



Understanding the Network-level Behavior of Spammers

Anirudh Ramachandran and Nick Feamster
{avr, feamster}@cc.gatech.edu
College of Computing, Georgia Tech



Introduction and Motivation

- Most studies on spam focus on analyzing its **contents**
- However, **network-level** properties tend to be much less variable
- Some properties investigated by this study:
 - IP Address ranges sending the most spam
 - Use of Botnets and BGP Route-Hijacking for spamming
 - Effectiveness of DNS Blacklists
 - Techniques used in email harvesting
- Why current anti-spam techniques might be ineffective, and suggestions for improvement

Measurement Setup

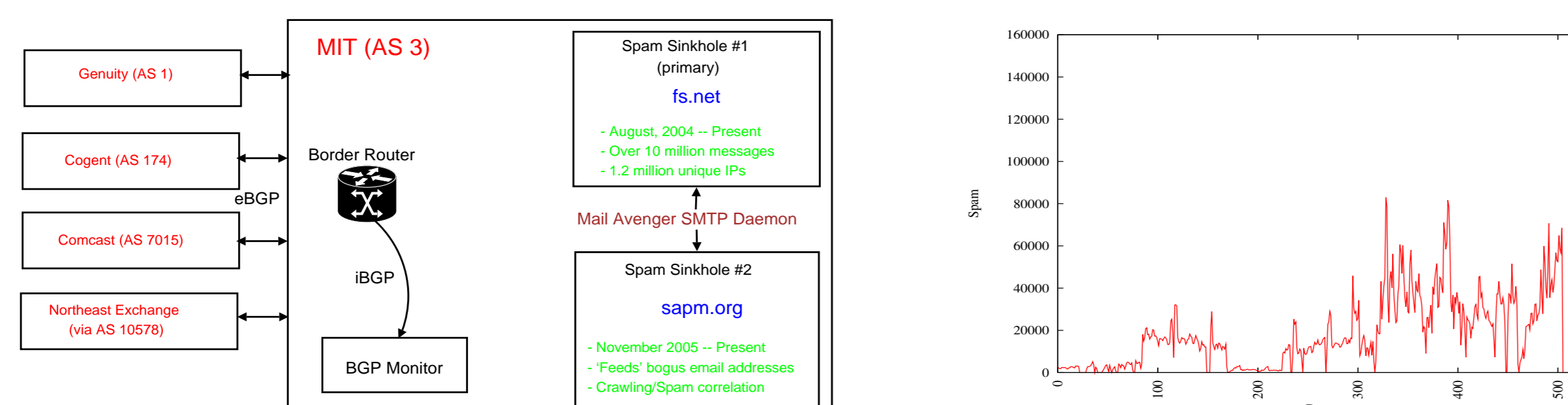


FIGURE 1: Data Collection: We use email logs from two spam sinkholes, BGP UPDATE logs from the BGP Monitor, and a list of IPs belonging to a spamming botnet (Bobax)

FIGURE 2: Email arrival at fs.net since August 2004

Spam Distribution across IP space

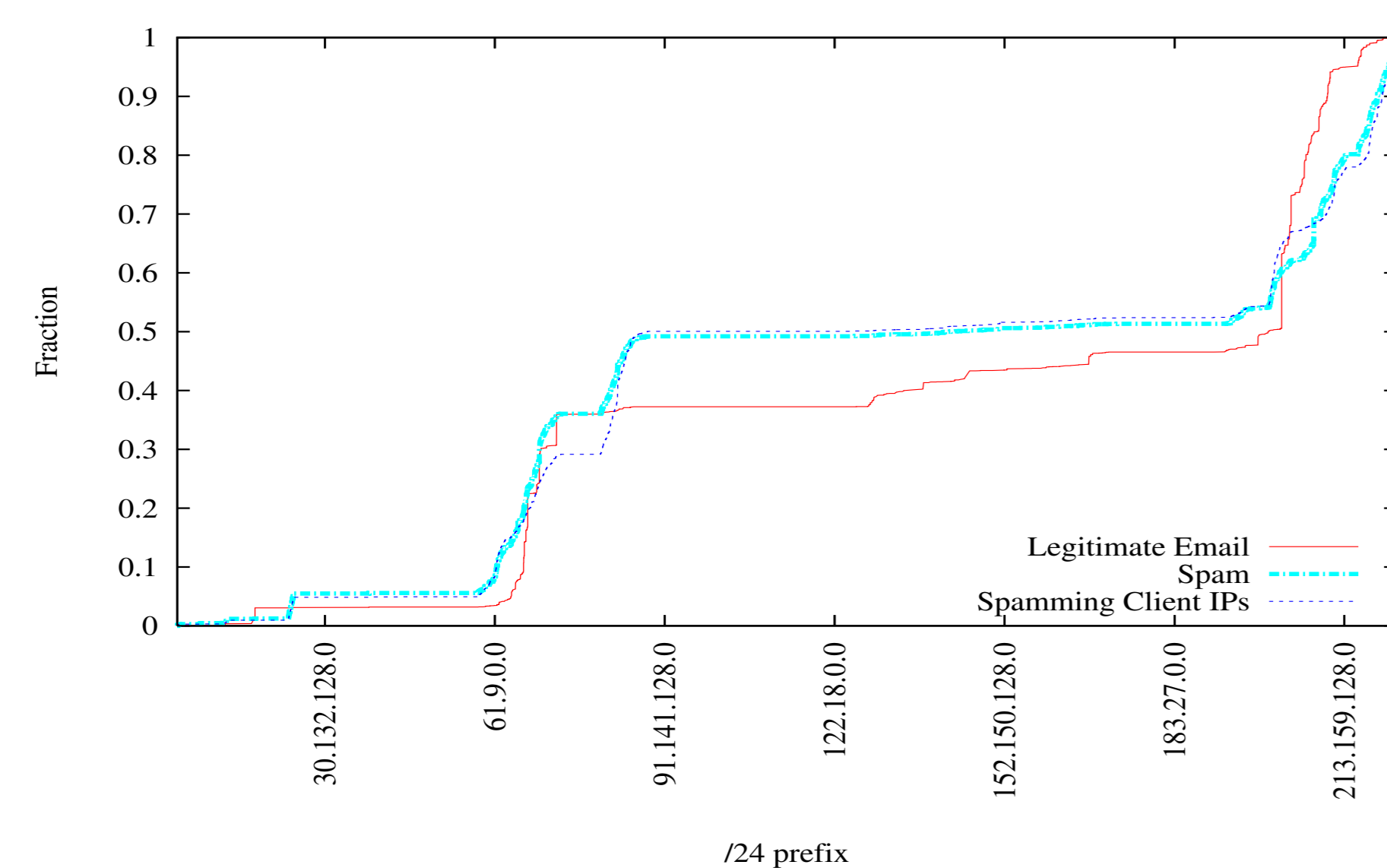


FIGURE 3: CDF of received spam over IP space, binned by /24. Also shown are the CDFs of legitimate email (from another domain) and the spammer IP distribution

The majority of spam is received from a relatively small fraction of IP address space

Spam from Botnets

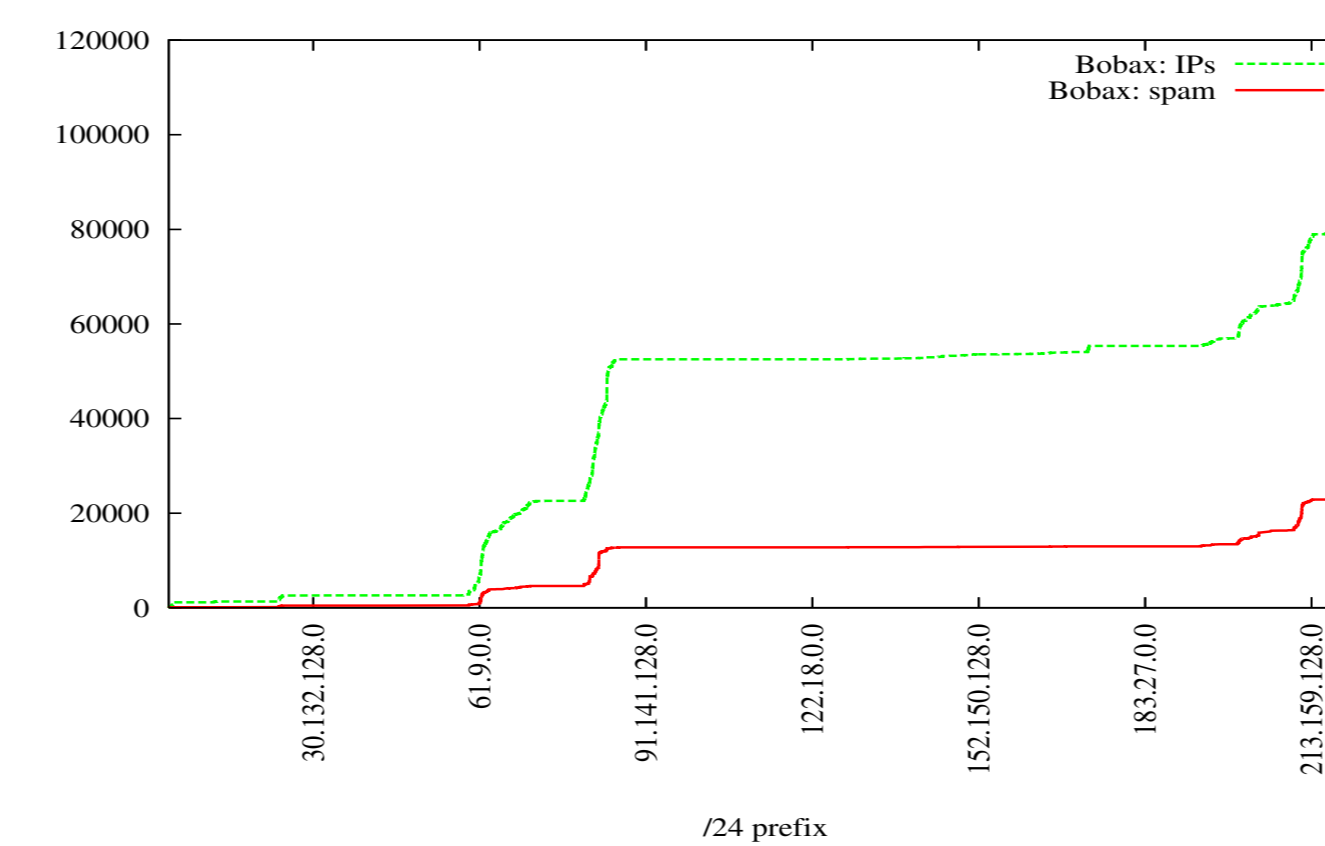


FIGURE 4: Spread across IP space of, and spam sent by, confirmed Bobax-infected hosts

Bobax host distribution is exceedingly similar to overall spam distribution
⇒ suggests that most spam is sent using botnets like Bobax

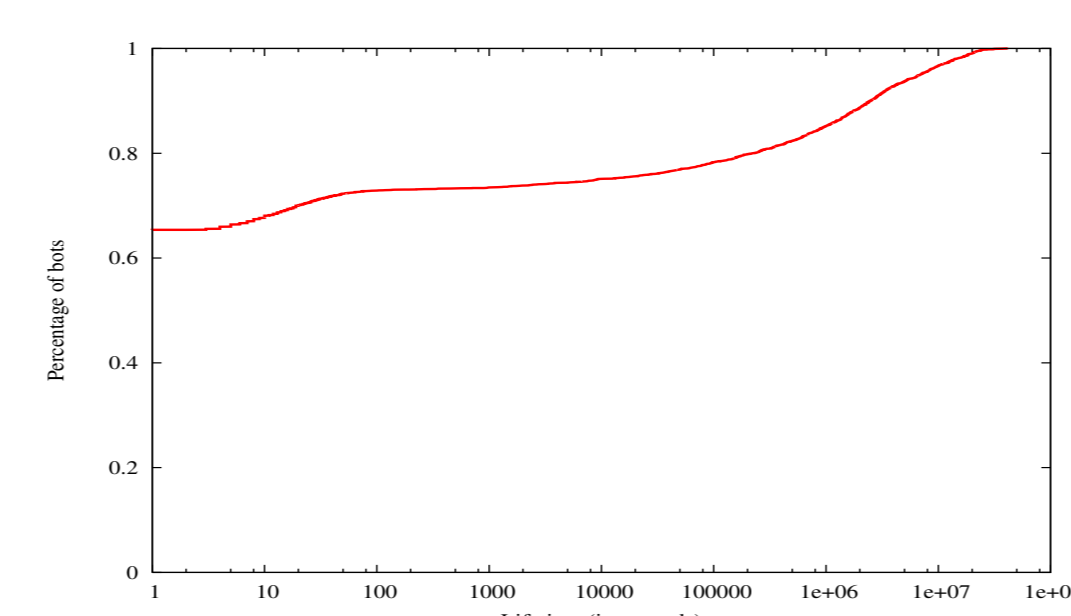


FIGURE 5: Bobax Drone Persistence

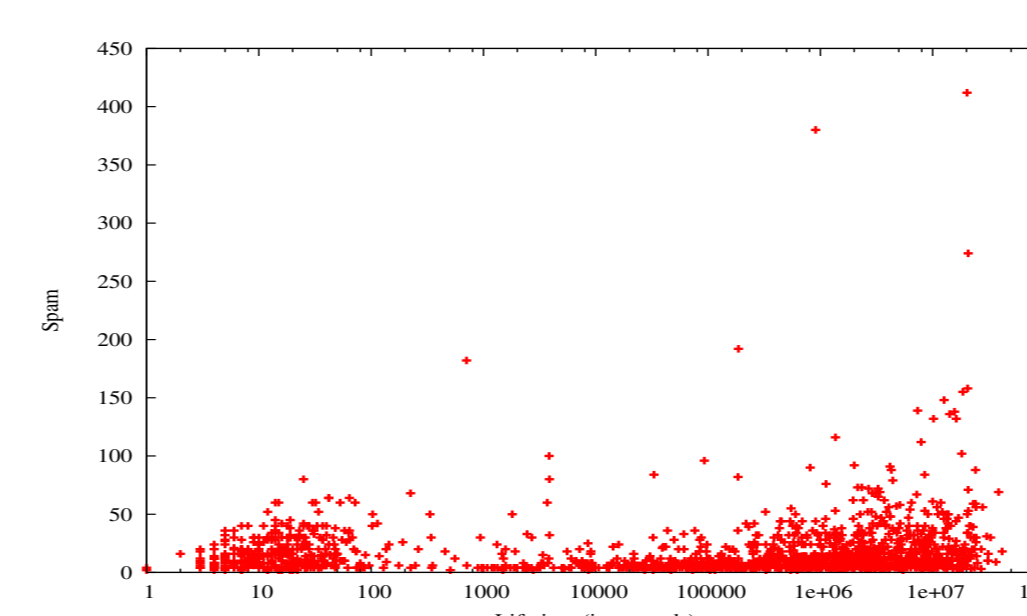


FIGURE 6: Number of spam received vs. Bobax drone persistence

- 65% bots send spam only once, though most of them send multiple pieces of mail in each session to our sinkhole
- Regardless of persistence, 99% of bots sent less than 100 pieces of mail

BGP Spectrum Agility

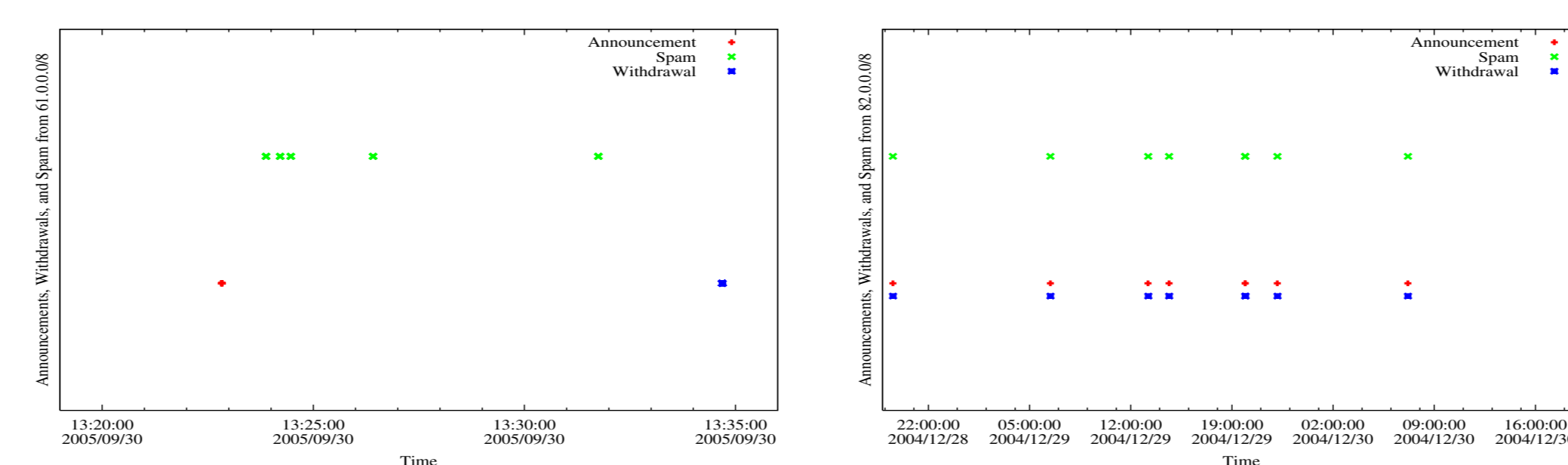


FIGURE 7: Short-lived BGP announcement from two large prefix blocks (61.0.0.0/8 and 82.0.0.0/8), spam arrival from relays in that prefix, and the subsequent withdrawal of the route

- A small, persistent group of spammers hijack large blocks of IP address space (i.e., /8s) to send spam (*Spectrum Agility*)
- The sinkhole received spam from IPs scattered over the hijacked prefix with each IP spamming only once

Efficiency of DNS Blacklists

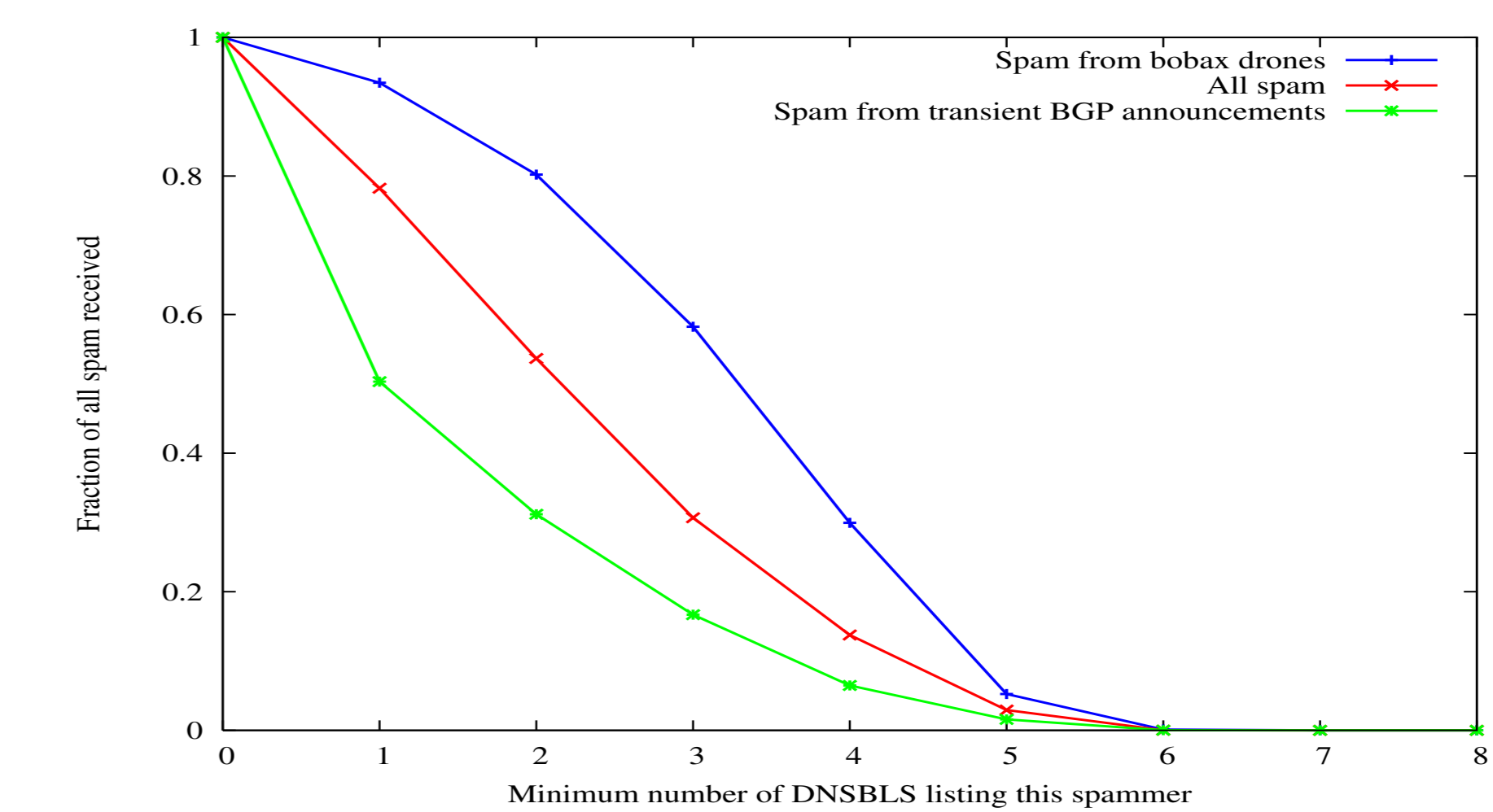


FIGURE 8: The fraction of mails that were listed in a certain number of blacklists or more, at the time each mail was received

- 80% spamming clients appear in atleast one blacklist
- Only 50%, for clients using transient BGP announcements to spam

Lessons for Better Spam Mitigation

- **Lesson 1:** Effective Spam filtering requires a *better notion of end-host identity*. Spam from one-shot bots and transient BGP announcements render the host IP address meaningless
- **Lesson 2:** Detection techniques should be based on *distributions and aggregate behavior*, rather than observations of a single IP address
- **Lesson 3:** *Distributions of spam and host IPs is highly skewed*. Filters could assign higher suspicions for email coming from certain regions of IP space
- **Lesson 4:** *Securing the Internet routing infrastructure* is critical, Spectrum-Agile spammers are to be countered.
- **Lesson 5:** Some *network-level detection* can be incorporated relatively easily into current spam filters; which could be quite effective at detecting spam missed by other techniques.

Reference

A. Ramachandran and N. Feamster. Understanding the Network-level Behavior of Spammers. In *Proc. ACM SIGCOMM*, Pisa, Italy, September 2006 (*To Appear*).