Correlating Spam Activity with IP Address Characteristics

Chris Wilcox, Christos Papadopoulos
CSU
John Heidemann
USC/ISI

IEEE Global Internet Symposium 2010
March 19, 2010

Introduction

- Common belief - spammers have specific address characteristics:
  - We confirmed and measured differences (help identify and mitigate spam)
- Common practice - blocking entire /24 subnet when spammer is present:
  - New result: We quantified collateral damage

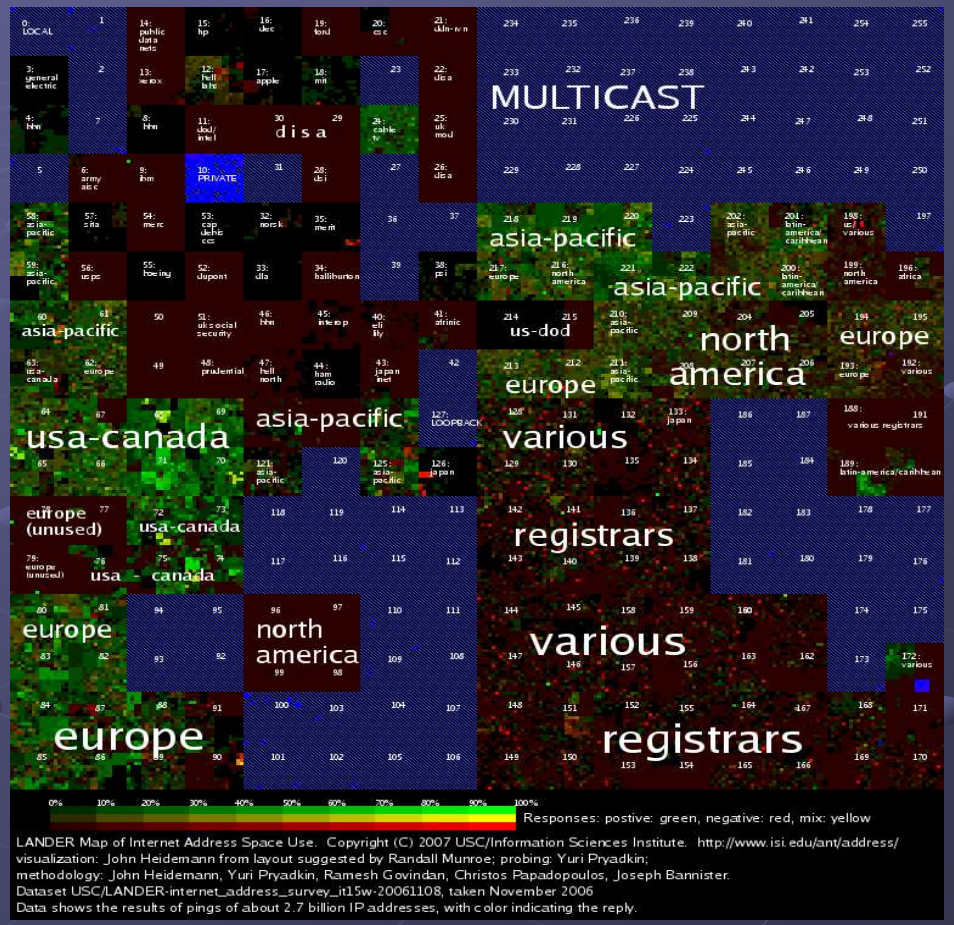
A. Church, "*DNS blacklists considered harmful*", Internet Draft, Work in Progress, Aug. 2005

Approach

- Correlate an IP blacklist with a study of IP visibility to distinguish spammers from non-spammers (control group):
  - Survey of visible /24 subnets (responsive to pings)
  - Commercial IP spam blacklist (reputation-based)

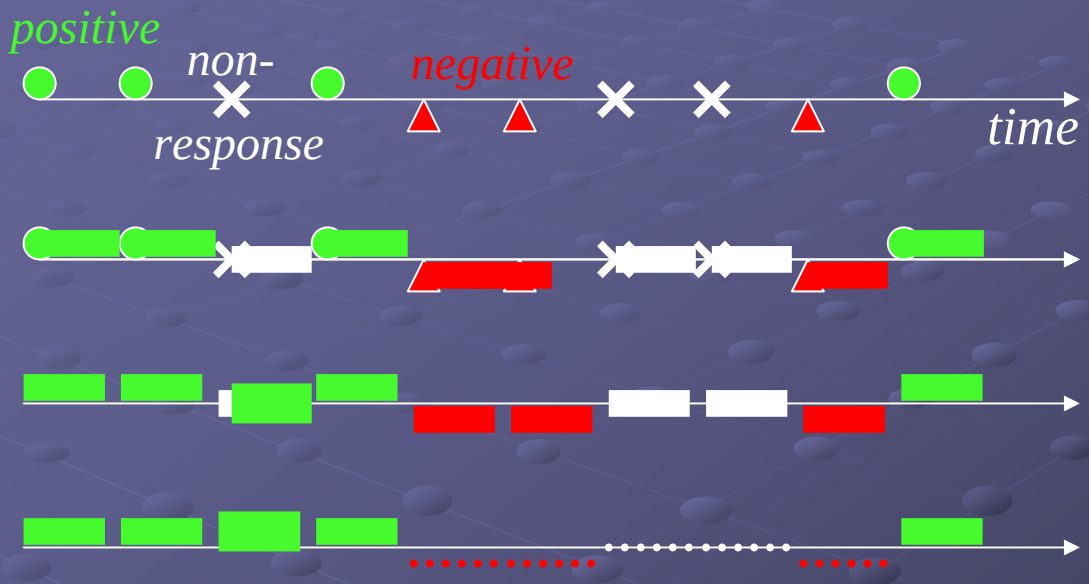
Visibility Study

- Periodic survey of ~24,000 /24 blocks or 1% of Internet
- Duration is ~14 days, hosts probed (ICMP) every 11 minutes



http://www.isi.edu/ant/lander/index.html

Visibility Metrics



Availability := positive replies / echo requests = 5 / 10 = 0.5

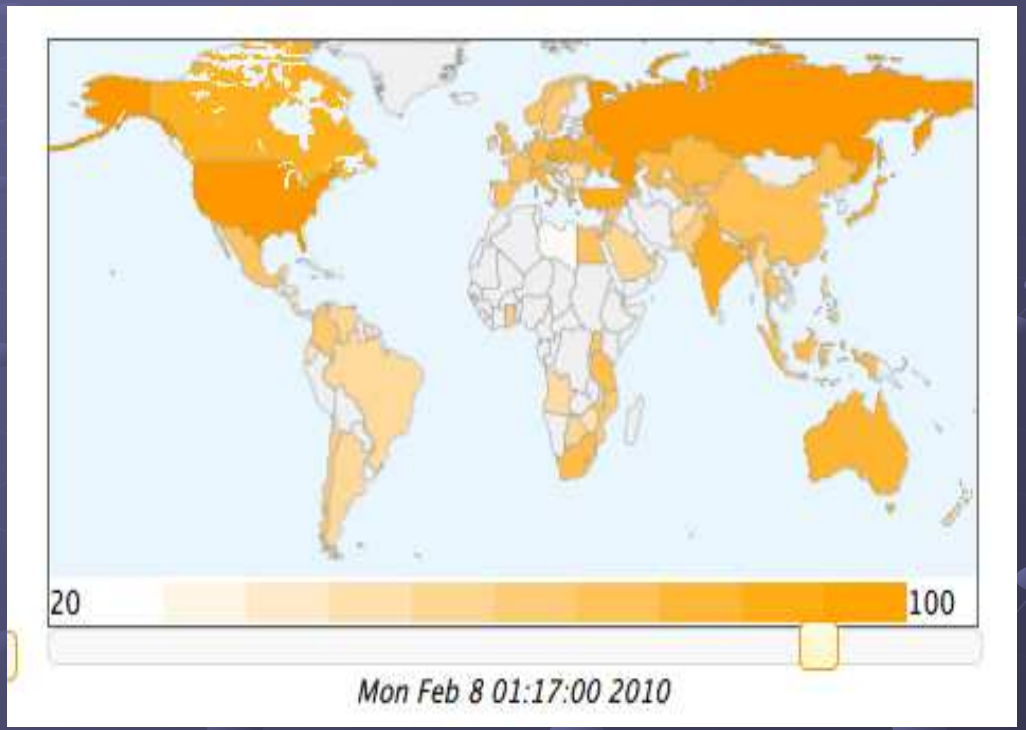
Volatility := responsive periods / (echo requests / 2) = 2 / (10 / 2) = 0.4

Uptime := median of responsive periods = ((44m + 11m) / 2) = 27.5m

5

eSoft Blacklist

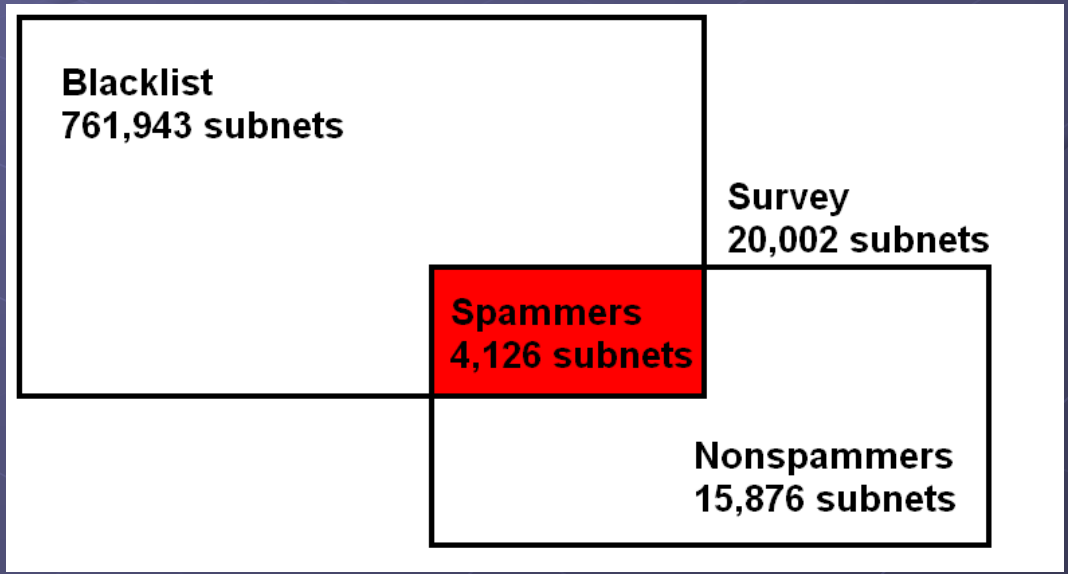
- Reputation based blacklist: <ip, score>
- Delivered by eSoft every 30 minutes
- Global coverage is very good
- Yes, we can share the data



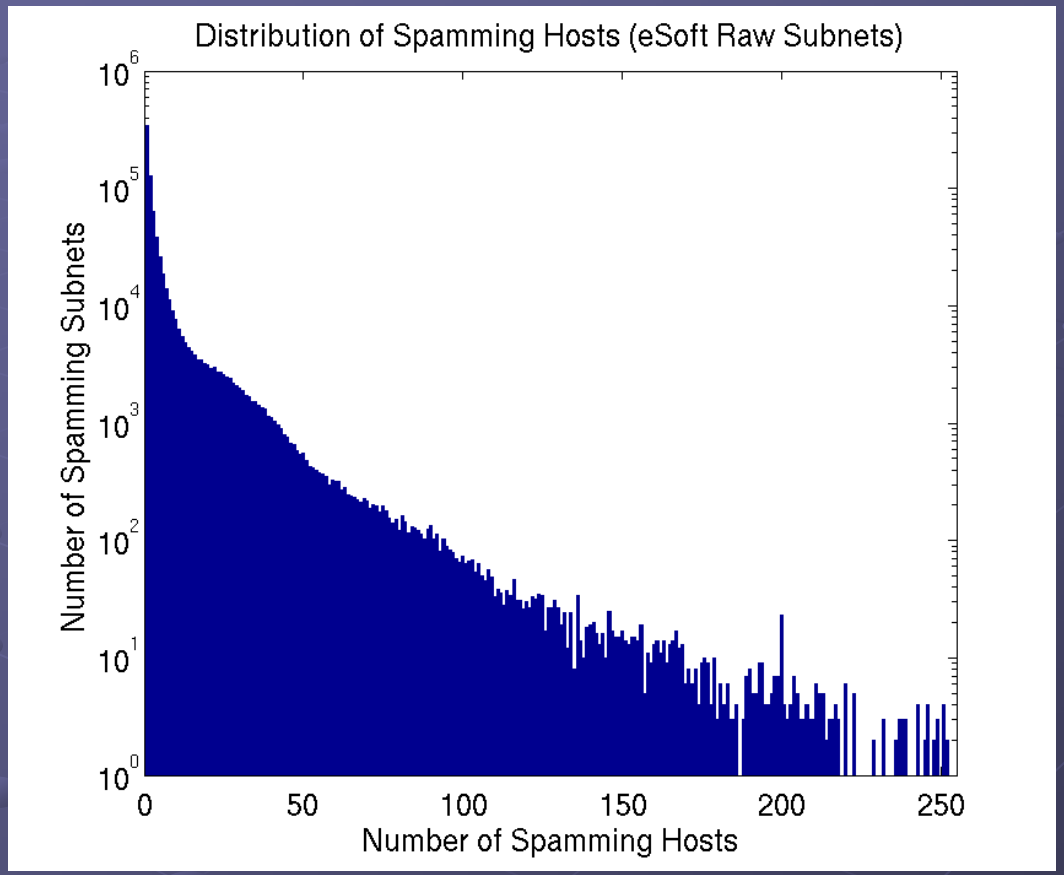
http://www.esoft.com/

Methodology

- Use eSoft scores to differentiate survey subnets into spammers from non-spammers
- Survey subnets divide into 4,126 (21%) spamming and 15,876 (79%) non-spamming

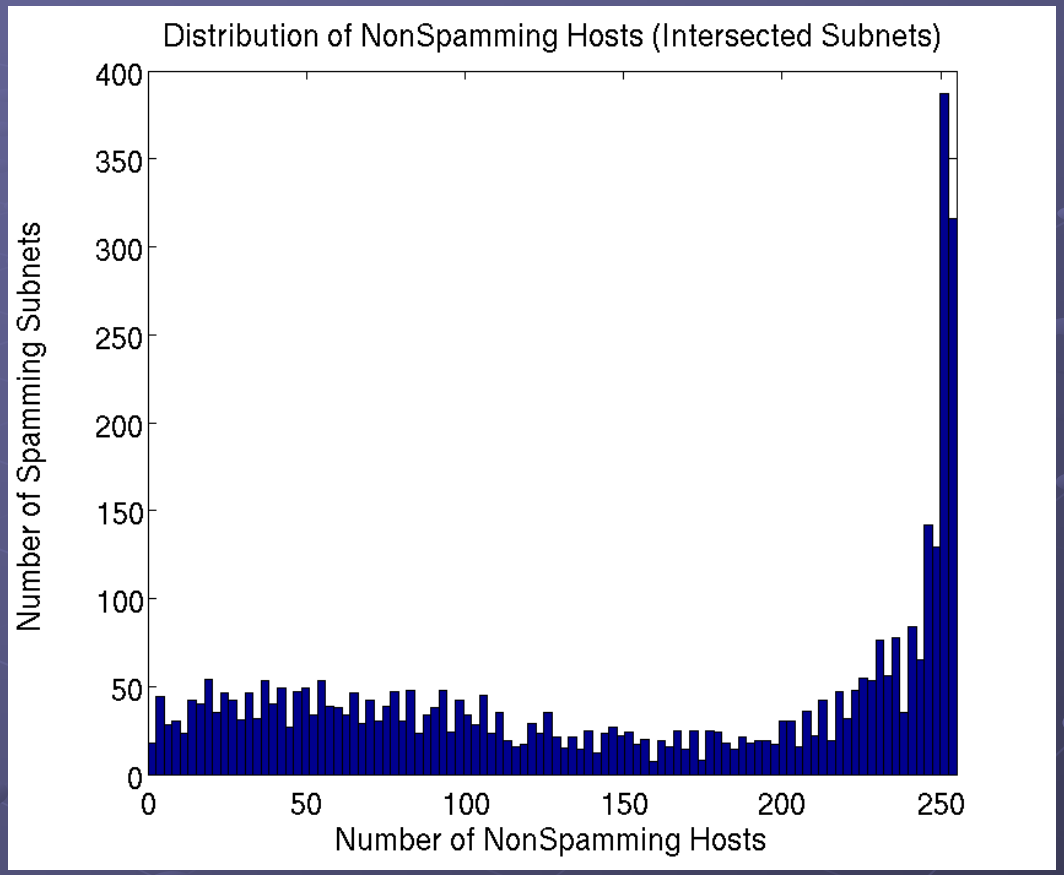


Spammer Distribution



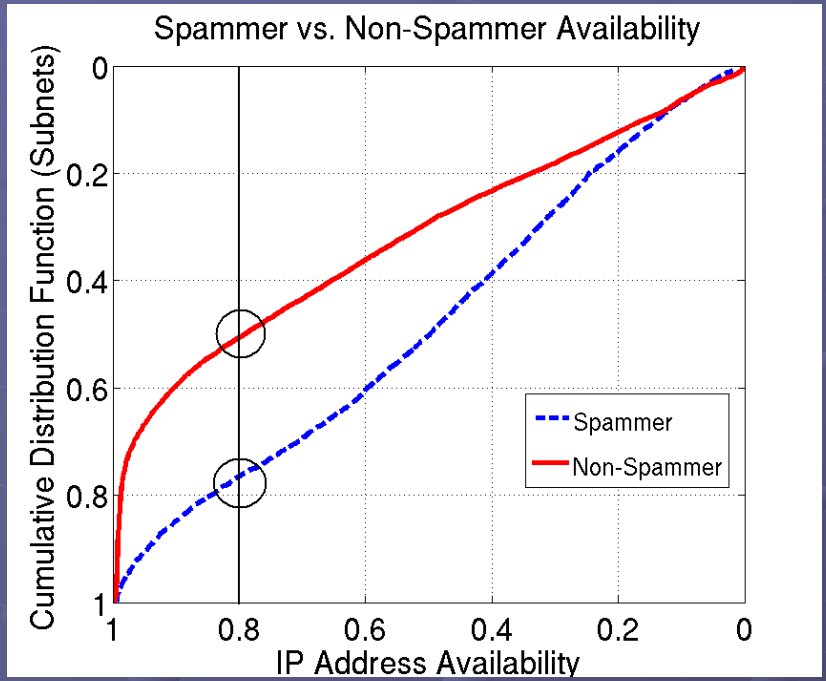
- Majority of subnets have few spamming hosts

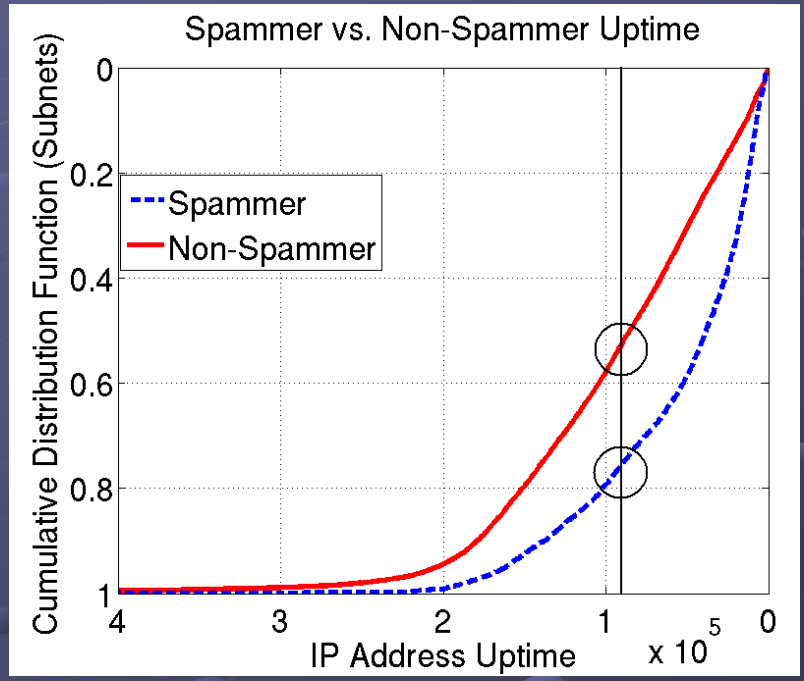
Non-Spammer Distribution



- Non-spamming distribution is more uniform

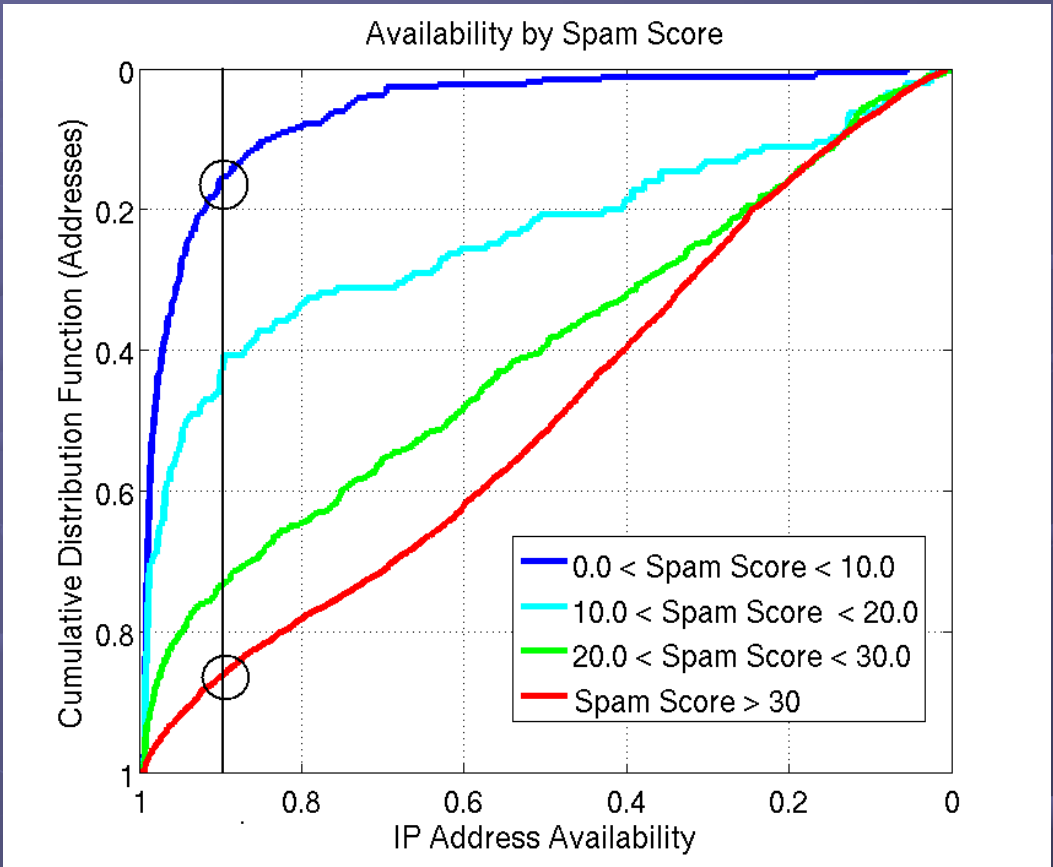
Question 1: Address Characteristics





- 50% of non-spammers, 24% of spammers have >0.8 availability
- 44% of non-spammers, 22% of spammers have > 24 hour uptime

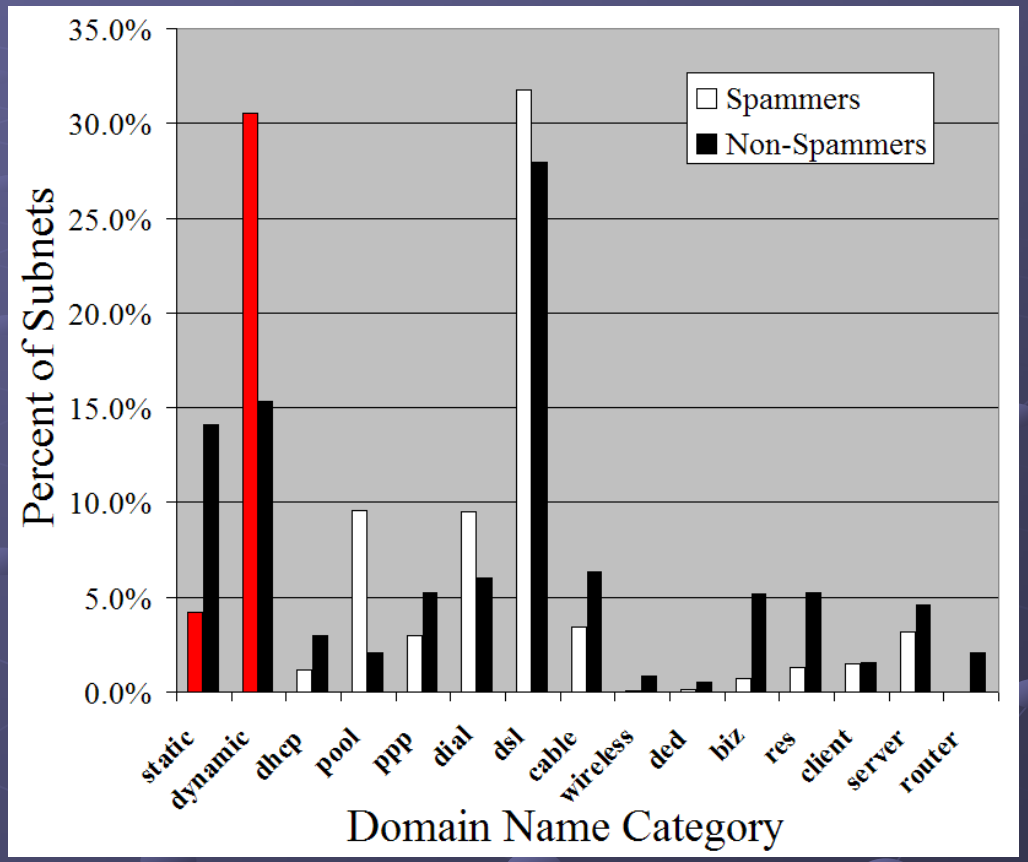
Availability by Spam Score



- 83% of low spammers have > 0.9 availability
- 16% of high spammers have > 0.9 availability

Question 2: Domain Names

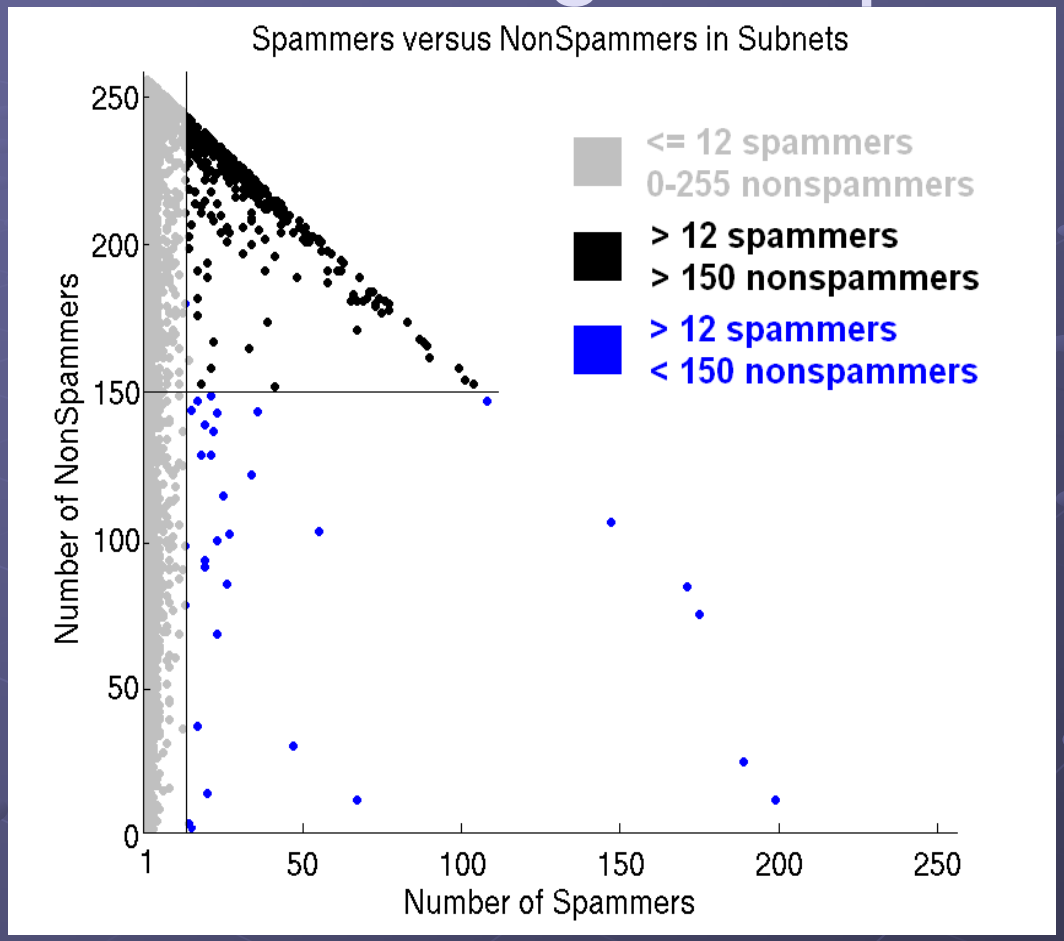
- Dynamic category has more spammers: 30.5% vs. 15.3%
- Static category has fewer spammers: 14.1% vs. 4.2%
- Results confirm previous research



Question 3: Collateral Damage

- Collateral Damage occurs when legitimate mail servers are blacklisted due to spammers on same subnet:
- How much collateral damage?
  - 1) Compute population of spamming and non-spamming hosts per subnet
  - 2) Quantify the number of legitimate mail servers in spamming subnets

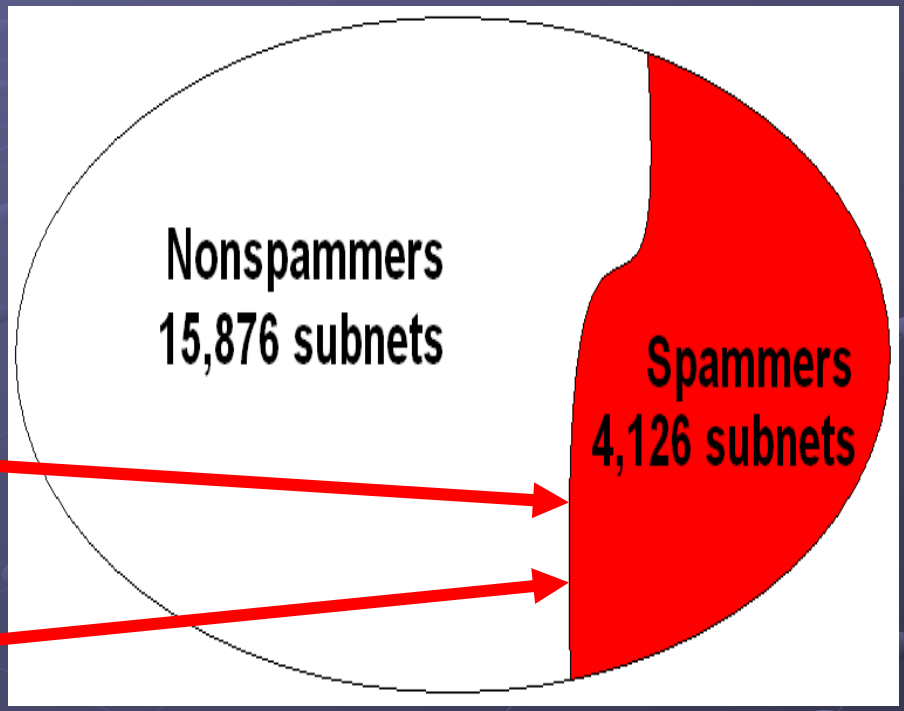
Collateral Damage: Population



- Non-spammers are potential collateral damage.

Collateral Damage: Results

- 3,872 unique mail servers found
- Collateral damage:
  - 1,377 out of 3,872 servers(36%)
  - 365 out of 4,126 subnets (9%)



Blocking /24 subnets disrupts legitimate e-mail!

Conclusions

- Our study confirms major differences in IP address characteristics and domain names between spammers and non-spammers:
  - Useful for identification and mitigation of spamming behavior
- Coarse-grained blacklisting of /24 blocks causes significant collateral damage (36% of mail servers), and should be avoided

https://wiki.netsec.colostate.edu/index.php/Correlate

Robustness

- Blacklists may be incomplete or incorrect
- MX record identification may be inaccurate
- Ping probes may undercount addresses
- Collateral damage should (ideally) consider e-mail volume