Revealing Botnet Membership using DNSBL Counter-Intelligence

Nick Feamster

Anirudh Ramachandran
David Dagon
Georgia Tech

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?

Detecting Reconnaissance

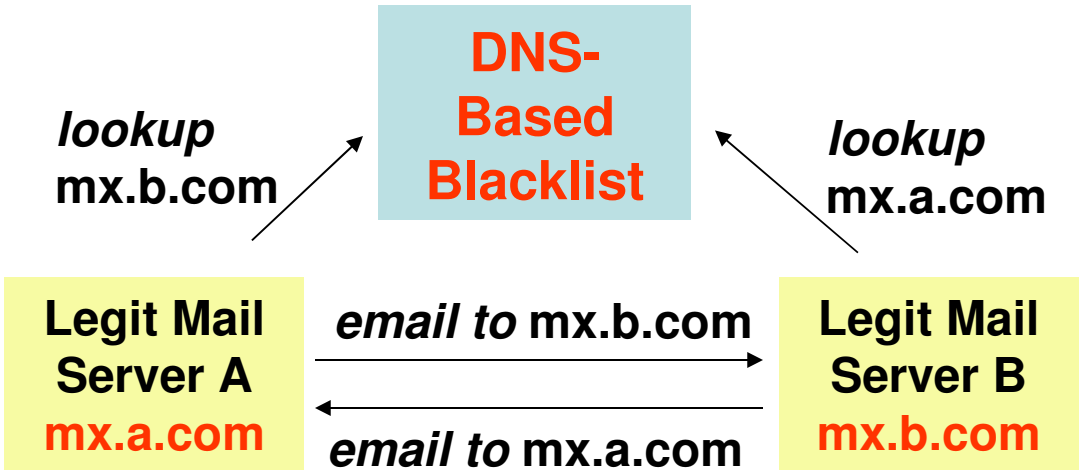
- *Key Requirement:* Distinguish reconnaissance queries from queries performed by legitimate mail servers

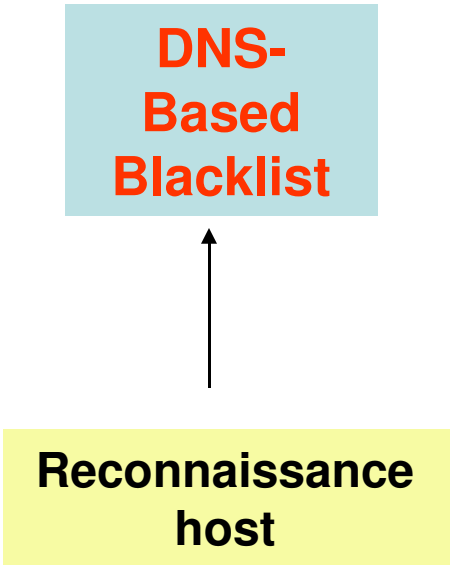
- *Our Approach:* Develop heuristics based on the spatial and temporal properties of a *DNSBL Query Graph*

- We focus (mostly) on spatial heuristics

Legit Queries vs. Reconnaissance

- Legitimate queriers are also the targets of queries
- Reconnaissance queriers are ususally not queried themselves



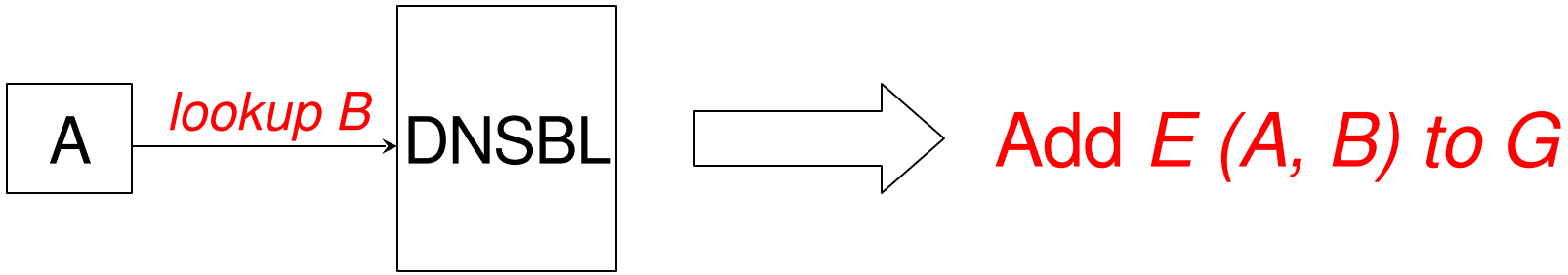


Measurement Approach

- Log Spamhaus queries
- Construct querier/queried graph
- Prune graph: only nodes in the Bobax trace
- Examine nodes with high out-degree
  - **Hypothesis:** targets of nodes with high out-degree likely bots

Applying the Spatial Heuristic

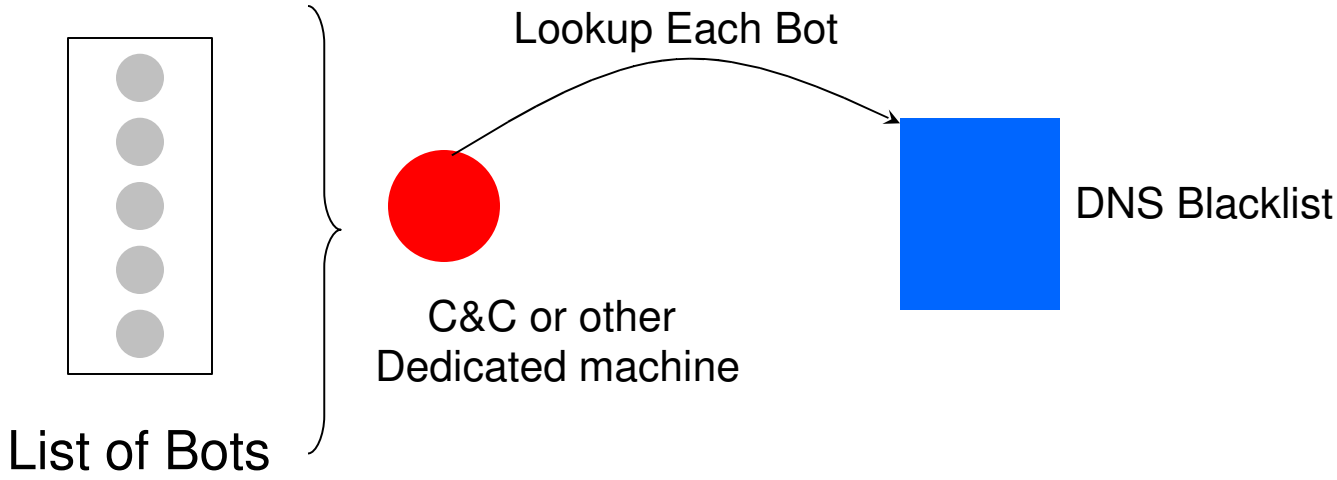
- Construct the directed *DNSBL Query Graph G*



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

Third-Party Reconnaissance

- *Third-party performs reconnaissance query*



- Relatively easy to detect using the spatial metric

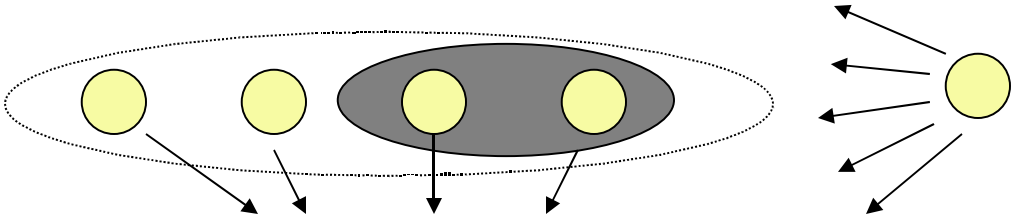
Other Techniques

- 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*

Distributed Reconnaissance

- The botmaster, on behalf of the bots
- The bots, on behalf of themselves
- **The bots, on behalf of each other**

ASN of Node	Out-degree
Everyone's Internet (AS 13749)	36,875
IQuest (AS 7332)	32,159
UUNet (AS 701)	31,682
UPC Broadband (AS 6830)	26,502
E-xpedient (AS 17054)	19,530

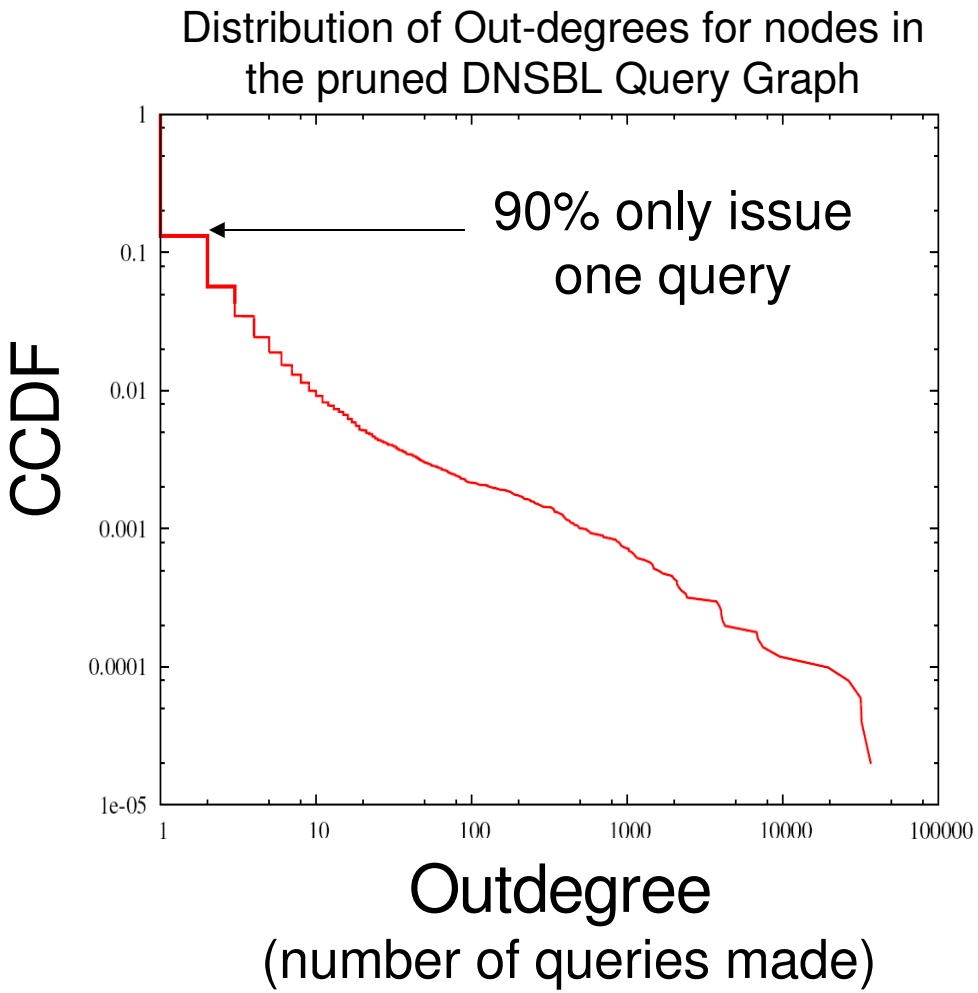


Known bobax drone!

Spam Sinkhole

Implication: Use a "seed" set to bootstrap?

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"

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.

Some Problems with Counter-Intel

- Constructing the query graph is intensive
  - Computationally
  - Storage-wise

- Initially pruning the graph with IP addresses of known suspects (e.g., spammers) could help