Domain names abuse and TLDs: from monetization towards mitigation

Giovane C. M. Moura, Moritz Müller, Marco Davids, Maarten Wullink, and Cristian Hesselman
SIDN Labs
{first.lastname}@sidn.nl

DNS OARC Meeting
May 15th, 2017
Madrid, Spain
Introduction

- DNS provides a simple label for hosts, services, applications on the Internet
- Often, it is misused in malicious activities such as:
  - phishing campaigns
  - malware
  - spam
- Underlying each type of abuse, a different business model
  - provides the incentives for the crooks to keep on
Introduction

- Plenty of research work in curbing DNS-related abuse [1, 2, 3, 4, 5, 6]
  - With a clear contribution
- But, they suffer from similar issues:
  1. Bound by dataset type/duration
  2. Cover specific attacks; missing broader view on all abuses
- This paper:
  1. Cover first issue with longitudinal measurements and registration (.nl)
  2. Present a survey on domain abuses from the point of view of a TLD operator (centralized view)
Motivation: why doing this?

Came from a situation we faced:

- There’s no one size fits all
- we have all this data
- how to better use it?
- where to begin with?
  - e.g.: malicious registered phishing or compromised phishing?
  - or other sort of abuse?
  - how to prioritize it?
  - **Which datasets too look first?**
- Other TLD operators may be facing the same problem
Understanding business models

- Helps you to understand how money is made
- And how it impact your datasets
- It’s been done many times in Internet abuse. E.g.: PharmaLeaks[7].
- Business model → abuse → money
TLD Operations and Datasets

Figure: TLD Operations: registration (left), domain name resolution (right), and derived datasets.

- **RegDB**: your registration DB
- **Zone File Scans**: in our case, the OpenIntel.nl project
- **AuthDNS**: data from auth servers, we use ENTRADA [8]
<table>
<thead>
<tr>
<th>Business</th>
<th>Spam</th>
<th>RegDB</th>
<th>AuthDNS</th>
<th>Records</th>
<th>Lit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phishing(0-day)</td>
<td>Yes</td>
<td>Weak</td>
<td>Strong</td>
<td>Weak</td>
<td>[3, 6]</td>
</tr>
<tr>
<td>Phishing(comp.)</td>
<td>Yes</td>
<td>None</td>
<td>Strong</td>
<td>Weak</td>
<td>[9]</td>
</tr>
<tr>
<td>Parking (Ads)</td>
<td>No</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>[10, 11]</td>
</tr>
<tr>
<td>Parking (Mal)</td>
<td>No</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>[10, 11]</td>
</tr>
<tr>
<td>Fake Goods</td>
<td>Yes</td>
<td>Weak</td>
<td>Weak</td>
<td>Medium</td>
<td>[6, 7]</td>
</tr>
<tr>
<td>Drop-Catch</td>
<td>No</td>
<td>Medium</td>
<td>Medium</td>
<td>Weak</td>
<td>[12]</td>
</tr>
<tr>
<td>Botnet C&amp;C</td>
<td>No</td>
<td>Medium</td>
<td>Strong</td>
<td>?</td>
<td>[13]</td>
</tr>
<tr>
<td>Blackhat SEO</td>
<td>No</td>
<td>Medium</td>
<td>Medium</td>
<td>Strong</td>
<td>[14, 15]</td>
</tr>
</tbody>
</table>

Table: Business Models and Datasets/signal “strength”, and research works that cover those.
Phising (0-day)

- Two types of phishing: compromised and 0-day (newly registered)
- 0-day phishing business model:
  1. Registered domain(s)
  2. Large spam campaign at the same time
  3. ID theft (ID, credit card, etc).
  4. Money: selling the data, using it themselves
Phishing (0-day)

- **Datasets:**
  1. **Records:** harder to detect, IP/registrar reputation
  2. **RegDB:** hard but possible to detect (it’s been done for spamming domains [16])
  3. **AuthDNS:** strongest signal, but after attack has started [3, 6]

---

**Figure:** .nl Random vs Phishing new domains average daily queries [6]
Phishing (compromised)

- Most common sort of phishing
- Typically on hacked CMSes, instead of newly registered
- Business model:
  1. Hack a website
  2. Sam campaign at the same time
  3. ID theft (ID, credit card, etc).
  4. Money: selling the data, using it themselves
Phishing (compromised)

- Datasets:
  1. Records: harder to detect, typically no changes
  2. RegDB: also, usually no changes in here
  3. AuthDNS: possible to detect, very hard to tell false positives

Figure: Median daily queries for 1,374 compromised phishing sites on .nl, before and after Netcraft's notification
Parking

- Parking is a big industry

- Business model:
  1. Register many domains (bulk)
  2. Wait for traffic to come in
  3. Redirect to ad networks
  4. Money:
     - Legal: ad networks
     - Illegal: malicious, ID-theft
Parking

Datasets:

1. Records: can be done, same ASes, IPs, etc
2. RegDB: Yes, bulk registrations, same registrar, etc.
3. AuthDNS: usually not the case

Figure: Number of domains registered for one registrar, in every 30min – spikes indicates anomalies
Datasets:

1. Records: can be done, same ASes, IPs, etc.
2. RegDB: Yes, bulk registrations, same registrar, etc.
3. AuthDNS: usually not the case

Figure: Anomalous registrations for Registrars and Top 1 registrant – most of registrations are done in bulk by 1 registrant.
Key aspect: tell ad networks from malicious (e.g.: malicious redirection)

Malicious redirection type has more incentives to use a new e-mail addresses during registrations (and no reuse)

We’ve seen that for ad nets

Need to develop a solution that address this (open)
Fake Goods

- When we develop nDEWS [8] to detect 0-day phishing, we notice a lot of domains were neither phishing neither false positives
- Their frequency and continuity suggested a profitable business model
- Just like phishing (0-day) business model, and detection too
- This sort of abuse falls into a “gray area”:
  - not as bad as phishing
  - still bad because of ID theft
  - hard to tell if it’s fake or not
- Detection: similar to 0-day phishing
Botnet C&C

- Domains can be used also for botnet command-and-control channels
- Domain generation algorithms (DGA) typically used for that
- Bots are supposed to contact a new domain every $x$ time
- DGAs generated many, but only a few are registered, to avoid detection
Business model: registration

Datasets:
1. RegDB: registration of “weird” looking names
2. Records: in combination with the previous
3. AuthDNS: NXdomain queries for non registered DGAs
Summary

- DNS abuse has been active for many years
- There are many types, which different business models
- Business models of DNS abuse impact datasets differently
- TLDs ops should develop applications according to business models
  - no one-size fits all
- Which one first?
  - that depends on the frequency of the abuse on their zone
- This paper presents a survey and a discussion on which datasets can be used
  - And some of our experience with these abuses on .nl
Questions?

▶ Contact:
  ▶ http://sidnlabs.nl
  ▶ giovane.moura@sidn.nl
  ▶ Twitter: @giomourasec

▶ Thank you for your attention

Download our paper at:

Bibliography I


