Revealing Botnet Membership Using DNSBL Counter-Intelligence

David Dagon

dagon@cc.gatech.edu

Anirudh Ramachandran, Nick Feamster,

College of Computing, Georgia Tech

Georgia Tech

From the presses…

- *“Botnets send masses of spam until they are blacklisted by anti-spam firms. Once blacklisted, the owner sells the botnet to people who launch denial-of-service (DDOS) attacks.”*

- *“Spam clubs also advertise lists of botnets on hire and fresh proxies -- computers that have recently been taken over.”*

-- Steve Linford, CEO, Spamhaus

ZDNet UK News, September 2004

Motivation for this work

- Fact: Bot-herds advertise and sell their "clean" bots at a premium
- Insight: If the claims are true, they must be looking up their bots' status in some blacklist!
- Opportunistic Application: Might it be possible to mine DNS Blacklist *queries* to reveal such *reconnaissance* activity?

DNS Blacklists – How they work

- First: Mail Abuse Prevention System (MAPS)
  - Paul Vixie, Dave Rand -- 1996
- Today: Spamhaus, spamcop, dnsrbl.org etc.

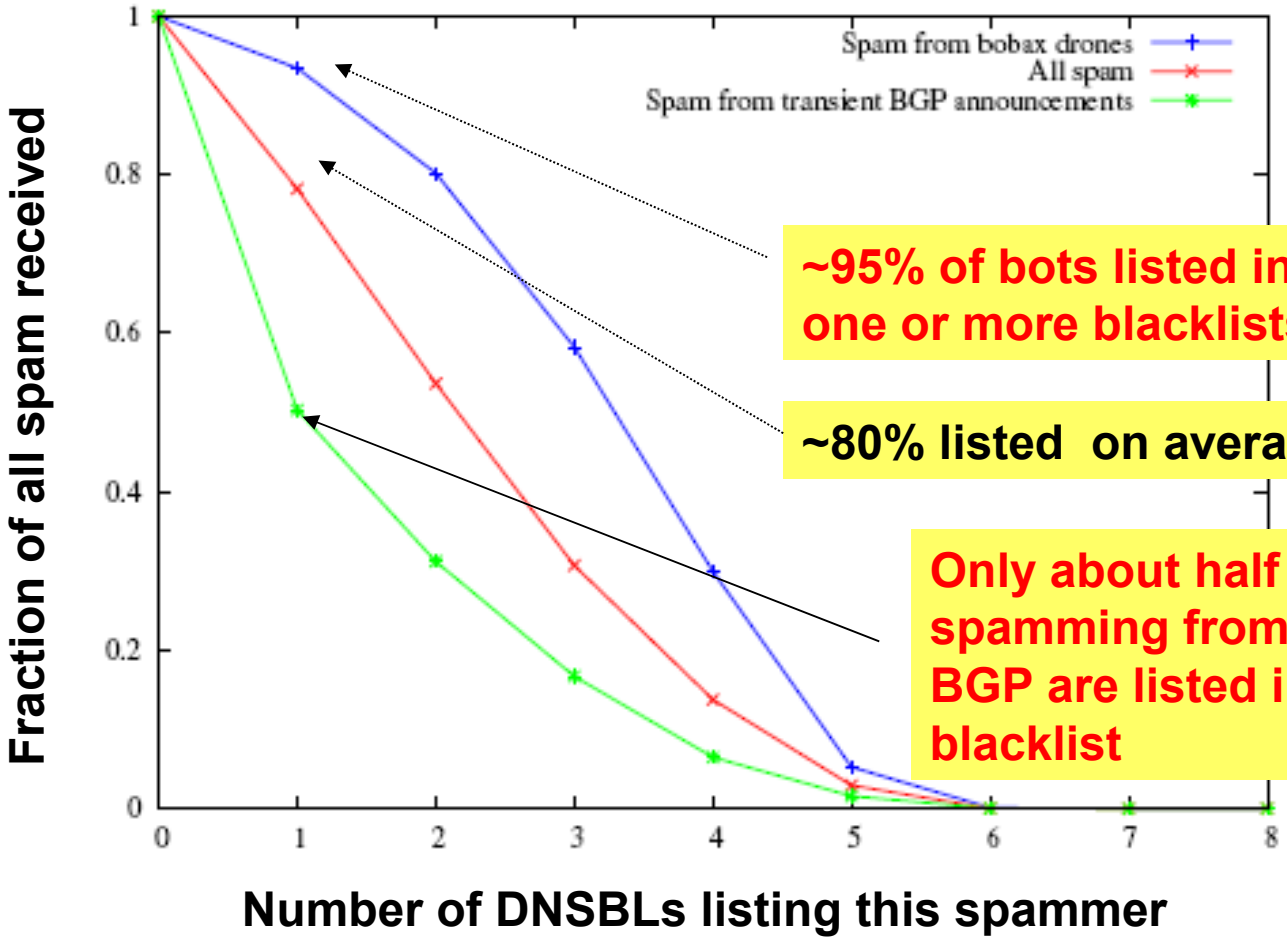
Different Addresses refer to different reasons for blocking

$ dig 91.53.195.211.bl.spamcop.net

;; ANSWER SECTION:
91.53.195.211.bl.spamcop.net. 2100 IN A 127.0.0.2

;; ANSWER SECTION:
91.53.195.211.bl.spamcop.net. 1799 IN TXT "Blocked - see http://www.spamcop.net/bl.shtml?211.195.53.91"

DNSBLs: Useful



~95% of bots listed in one or more blacklists

~80% listed on average

Only about half of the IPs spamming from short-lived BGP are listed in any blacklist

DNSBLs have some value; spam from IP-agile senders tend to be listed in fewer blacklists

Outline

- Motivation
- Detecting Reconnaissance
- Reconnaissance Techniques
- Analysis and Results
- Mitigation and Countermeasures
- Conclusion

Detecting Reconnaissance

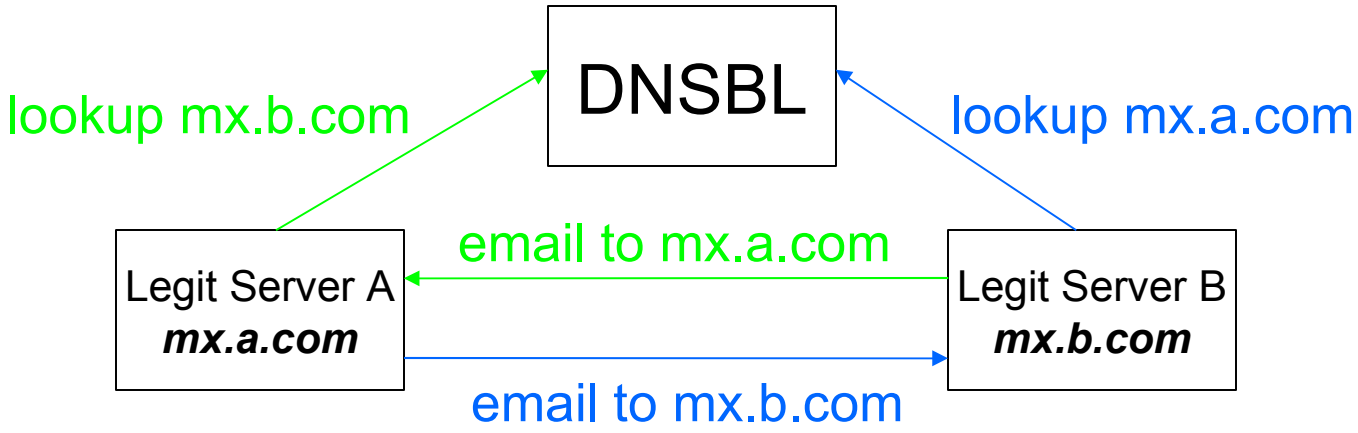
- *Key Requirement:* Distinguish reconnaissance queries from queries performed by legitimate mail servers

- *Our Solution:* Develop heuristics based on the spatial and temporal properties of a *DNSBL Query Graph*

- We focus (mostly) on spatial heuristics

Heuristics

- *Spatial Heuristic:* Legitimate mail servers will perform queries *and be the object of queries.*

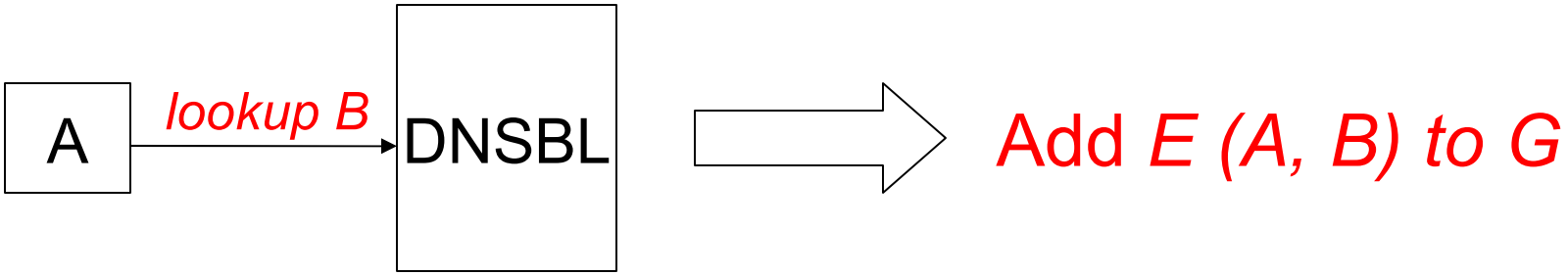


  - *Hosts issuing reconnaissance queries usually will not be queried*

- *Temporal Heuristic:* Legitimate lookups reflect arrival patterns of legitimate email

Applying the Spatial Heuristic

- Construct the directed *DNSBL Query Graph G*



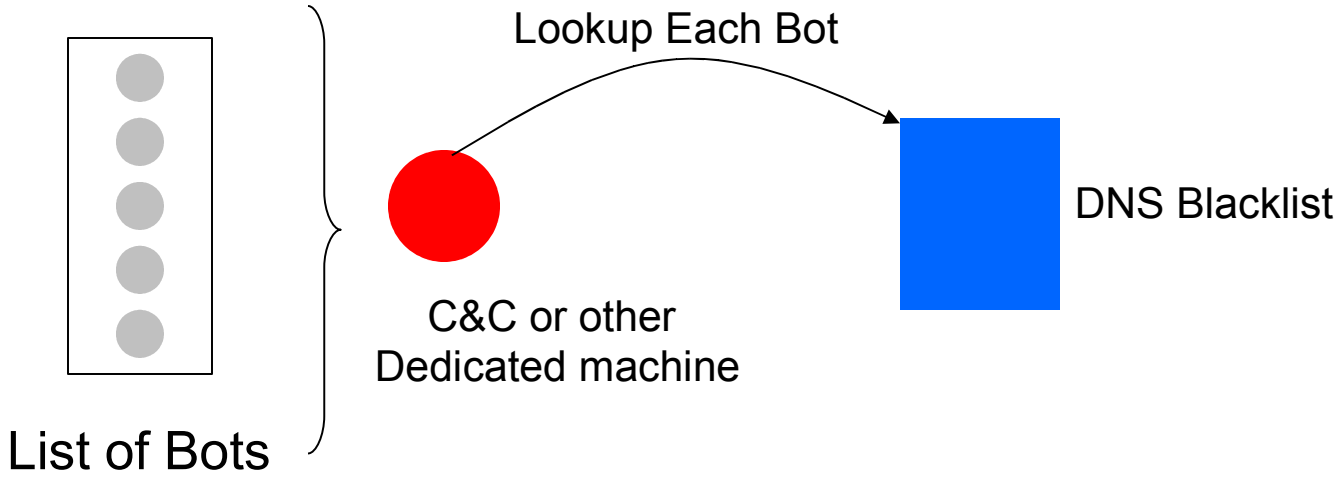
- *Extract nodes (and their connected components) with the highest values of the spatial metric λ, where λ = (Out-degree/In-degree)*

Outline

- Motivation
- Detecting Reconnaissance
- Reconnaissance Techniques
  - Third-party reconnaissance
  - Self-reconnaissance
  - Distributed reconnaissance
- Analysis and Results
- Mitigation and Countermeasures
- Conclusion

Third-Party Reconnaissance

- *Third-party performs reconnaissance query*



- Relatively easy to detect using the spatial metric

Other Techniques

- *Self-Reconnaissance*
  - Each bot looks itself up
  - This should not happen normally (at least, not *en-masse*) – thus, easy to detect
- *Distributed Reconnaissance*
  - Bots perform lookups for other bots
  - Complex to deploy and operate
  - *We witnessed evidence of this technique*

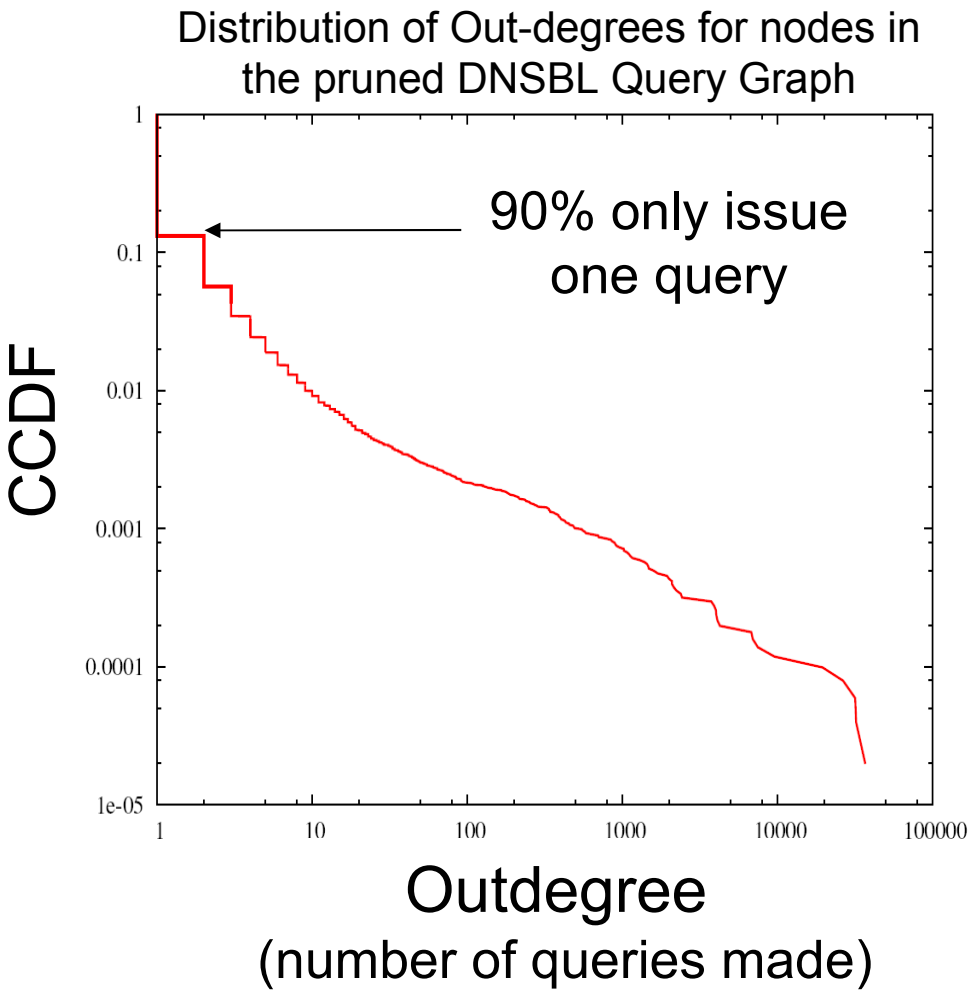
Outline

- Motivation
- Detecting Reconnaissance
- Reconnaissance Techniques
- Analysis and Results
- Mitigation and Countermeasures
- Conclusion

Analysis

- Data
  - Two days' worth of pcaps of DNSBL lookups from a mirror of a large DNS Blacklist Provider
  - A 'seed' list of known spambots (Bobax) active around the same time-period, to prune lookup logs
- Analysis
  - Extract nodes with highest values of λ, with their respective connected components

Prevalence of Reconnaissance



- *Long tail* – Bot-herds might already have the capability to distribute reconnaissance among many bots
- *A few high out-degree nodes* – multiple vantage points might help identify "prominent players"

Findings

- Many of the nodes with highest values of λ were *known bots*
- Nodes being looked-up were unlisted, possibly newly compromised bots
  - Our spam trap had already captured spam from a few of these nodes

Node #	Out-degree	Known Spammers (observed at spam trap)
1 (Everyone's Internet, AS13749)	36,875	12
2 (IQuest, AS7332)	32,159	7
3 (UUNet, AS701)	31,682	5
4 (UPC Broadband, AS6830)	26,502	8
5 (E-xpedient, AS17054)	19,530	4

Implications

- Bad news! Bot reconnaissance techniques are pretty advanced

- Good news, too
  - Can use these spatial dependencies to opportunistically identify new bots

Opportunistic Bot Detection

- Many sources of data for *bootstrapping* passive botnet detection (*i.e.,* to compile a 'seed' list) like
  - SMTP/Spam logs,
  - Portscan logs from Intrusion Detection Systems
- Knowledge of botnet membership → ability to stop attacks closer to the source
- Multiple vantage points increase confidence and reduce risk of false positives.

Countermeasures

- Proactive blacklisting of newly-identified bots
- Reconnaissance Poisoning
  - *Return 'listed' for an unlisted bot:* might prevent bot from being used
  - *Return 'unlisted' for a listed bot:* trick bot-herd into using blacklisted IPs
- Caveats
  - Risk of false positives

Summary

- DNSBL logs provide a means for *passive botnet detection*, by correlating with a set of known bots
- *More vantage points*, and other data sources containing bot activity (*e.g.,* spam logs), will help increase confidence
- Potential to identify bots *before they are used*, and also to mislead bot-herds into using blacklisted bots

Future Work

- The botnet use of DNSBLs is an abuse requiring future study
- Currently performing distributed data mining for DNSBLs
  - 3 sites; soon to be more
  - Focused on botnet recon
- Run a mirror (or want to)? Want to help a research project?
  - Please contact dagon@cc.gatech.edu

Questions?

- Thank you for your time
- Thanks to our hosts