



# **UNDERSTANDING THE NETWORK-LEVEL BEHAVIOR OF SPAMMERS**

**Anirudh Ramachandran  
Nick Feamster**

# SPAM

- Unsolicited commercial email
- **90%** of all email is spam and **30 billion** messages are sent through the internet everyday
- Recent researches indicate that spam costs businesses worldwide **\$50 billion** every year
- Common spam filtering techniques:
  - Content-based filtering
  - DNS Blacklist Lookups

# PROBLEMS WITH CURRENT FILTERING TECHNIQUES

- Content-based filtering
  - **Low cost to evasion:** Spammers can easily alter the features and content of spam e-mails
  - **High admin/user cost:** Filters must be updated continuously and frequently as new type of emails are captured
  - **Applied at the destination:** Wasted network bandwidth and storage,etc...
- DNS Blacklist Lookups
  - Significant fractions of today's DNS queries are DNSBL lookups.

# NETWORK-LEVEL SPAM FILTERING

- Instead of highly variable content properties, we focus on **network-level properties** which are **more fixed** such as:
  - ISP or AS hosting spammers
  - Location
  - IP address block
  - Routes to destination
  - Botnet membership
  - Operating system
  - ...
- Using network-level properties, spams could be filtered **better** and stopped **closer** to the source.

# OUTLINE

- Background
- Data Collection
- Network-level Characteristics of Spammers
- Botnets
- BGP Spectrum Agility
- Lessons
- Conclusion

# SPAMMING METHODS

## ○ Direct Spamming

- Spammers purchase upstream connection from **spam-friendly ISPs**
- They buy connectivity from **non spam-friendly ISP** and after spamming, switch to another ISP.
- They sometimes obtain a **pool of dispensable dialup IP addresses** and proxy traffic through these connections

## ○ Open Relays and proxies

- They use **mail servers** which allows **unauthenticated** internet hosts to send emails through them

# SPAMMING METHODS

## ○ Botnets

- Collections of software robots(**worms**) under one centralized controller
- **Infected** hosts are used as a mail relay

## ○ BGP Spectrum Agility

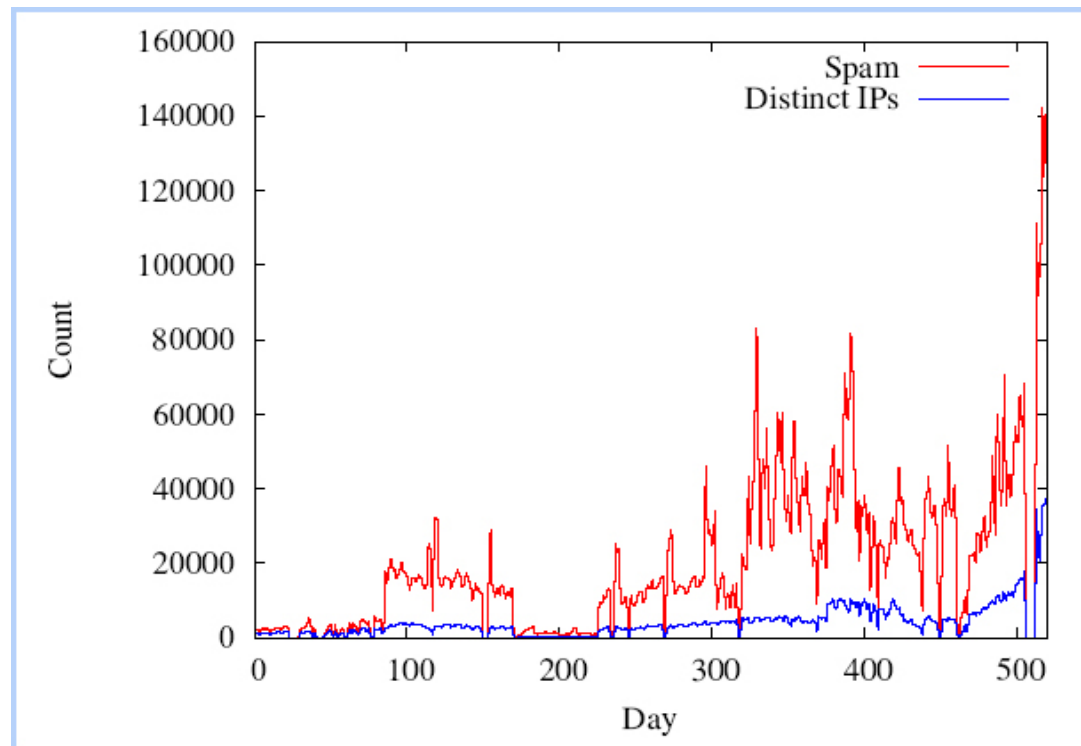
- A hijacked IP address range is briefly **advertised** via BGP and used to send spam
- Once mails are sent, they **withdraw** the route from the network

# DATA COLLECTION

- Spam Email Traces
  - Registered a domain with no legitimate email address
  - Almost 10million spams are collected between Aug,2005 and Jan, 2006
  - Collects following information about each mail:
    - The IP address of the relay
    - Traceroute -> location
    - Passive OS fingerprint
    - Result of DNSBL lookups

# DATA COLLECTION

- Spam Email Traces



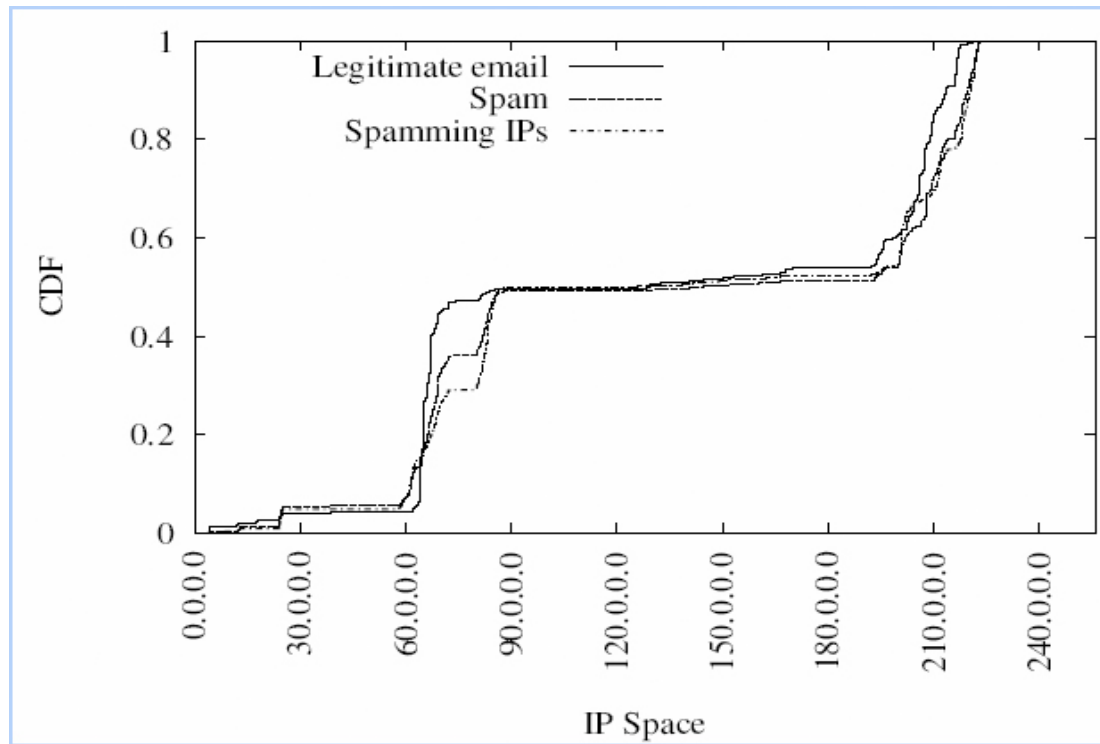
*The amount of spam received per day at our sinkhole from August 2004 through December 2005.*

# DATA COLLECTION

- Legitimate Email Traces
  - Obtained a huge amount of mail logs from a large email provider (700K)
  - Logs includes:
    - The timestamp of connection attempt
    - IP address of the host
    - Whether rejected or not
    - Reason of rejection
- Botnet Command and Control Data
  - Used a trace of hosts infected by [W32.Bobax](#) worm
- Monitoring BGP route advertisements from same network

# NETWORK-LEVEL CHARACTERISTICS OF SPAMMERS

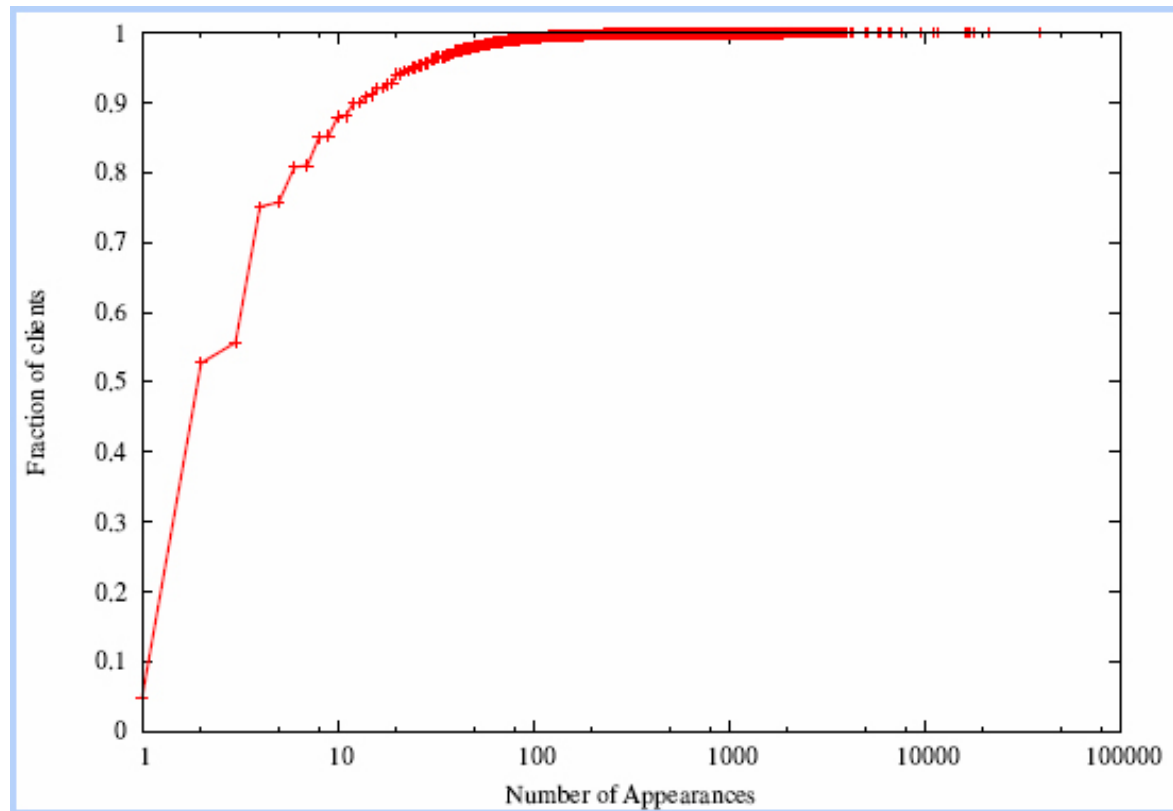
- The majority of spam is sent from a relatively **small** fraction of IP address space



Distribution of spam across IP space

# NETWORK-LEVEL CHARACTERISTICS OF SPAMMERS

- **%85** of client IP addresses sent less than **10 emails** to the sink



The number of distinct times that each client sent mail to the sinkhole

# NETWORK-LEVEL CHARACTERISTICS OF SPAMMERS

<i>AS Number</i>	<i># Spam</i>	<i>AS Name</i>	<i>Primary Country</i>
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

The amount of spam received from mail relays in top 20 ASes

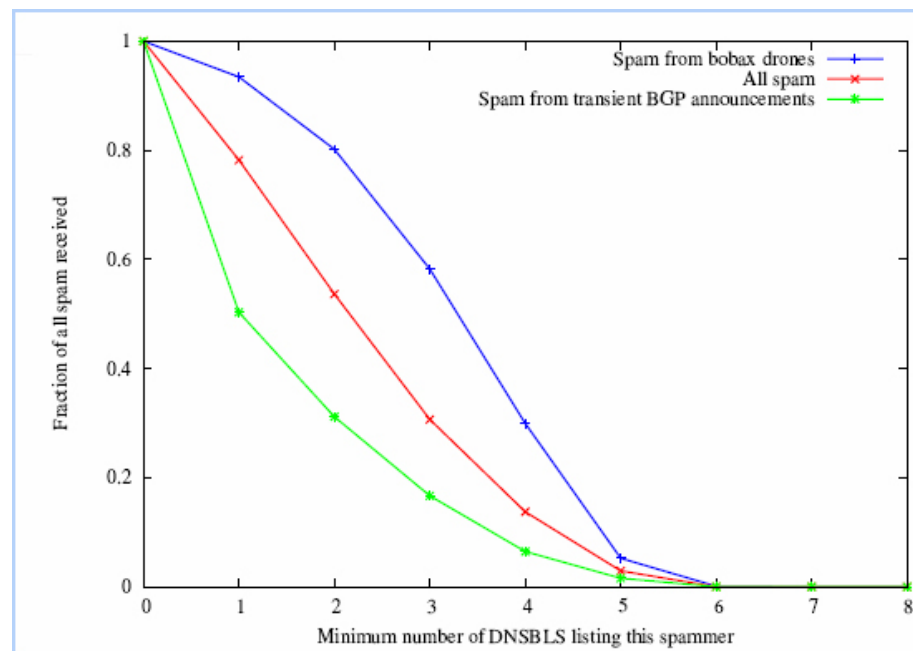
# NETWORK-LEVEL CHARACTERISTICS OF SPAMMERS

- More than **%10** of spam received at the sinkhole originated from mail relays in **two ASes**
- **%36** of all received spam originated from **only 20 ASes**
- With a few exceptions, ASes containing hosts responsible for sending large quantities of spam differ from those sending large quantities of legitimate email
- Although the top two ASes from which the sinkhole received spam were from Asia, 11 of the top 20 ASes were from the United States and **40%** of all spam from the top **20 ASes**.
- An email's country of origin may be an effective filtering technique for some networks

# THE EFFECTIVENESS OF BLACKLISTS

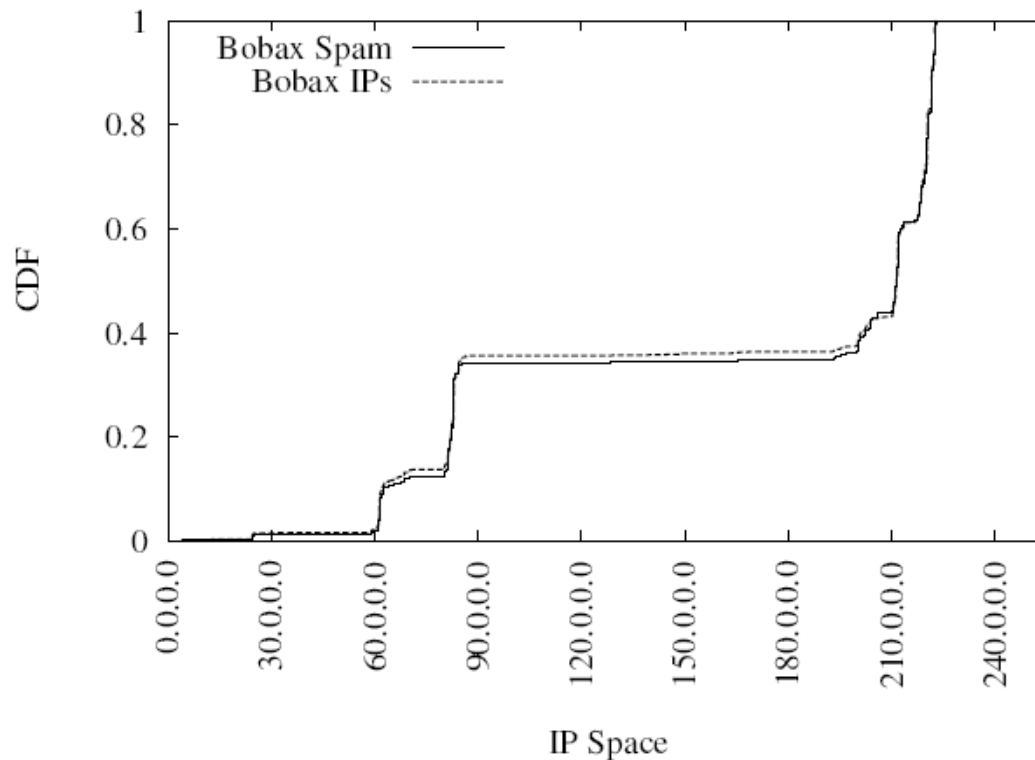
- Effectiveness of blacklists: Nearly **80%** of all spam was received from mail relays that appear in **at least** one of eight blacklists.
- A **high** fraction of Bobax drones were blacklisted, but relatively **fewer** IP addresses sending spam from short-lived BGP routes were blacklisted. Only half of these mail relays appeared in any blacklist.

The fraction of spam emails that were listed in a certain number of blacklists or more



# SPAM FROM BOTNETS

- Spamming hosts and Bobax drones have similar distribution across IP address space
- Much of the spam received at the sinkhole may be due to botnets such as Bobax



Distribution of Bobax drones and the amount of spam received from those drones

## OPERATING SYSTEMS OF SPAMMING HOSTS

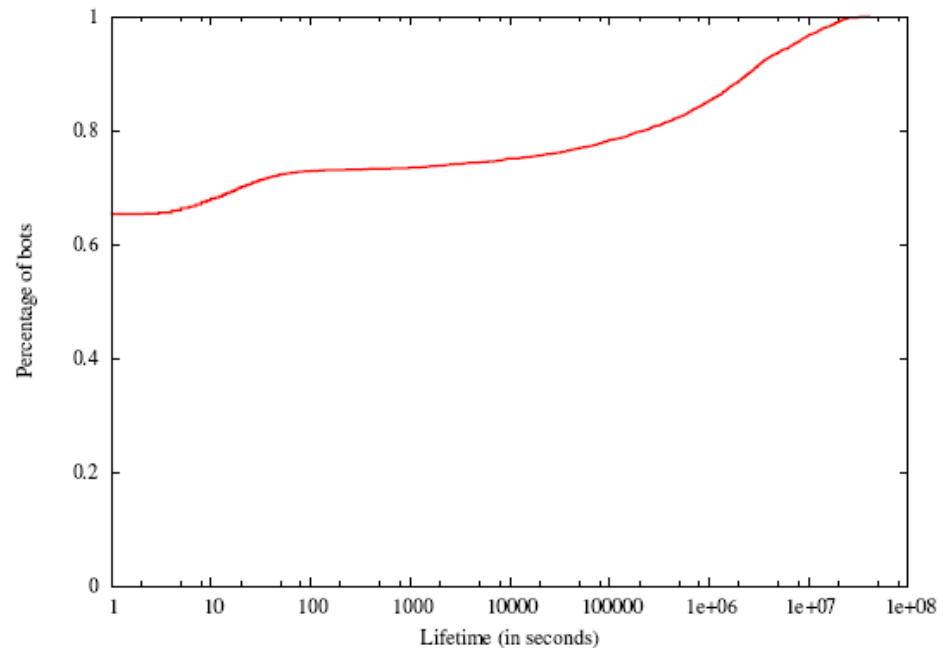
- 75% of all hosts could be identified for their OS
- 4% of hosts are not Windows but are responsible for 8% of all spam

<i>Operating System</i>	<i>Clients</i>	<i>Total Spam</i>
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

Operating systems of hosts determined by passive OS fingerprinting

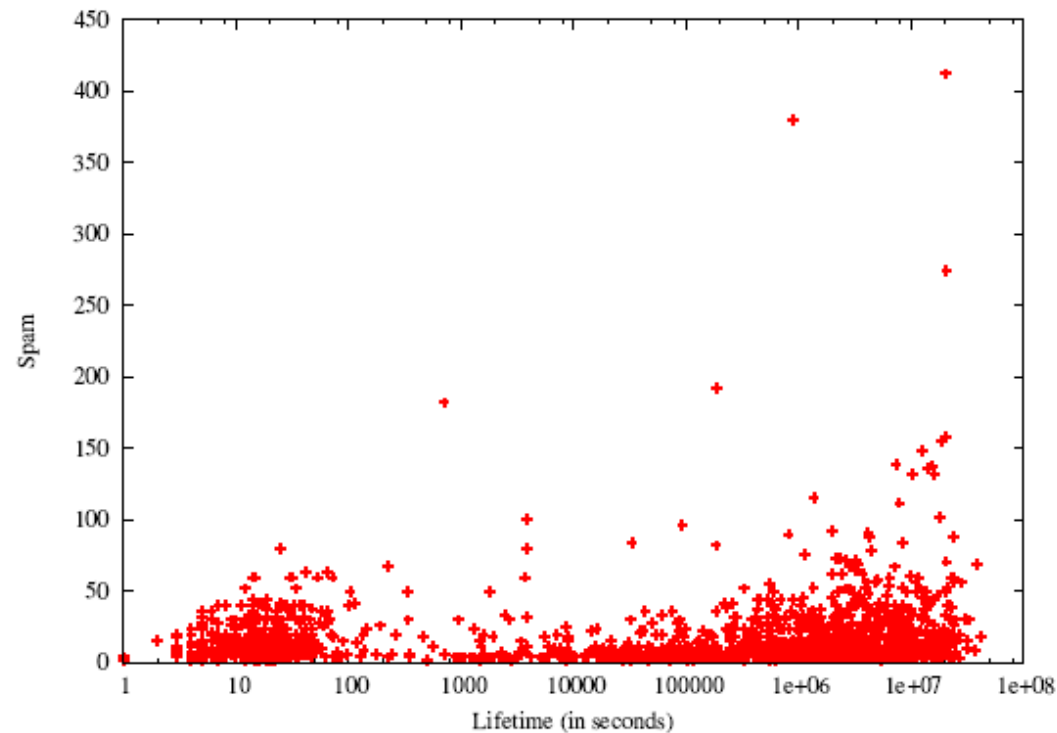
# SPAMMING BOT ACTIVITY PROFILE

- *How many of the known Bobax drones send spam to sinkhole?:*
  - Only **4693** of 117,268 Bobax infected hosts sent email to the sinkhole -> spammers use different spamming patterns
- For how long does any Bobax drone send spam?:
  - **65%** of hosts infected with Bobax send spam only once and **75%** of them persisted less than two minutes
  - Single-shot bots
  - How about blacklisting?



# SPAMMING BOT ACTIVITY PROFILE

- Volume:
  - Spams arrives from bots at very **low rates**
  - **%99** of the bots sent fewer than **100** pieces of spam over entire trace



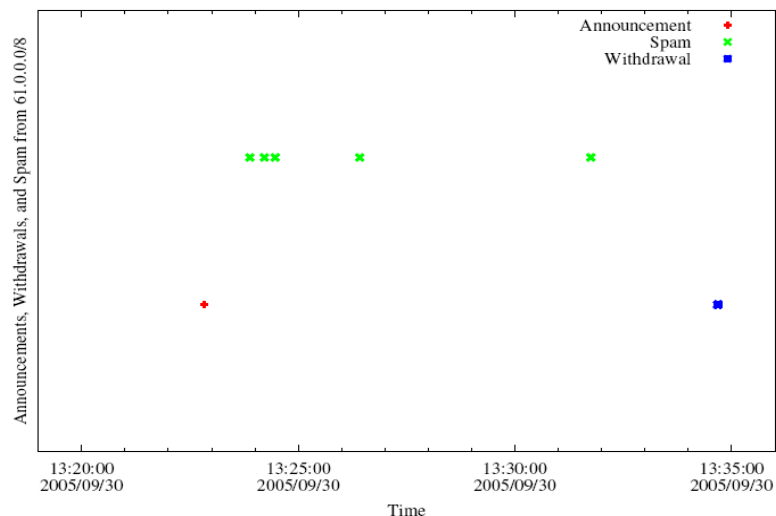
Amount of spam mail and Bobax drone persistence

# SPAM FROM TRANSIENT BGP ANNOUNCEMENTS

- One of the most sophisticated techniques and difficult to track spam to the sources
- Not a dominant technique that spam is sent today (at most %10)
- How it works
  - Briefly advertise portions of IP space
  - Send spam from mail relays with IP addresses in that space
  - Subsequently withdraw the routes for that space after spam is sent

Common short-lived prefixes and ASes

61.0.0.0/8	4678
66.0.0.0/8	21562
82.0.0.0/8	8717



# BGP SPECTRUM AGILITY

- Discovered patterns and locations to sent spam:

<b>AS 21562</b>	<b>an ISP in Indianapolis</b>	<b>66.0.0.0/8</b>
<b>AS 8712</b>	<b>an ISP in Sofia, Bulgaria</b>	<b>82.0.0.0/8</b>
<b>AS 4678</b>	<b>Conan Netw. Comm., Japan</b>	<b>61.0.0.0/8</b>

- A small, but persistent, group of spammers appear to send spam by:
  - advertising (in fact, hijacking) large blocks of IP address space (i.e., /8s)
  - sending spam from IP addresses that are scattered throughout that space
  - withdrawing the route for the IP address space shortly after the spam is sent.

# WHY SUCH BIG PREFIXES?

## ○ **Flexibility:**

- Client IPs can be scattered throughout dark space within a large /8
- Same sender usually returns with different IP addresses

## ○ **Visibility:**

- Route typically won't be filtered by ISPs

# LESSONS FOR BETTER SPAM FILTERING

- Spam filtering requires a better notion of **host identity**
- Detection techniques based on **aggregate behaviour(IP space)** are more likely to expose spam behavior than techniques based on observation of a single IP address
- **Securing the Internet routing** infrastructure is a necessary step for traceability of email senders
- Some network level properties of spam can be **integrated** into the spam filters easily and detect spam which can not be caught with other techniques
  - Recently announced BGP announcement

## CONCLUSION

- Network-level behavior of spammers are presented using a analysis of **four** datasets as result of **17** months study
- **Bobax** drones are used to better understand the spamming botnets
- Although most of the drones doesn't send spam more than **twice**, blacklists works quite well at detecting them
- **BGP** spectrum agility technique makes tracebility and blacklisting more difficult
- Spam filters using network-level behaviors could be more **effective** than regular content-based filters