Creating a "long-term memory" for the global DNS

in numbers:

- 216 million domains measured on a daily basis
- 2.3 billion data points collected daily
- 2.9 trillion data points collected since the start in 2015

OPENINTEL IS A JOINT PROJECT

UNIVERSITY OF TWENTE, SURFNET, SDN LABS, NLNET LABS
Introduction

• Almost **five years ago**, we started with **an idea**:

"Can we measure (large parts of) the global DNS on a daily basis?"

• In this talk, we will discuss:
  • **Why** we wanted to do this
  • **How** we do it
  • And examples of **what we have learned** so far
Why measure the DNS?

- **DNS translates** from the **human** world **to** the **machine** world (and also helps in machine-to-machine interaction)

- (Almost) every networked service relies on the DNS

- Consequently, measuring what is in the DNS tells a story about the evolution of the Internet and its protocols
Hasn't someone tried this before?

- You may be familiar with **passive DNS** (popular in the security community)

- Has **two downsides**:
  1. Only sees what clients ask for (and is thus **biased**)!
  2. No control over query timing, so **unsuitable for time series**
How we measure

- **OpenINTEL performs an active measurement**, sending a fixed set of queries for all covered domains **once every 24 hours**

- We do this **at scale**, covering **over 218 million domains** per day:
  - **gTLDs:**
    - .com, .net, .org, .info, .mobi, .aero, .asia, .name, .biz, .gov
    - + almost 1200 "new" gTLDs (.xxx, .xyz, .amsterdam, .berlin, ...)
  - **ccTLDs:**
    - .nl, .se, .nu, .ca, .fi, .at, .dk, .ru, .рф, .us, .na, .gt,
    - <your ccTLD here?>
Grab your bingo cards folks!

- On the next slide, I am going to call this:

  (a) A blockchain
  (b) "Agile" and "lean"
  (c) Big data
  (d) Cyber!!!
Big data? Big data!

• Calling your research big data is all the rage -- **research funders love it**!

• So would our work qualify as big data?

• One **human genome** is about $3 \cdot 10^9$ DNA base pairs

• We collect **over $2.3 \cdot 10^9$ DNS records each day** (about $3/4$ of a human)

• **Since February 2015** we collected **over $3.4 \cdot 10^{12}$ results** (3.4 trillion) or: **over 1129 human genomes** (I bet there's fewer people in this room)
We think we measure responsibly

- We have **clearly marked** the **address space** from which we measure (including **reverse DNS**)

- We have **reached out to large operators** in our datasets

- Very **few complaints** received (less than 5 since February 2015)
What can we do with all this data?

- We will illustrate the use of OpenINTEL with **three examples**:
  - Example 1: DNSSEC operational practices
  - Example 2: Improving DNS resilience
  - Example 3: The stupidest thing you can put in a TXT record
Example 1: DNSSEC

- (Hopefully) it is well known that .nl and .se have a high level of DNSSEC deployment, due to financial incentives.

- (Small) financial incentives economically only benefit large DNS operators.

- We hypothesised that the incentives would encourage deployment en masse but that deployments would not necessarily follow security best practices.
**Large operators**

<table>
<thead>
<tr>
<th>TLD</th>
<th>#Domains</th>
<th>#Signed</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>93,464,626</td>
<td>712,162</td>
<td>0.76%</td>
</tr>
<tr>
<td>.net</td>
<td>10,412,605</td>
<td>114,687</td>
<td>1.10%</td>
</tr>
<tr>
<td>.org</td>
<td>7,501,310</td>
<td>85,166</td>
<td>1.14%</td>
</tr>
<tr>
<td>.nl</td>
<td>4,353,518</td>
<td>2,736,393</td>
<td>62.85%</td>
</tr>
<tr>
<td>.se</td>
<td>1,153,129</td>
<td>723,532</td>
<td>62.75%</td>
</tr>
</tbody>
</table>

**Small operators**

<table>
<thead>
<tr>
<th>TLD</th>
<th>#Domains</th>
<th>#Signed</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>23,349,922</td>
<td>224,251</td>
<td>0.96%</td>
</tr>
<tr>
<td>.net</td>
<td>2,598,823</td>
<td>26,400</td>
<td>1.02%</td>
</tr>
<tr>
<td>.org</td>
<td>1,871,904</td>
<td>20,342</td>
<td>1.09%</td>
</tr>
<tr>
<td>.nl</td>
<td>1,087,457</td>
<td>92,791</td>
<td>8.53%</td>
</tr>
<tr>
<td>.se</td>
<td>287,115</td>
<td>13,794</td>
<td>4.80%</td>
</tr>
</tbody>
</table>

*Just 14 operators responsible for over 80% of signed domains*

*Just 3 operators responsible for over 80% of signed domains*
Example 1: DNSSEC

- We **checked** DNSSEC practices against guidelines from NIST
- **Result:** operators use (too) small ZSKs (1024-bit) they never roll
- **Similar results for all large operators** in .se and .nl

<table>
<thead>
<tr>
<th>DNS operator</th>
<th>Master NS†</th>
<th>#Signed</th>
<th>Algorithm</th>
<th>KSK size</th>
<th>ZSK size</th>
<th>ZSK Rollover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loopia AB</td>
<td>*.loopia.se.</td>
<td>282,604</td>
<td>✓</td>
<td>✓</td>
<td>△+</td>
<td>×</td>
</tr>
<tr>
<td>One.com</td>
<td>*.one.com.</td>
<td>221,372</td>
<td>✓</td>
<td>△</td>
<td>△+</td>
<td>×</td>
</tr>
<tr>
<td>Binero AB</td>
<td>*.binero.se.</td>
<td>123,131</td>
<td>✓</td>
<td>✓</td>
<td>△+</td>
<td>×</td>
</tr>
</tbody>
</table>
Example 1: DNSSEC

- **Impact:** IIS (.se operator) decided to change their incentive policy and set **explicit security requirements.** This is already having an effect!

![Graph showing DNSSEC usage over time](image-url)
Example 2: DNS resilience

- The **attack on Dyn in 2016** shows the risk of sharing DNS infrastructure
- **Data** from OpenINTEL **shows** that many **key customers switched to using two DNS providers**

![Graph showing percentage of traffic](chart.png)
Example 2: DNS resilience

- Recently started a collaborative project on DNS resilience against DDoS attacks called "MADVIPR"
- Collaboration between UTwente (NL) and CAIDA/UCSD (US)
- Makes extensive use of OpenINTEL to map points of failure, e.g.:
  - Parent/child delegation mismatches
  - Parent/child delegation TTL mismatches
  - Shared infrastructure
  - Topological bottlenecks
Example 2: DNS resilience

- We are currently studying parent/child delegation TTL mismatches.
- These impact resilience under DDoS (time to change) and how long a DNS hijack lingers.
Example 2: DNS resilience

- **Topological diversity** is important to **protect against denial-of-service**
- Vast **majority of .com domains** has **name servers** located in a single AS
- For **.nl almost half of domains** have **name servers in at least two AS-es**

![Pie chart showing distribution of ASNs for .com and .nl domains](chart.png)
Example 2: DNS resilience

- **Majority** of `.com` and `.nl` have name servers in multiple prefixes, yet **15% only** have name servers in a single prefix (IPv4)

- **Student project**: use RIPE Atlas to check if name servers share a location (using speed-of-light triangulation)
Example 3: put it in a TXT record

- In TXT records we find:
  - HTML snippets
  - JavaScript
  - Windows Powershell code
  - Other scripting languages (bash, python, ...)
  - PEM-encoded X.509 certificates
  - Snippets of DNS zone files
  - … (you literally can’t make this stuff up)

→ Studying these closely, as they appear (partly) malicious
Hanlon's maxim

“Never attribute to malice, that which can adequately be explained by stupidity”
in numbers:

- 216 MILLION domains measured on a daily basis
- 2.3 BILLION data points collected daily
- 2.9 TRILLION data points collected since the start in 2015

Contact in case of problems

In case you are a DNS operator and you think our measurement is impacting your infrastructure, please go to our Problems page, where you can read more about what traffic you can expect to see from us, how you can contact us and if necessary, how you can block our traffic.

News overview

Blog: a Brief History of OpenINTEL

In honour of OpenINTEL winning the Research Data NL Prize, we have published a short blog that tells the story of OpenINTEL so far. Read the blog here.

© Copyright 2015-2019 by Drum roll...
And the winner is...

-----BEGIN RSA PRIVATE KEY-----
MIICXwIBAAKBgQC36kRNc50wG3uDLry90OxU+9X5LYlhdj0D+ax6BiC27W7iweVwfwupxsMvLBhhgtegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoRthy07SSLsFACkoXP++JxZ7bIakqdj5wAyIJ53zSJU7wKlmH1Eha7+Myip9LG8HPfsZtY3wIDAQAB...

... <- I left this part out...
-----END RSA PRIVATE KEY-----

• Why, oh why, oh why...

• And this is just one example, we’ve seen quite a few of these.

• What on Earth are these people doing?!
And the winner is...

-----BEGIN RSA PRIVATE KEY-----
MIICXwIBAAKBgQC36kRNc50wG3uDLry00xU+9X5LYlhdj0D+ax6BiC7W7iweVwf
wupxsMvLBhhgegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoRthy07SSLsFAC
koXP++JxZ7bIakqdj5wAyIJ53zSJu7wK1mH1Eha7+Myip9LG8HPfsZtY3wIDAQAB
...

-----END RSA PRIVATE KEY-----

- Why, oh why, oh why... oh wait, someone's trying to configure DKIM --- D'oh!

<redacteddomain.tld> IN TXT "v=DKIM1; k=rsa;
p=MIGfMA0GCSqGSIb3DQEBAQUA4GNADCBiQKBgQC36kRNc50wG3uDLry00xU+9X5LYlhdj
0D+ax6BiC7W7iweVwfwupxsMvLBhhgegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoR
thy07SSLsFACkoXP+JxZ7bIakqdj5wAyIJ53zSJu7wK1mH1Eha7+Myip9LG8HPfsZtY3wIDAQAB"
Future of the project

- **Short term** challenges:
  - Ensure **robust data archival**
  - **Expand** the number of ccTLDs we **cover** ← can you help us?

- **Long term** goals:
  - Be the "**long-term memory" of the DNS" -- if someone in 2025 wants to know what DNS looked like in 2015, we have the answer
  - **Have real-world impact**, by improving the performance, resilience and security of the DNS
Questions? Talk to the team

Anna Sperotto
Anna procures research funding for PhD and postdoctoral research projects that use OpenINTEL data. She supervises PhD students that work with OpenINTEL data for their research.

Mattijs Jonker
Mattijs manages the development on the Big Data side of OpenINTEL, which ranges from having designed the data schema, to building data pipelines to collaborating institutions. He also administers the OpenINTEL Hadoop cluster, oversees day-to-day operation with the rest of the team, and tutors colleagues and collaborators in data use and analyses.

Olivier van der Toorn
Olivier takes care of the monitoring of the OpenINTEL measurement infrastructure, when a measurement stalls he is the first to know. Additionally, Olivier is closely involved in maintaining this infrastructure.

Roland van Rijswijk-Deij
Roland designed the architecture of OpenINTEL, writes most of the core measurement code of OpenINTEL and takes care of continuously expanding the measurement with new TLD data sources. Next to this, he manages the funding of the OpenINTEL measurement cluster and Hadoop cluster.

© Copyright 2015-2019 by LACTLD
Questions?

Thank you for your attention!

Visit our webpage for more information: https://openintel.nl/