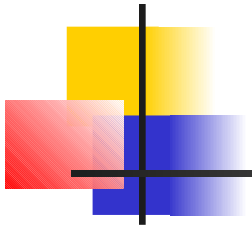
“E-mail Forensics: Eliminating Spam, Scams and Phishing"

Les Hatton

Professor of Forensic Software Engineerng
SEC, Kingston University
Les.Hatton@kingston.ac.uk

Version 1.1: 25/Jan/2013

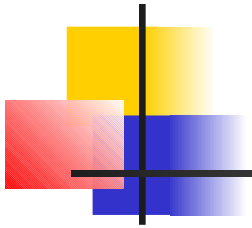
Types of email forensics



- Investigative forensics
  - Generally poring over immense logfiles trying to determine the provenance of an email
- 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

- My servers receive anywhere between 10,000 and 150,000 emails a day. 99.97% are junk.
- 30 scams and about 150 spam messages a month are good enough to make it into my operating theatre to be dissected for further analysis. Detection efficiency is currently 4 mistakes per million messages received.
- BCS (January 2013) reports 20,000 malicious emails a month of which 5% are cyber attacks on UK government networks, (which means they are missing quite a few).

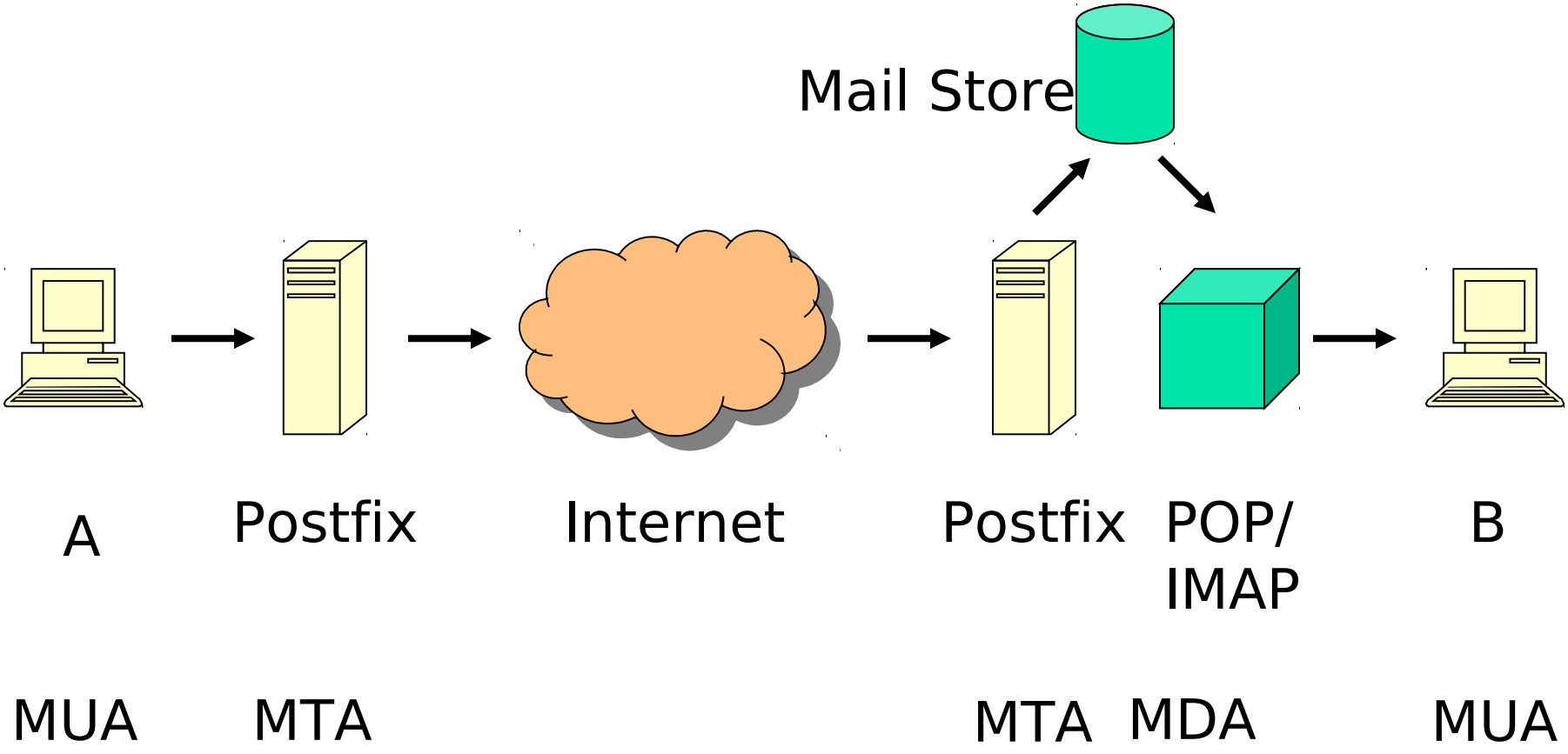
Christmas joy – my operating theatre

	/19.68/[DANGER: SCAM]/10.1/ Tesco Voucher Luc...	Tesco Personal Finance	05/11/12 14:22
	/19.68/[DANGER: SCAM]/10.9/ BELOVED	Mrs. Vera Shumejda	06/11/12 23:58
	/19.69/[DANGER: SCAM]/8.1/ Contract Deal	NNPC OIL	07/11/12 12:32
	/25.98/[DANGER: SCAM]/4.6/ Promotional Product...	4imprint	07/11/12 15:00
	/41.45/[DANGER: SCAM]/3.7/ 15% Off Bags & Cas...	Laptopshop	09/11/12 00:00
	/19.69/[DANGER: SCAM]/7.7/ Member, I added the...	Bonus Team	09/11/12 23:35
	/19.70/[DANGER: SCAM]/7.6/ Sub: Urgent & Confi...	George Mbali	12/11/12 20:41
📎	/19.70/[DANGER: SCAM]/6.9/ Investment Assista...	From:Dr.A. T.Ntsaluba	13/11/12 16:34
	/19.68/[DANGER: SCAM]/10.5/ Congratulation Fro...	ONLINELOTTO	13/11/12 22:41
	/19.68/[DANGER: SCAM]/10.9/ Congratulation Fro...	ONLINELOTTO	14/11/12 12:27
	/26.99/[DANGER: SCAM]/4.5/ Your Amazon.co.uk S...	Amazon Services	15/11/12 15:48
	/19.70/[DANGER: SCAM]/6.7/ REQUEST	Hussani Ali	17/11/12 17:14
📎	/19.68/[DANGER: SCAM]/6.4/ Tax Refund Notification	HM Revenue & Customs	19/11/12 05:45
📎	/19.68/[DANGER: SCAM]/7.7/ Congratulations Che...	Australian International Lotte...	20/11/12 21:15
📎	/19.71/[DANGER: SCAM]/9.9/ Your account is limit...	security@paypal.co.uk	21/11/12 08:59
	/19.69/[DANGER: SCAM]/8.0/ Winning No: MSP798...	Microsoft Corporation®	23/11/12 13:24
	/19.70/[DANGER: SCAM]/8.0/ Winning No: MSP798...	Microsoft Corporation®	23/11/12 13:25
	/19.68/[DANGER: SCAM]/5.4/ Partner Required.	Mr. Michael Woo Lee.	26/11/12 04:45
	/19.70/[DANGER: SCAM]/8.5/ Winning No: MSP798...	Microsoft Corporation®	26/11/12 11:35
	/19.68/[DANGER: SCAM]/8.8/ You have an importa...	Coca-Cola UK.	27/11/12 07:58
	/19.69/[DANGER: SCAM]/8.8/ YOUR ATM PAYMENT ...	United Nations Office	29/11/12 21:57

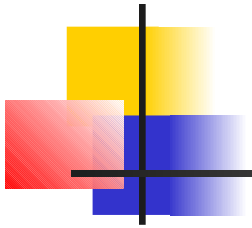
Where does it all come from ? (this weeks top ten)

1. China
2. Iceland
3. USA
4. Russia
5. Israel
6. Germany
7. Indonesia
8. Australia
9. South Korea
10. Canada

The basics: how mail goes from 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)

Date: ...
DATA (an MTA responds)

From: alice@send.com

To: bob@receive.com

Reply-To: ...

Message-ID: ...

Subject: ...

Blah blah blah

Disconnect

MTA

MUA

The basics: Headers and Content

Send MTA

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)

....

Receive MTA

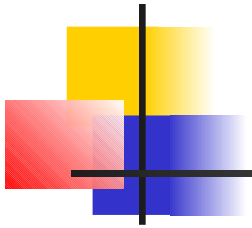
Accept (loses HELO)

Reject or Discard via HELO, …

Backscatter ! Bounce via From:

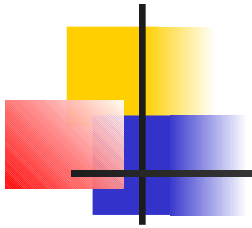
*** Reject as early as possible ***

The basics: anybody can read it



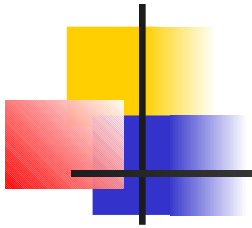
- Nearly all e-mail is sent in clear as if you had written it on a postcard and asked a complete stranger to post it for you.
- It can be arbitrarily spoofed

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

Recent load

- For a single mail-server handling mail for 8 domains in 12-19 November, 2012

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**

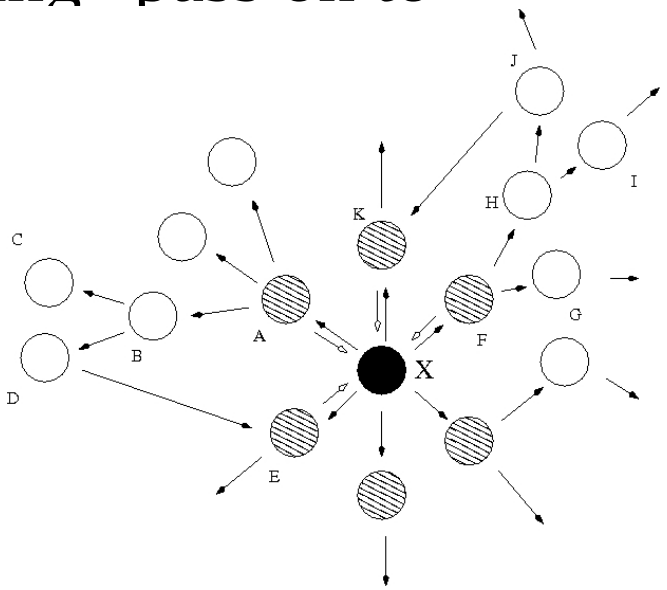
Heya,

V I A G R A

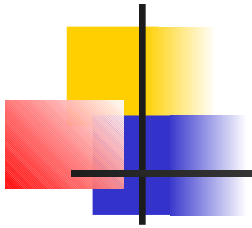
WWW . NEVOB . COM

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 recently, (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 many spammers are not very sophisticated.

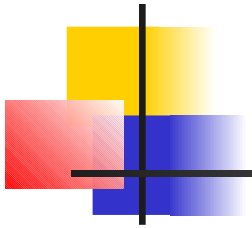
Bayes theorem

- 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.

The threat landscape – e-mail borne toxins



- Spam
- Scams
  - Nature
  - Patterns
- Intrusions

The threat landscape – Nature of scams

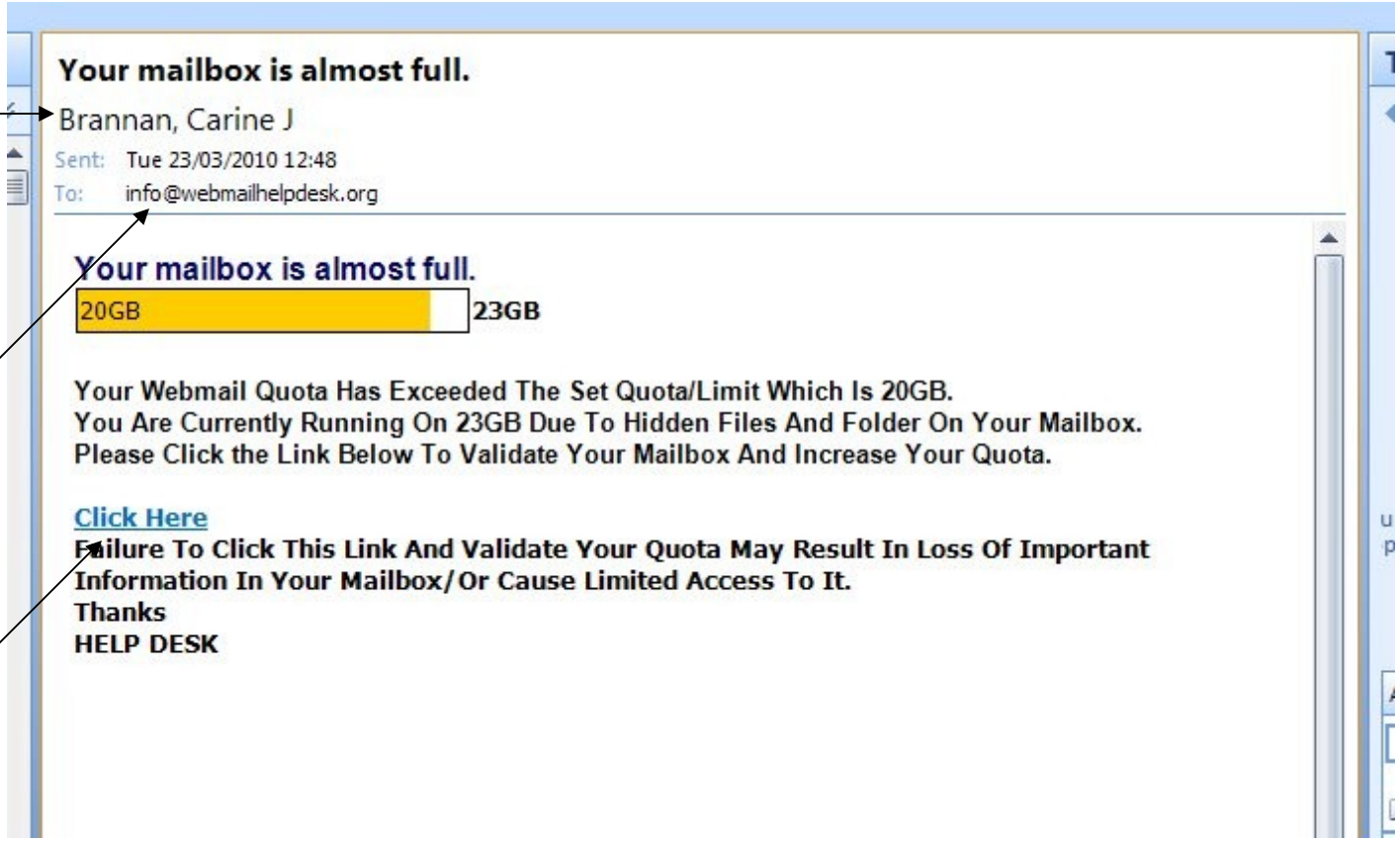
- Scams can be much more of a challenge
  - Lotteries (easy)
  - “My left leg has been biten off by a mad cheetah and I have $4 million in da trouser leg” (Nigerian 419) (easy)
  - Phishing for account details (increasingly convincing)
  - Pharming, DNS hijacking, … (ditto)

The threat landscape – account hijacking attack

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

From: address does not exist

Link –zyef.9hz.com does not exist



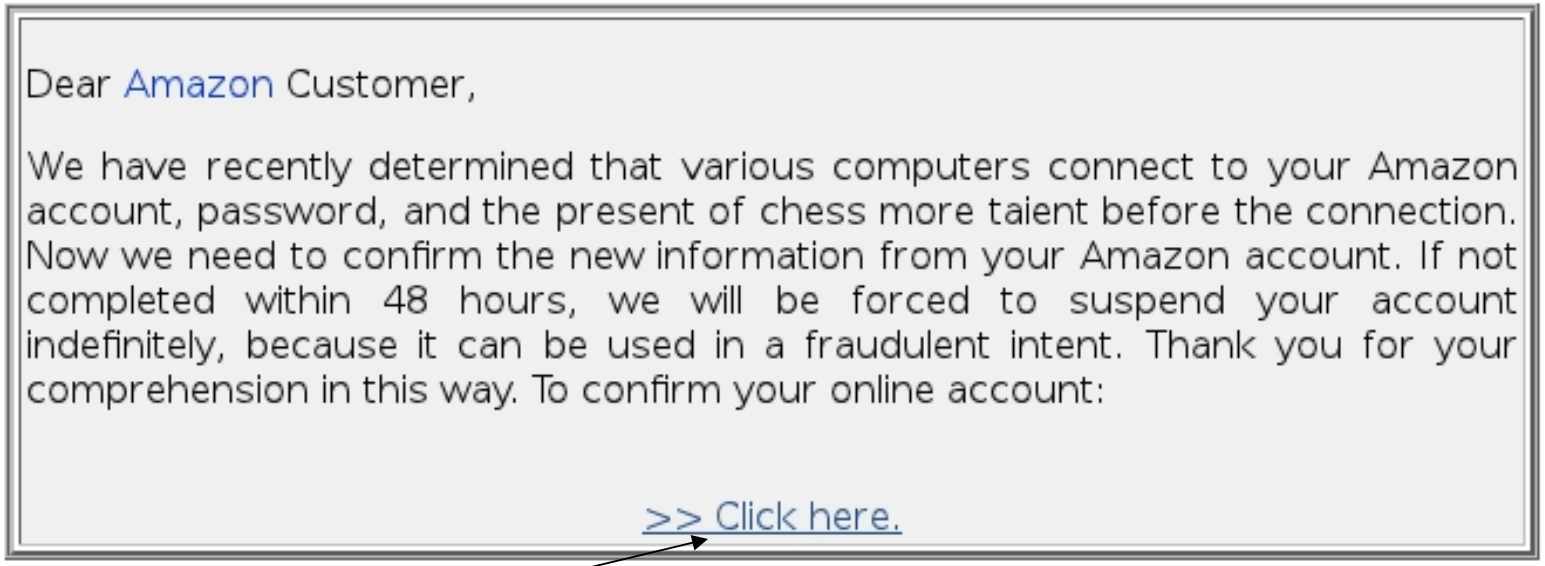
The threat landscape – friend in trouble scam

From: [redacted] <[redacted]@yahoo.co.uk>
Date: 21 January 2013 05:00:53 GMT
To: undisclosed recipients: ;
Subject: /47.02//3.4/ Urgent Assistance
Reply-To: [redacted]@yahoo.co.uk

Hope you read this soon, I'm in Limassol, Cyprus and i lost my bag with passport and credit card. The embassy is willing to assist me fly back without my passport but I must pay for my ticket and hotel bills. Unfortunately I have no money left, my credit card would have helped, but it was also in the bag i lost. I have notified my bank but they need more time to give me a new credit card so for now my bank account as been blocked for security reason.
I wonder if you can lend me some money as soon as possible. I'll give you back as soon as I get back. But right now I definitely need to get on the next flight.Please contact me by e-mail,i lost my cell phone as well.I'm waiting for your reply.Thanks
[redacted]

Solution: Switch your email account to somebody else
(Try googling “Hacking yahoo mail accounts”).

The threat landscape
- supplier scams, Jan 2013



Dear Amazon Customer,

We have recently determined that various computers connect to your Amazon account, password, and the present of chess more taient before the connection. Now we need to confirm the new information from your Amazon account. If not completed within 48 hours, we will be forced to suspend your account indefinitely, because it can be used in a fraudulent intent. Thank you for your comprehension in this way. To confirm your online account:

>> Click here.

Link points to http://222.66.64.165:5800/uk.html

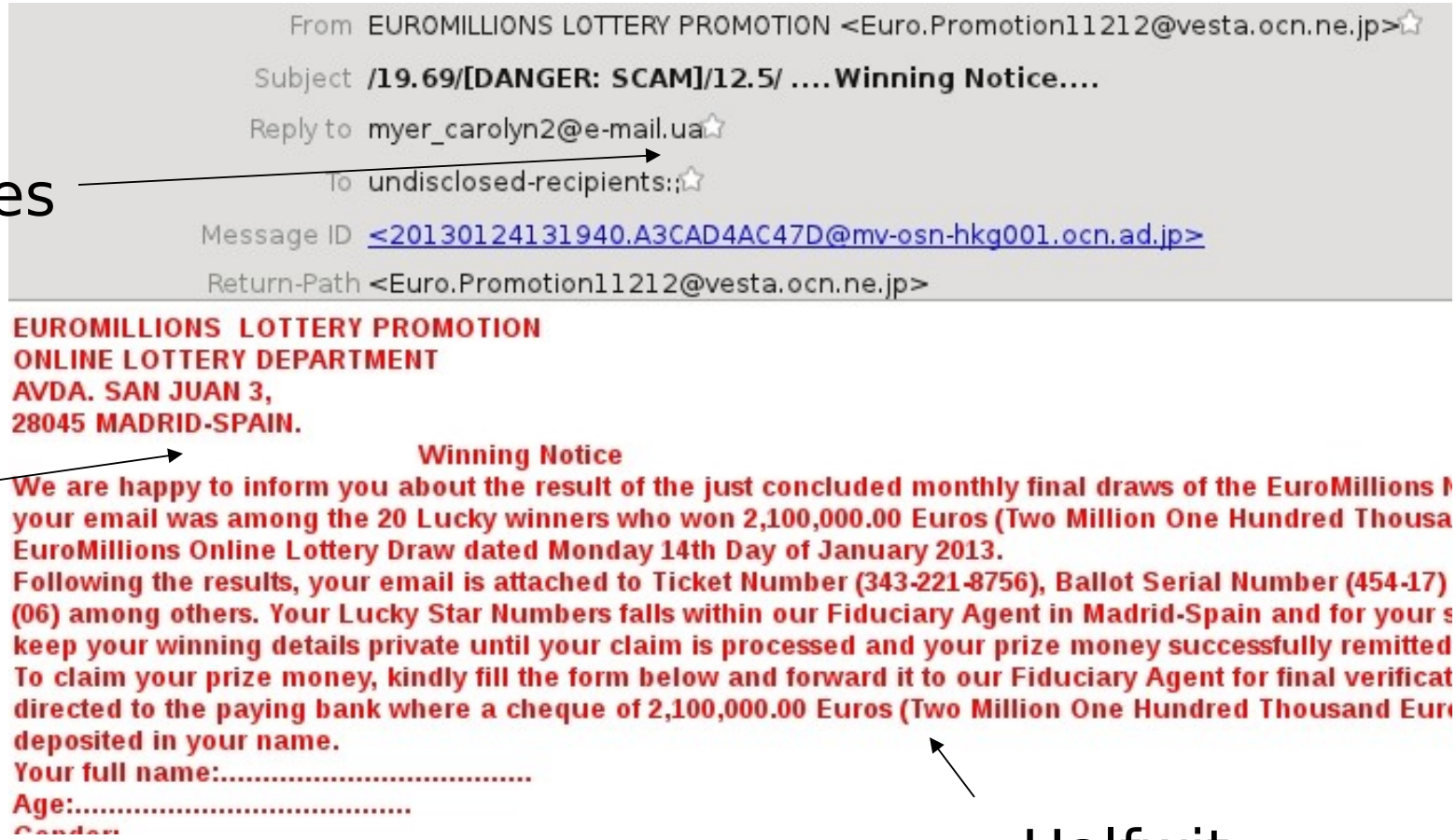
Link in Shanghai

The threat landscape - lottery scams, Jan 2013

Japan

Emirates

Spain



From EUROMILLIONS LOTTERY PROMOTION <Euro.Promotion11212@vesta.ocn.ne.jp>
Subject /19.69/[DANGER: SCAM]/12.5/ ....Winning Notice....
Reply to myer_carolyn2@e-mail.ua
To undisclosed-recipients:;
Message ID <20130124131940.A3CAD4AC47D@mv-osn-hkg001.ocn.ad.jp>
Return-Path <Euro.Promotion11212@vesta.ocn.ne.jp>

EUROMILLIONS LOTTERY PROMOTION
ONLINE LOTTERY DEPARTMENT
AVDA. SAN JUAN 3,
28045 MADRID-SPAIN.

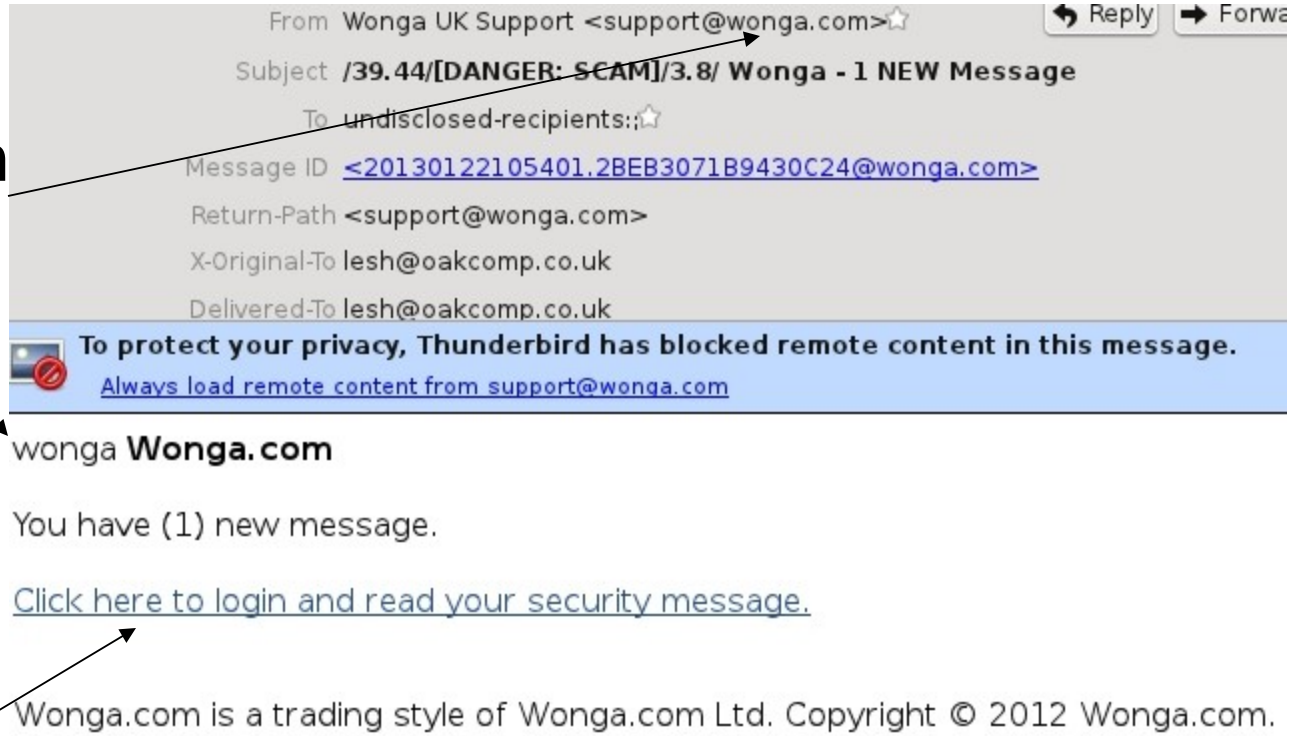
Winning Notice

We are happy to inform you about the result of the just concluded monthly final draws of the EuroMillions M
your email was among the 20 Lucky winners who won 2,100,000.00 Euros (Two Million One Hundred Thousa
EuroMillions Online Lottery Draw dated Monday 14th Day of January 2013.
Following the results, your email is attached to Ticket Number (343-221-8756), Ballot Serial Number (454-17)
(06) among others. Your Lucky Star Numbers falls within our Fiduciary Agent in Madrid-Spain and for your s
keep your winning details private until your claim is processed and your prize money successfully remitted
To claim your prize money, kindly fill the form below and forward it to our Fiduciary Agent for final verificat
directed to the paying bank where a cheque of 2,100,000.00 Euros (Two Million One Hundred Thousand Euro
deposited in your name.
Your full name:.....
Age:.....

Halfwit

The threat landscape - financial scams, Jan 2013

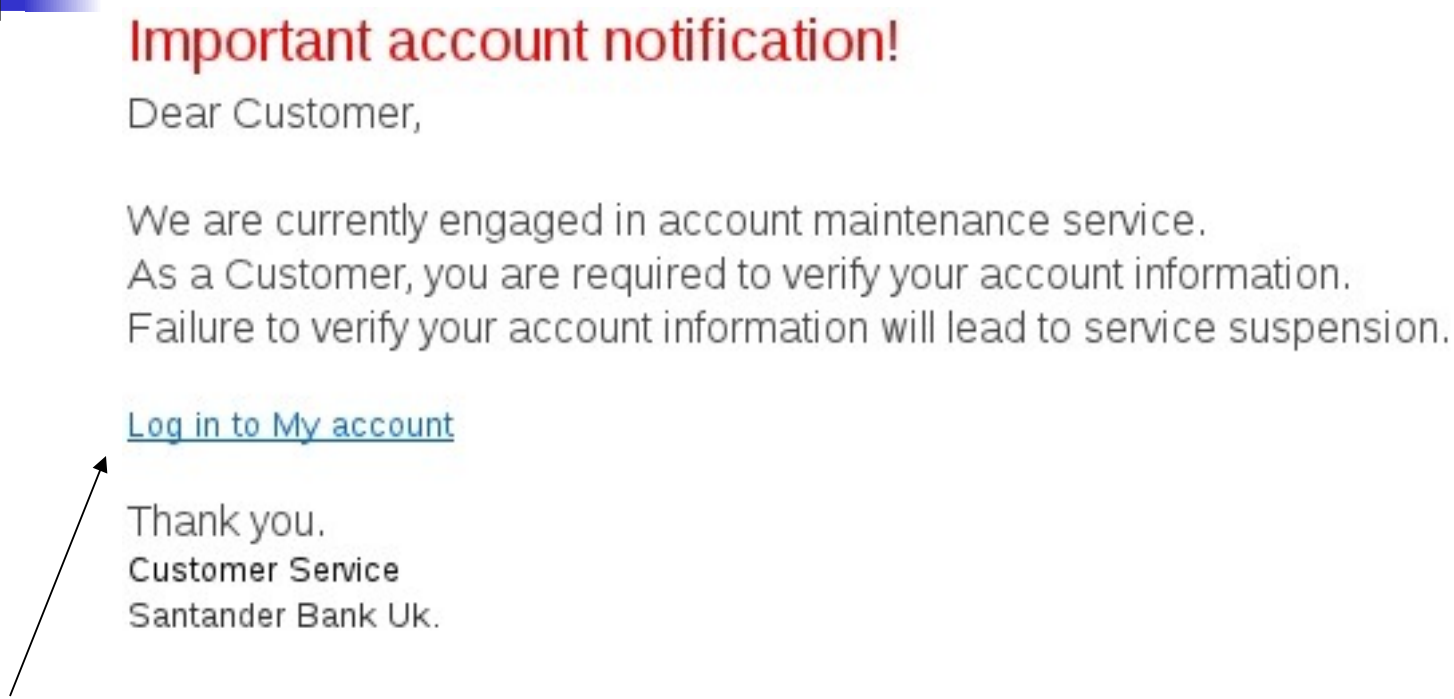
Based in London



Link points to http://freco.com/log/wonga/index.html

Domain (202.78.200.175) registered in Jakarta

The threat landscape
- bank scams, Oct 2012



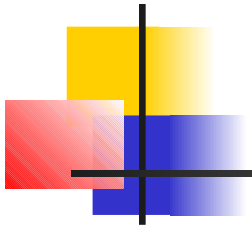
Link points to http://unisonlubricant.com/skin/frontend...

Domain in India, hosted on server in Los Angeles

The threat landscape – scam patterns

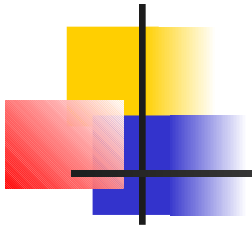
- Need to recognise structure
  - Word frequency sometimes useful – e.g. “codicil”
  - Recognition of phases better
    - The tease (419) or threat (account hacking)
    - The link to contact, (always bogus)
    - The sign-off
- Link hoovering

The threat landscape – mitigating the attacks



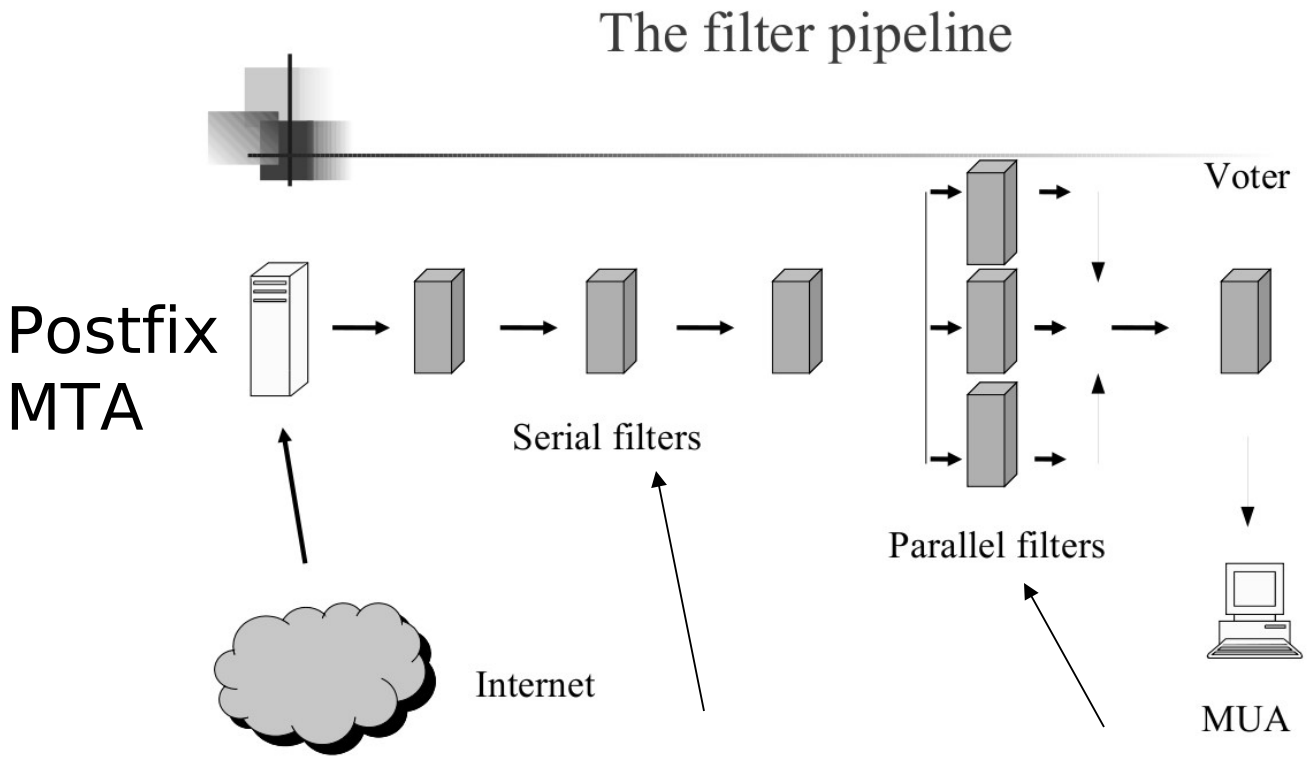
- Use an Internet Service Provider which is reasonable at filtering junk. Most are not.
- NEVER give details away in an email. No responsible company would ever ask.
- Do not click on links in emails
- Turn HTML rendering off in your mail program
- Configure your browser not to download remote content, (starve toxic mailers from any feedback)

Overview



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

Architectures



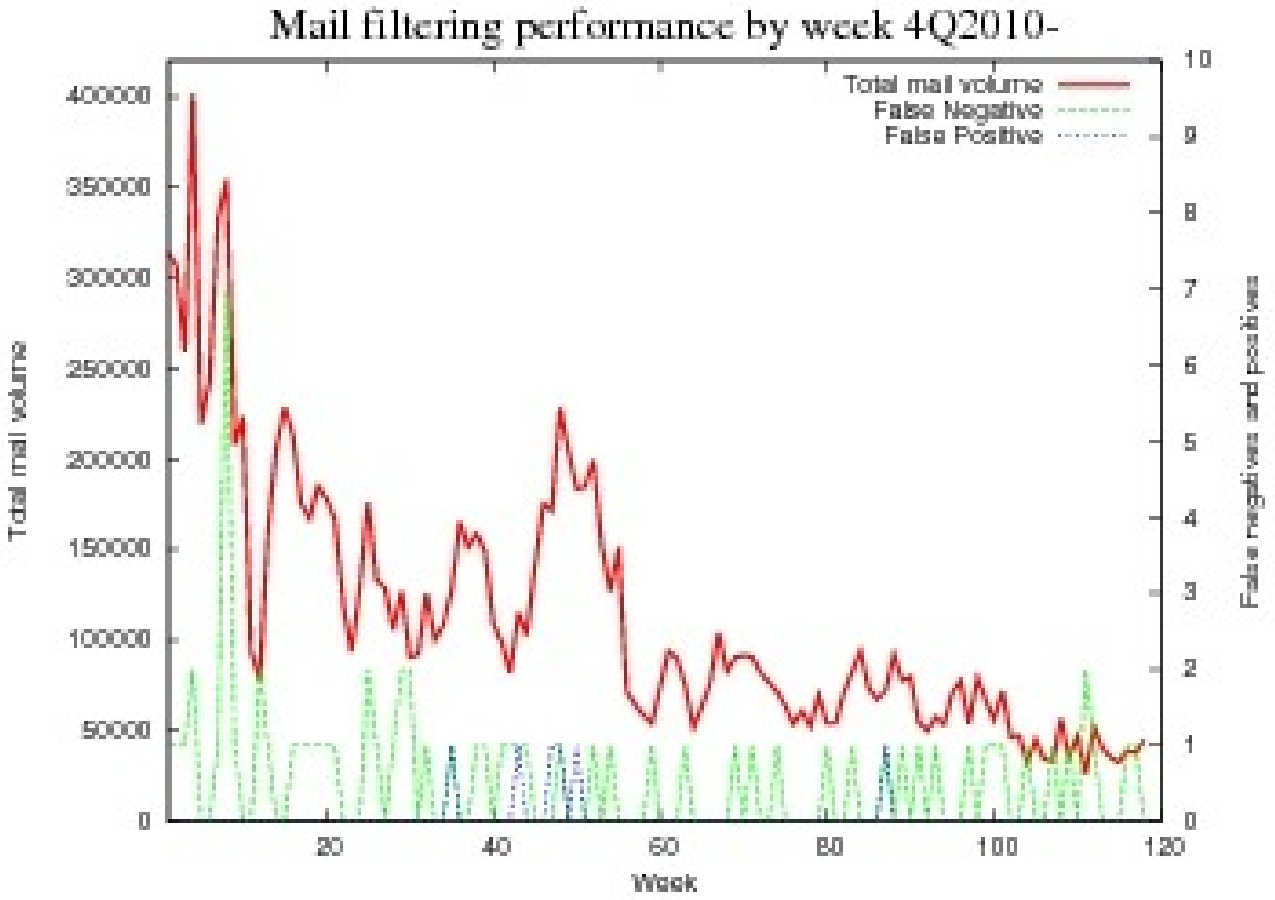
Must be *individually* very accurate

Must be *collectively* very accurate

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.

Summary of last 120 weeks

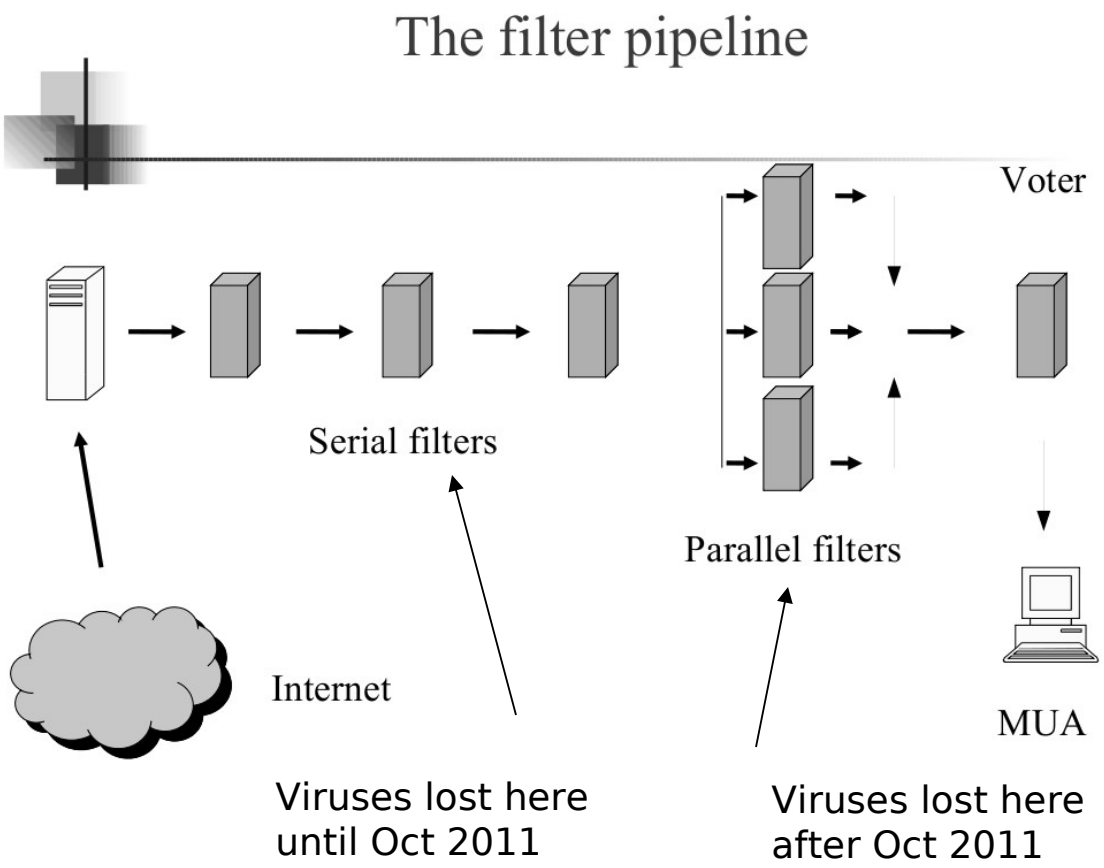


Viral penetration: evidence of increasing sophistication

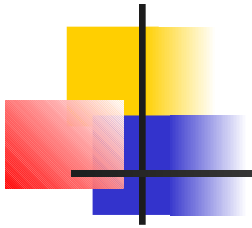
	2008	2011	2012
Viruses hitting virus filters per month*	299	0.5	21

- 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

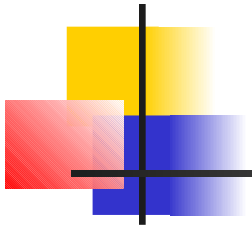


- Overview
- Defence in Depth
- Wrap-up

A last word

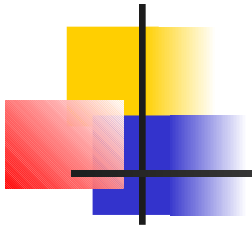
- Some legislation currently makes things potentially much worse
  - Freedom of Information Act (2000). The Data Protection Act (1998) does not give you immunity from having to release e-mail addresses as a public body.
  - Use section 36 of the FOIA instead.

Conclusions



- It is possible to detect and remove nearly.all toxic email - 99.9996% accuracy.
- Much of the really unsophisticated junk has disappeared.
- Some scamming attacks are getting very sophisticated and are a bigger percentage of all junk.
- Viruses almost always have something else wrong with them allowing early rejection, (so far).
- 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/