Understanding the NetworkLevel Behavior of Spammers

The work of:

Anirudh Ramachandran and Nick Feamster

ACM SIGCOMM 2006

Sending Techniques

- Direct spamming
  - Use spam friendly ISP
  - High bandwidth out going, low bandwidth incoming
- Open relays
  - Unauthenticated host relay mail server traffic
  - Solved with blacklisting techniques
- Botnet
  - Suggested as a major source of spam
- Hijack
  - Temporary takeover of IP address space

Anti-spam Techniques

- Content filtering
  - Email header or body
  - Content is easy to alter
  - Extensive filtering rules
  - Last remote mail relay
- Blacklist
  - Sending IP address, open relays, open proxies
  - Many lists, each maintained separately
  - Effective (80%), but high FN rate (50%)

Spam Collection

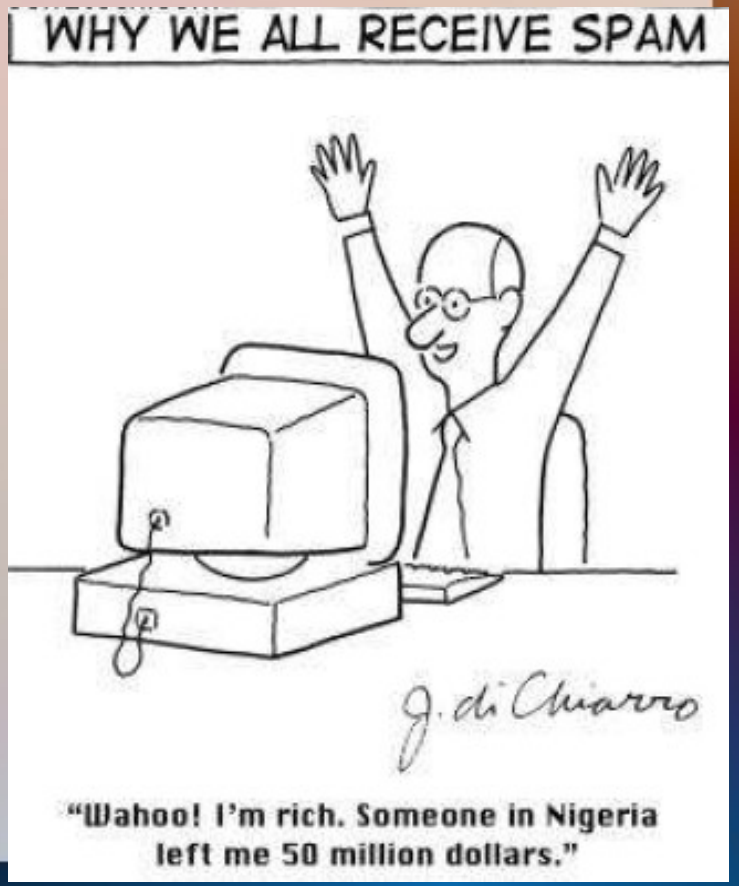
- Wild spam
  - Domain w/o associated Email addresses
  - Create a DNS mail exchanged record
  - Sender properties
    - IP of the mail relay (difficult to forge)
    - Traceroute of the relay
    - OS used by the relay
    - Blacklist lookup results

Filtered Spam Collection

- Large Email provider (millions of Email boxes)
  - SMTP connection attempts
  - Time of connection
  - IP of connection host
  - Accepted or rejected
  - 700000 accepted Emails on a single day

Botnet Spam

- Hijack the command and control of the Bobax worm
- Compare IP from the botnet to the sinkhole Email
  - Dynamic addressing problems

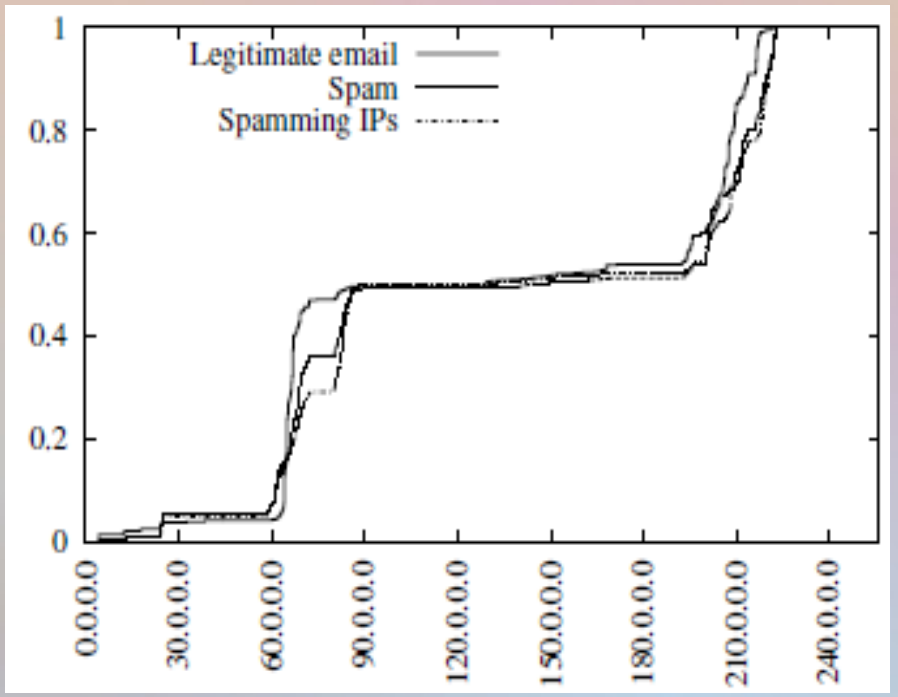


Border Gateway Protocol Exploit

- BGP – routing protocol
- Is the mail relay reachable?
- How long is it reachable?
- How much spam is coming from short duration routes?
  - Wild spam receive time
  - BGP open window

Spam IP's

- Distribution of Email traffic by address range
- Certain ranges of IP are responsible for higher concentration of spam
- Not useful for individual IP detection
- Consistent over time



Spam by Autonomous System

- Which AS's generate the most spam
- Korea & China: 10%
- US: 40%
- Concentration

AS Number	*# Spam*	*AS Name*	*Primary Country*
766	580559	Korean Internet Exchange	Korea
4134	560765	China Telecom	China
1239	437660	Sprint	United States
4837	236434	China Network Communications	China
9318	225830	Hanaro Telecom	Japan
32311	198185	JKS Media, LLC	United States
5617	181270	Polish Telecom	Poland
6478	152671	AT&T WorldNet Services	United States
19262	142237	Verizon Global Networks	United States
8075	107056	Microsoft	United States
7132	99585	SBC Internet Services	United States
6517	94600	Yipes Communications, Inc.	United States
31797	89698	GalaxyVisions	United States
12322	87340	PROXAD AS for Proxad ISP	France
3356	87042	Level 3 Communications, LLC	United States
22909	86150	Comcast Cable Corporation	United States
8151	81721	UniNet S.A. de C.V.	Mexico
3320	79987	Deutsche Telekom AG	Germany
7018	74320	AT&T WorldNet Services	United States
4814	74266	China Telecom	China

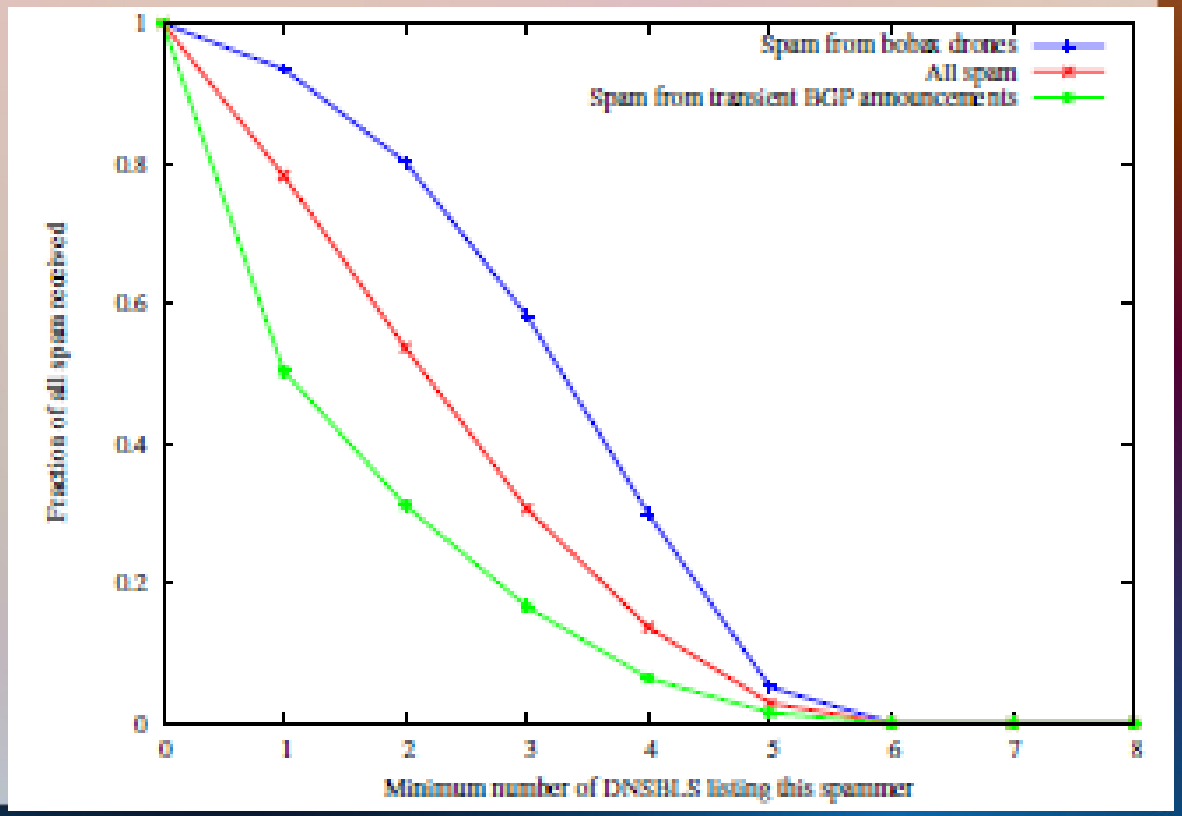
Email Providers

- Who are the biggest Email providers?
  - From the legitimate Email list

AS Number	*# Email*	*AS Name*	*Primary Country*
15169	49500	Google Inc.	United States
5731	38238	AT&T WorldNet Services	United States
26101	30406	Yahoo	United States
3561	22730	Savvis	United States
4355	17381	Earthlink, Inc	United States
8560	16666	Schlund Partner AG	Germany
8075	14699	Microsoft Corp	United States
14779	13115	Inktomi Corporation	United States
6541	12493	GTE.net LLC	United States
14780	11597	Inktomi Corporation	United States

Blacklisting

- Is it effective?
  - Works better on some spam approaches
  - Need a better BGP detection
  - 80% on a list
  - 50% single list



Spam Bots

- Specific to Bobax worm
- 70% of spam bots were Windows machines
- Each bot is a low volume spammer
  - Shutting down any one bot drone has little effect
  - Large number of machines are involved
- IP address distribution suggests bot spam is responsible for most spam (inconclusive)
- Could be from just a few machines with changing IP addresses

OS of choice

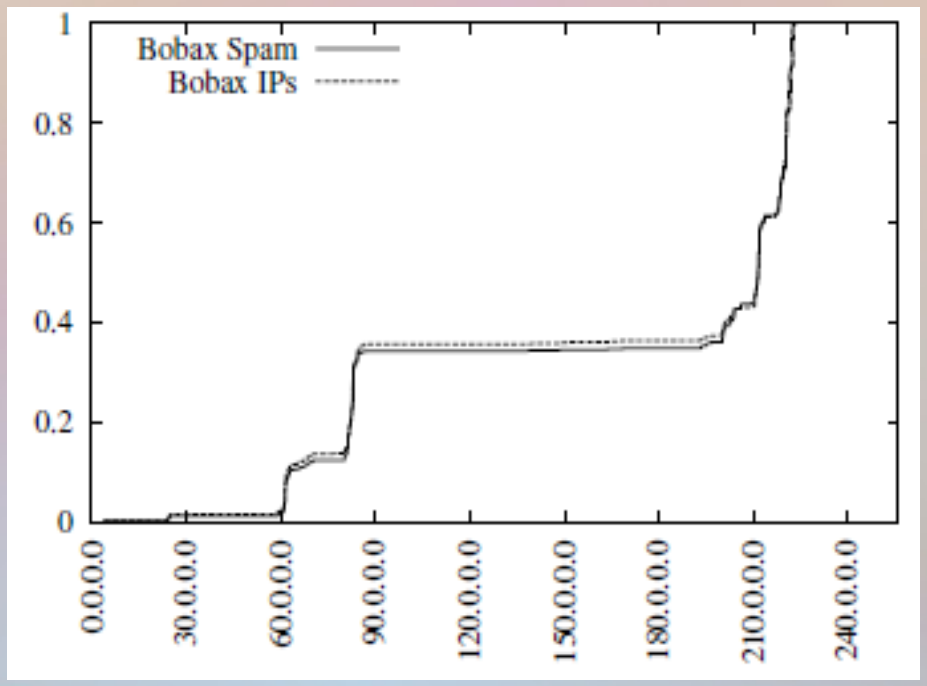
- OS identified by tool in Mail Avenger
- Most mail senders use Windows
- NonWin senders (4%)

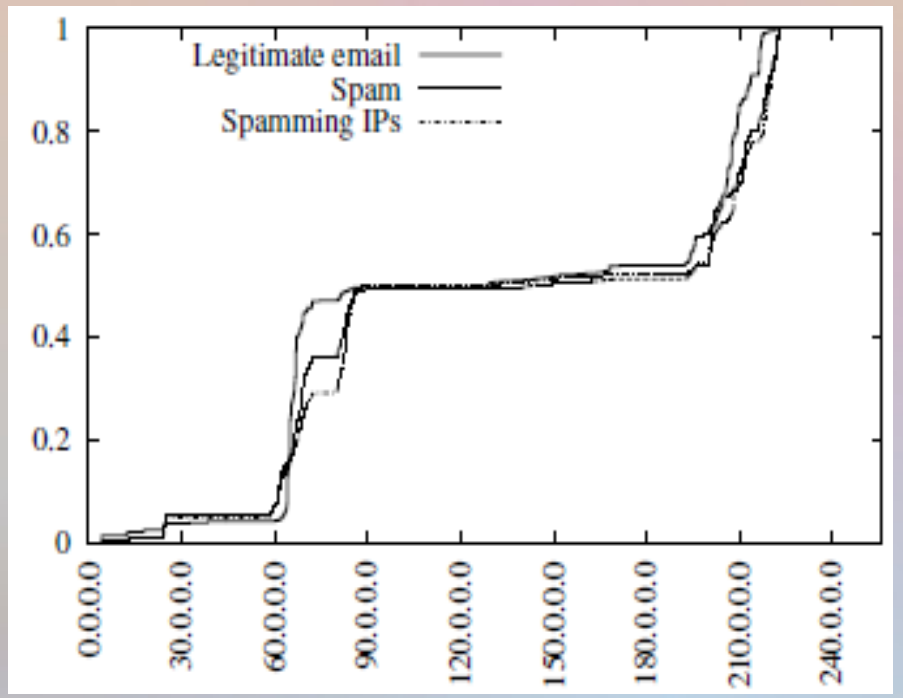
  Suspect? (8%)

Operating System	*Clients*	*Total Spam*
Windows	854404 (70%)	5863112 (58%)
- Windows 2000 or XP	604252 (49%)	4060290 (40.2%)
- Windows 98	13727 (1.1%)	54856 (0.54%)
- Windows 95	559 (<0.1%)	2797 (<0.1%)
- Windows (other/unconfirmed)	235866 (19%)	1745169 (17.2%)
Linux	28132 (2.3%)	557377 (5.5%)
FreeBSD	6584 (0.5%)	152456 (1.5%)
MacOS	2944 (0.2%)	46151 (0.4%)
Solaris	1275 (< 0.1%)	18084 (0.2%)
OpenBSD	797 (< 0.1%)	21496 (0.2%)
Cisco IOS	736 (< 0.1%)	5949 (<0.1%)
NetBSD	44 (< 0.1%)	327 (<0.1%)
HP-UX	31 (< 0.1%)	120 (<0.1%)
Tru64	26 (< 0.1%)	143 (<0.1%)
AIX	23 (< 0.1%)	366 (<0.1%)
OpenVMS	18 (< 0.1%)	62 (<0.1%)
IRIX	7 (< 0.1%)	62 (<0.1%)
Other/Unidentified	128580 (10.4%)	1212722 (12%)
No Fingerprint	204802 (16.7%)	2225410 (22%)
Total	1228403	10103837

Bobax Behavior

- Received ~4700 spam Emails from the Bobax botnet
  - out of ~117000 sent by the botnet
- IP address distribution is consistent with typical Email use





One shot Bots

- 65% of the bots send spam only once
- Of these, 75% send spam to the sinkhole
- 75% of the bots send spam for less than 2 minutes
  - Handful sent spam for 6 months
  - A smaller number never stopped sending spam
- Suggests that a blacklist approach will not be effective for this type of behavior
- Collaborative spam filtering
  - Identify which spammers only send a single Email to many different domains

Bobax Spam Rate

- Most bots don't send a high volume of spam
  - Whether the bot is active for a long period or not
  - Typically less than 100 pieces
- Large number of hosts
- Used for short periods of time
- Each sends a few pieces of spam
- Typical rate: < 1 piece of spam per bot per day

Transient BGP

- Here and gone
  - Briefly advertise a route to a section of IP address space
  - Send the spam
  - Remove the routes
- Difficult to trace
  - Routes are not monitored in real time
  - When looking for the spam source, it is unreachable
- Accounts for ~<10% of spam received
- Something new (2005)

BGP Exploit Behavior

- Spam sent on regular basis
- Short lived route ~ 12 minute
- Hijacked IP addresses
  - Allocated IP space
  - Unused IP address space
- Large address blocks (66.0.0.0/8)
  - Various individual IP addresses (Not fixed)
  - Not reused (Continually changing)
  - Less likely to be blocked by ISP's
- Route typically advertised continuously for a day

- Blacklist approach is not effective
- Popular sources: Malaysia, Japan
  - Other: Bulgaria, US
- Currently a low volume source of spam (2005)

Observations (2005)

- Spam volume is increasing over time
- Number of spam IP's is increasing
- High volume spammers identified by autonomous system (AS) prefix – Korea & China (10%)
- Content filtering is a typical mitigation technique
- Blacklist approach is effective on some techniques
- Network approach to spam mitigation appears viable

Network Level Mitigation

- Difficult to forge network properties
- Middle of the network
  - Allows quarantine mitigation
  - Destination mail server never sees it
  - Bandwidth savings

Lessons

- Spam filtering requires a better notion of host identity
  - Blacklisting is ineffective
    - One shot bots
    - Short lived bots
    - BGP exploits
- Detection techniques based on aggregate behavior are more likely to expose nefarious behavior than techniques based on observations of a single IP address
  - Observe an IP address range, rather than an individual IP address

Lessons

- Securing the Internet routing infrastructure is a necessary step for bolstering identity and traceability of email senders
  - Close the BGP loophole
  - Prevent route hijacking
- Some network-level properties of spam can be incorporated relatively easily into spam filters and may be quite effective at detecting spam that is missed by other techniques
  - Monitor recently BGP announcements
  - Tracking Email behavior across multiple domains

Questions?