Anti-Spam & Anti-Phishing a Reality Check

Richard Clayton

UCL

16th March 2005

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Demon

Structure of this talk

- Proof-of-work as an anti-spam device
  - why “hash-cash” isn’t a silver bullet
- Extrusion Detection for spotting spam
  - a practical & effective approach for ISPs
- Tackling Phishing
  - are we upping the ante fast enough ?

Structure of this talk

- Proof-of-work as an anti-spam device
  - why “hash-cash” isn’t a silver bullet
- Extrusion Detection for spotting spam
  - a practical & effective approach for ISPs
- Tackling Phishing
  - are we upping the ante fast enough ?

Is spam an Economics problem?

- Many argue that problem is “Economics”
  - no charge for sending email
  - hence “one in a million” response is profitable
- Hence the fix is to charge for email ?
  - real money? 1p/email => $180 billion annually
    - phone companies would love this -- would we ?
  - eCash? doesn’t seem to have happened yet !

Proof-of-work schemes I

- Idea is to show that you care enough about your email to have expended effort in doing a (rather pointless) calculation first
  - there are ideas for useful calculations eg “Bread Pudding Protocols” (Jakobsson & Juels 1999) but generally just warms up the planet ☹
- Original idea: Dwork & Naur : Crypto 1992
  - used central server ☹☹☹

Proof-of-work schemes II

- Reinvented as HashCash (Adam Back, 1997)
  - compute HASH(destination, time, nonce) such that result has “n” leading zeros
    - 2^n hard for sender, but trivial check for receiver
- Dwork, Goldberg, Naor (Crypto 2003)
  - analyse a function limited by memory speed
  - small variation between systems (factor of 4)
  - so this is much better than using classic HASH

Email statistics

- November 2003 (consistent stats available)
  - 2.30×10^8 Internet hosts (ISC)
  - 5.13×10^8 Internet users (Radicati)
  - 5.70×10^{10} emails sent daily (Radicati)
  - 56% of all email is “spam” (Brightmail)
- Hence the average situation is
  - 60 spam (& 50 real) emails per person per day
  - 125 real emails per host per day

What about “mailing lists” ?

- Expect to delegate proof-of-work analysis
- Lists common, but no published figures
- Inspected logs at Demon (200K users)
  - this was after a spam filtering stage
  - consider identical source but >10 destinations
  - approximately 40% are of this form
- ie: reduce total to 75 emails per host per day
  - “back of envelope”, but only magnitude matters

How much work must we prove?

- Legitimate hosts must be able to send 75 emails per day (best case situation)
- Must reduce spam from 3.2×10^{10} per day
- Must allow for factor of 4 in capabilities
- Must assume spammers work 24 hours per day, but legitimate hosts may be switched off when not being actively used

… so all we need to do is to pick "n"

Economic analysis I

- Spammers charge 0.001 to 0.030¢ per email
  - survey in Goodman & Rounthwaite, 2004
- PC costs $500 / three years 50¢ per day
  - and pay electricity bill! 25¢ per day
- Spammer invests $50K and buys 100 PCs:
  - Salary $30K/annum 100¢ per day
  - So break-even at 35,000 emails/day/PC if can charge 0.005¢ each (ie: total 3.5 million /day)

[Scott Richter does 21 million/day @ 0.020¢]

Economic analysis II

- But spammers used to charge 0.1¢ per email (which leads to a break even rate of 1750)
- Spam response rates badly documented
  - Ms Betterly (WSJ Nov 2002) : 0.0023%
  - 0.0126% Iraqi Cards ("four times normal")
- If 0.003% and 0.1¢ then cost of ads is $33/sale. Only viable for some products
  - $50/mortage lead; $85/cellphone, $60/pills


A
A
A
A

Economic analysis III

- Iraqi cards article (NYT 9 July 03) goes on:
  - best days: $5000 profit per million emails
    ie: half a cent per email in commission
  - printer ink: $500 to $1200 per million emails
    ie: 0.05¢ to 0.12¢ per email in commission
- BUT note that legitimate email response rates are expected to be 0.7 to 1.6%
- Obviously wise to own more of value chain

Economic conclusion

- Good guys
  - 75 emails/host (best case)
- Bad guys
  - 1750 emails/host (if price returns to 0.1¢)
  - but this will exclude low margin products ☺
- BUT bad guys have “factor of 4” advantage
- So some headroom here, but not lots & lots

AT CURRENT RESPONSE RATES

Security analysis I

- Lots of ***0wned*** machines out there
  - SORBS: 1.2M HTTP, 1.4M SOCKS proxies
  - Recent viruses have hit million+ machines each
- Currently easy to spot ***0wned*** machines
  - they send a lot of email!
- But what if they computed "proof-of-work"
  - quietly giving results to sender systems
  - hard to spot and so likely to be long-lived

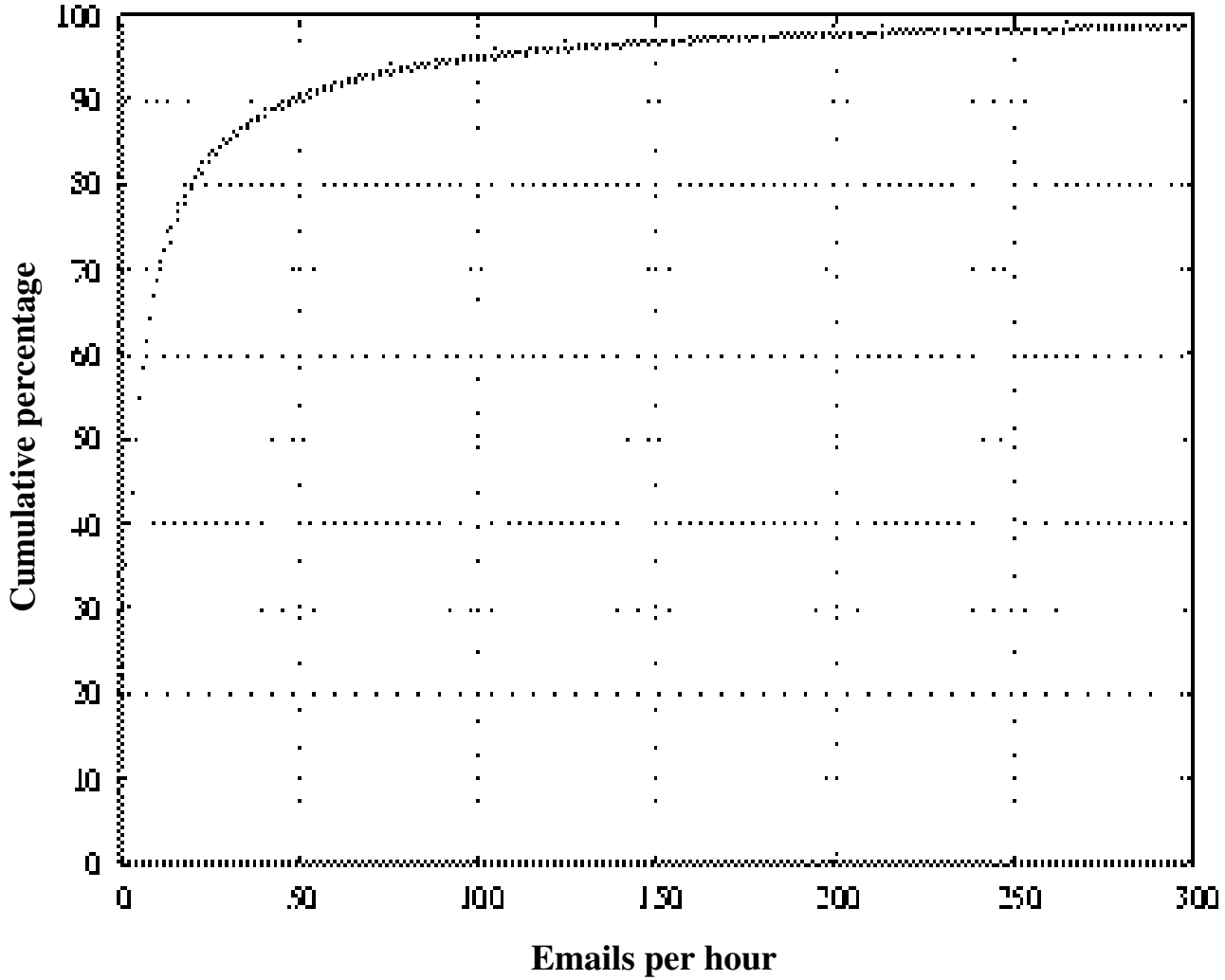
Security analysis II

- Nov 2003, 3.2 x 10^{10} spam emails
- Suppose one million machines hijacked for proof-of-work (spammers share them out!)
- So, they only need to do 32,000 each
  - consistent with ISP figures for abused hosts
- If want 99% of our mailboxes to be "real" then must restrict spam to 250/host per day
- & for just 0.1% to be spam, then 25 per day

Security conclusion

- Good guys
  - 75 emails/host (best case)
- Bad guys
  - 250 emails/host (if spam is just 1% of mailbox)
- No “factor of 4” advantage this time
  - unless spammers can choose *0wned* machines
- So **very** limited headroom
  - & impossible to reach “one in a thousand” level

Real hosts : daily rates



93.5% < 75

BUT

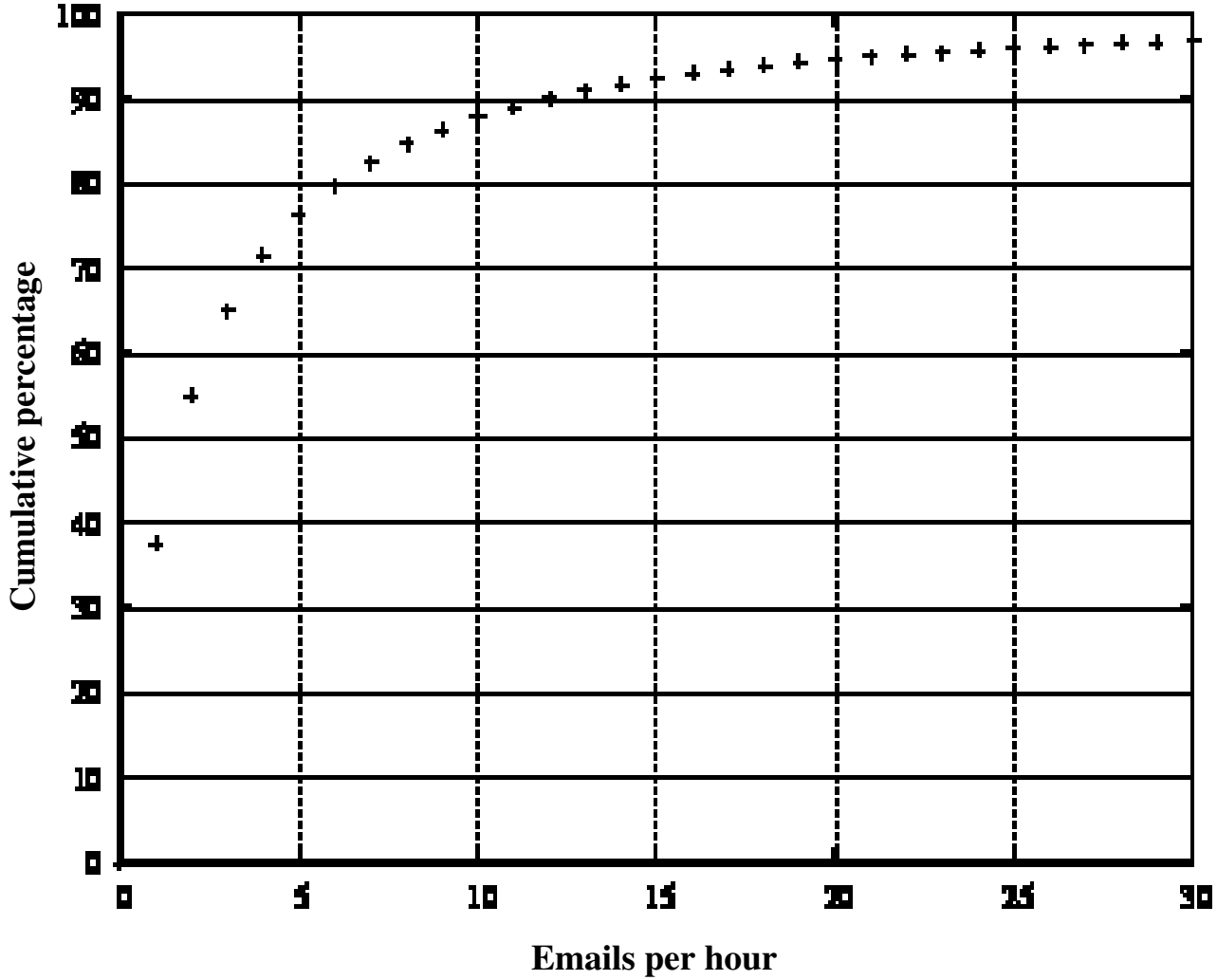
0.13% > 1750

1.56% > 250

viz: this impacts real senders

albeit some are just [exempted] mailing lists

Real hosts : hourly rates



Spammers run
24 hours/day,
real users don't!

1% > 73/hour
i.e.1750/day

13% > 11/hour
i.e. 250/day

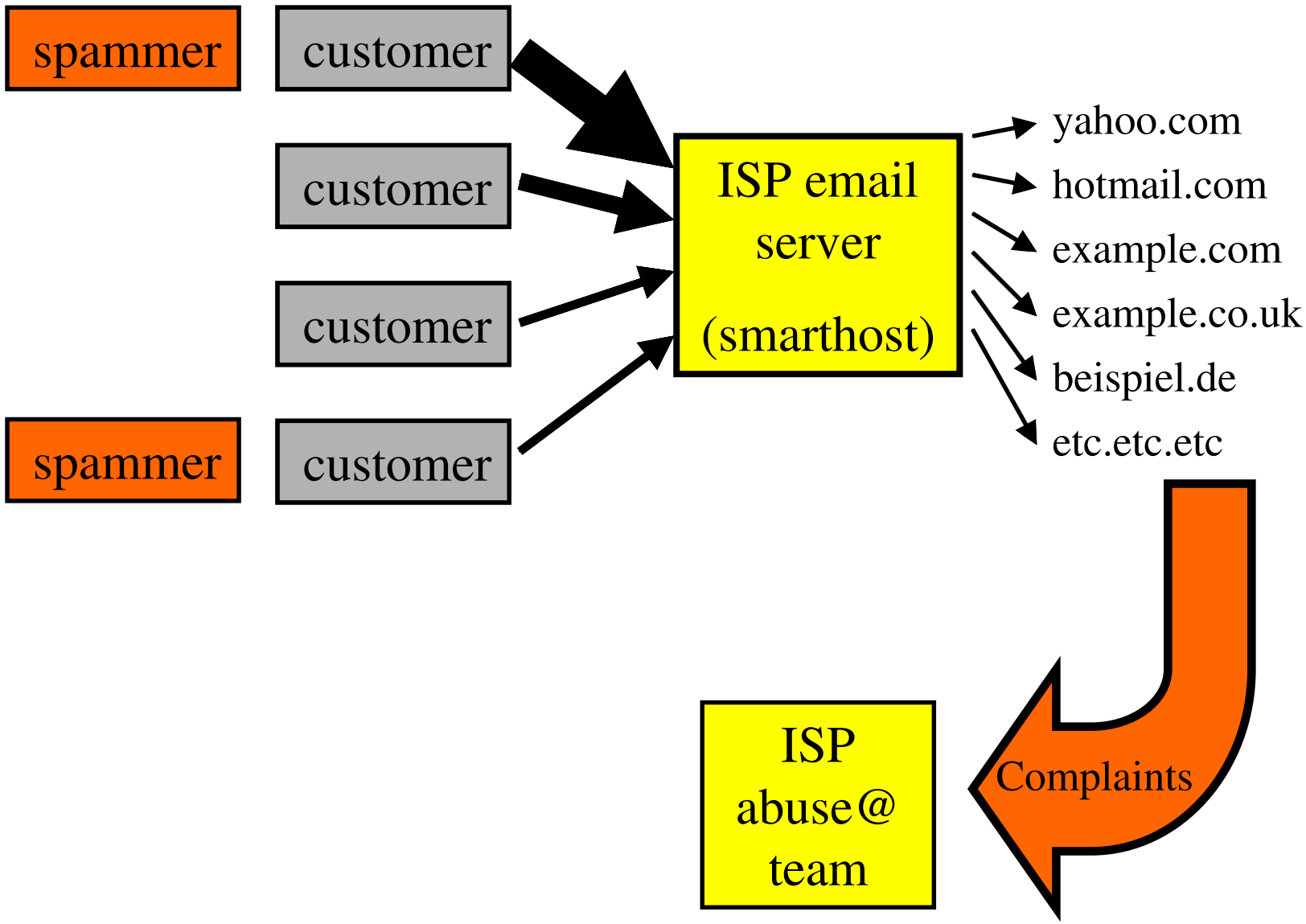
viz: this impacts
lots of people

Proof-of-Work conclusions

- HashCash payment for email is attractive
- BUT spammer profit margins per sale mean that some will be able to afford the PCs to do the proof-of-work required
- BUT hijacking of end-user machines means impractical to restrict them to 1% of email
- Simplistic proof-of-work just doesn't work!

Structure of this talk

- Proof-of-work as an anti-spam device
  - why “hash-cash” isn’t a silver bullet
- Extrusion Detection for spotting spam
  - a practical & effective approach for ISPs
- Tackling Phishing
  - are we upping the ante fast enough ?

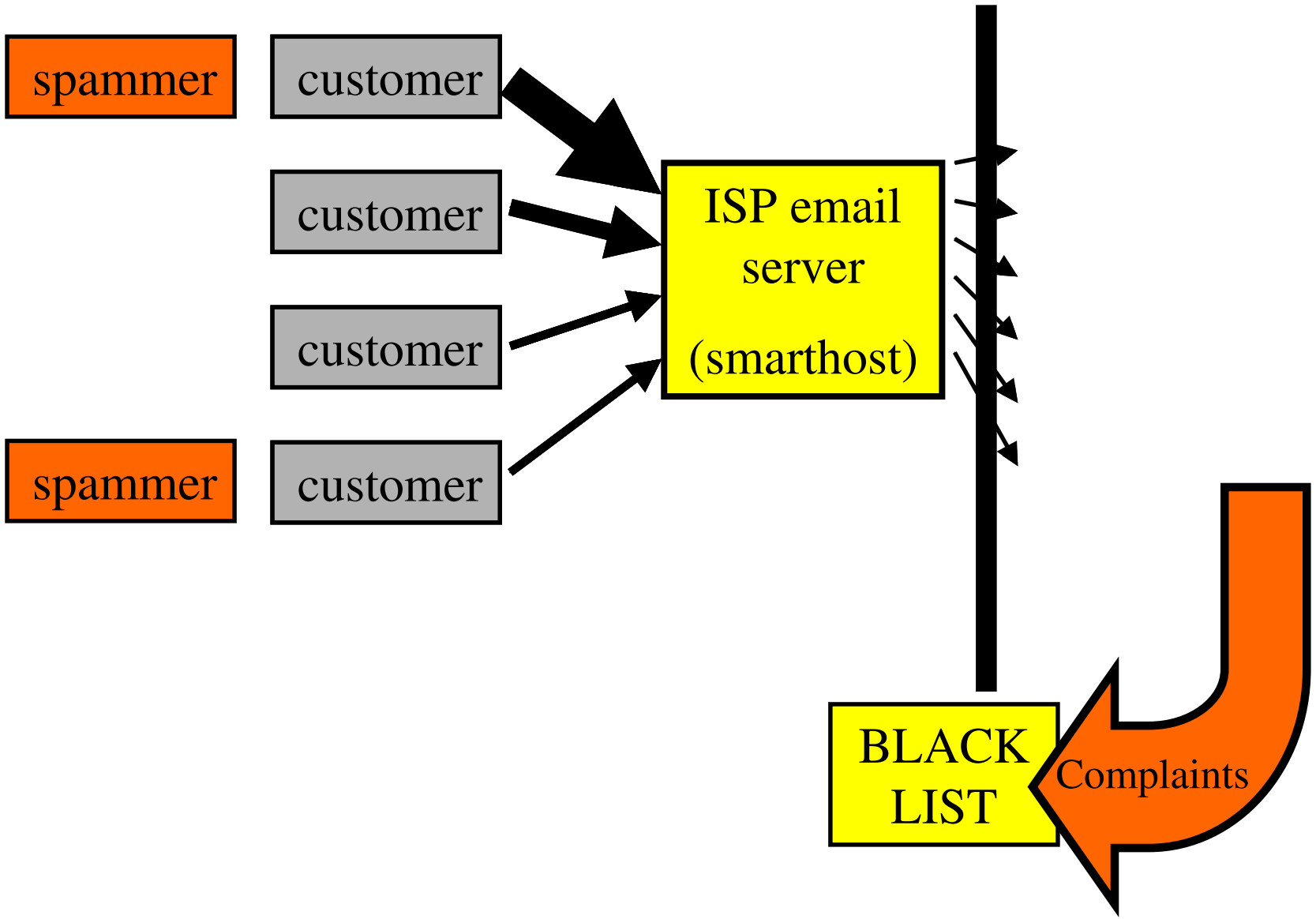

spammer
customer
customer
customer
spammer
customer
ISP email server (smarthost)
yahoo.com
hotmail.com
example.com
example.co.uk
beispiel.de
etc.etc.etc
ISP abuse@ team
Complaints

Current (Mar 05) problems for ISPs

- Insecure customers
  - very few real spammers in the UK!
- Open proxies
  - mainly “trojans on non-standard ports”
- SMTP AUTH
  - Exchange “admin” accounts + *many others*
- Systems still insecure “out of the box”
  - XP SP1 is compromised before secured

ISP's real problem

- Blacklisting of IP ranges & smarthosts
  - **`listme@listme.dsbl.org`**
- Rapid action necessary to ensure continued service to all other customers
- But reports may go to the blacklist and not to the ISP (or will lack essential details)


spammer
customer
customer
customer
spammer
customer
ISP email server (smarthost)
BLACK LIST
Complaints

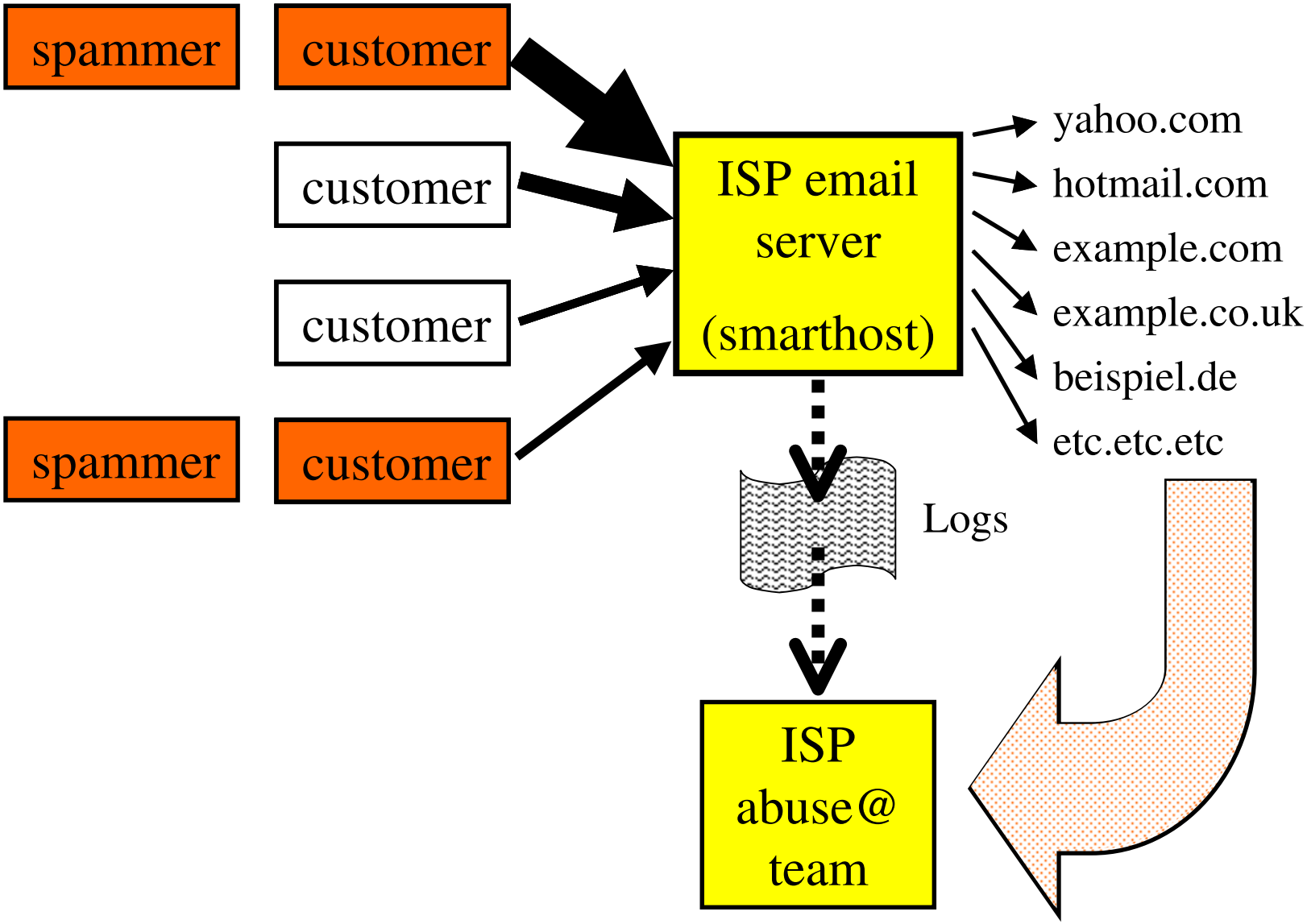
Why spotting spam is hard

- Expensive to examine outgoing content
- Legal/contractual issues with blocking
  - and “false positives” could cost you customers
- Volume is not a good indicator of spam
  - many customers with occasional mailshots
- “Incorrect” sender doesn’t indicate spam
  - many customers with multiple domains
  - many customers forward email to others

Key insight

- Lots of spam is to ancient email addresses
- Lots of spam is to invented addresses
- Lots of spam is blocked by remote filters

- Can process server logs to pick out this information. Spam has delivery failures whereas legitimate email mainly works


spammer
customer
customer
customer
spammer
customer
ISP email server (smarthost)
yahoo.com
hotmail.com
example.com
example.co.uk
beispiel.de
etc.etc.etc
Logs
ISP abuse@ team

My log processing heuristics

- **Report “too many” failures to deliver**
  - more than 40 works pretty well
- Ignore “bounces” !
  - have null “< >” return path, these often fail
  - detect rejection daemons without < > paths
- Ignore “mailing lists”
  - most destinations work, only a few fail
  - more than one mailing list is a spam indicator!

Bonus! Also detects viruses

- Common for mass mailing “worms” to use address book (mainly valid addresses)
- Many now scan the browser cache and (Swen) accessing Usenet servers
  - so many addresses now invalid or badly formed
  - plus remote sites may reject incoming malware
- Many different HELO strings
- **So virus infections are also detected**

Evaluation at large UK ISP

- 28 day period (1-28 March 2004)
- No public holidays (ie 20 working days)
- 85K active customers (of 200K total)
- 33.4 million emails (51.8 million destinations)
- System had been in production 6 months
  - hence there are no edge effects (initially was spotting dozens of problems per day)
- No major virus events occurred

Evaluation methodology

- Manually check all reports from system
  - spamming patterns are very obvious
- False positive occurs when report is wrong!
- False negatives assessed by comparison of results with manual inspection of results from a far more sensitively tuned version.
  - also examined all other reports of viruses etc

Results (total over 28 days)

Abuse Type	total detected	false positive	false negative
Real Spammers	0	0	0
Open Servers	56	69	10
Virus Infection	29	6	4
Email loops	14	3	0

Looking more closely

Abuse type	total	False+ve	False -ve
Open Servers	56	69	10

FALSE POSITIVES:

36 customers running multiple genuine mailing lists

22 customers with >40 delivery failures during one day

11 assorted other reasons (see paper)

FALSE NEGATIVES:

7 (of the 10) were one “cutecandy” spammer (using a fixed sender string & remote sites accepted a dictionary attack)

Future work

- Spammers will evolve!
  - Spam resembling bounces will be hard to spot
  - Valid MAIL FROM will be harder to detect
  - Reducing the volume will be harder to spot
- Viruses will evolve!
  - Changing HELO isn't doing them much good
  - May begin to avoid nonsense destinations

Conclusions

- Spammers & viruses that hide a pattern at the destination make a pattern at the source
- Some simple heuristics currently spot these patterns : with delivery failures being key
- False positives mainly caused by software & users that are being especially clueless ☹

Structure of this talk

- Proof-of-work as an anti-spam device
  - why “hash-cash” isn’t a silver bullet
- Extrusion Detection for spotting spam
  - a practical & effective approach for ISPs
- Tackling Phishing
  - are we upping the ante fast enough ?

Why does phishing work?

- Con artists are really, really good at persuading people to do dumb things.
- Almost no context to an email, or a website; so you no longer need an Intaglio-capable printing press to produce plausible props.
- The underlying protocols and procedures are pretty rubbish…

Authentication(?) protocols

- Password (or 1-time password, or SecurID)

  A→B: A, S_n

  & hence Man-In-The-Middle attack

  A→P: A, S_n

  P→B: A, S_n

- Even if Alice proves her identity (& liveness) in every message there is no binding of that to the type of transaction (or the amount)

Surely, we can fix it with Crypto?

$A \rightarrow B: \quad \{A, B, nonce, Transaction_n\}_{K_A^{-1}}$

This is fine if Alice trusts the program she is using to do the crypto. So what if the phisher invites her to download a new improved version from **`www.bankname.newsoftware.com`** ?

note that *bankname* doesn't see this being registered! So policing the DNS won't help

What about Client Certificates ?

- Client Certificates fix Man-in-the-Middle
  - also kills off account aggregation, and stops you doing your banking from a cybercafe…
- and if phishers now offer you an updated Client Certificate (“and please email back the previous copies for secure destruction”)
  - or if the next virus targets Certificates ?
  - exactly what is the binding to the Certificate ?

What about browser pop-ups?

- Phishers already overwrite padlocks, the URL being visited and the URL asked for…
  - with current browser "security" models you cannot really rely on *anything* on the screen being in the least bit valid
  - it is not credible to insist consumers check for the browser patches every day **and** also turn off Java, JavaScript, ActiveX and Flash…

    ...besides, the banking site probably needs them!

So what will work ?

- Lots of small improvements are possible
  - One-time passwords
  - Client Certificates
  - Real-time browser checks on websites
  - Validation of incoming IP address
  - Multiple levels of authentication by the bank
  - etc etc
- All can be overcome one-by-one, but if introduced all at once they may be daunting!

Who’d climb Kilimanjaro just to go phishing ?

Anti-Spam & Anti-Phishing
A Reality Check

http://www.cl.cam.ac.uk/~rnc1/

THE END : Any questions ??

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Demon