Unwanted traffic: Important problems, research approaches

Balachander Krishnamurthy
http://www.research.att.com/~bala/papers

Internet: A wholly owned subsidiary of....

Unwanted traffic: Important workshop

SRUTI: Steps to Reducing Unwanted Traffic on the Internet

Usenix (Corporate sponsor: AT&T)

2nd workshop to be held July 6-7 2006, San Jose, CA

6 pages, submission deadline April 20

PC chair: smb

http://www.usenix.org/sruti (has proceedings of SRUTI 2005)

Talk outline

- Problems that are most important
- Some research approaches/potential solutions

Key problems

1. Spam (including pop-up spam)
2. IP block theft, IP spoofing
3. Botnets (popular attack vehicle $today$), DDoS
4. Phishing: from large companies to (soon) small credit unions
5. Worms/virus, Web exploits, hot networks, wireless attacks...

Whose problems?

- Internet infrastructure: lack of authorization in routing or networks running hot (traffic concentration leads to magnification of attack's impact)
- ISP problems: Large DDoS may kill but not spam (not to mention conflict of interest..)
- End-user problems (business vpn customer, home user) - stop spam and phishing, reduce ads

Dollar dictates. Cui bono?

Solution vectors

- Spam: architectural, filtering/blackholing, throttling, economics-based
- IP/DNS: characterization, monitoring, detection
- Botnets: characterization, monitoring, some defenses
- DDoS: traceback, prevention/mitigation, tolerance
- Phishing: reporting, filtering (via toolbars), early detection

Some serendipitous help can also be harnessed

Pop-up spam: Not discussed thus far

- Traffic sent to UDP ports 1025-1030 (mostly), causes a Windows messenger service pop-up
- Occasionally phish variant: "error occurred" "machine compromised" Download software for "fix"
- Businesses often block such ports
- Consumers (DSL, cable modem) are vulnerable
- Hundreds of millions of these messages sent/hour.
- Erodes trust needed to encourage financial transactions

This started at least 3 years ago

Spam: What works and what's new

ML: Around 68.6% in '05 (60.6% in 2/06)
Malware (virus or trojan) attacks in email is 2.8% in '05 (2.3% in 2/06)

- Spammers use disposable domains that may last for less than a day to a few days. ML claims 10% of disposable domains had a lifetime less than 3 hours (no traffic goes to it)
- Filtering working (people ignore false +ves) but spammers run SA, BM
- High but varying block rates; not cheap, has not stopped spam origination
- Authentication: DK (some penetration: yahoo, google but not enough)

ML numbers Feb '06

- Traffic Management: throttling unwanted senders
- Connection Management: at SMTP level, verify legit conns to server

  SMTP Validation: Id's known bad sending srcs (open proxy/botnet)
  Registered Users Adddress Validation: Valid id list updated daily

Region	SMTP Validation (botnet sources)	User Validation (directory attacks)
USA	3.6%	13.6%
UK	5.2%	12.0%
Europe	4.7%	17.8%
Asia-Pac	4.2%	3.3%
Worldwide	4.3%	13.4%

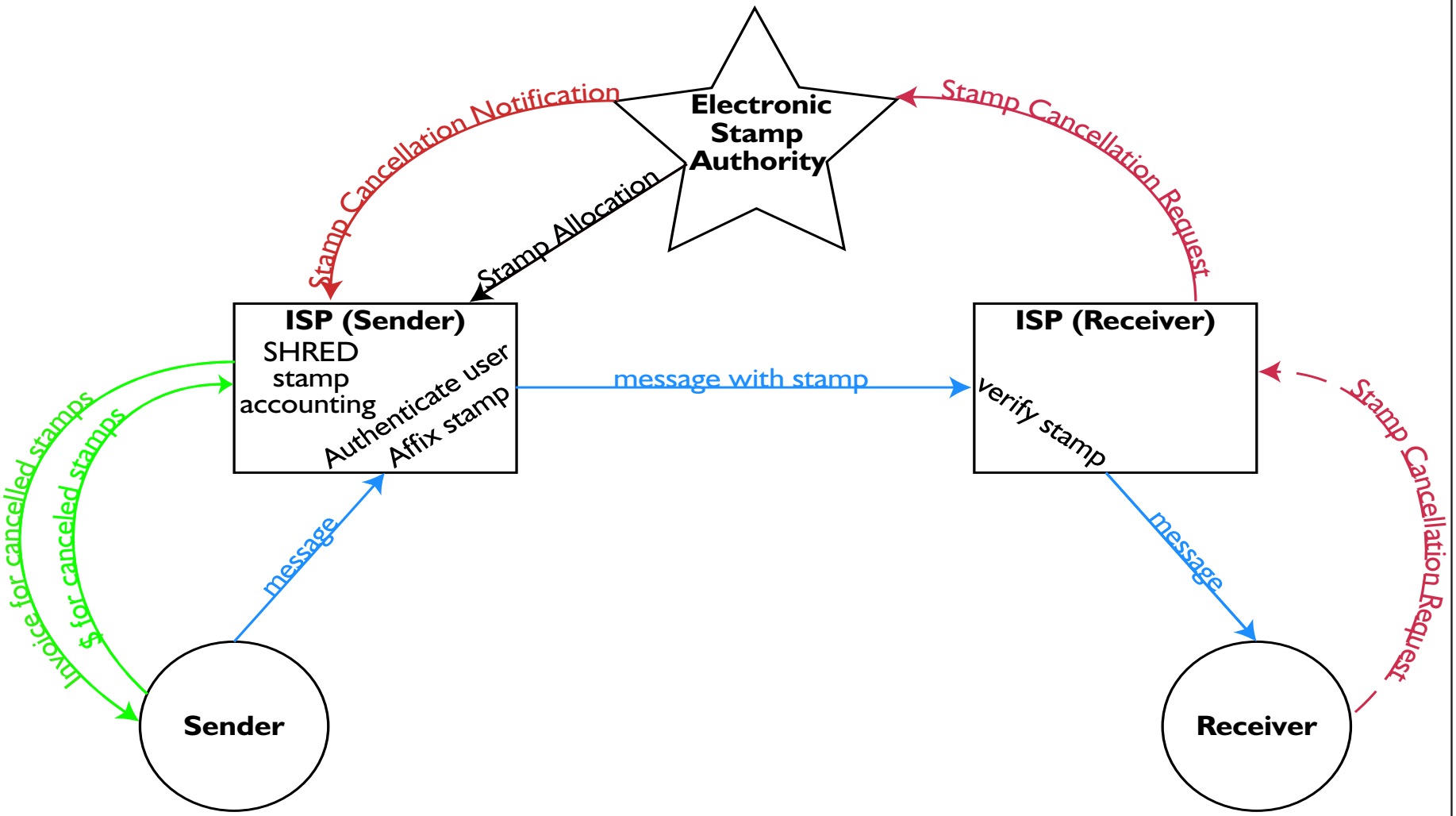
Spam economics

- Requirements are hard: wanted email flows exactly as today without added monetary cost/new protocol/losing features
- C Dwork/ M Naor: Pricing via processing or combating junk mail
- 1992
- Charge senders computationally (less for good guys, more for bad guys, wasted with zombies?)
- Bad? Goodmail (AOL: 'bulk senders' pay; soon, all?)
- Good: SHRED (good guys don't pay; bad guys do)

SHRED: Spam Harassment Reduction via Economic Disincentives

- Economic disincentives complementary to filtering schemes
- Contingent limited liability (not necessarily translated to cost)
- Expression of liability is "stamp" with associated expiry time.
- Credit Limit: number of stamps available to user at any given time. Varies between classes of users, set by ISP.
- Several Electronic Stamp Authorities (ESA): stamp managing entities
- ESA's customers are ISPs; subscribe and pay cancelled stamp charge
- One time stamps, single or multi-valued stamps
  In practice, cryptographically strong header with expiry time, ISP to which it was issued etc. encoded

SHRED Architecture



IP block theft

- NANOG 36: between 26 and 95 successful prefix hijackings in December 2005 (Boothe et al.)
- Tier-1 ISPs see evidence of this (e.g. blocks used only privately)
- One or more chunks allegedly used for sharing pirated software
- Often first one may hear may be through lawsuits
- Internal solution: closely monitor advertisements, alert affected customers

Web exploits

CVE (common vulnerabilities and exposures, cve.mitre.org) from '99-'05 says 25% of security flaws are web exploits (robertson et al. ndss'06). Common exploits:

- Reading entire db of a e-commerce site (mangling url)
- Editing cookies to get higher privilege
- Looking for math bugs (-ve dollar amount)
- Storing code in the comments section (a la blog poisoning)
- Access soi disant hidden modules via 'forceful browsing' (demo at recent rsa conf by imperva)
- Reverse engineering

IP spoofing

- Spoofer project at MIT (Beverly/Bauer) continues to measure filtering ability in various address blocks
  http://spoofer.csail.mit.edu/summary.php
- Set src to be {bogon, valid, martian, neighbor}
- Partial/full spoofing seen in over 20% of addresses/IP blocks
- With botnets spoofing may not be needed; study shows some known problems rarely get fixed
- Spoofed TCP RST packets (Touch ID) - port hiding may not be clever enough, connection times can be large (think BGP) - not often seen?

Botnets: some numbers

- Numbers range from 1M (Cooke et al. SRUTI '05) to 2M (Symantec)...
- ... to 100M (Merrick Furst, Ga. Tech) w/ conscription rate of 7K/day with AOL+MSN comprising a third, 6K C+C points per month
- Believability of this number depends on

  Filtered by dynamic IPs?
  Handles targets that move?
  Factors possible recounting? Same host gets infected again

- No public methodology information available

Botnets: In terms of dollars

- Botmasters can make $15K/month easily just through clickfraud (Google agreed to pay $90M for bogus referrals)
- Ancheta made $60K by controlling 400K zombies
- Cheaper than human clickers in Patparganj who make INR 9K/month (USD 200) although they can handle Turing tests
- Rent, don't buy: .gov .25, .edu .30, broadband: .40, corporate: .80/bot
- Minimum lot size: 100/hour but available in 500, 1000, 5000, 10000; comes with estimated bandwidth
- server: 200/bot (more CPU, better connectivity, transient)
- Generally, higher the cost. more pps each bot can generate

Botnet: research

- Identify botmasters than bots by watching how they communicate
- Identification at army formation times; armies range from 10K to 100K
- Cooke et al. SRUTI'05 Zombie Roundup': behavioral methodology for analyzing IRC traffic from end-hosts to detect bot chat
- Transient BGP ads used by spammers (Ramachandran et al., NANOG 36) (hide in a large /8 space, gone by the time checked, in allocated unannounced space) able to bypass blacklists

DDoS: research solutions

- Techniques: traceback, prevention/mitigation, tolerance
- Prevention via rate limiting or packet filtering (route-based distributed using topology knowledge - Park/Lee sigcomm '01)
- Audit trails as traceback
- Tolerance: common technique is buying bandwidth
- For in-network detection monitoring thousands of interfaces hard
- SNMP-based anomalies trigger netflow records gathering. Flow records using uni-dimensional aggregation and clustering techniques. Layered detection mechanism achieves accuracy (Sekar et al. Usenix '06)

Phishing

- 0.3% in '05 or 1 in 304 of email traffic (ML), same in 2/06
- Targets: Top-n banks and other institutions (amazon, ebay, paypal, visa)
- Countries where phish sites are hosted: (netcraft)
  South Korea, Romania, Taiwan, India, Hong Kong
  Thailand, Mexico, Malaysia, Philippines, Lithuania
- Phishkits with copies of websites of top N sites, email list segmented by target and exploit, 50K chunks of email addresses
- Scripts in tcl, python, bash; browser sniffers and form validators in js
- c code (ssl stuff), port scan, ssh scan

Phishing contd.

- password files with 700 uid/password checks, password generation scripts
- common names including my favorites:
  'balakris' 'balas' 'balasubr' 'balkrish'
- popular exploits: myptrace kmexp modprobe, adding stuff to cron,
  various buffer exploits openssl remote exploits (spawning a
  nobody/apache shell on apache, root on other web servers)

Phishing: early detection

- Unlike spam, phishers have to stick around to get information
- Phishers spreading of URL cannot be staggered over time
- Use the relatively long time between spam and HTTP connection
- Watch for increases in incoming spam followed by outgoing HTTP to hitherto rare destination

Serendipitous help: popular software

- Firefox extensions: Outsourcing security to browser - entry point to the Internet for many (ok 1 in 5 in europe, 11% worldwide)
- Passwordmaker extension: always generate passwords
- One-way hash algorithm calculate a message digest that is opaque about input used to generate; master pwd cannot be reverse engineered.
- Even with master pwd 10 variables are needed to id other passwords
- Like many firefox extensions, easy-to-use UI increases deployment/use. Passwords are auto-pasted in password boxes (a la BugMeNot) frustrating keyloggers, defeats phishing with syntactically close URL variants
- Firefox-2 will have anti-phishing builtin

Combination attack

Metasploit: courtesy Marcus Sachs SRI

- 57 exploits, 66 payloads
- Targeting BSD, Linux, Solaris, MS
- GUI
- http://metasploit.com/projects/Framework/downloads.html

What happened to this?

What's needed at high level

- Conflict between constraints and new applications (skype)
- Coordination: useful information is still diffused
- A few different updaters in March 2006: AntiVir/AVPE, Avast, AVG 7, Bit Defender (Web/FTP), Dr. Web, eTrust EZ, F-Prot (Web/FTP), F-Secure, KAV (8 updates daily), McAfee Daily DAT, NAV LU, Nod-32, Norman Virus Control, Panda, Sophos,TrendMicro
- Above is just for anti-virus. Additional ones for anti-trojans, privacy, phishing etc.

Phighting back?

http://www.phishfighting.com

Asks people to enter target of phish mail
Sends multiple submissions to the phisher site with fake data
Wastes their time trying to use/cash in on fake userid/pwd/cc info
As of 2/15/06 site claims to have received 19,662 phish URLs and sent 5,716,494 fake entries

Inverse bugmenot without distributed approach. Easy to abuse.

Missing focus in solution space – Economics

- Counting cost due to problems
- (Correctly) Leveraging economics for solutions
- Costs not often known outside business circles
- Estimates vary (even for renting botnets)
- Not often understood and rarely attempted as solution
- Ignoring economics: we still have spam, click fraud

Personal wanderings in this space

- Spam: Increase cost for senders
  `http://www.research.att.com/~bala/papers/shred-ietf56-talk.ps`
- Attacks: Frustrate reverse blacklist; instead of hiding honeypots try to find sources closer to sender by advertising dark prefixes (see Mohonk, `http://www.research.att.com/~bala/papers/mohonk.pdf`
- Stress testing traffic to infer its legitimacy
  `http://www.research.att.com/~bala/papers/tramp.pdf`
- Saving unwanted traffic (a third of bytes!) by blocking ads
  `http://www.research.att.com/~bala/papers/cam.pdf`
- Phish: Use the time between set up, broadcast, and access (ongoing)