Monitoring the Initial DNS Behavior of Malicious Domains

Shuang Hao (Gatech), Nick Feamster (Gatech), Ramakant Pandrangi (Verisign, Inc.)
DNS: A Critical Internet Service

- A distributed database mapping host names to IPs
  - Most network connections are preceded by DNS lookups

More than 215 million domain name registrations across all top-level domains (TLDs) (Source: Zooknic, Verisign, July 2011)
Why Monitor DNS Activities?

• Domains are registered to host malicious content
  – Direct to scam, phishing or malware sites
    Hey, you look funny in that video... http://bad-domain.com/bcddf
  – > 56% malicious domains are second-level domains
    (source: SIE)

• Monitor domains’ behaviors to mitigate threats
  – Investigation is usually triggered after attacks take place

• Domain registration grows quickly
  – ~150 thousand new .com and .net domains every day

*It is challenging to monitor DNS activities!*
Highlights of Our Study

“Monitoring the Initial DNS Behavior of Malicious Domains”

Start monitoring as soon as a new domain is registered

1) Active queries to authoritative servers periodically to fetch resource records
2) DNS lookups collected from Verisign top-level domain servers

Domains identified by appearance in spam traps
Motivation

Questions

– *When does a malicious domain start to be used in attack after registration?*
  Purpose: The potential time window to prevent attack happening

– *What networks are the resource records mapped to?*
  Purpose: Re-used IPs or ASes to identify bad domain registration

– *Who looks up which domains?*
  Purpose: Global DNS traffic to find patterns across malicious domains
Talk Outline

• Motivation

• **DNS Data Monitoring**
  – Categorizing malicious and legitimate domains
  – Collecting snapshots of resource records
  – Monitoring DNS lookups

• Findings in the DNS Characteristics

• Conclusion
Categorizing Malicious & Legitimate Domains

• Target domains
  – Newly registered second-level domains (2LDs) under .com and .net during March 2011
    • On average, 150 thousand 2LDs get registered everyday
    • Continuous monitoring throughout the month

• Define as “malicious”
  – 5,988 2LDs identified in spam trap (including spamhaus) during March 2011

• Legitimate domain samples
  – Sample 6,000 new domains that have not appeared in any blacklist
Collecting Snapshots of Resource Records

• Resolved IPs from resource records (RRs)

<table>
<thead>
<tr>
<th>record type</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS</td>
<td>the authoritative name server</td>
</tr>
<tr>
<td>MX</td>
<td>a mail server for the domain</td>
</tr>
<tr>
<td>A</td>
<td>IP address of a host</td>
</tr>
</tbody>
</table>

• Collection process
  – Zone update logged at TLD servers (NS-type RRs)
    • Include alerts of new domain registration
      
    add-new example.com NS ns1.example.com
  – Continuous active querying (NS, MX, A types of RRs)
    • Daily queries dispatched from PlanetLab
Monitoring DNS Lookups

• Collection process *
  – Querying /24 subnets aggregated every day

  example.com  111.111.111.0 , 222.222.222.0

* Similar monitoring point used in “Detecting Malware Domains at the Upper DNS Hierarchy”. In USENIX Security (2011).
Talk Outline

• Motivation

• DNS Data Monitoring

• Findings in the DNS Characteristics
  – How long is the delay until attack?
  – What networks are the resource records mapped to?
  – Who looks up which domains?

• Conclusion
Time Between Registration and Attack

- Time when first observing records about the malicious domains, to the earliest time when the domains appeared in the spam messages.

**Finding**: About 55% of the malicious domains showed in spam more than one day after they were registered.
Resolved DNS Records across IP space

- The A records of 2.6 million 2LDs registered in March 2011 were mapped to 300 thousand IPs (similar statistics for NS and MX records)

Finding: A small fraction of IP space is heavily used to host malicious domains, even within the pre-attack period.
Lookup Patterns across Networks

- If two domains are queried by the same set of recursive DNS servers, they may be the same type of domains.
- Intuition: A user clicking a URL in spam might click on other spam URLs.
Lookup Patterns across Networks

- If two domains are queried by the same set of recursive DNS servers, they may be the same type of domains
- Intuition: A user clicking a URL in spam might click on other spam URLs

\[
S(D_A, D_B) = \frac{(J_1 + J_2 + \ldots + J_n)}{n}
\]

* Jaccard index of two sets = the size of the set intersection divided by the size of union
Lookup Patterns across Networks (Cont.)

- Clustering based on initial querying /24s (5-day from March 1--5, 2011)

<table>
<thead>
<tr>
<th></th>
<th>total</th>
<th>malicious</th>
<th>legitimate</th>
<th>% spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>1404</td>
<td>463</td>
<td>941</td>
<td></td>
<td>33.0%</td>
</tr>
<tr>
<td>157</td>
<td>156</td>
<td>1</td>
<td></td>
<td>99.4%</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>0</td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td></td>
<td>100.0%</td>
</tr>
</tbody>
</table>

- **Finding**: Malicious domains in the same campaign are looked up by similar group of recursive servers
Conclusion

- *How long is the delay until attack?*
  Purpose: The potential time window to prevent attack happening
  Finding: 50% malicious domains have more than one day inactivity before attack

- *What networks are the resource records mapped to?*
  Purpose: Re-used IPs or ASes to identify bad domain registration
  Finding: Some networks have more IPs pointed from bad domains' RRs

- *Who looks up which domains?*
  Purpose: Abnormal lookup patterns indicating malicious activities
  Finding: Similar groups could query multiple malicious domains

http://www.cc.gatech.edu/~shao