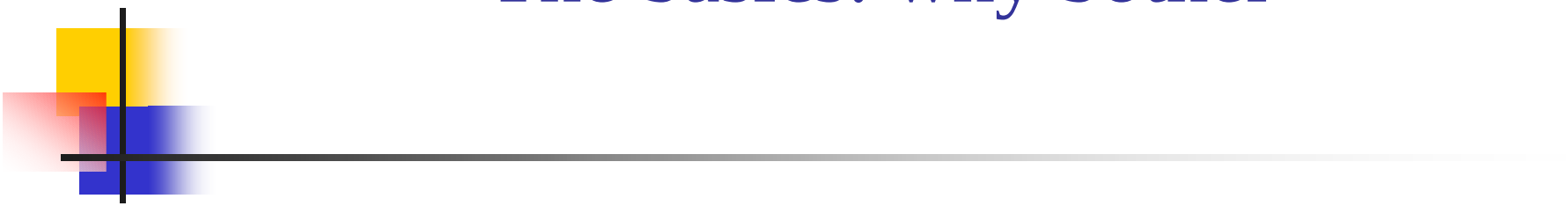
Both involve criminal activity, hence forensics.

Overview



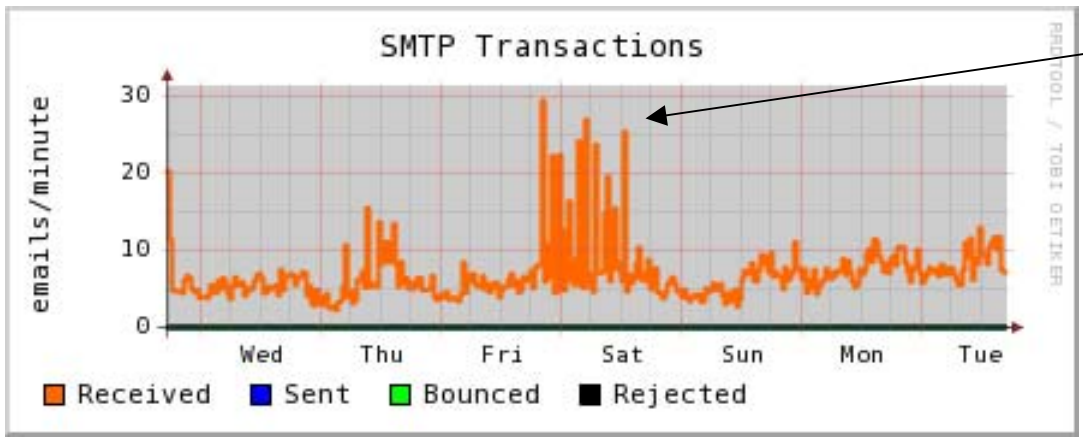
- Overview
  - The basics: why bother
  - The threat landscape
- Defence in Depth
- Wrap-up

The basics: why bother



- By 2008, I was receiving anywhere between 80,000 and 150,000 junk mails a day, (example, 2-9 November, 2008, a total of 836,000 arrived).
- ISP refused to handle it any more. Ultimatum - either I change or take it elsewhere.
- Acquired dedicated server, (running Centos Linux) and started researching.
- Goal – 6 sigma, or less than 4 mis-categorised messages per million messages received, (> 99.9996% accuracy)

The basics: learning on the job

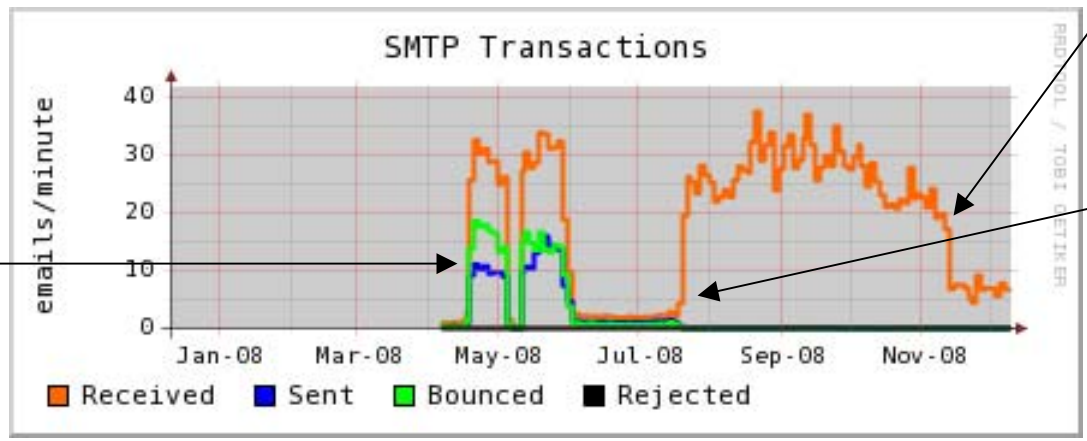


Weekend rubbish

Typical week

Attivo site closed in USA

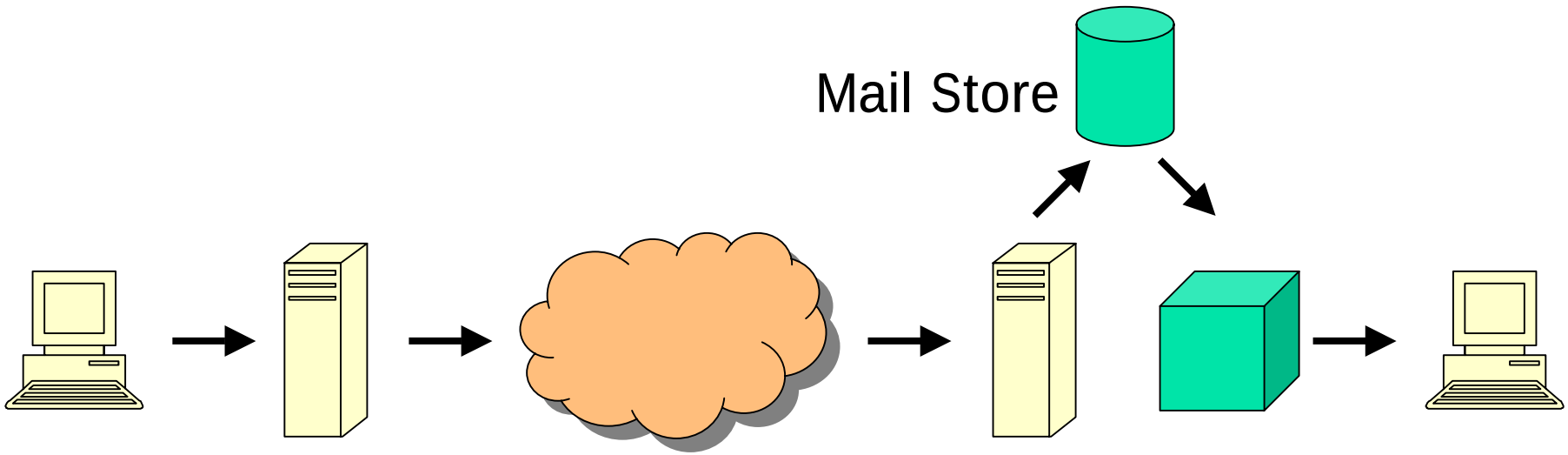
LH failing to understand the FormMail relay injection loophole



Silent discard, greylisting, RBL

Year to date

The basics: A ➔ B



The basics: Headers and Content

Envelope

Content

Connect mail.receive.com (an MTA responds)

HELO mail.send.com (an MTA responds)

MAIL FROM: alice@send.com (an MTA responds)

RCPT TO: bob@receive.com (an MTA responds)

DATA (an MTA responds)

Date: ...

From: alice@send.com

To: bob@receive.com

Reply-To: ...

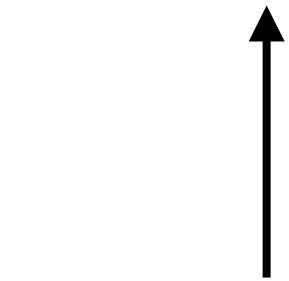
Message-ID: ...

Subject: ...

Blah blah blah

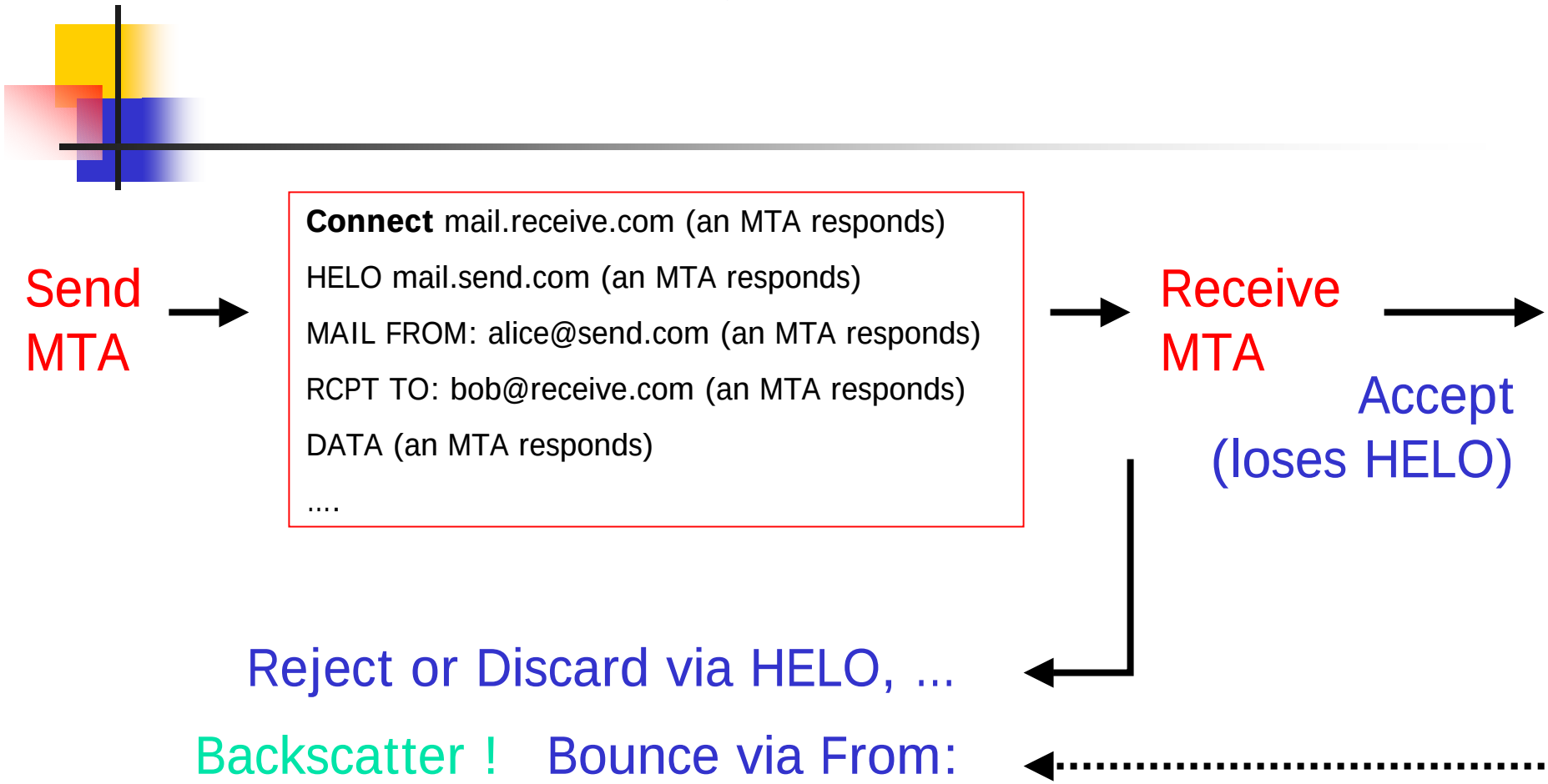
Disconnect

MTA



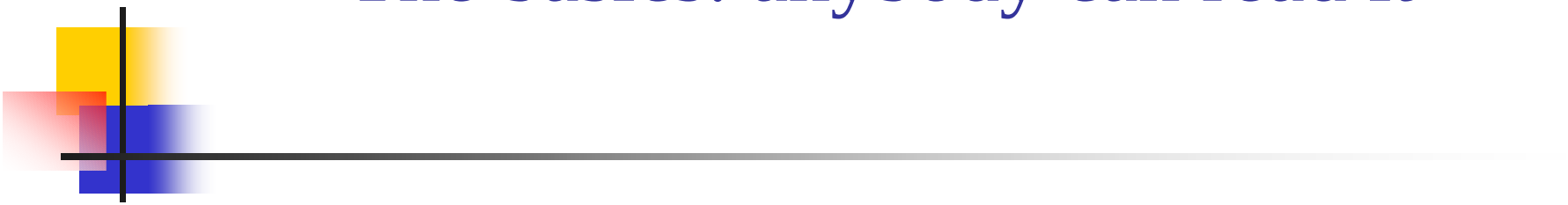
MUA

The basics: Headers and Content



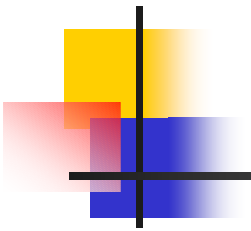
*** Reject as early as possible ***

The basics: anybody can read it



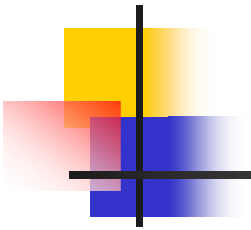
- Nearly all e-mail is sent in clear as if you had written it on a postcard and asked a complete stranger to post it for you.
- It can be arbitrarily spoofed

The basics: what can be forged ?



- Headers that can be forged
  - Subject, Date, Message-ID
  - Recipients: From, To, CC, BCC
  - Content body
  - Any arbitrary headers, X-Mailer …
  - All but the last Received header
- Headers that can ***not*** be forged
  - Last (top most) Received header
  - Originating mail server, specifically
    - IP address
    - Subsequent timestamps

The threat landscape – e-mail borne toxins



- Spam
  - Density
  - Works of Art
  - Harvesting
  - Patterns
- Scams

Load by 2010

- For a single mail-server handling mail for 8 domains in 13-20 November, 2010

Total received	401,975	100.00%
Discarded or rejected	401,784	99.952%
Rejected by deep content filtering	72	0.018%
Delivered to users	119	0.03%
(Missed spam FN / lost mail FP)	(2/0)	0.0005/0.00%

Load by 2011

- For a single mail-server handling mail for 8 domains in 12-19 November, 2011

Total received	64,764	100.00%
Discarded	64,526	99.63%
Rejected by deep content filtering	43	0.066%
Delivered to users	195	0.3%
(Missed spam / lost mail)	(0/0)	0.00/0.00%

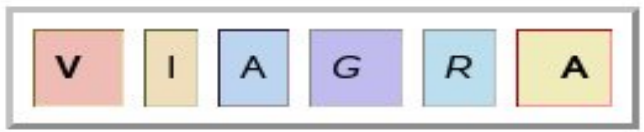
The threat landscape – HTML works of art

```
<body>=09
<p>=09What's up?<a name=3D"#tprw"></a></p><a name=3D"#qpqr"></a><span name=3D"#twqp">=09</span><br><a
name=3D"#rtwp">  </a><table border=3D="6" cellspacing=3D"7" cellpadding=3D"1" width=3D"199">
<tr><td bordercolor=3D"#4B41CE" nowrap=3D"nowrap" valign=3D"baseline" bgcolor=
=3D"#D9F0BC"><strong>V</strong><font color=3D"#D9F0BC">b</font>  </td>
  <td nowrap=3D"nowrap" valign=3D"middle" bordercolor=3D"#69FA49" bgcolor==3D"#BCB6F0"
align=3D"center">   I </td><td bordercolor=3D"#67DA87" valign=3D"baseline" bgcolor=3D"#F0BCC3" align=
=3D"left" nowrap=3D"nowrap">   <b>A</b></td>
<td align=3D"center" bgcolor=3D"#F0C3BC" bordercolor=3D"
<td nowrap=3D"nowrap" bgcolor=3D"#D4F0BC" bordercolor=3D
color=3D"#D4F0BC">y</font>  </td>
<td align=3D"left" bordercolor=3D"#6852DC" valign=3D"top
<font color=3D"#F0B9BC">3</font>A=09</td></tr></table>
=09<br><strong></strong><table><tr><td>WWW</td><span>
</td><br><td>.</td><span name=3D"#rrtp"></span><td>COM</
</b><br><strong>=09</strong><p><span></span></p><span na
name=3D"#qqtp"></a></p>
</body>
```

From: "Riso Nuzzi" <cohered@nda.co.nz>
To: gundalf@oakcomp.co.uk
Date: 2008-08-07 00:10

Spam Status: Spamassassin

Heya,



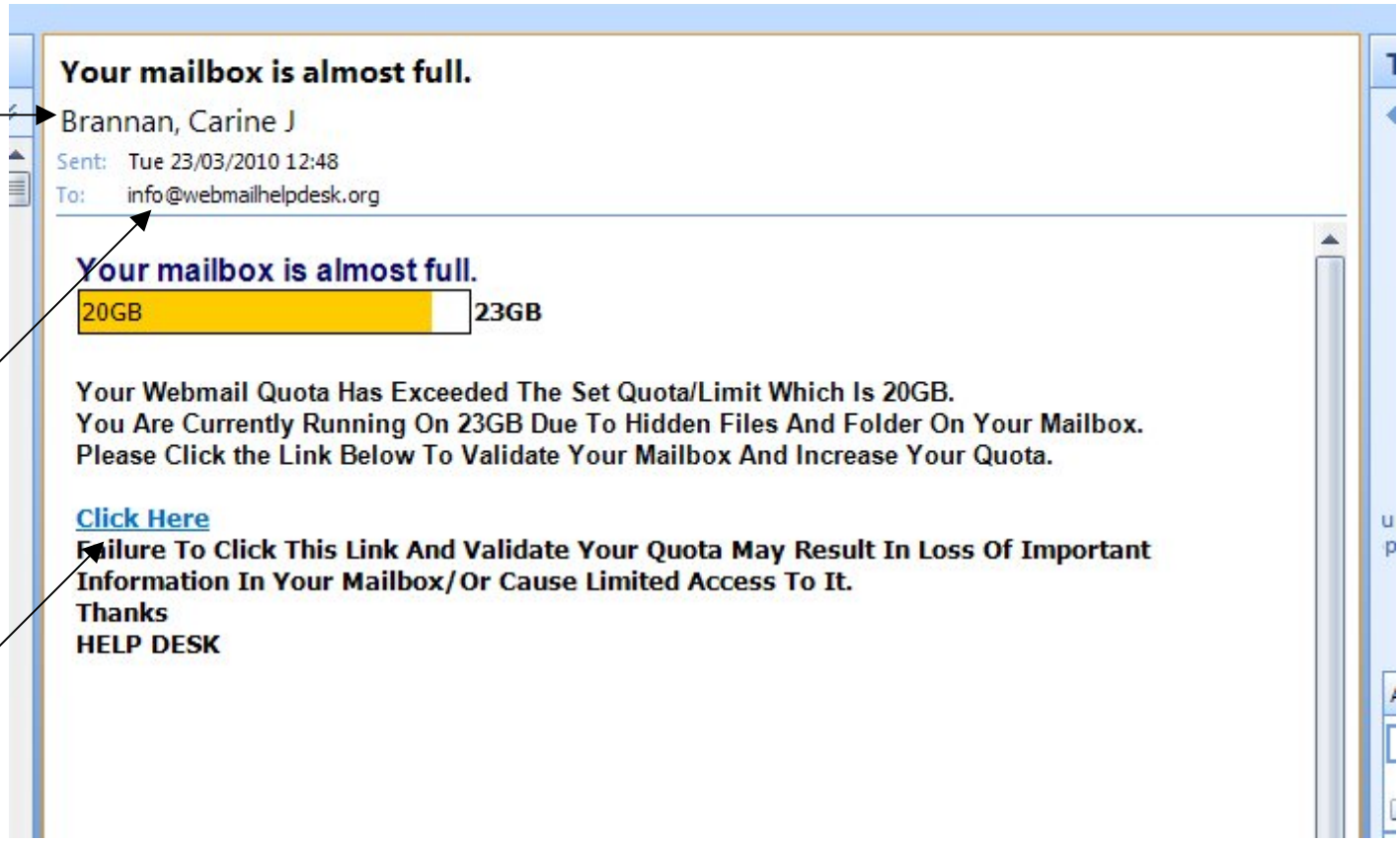
WWW . NEVOB . COM

The threat landscape – recent example

Came from compromised mailbox in KU. User does not exist.

From: address does not exist

Link –zyef.9hz.com does not exist



Your mailbox is almost full.

Brannan, Carine J

Sent: Tue 23/03/2010 12:48

To: info@webmailhelpdesk.org

Your mailbox is almost full.

20GB 23GB

Your Webmail Quota Has Exceeded The Set Quota/Limit Which Is 20GB.
You Are Currently Running On 23GB Due To Hidden Files And Folder On Your Mailbox.
Please Click the Link Below To Validate Your Mailbox And Increase Your Quota.

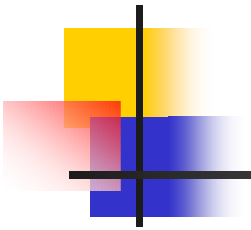
Click Here

Failure To Click This Link And Validate Your Quota May Result In Loss Of Important Information In Your Mailbox/Or Cause Limited Access To It.
Thanks
HELP DESK

The threat landscape – harvesting e-mails

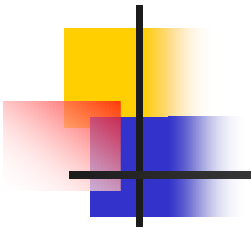
- “New virus coming – warn 25 of your friends …”
- “New speed camera – pass on to your friends”
- A beauty from yesterday, (07-02-2012):- “Advice on not passing on to your friends” saying “pass on to your friends” at the bottom.

The threat landscape – spam patterns



- More of a nuisance than a danger
  - Word frequency and *Bayes theorem* remain very effective – e.g. sales words
  - SpamAssassin still effective, largely because spammers are usually idiots.

The threat landscape – e-mail borne toxins



- Spam
- Scams
  - Nature
  - Patterns
- Intrusions

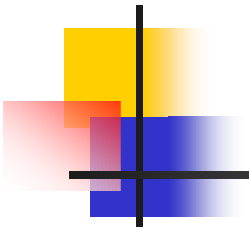
The threat landscape – Nature of scams

- Scams can be much more of a challenge
  - Lotteries (easy)
  - “My left leg has been biten off by a mad cheetah and I have $4 million in da trouser leg” (Nigerian 419) (easy)
  - Phishing for account details (can be convincing)
  - Pharming, DNS hijacking, … (ditto)

The threat landscape – scam patterns

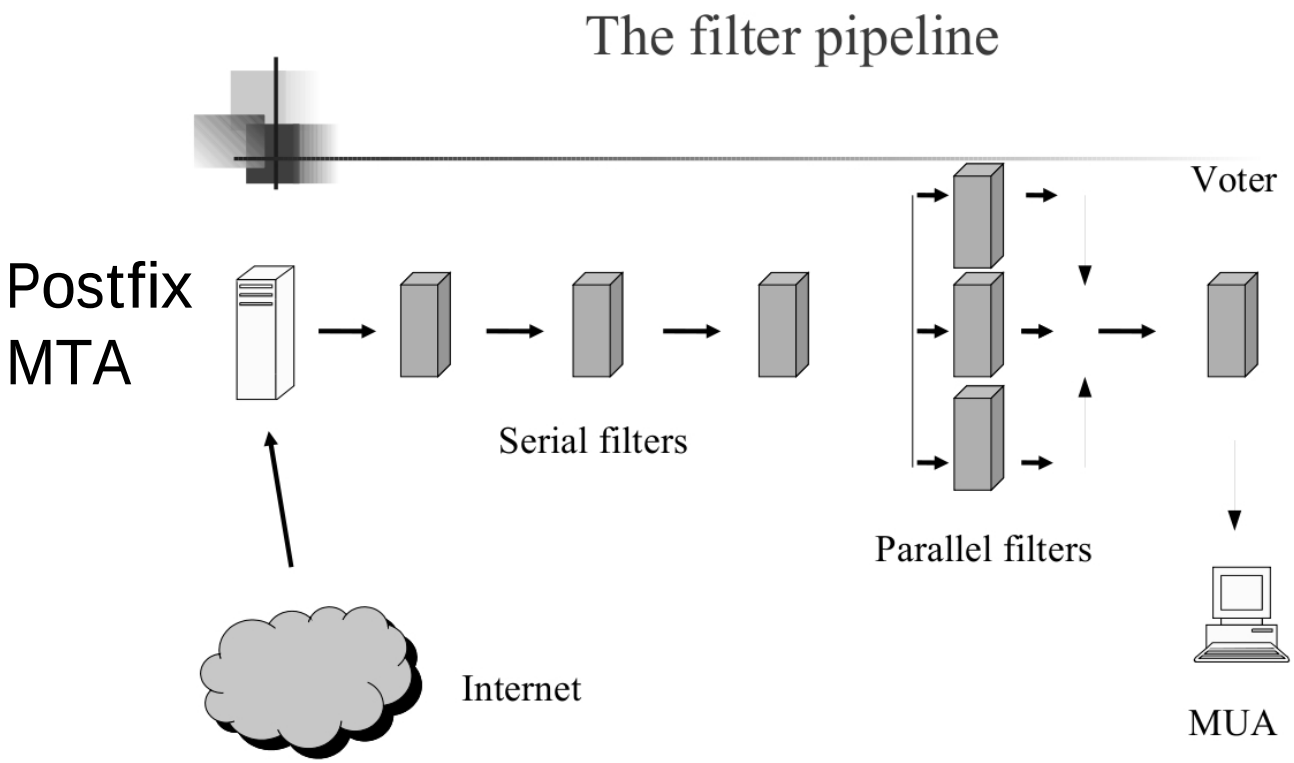
- Key to detection is to recognise structure
  - Word frequency sometimes useful – e.g. “codicil”
  - Recognition of phases better
    - The tease (419) or threat (account hacking)
    - The link to contact, (always bogus)
    - The sign-off

Overview



- Overview
  - The basics: why bother
  - The threat landscape
- Defence in Depth
- Wrap-up

Architectures



Layered protection and Postfix – typical installation with 2009 volumes

smtpd_client_restrictions	**None**
(618,000) smtpd_helo_restrictions	reject self_helo, reject_non_fqdn_hostname, reject_invalid_hostname
(neg.) smtpd_sender_restrictions	reject_non_fqdn_sender, reject_unknown_sender_domain, blacklist
(216,000) smtpd_recipient_restrictions	Reject_non_fqdn_recipient, reject_unknown_sender/recipient_domain, reject_unauth_destination, reject unknown users, reject_rbl_client (spamhaus etc.), **Greylist**, back scatter
(neg.) smtpd_data_restrictions	Reject_unauth_pipelining
(~ 200) header_checks	Different from/return-to, foreign character sets, various Windows skullduggery
(~ 200) body_checks	Other Windows skullduggery, tags, attachments, embedded .exe, ...
(~ 100) external content filtering	Viral filtering, sigs, creative html, SA, domain mismatch, embedded SURBL, cross-domain match

Defence in depth

- Serial filters
  - Must be 100% accurate, for example,
    - self-helo – MTAs pretending to be mine – one of 1,810 since Monday
    - daft addresses – gretchenlambbutch@oakcomp.co.uk, one of 15,181 received since Monday.
- Parallel filters
  - Will be less than 100% accurate but they only vote.
  - RBL, e-mail received trajectory, contact pattern, link hoovering, word / phrase / sequence content filtering and the Reverend Thomas Bayes.

Last line of defence – parallel content filtering

- We are looking for anything with some degree of independence
  - Bayesian word frequency filters
  - Bayesian pattern filtering, (patterns of higher abstraction than words)
  - Different Bayesian filters, (Cormack and Lynam(2007))
  - Prior link hoovering
  - Phrase checking and phrase order, (chance discovery and home-rolled filters)
  - Geographic information from envelope or embedded IP addresses

Bayes theorem

- Consider the appearance of the word “prize”

$$P(spam \mid prize) = \frac{P(prize \mid spam)P(spam)}{P(prize \mid spam)P(spam) + P(prize \mid nonspam)P(nonspam)}$$

We train Bayesian filters on an existing population of spam and nonspam and then use the above to predict. Individual Bayesian filters can reach around 99.5% accuracy.

Generalising Bayes theorem

- Consider the appearance of *pattern*

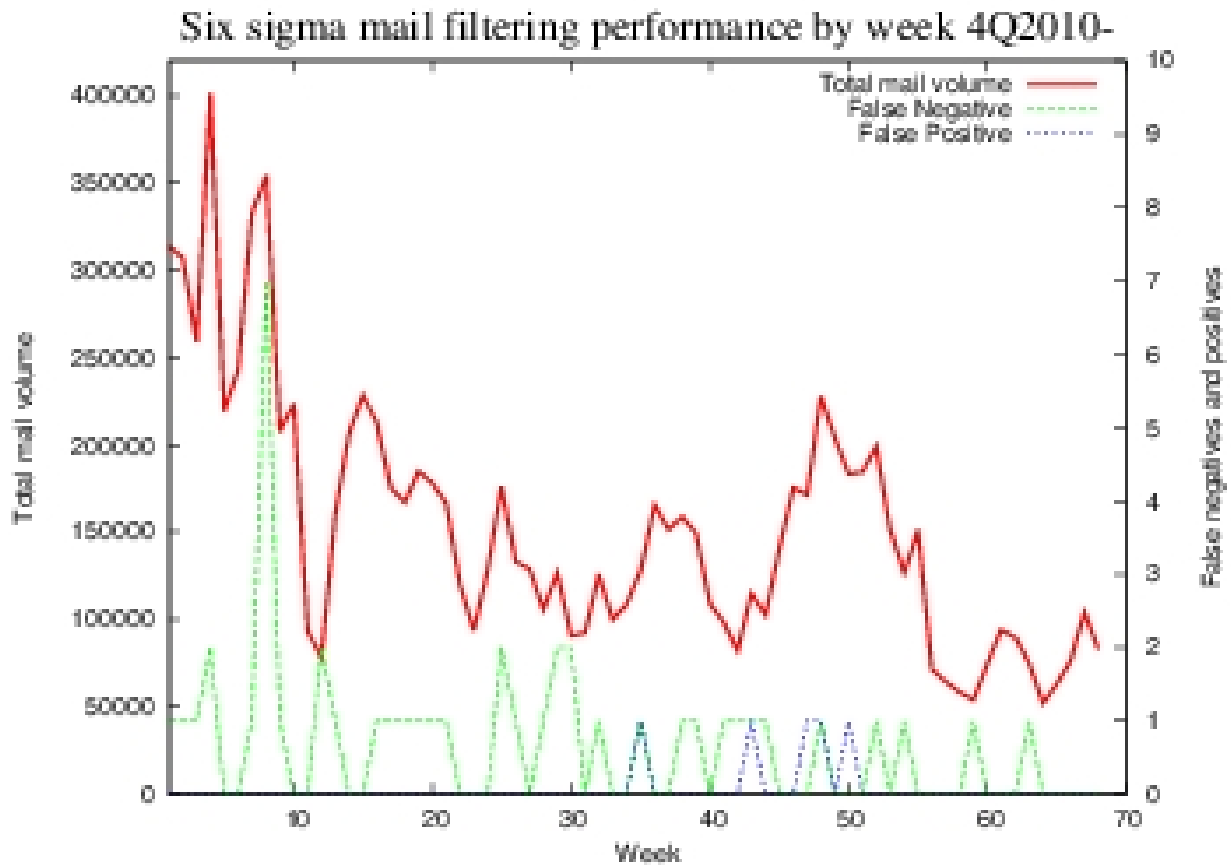
$$P(spam \mid pattern) = \frac{P(pattern \mid spam)P(spam)}{P(pattern \mid spam)P(spam) + P(pattern \mid nonspam)P(nonspam)}$$

Here *pattern* can be literally anything. For example, a mismatch between from:, reply-to: and return-path: addresses and a suspicious route.

Effectiveness of parallel semi-independent Bayesian filters

Number of filters and majority voting	Empirically measured accuracy	Errors per million
1	4σ	6,210
3	5σ	233
5	6σ	3.4
7	?	?

Summary of last 68 weeks



Binary classification of last 68 weeks

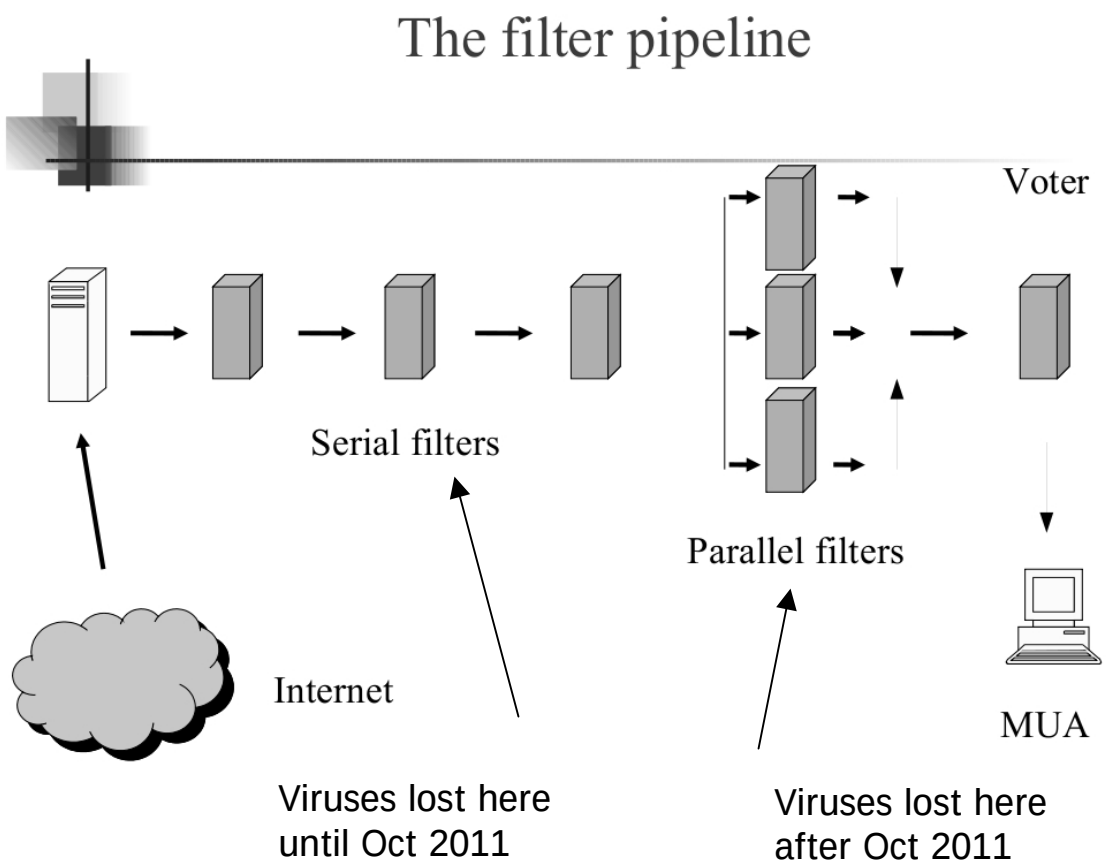
	Junk Mail	Good mail	
Flagged as Junk	10,745,337	5	PPV = 99.99995%
Flagged as Good	47	8,276	NPV = 99.44%
	Sensitivity = 99.99956 %	Specificity = 99.93962%	

Viral penetration: evidence of increasing sophistication

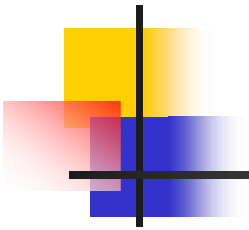
	2008	2011	2012
Viruses hitting virus filters per month*	299	0.5	21

- Note
  - No viruses have reached an end user since 2010 – they always have something else wrong with them.
  - Recent increase started in October 2011.
  - Latest all claim to be from Santander, Barclays and recently Paypal and HMRC

Viral penetration

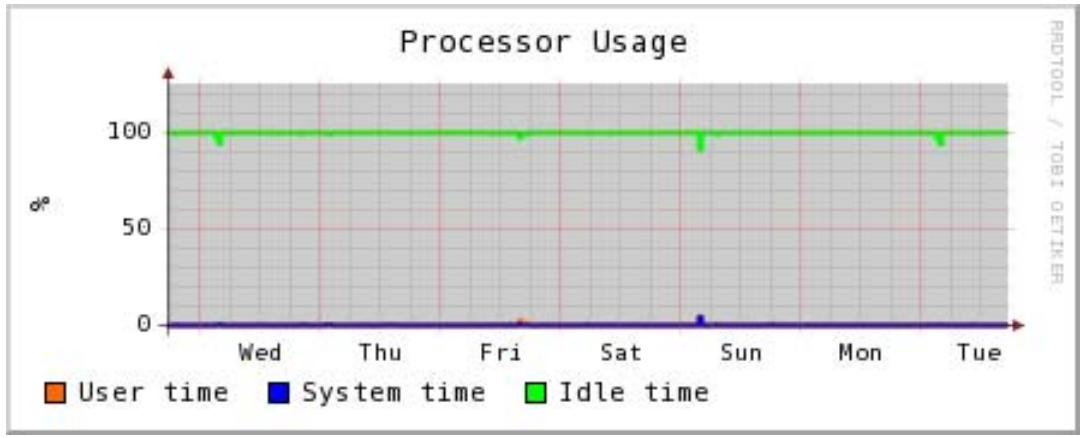


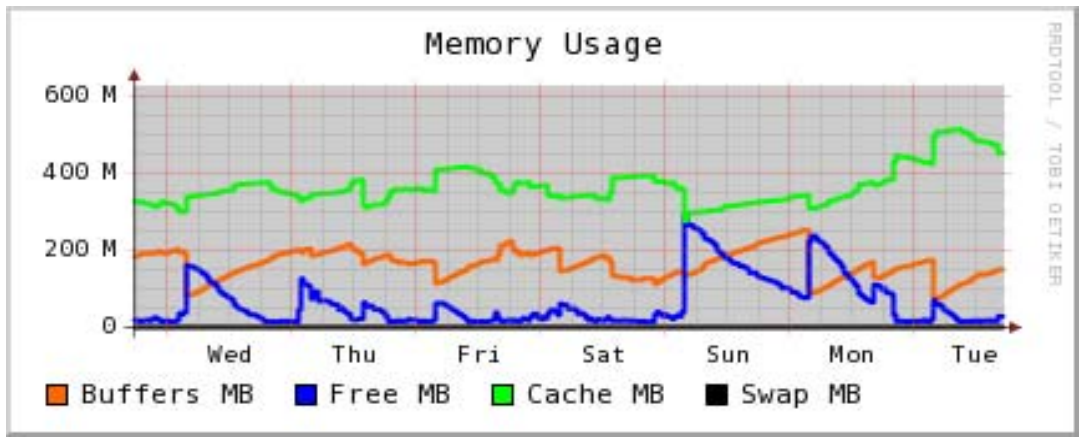
Overview



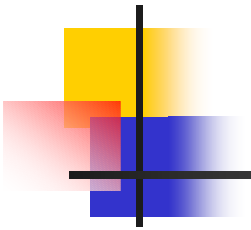
- Overview
- Defence in Depth
- Wrap-up
  - Load on server
  - Things still up our sleeve

Load on server





Things still up our sleeve

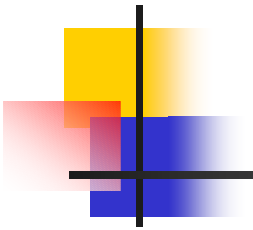


- Could use SPF / Domain Keys but botnets have undermined this. (Imagine receiving a Viagra advert in a bank's envelope by post).
- Still a lot you can do with content filtering
- Community networks – Vipul and so on.

A last word

- Some legislation currently makes things potentially much worse
  - Freedom of Information Act (2000). The Data Protection Act (1998) does not give you immunity from having to release e-mail addresses as a public body.
  - Use section 36 of the FOIA instead.

Conclusions



- It is possible to operate very close to 6 sigma.
- Most junk is still unsophisticated and can be rejected early.
- Some scamming attacks getting sophisticated and a much bigger percentage of all junk.
- Viruses almost always have something else wrong with them allowing early rejection.
- It remains an arms race with continuous evolution of attack and defence.

References

Loads of stuff on Wikipedia:-

http://www.wikipedia.org/

My writing site:-

http://www.leshatton.org/