Open INTEL

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

UNIVERSITY OF TWENTE.

SURF NET

SIDN LABS

NLNETLABS

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

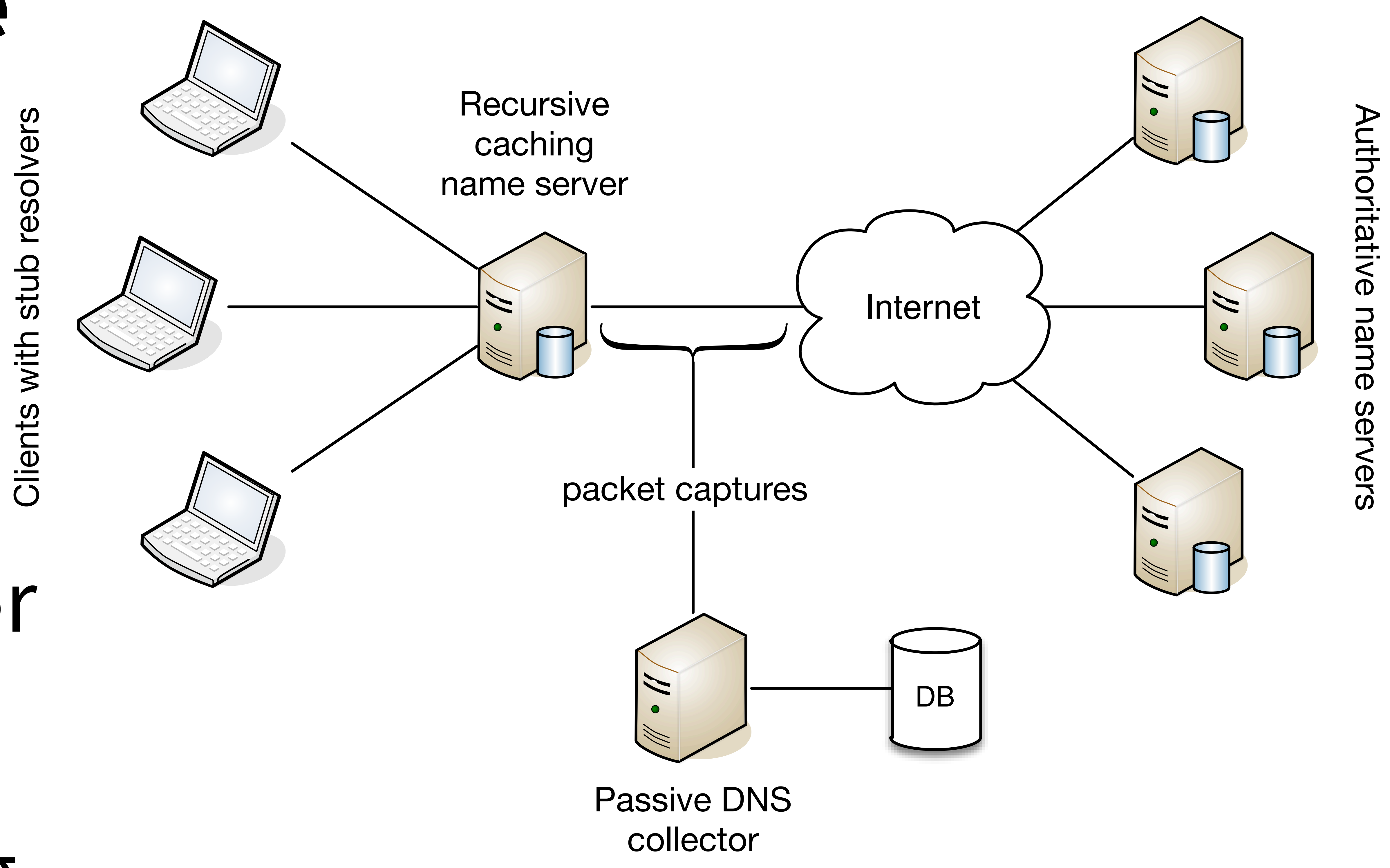
Why measure the DNS?

- **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**


Clients with stub resolvers
Recursive caching name server
Internet
Authoritative name servers
packet captures
DB
Passive DNS collector

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 216 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, **<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 $3.1 \cdot 10^{12}$ results** (3.1 **trillion**) or: **over 1047 human genomes** (I bet there's fewer people in this room)

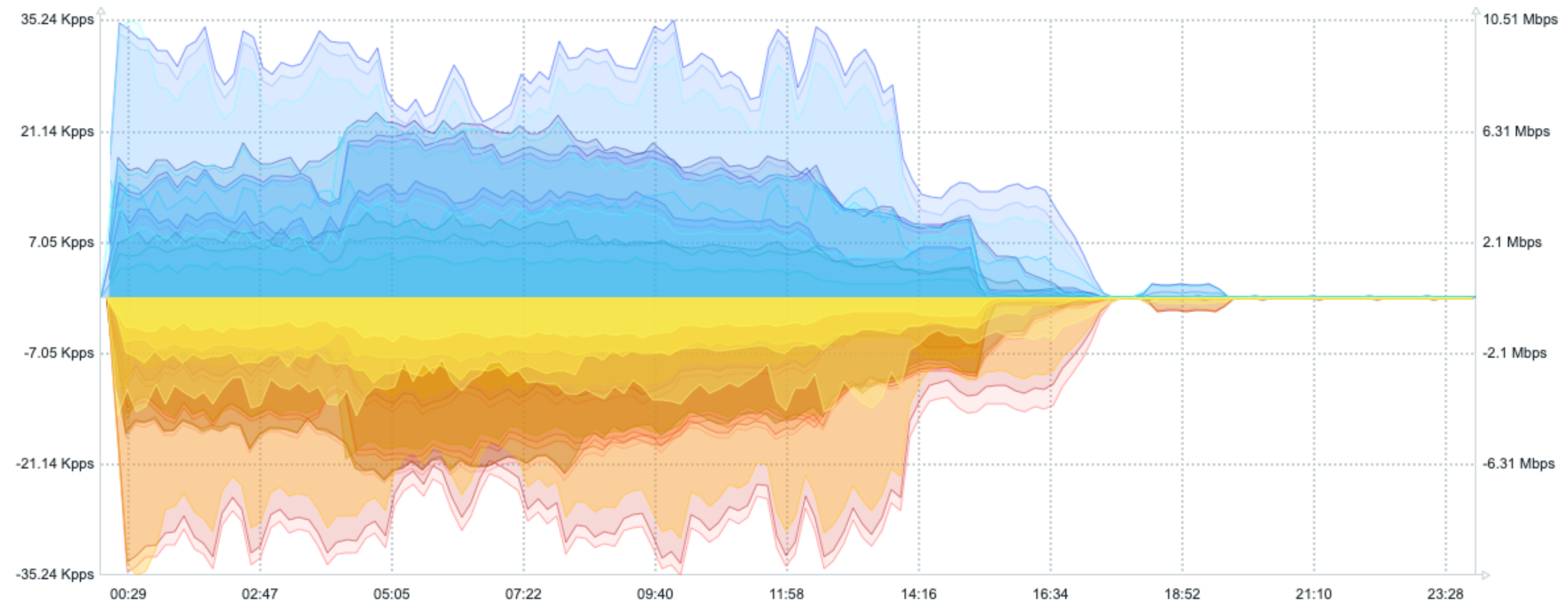
We think we measure responsibly

- 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)

```
inet6num:       xxxx:xxx:xxxx::/48
netname:        UTwente-OpenINTEL
descr:          University of Twente
descr:          Faculty EEMCS/DACS
descr:          OpenINTEL Active DNS Measurements
descr:          See http://www.openintel.nl/
                for more information
country:        NL
admin-c:        RVR180-RIPE
tech-c:         RVR180-RIPE
status:         ALLOCATED-BY-LIR
mnt-by:         SN-LIR-MNT
mnt-irt:        irt-SURFcert
created:        2018-06-26T08:53:10Z
last-modified:  2018-06-26T08:53:10Z
source:         RIPE
```



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

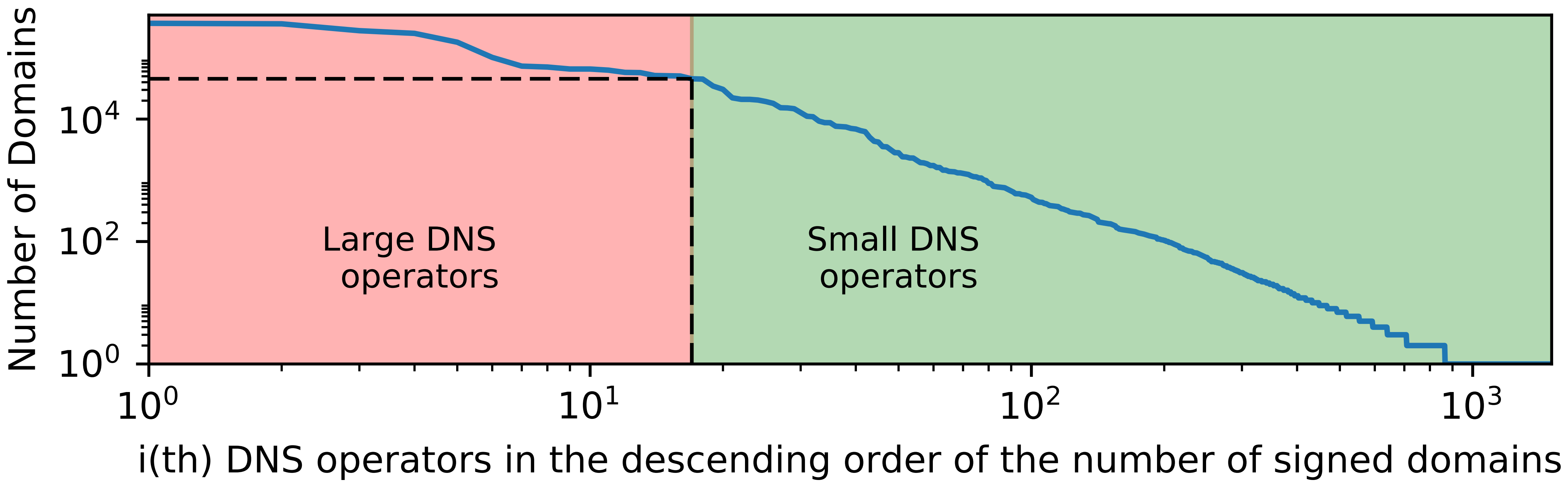
OpenINTEL

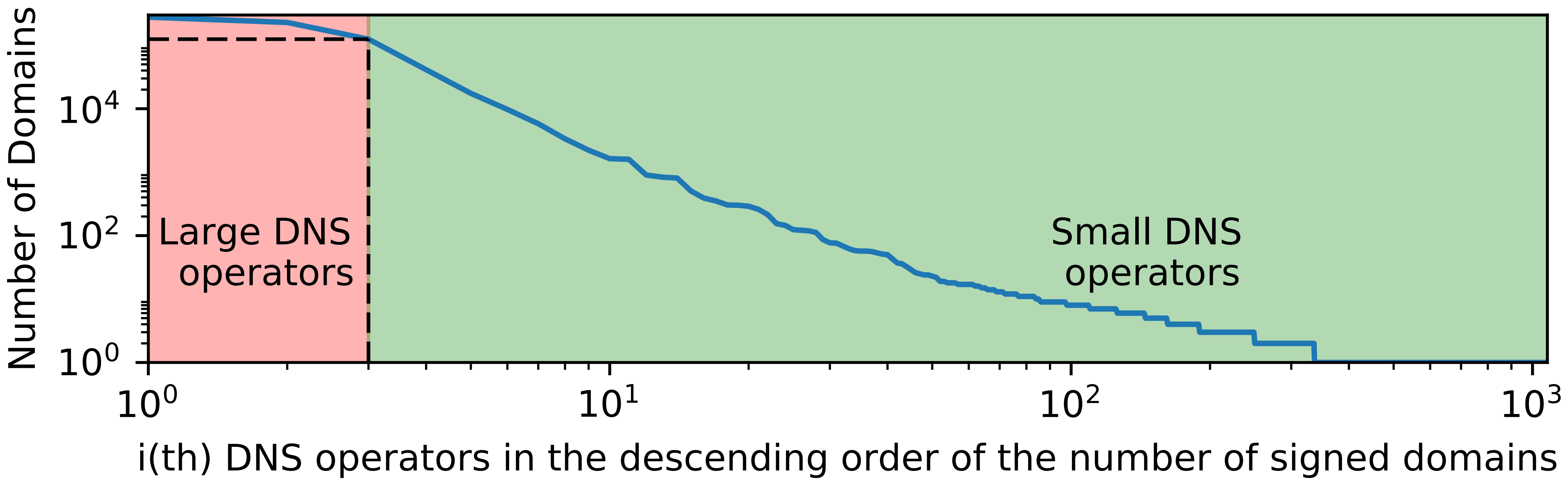
Example 1: DNSSEC

- (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**

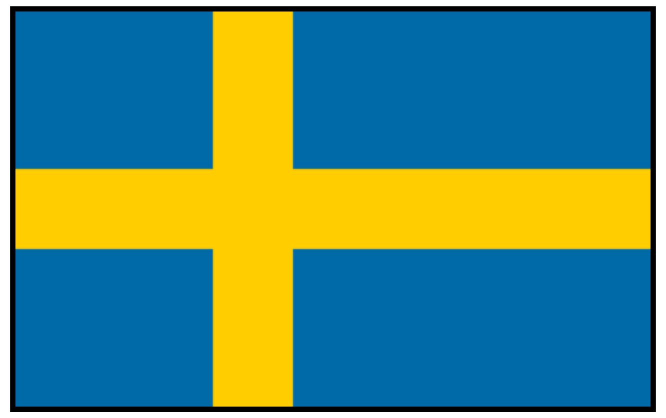
- We hypothesised that the **incentives** would **encourage deployment *en masse* but** that deployments would **not** necessarily **follow security best practices**





.nl

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



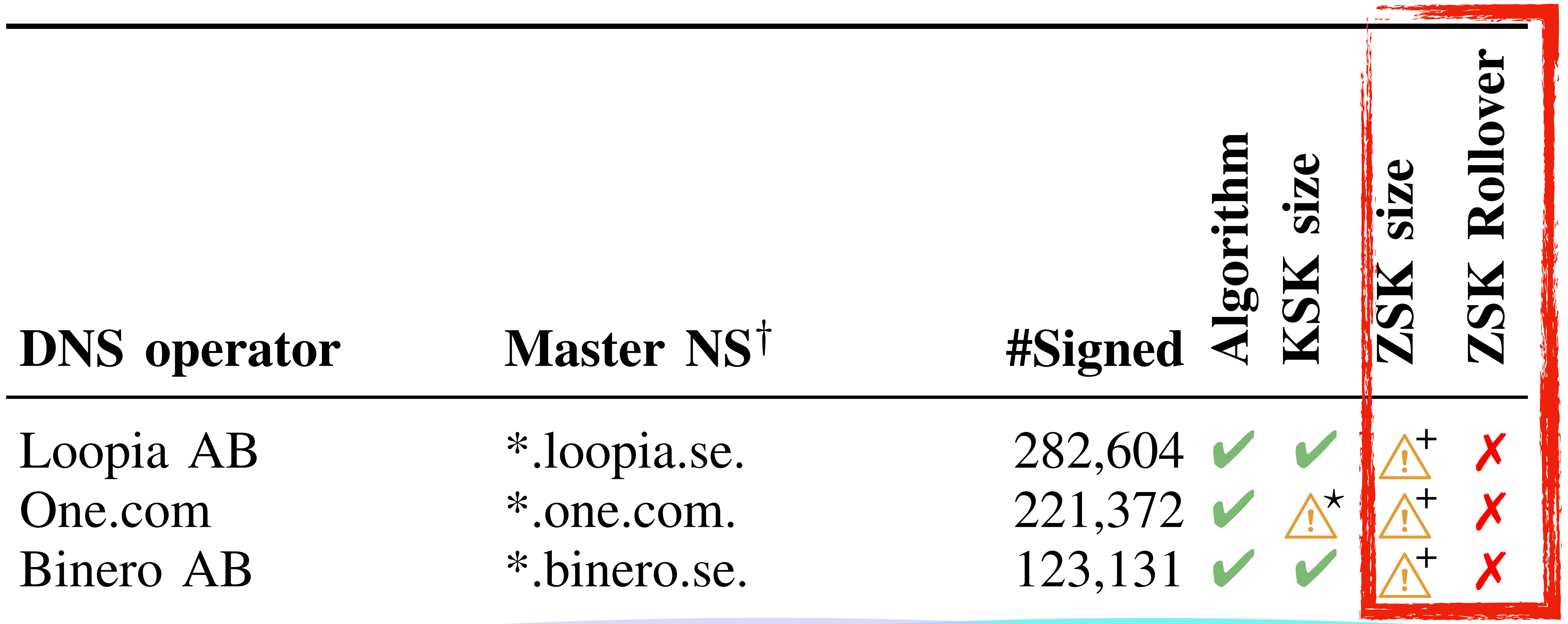
.se

Just **3 operators** responsible for **over 80% of signed domains**

	Large operators			**Small operators**		
TLD	**#Domains**	**#Signed**	**%**	**#Domains**	**#Signed**	**%**
.com	93,464,626	712,162	0.76%	23,349,922	224,251	0.96%
.net	10,412,605	114,687	1.10%	2,598,823	26,400	1.02%
.org	7,501,310	85,166	1.14%	1,871,904	20,342	1.09%
.nl	4,353,518	2,736,393	62.85%	1,087,457	92,791	8.53%
.se	1,153,129	723,532	62.75%	287,115	13,794	4.80%

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

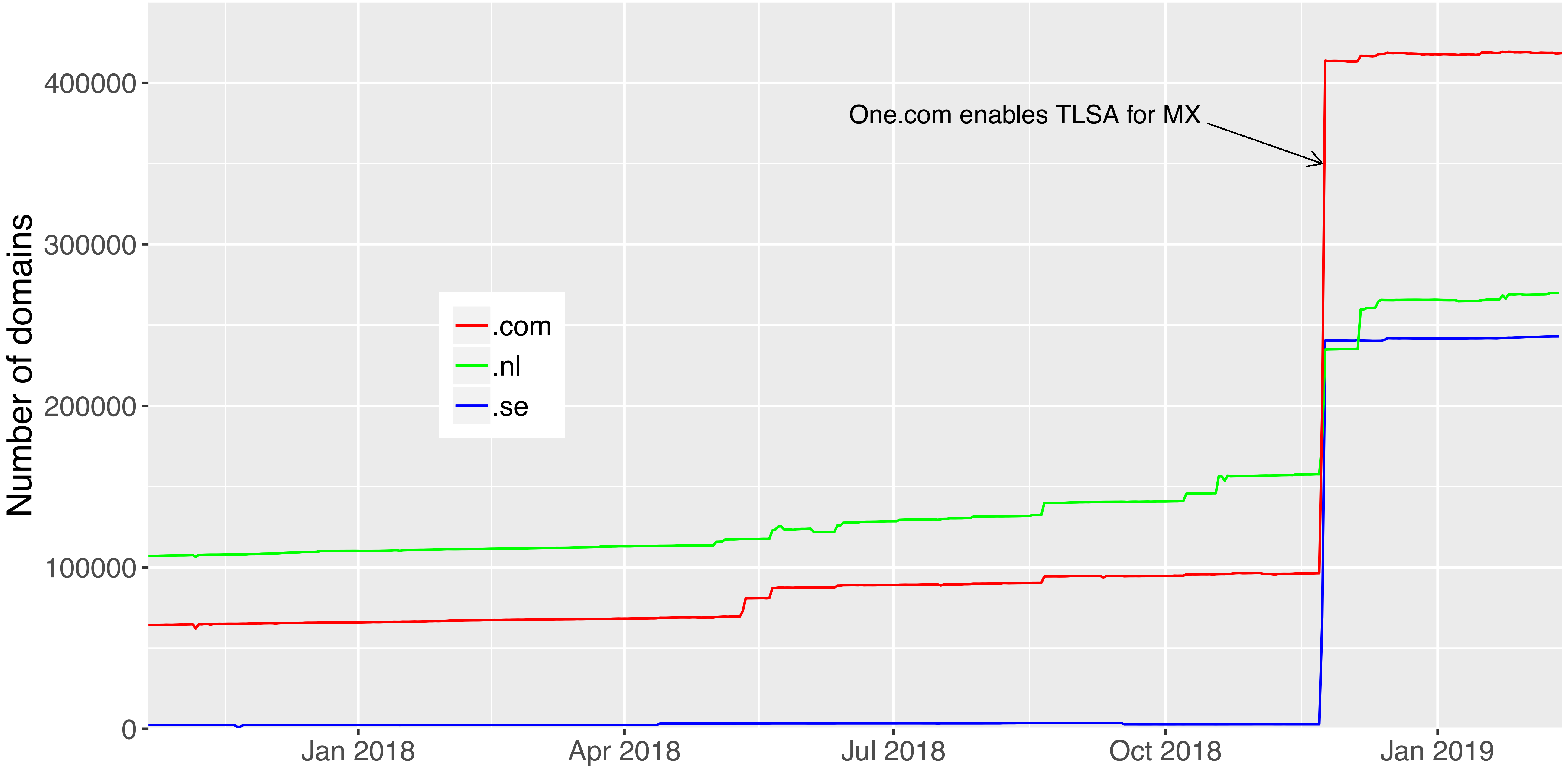


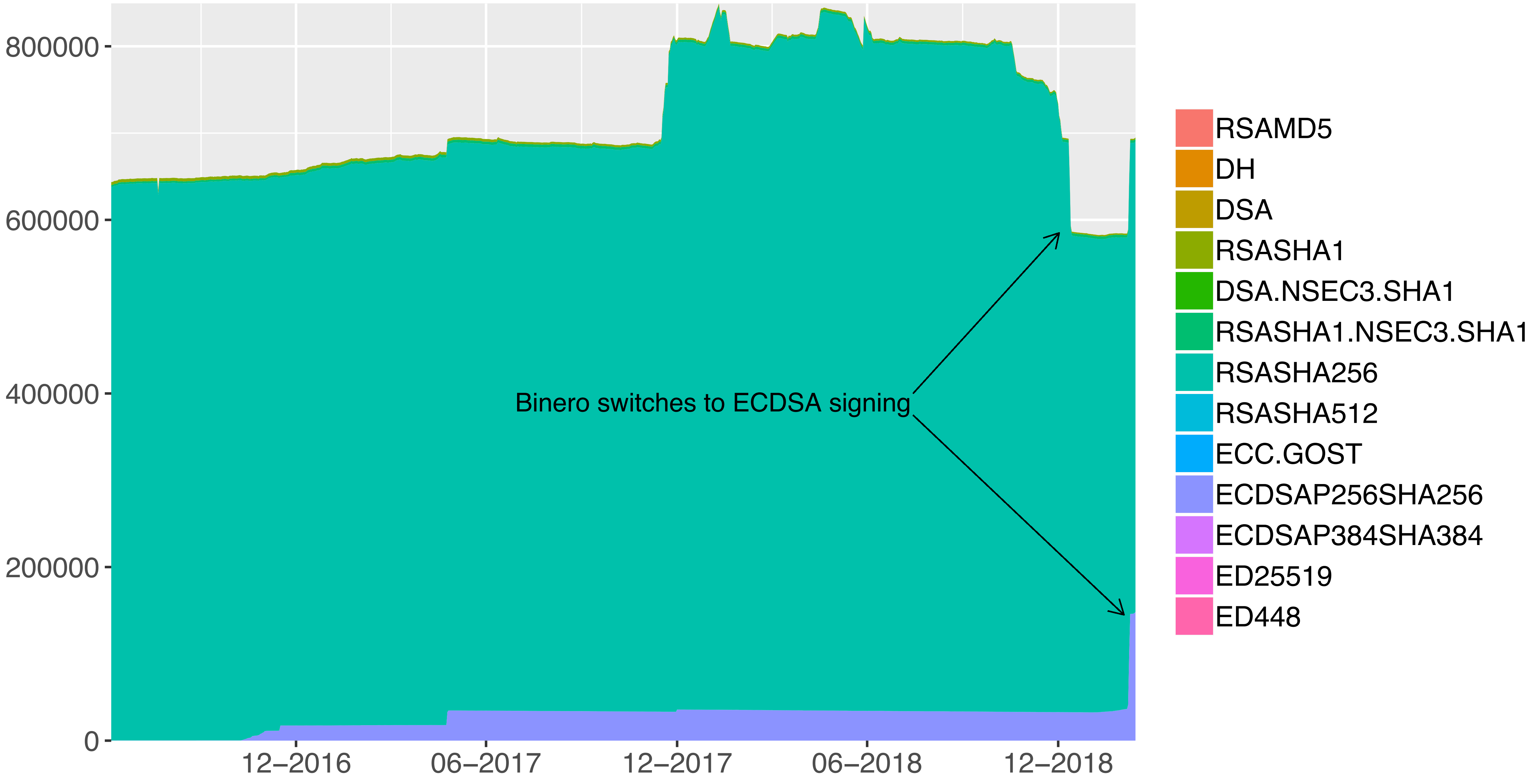
DNS operator	Master NS†	#Signed	Algorithm	KSK size	ZSK size	ZSK Rollover
Loopia AB	*.loopia.se.	282,604	✓	✓	⚠+	✗
One.com	*.one.com.	221,372	✓	⚠*	⚠+	✗
Binero AB	*.binero.se.	123,131	✓	✓	⚠+	✗

Open INTEL

Example 1: DNSSEC

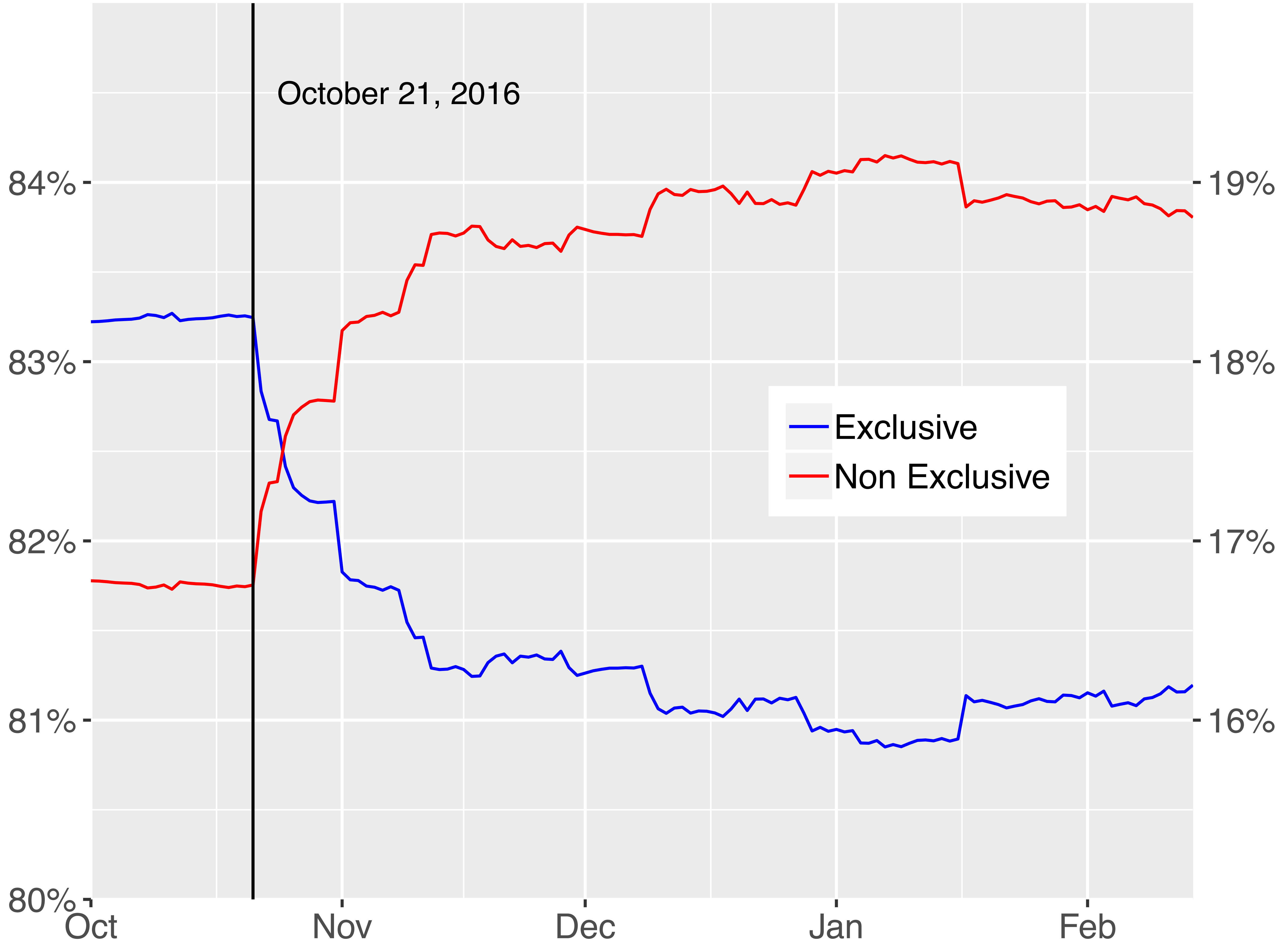
- **Impact: IIS** (.se operator) **decided to change** their **incentive** policy and set **explicit security requirements**. This is **already having an effect!**





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**



Example 2: DNS resilience

- Recently started a collaborative project on DNS resilience against DDoS attacks called "**MADDVIPR**"

- Collaboration between UTwente (NL) and CAIDA/UCSD (US)

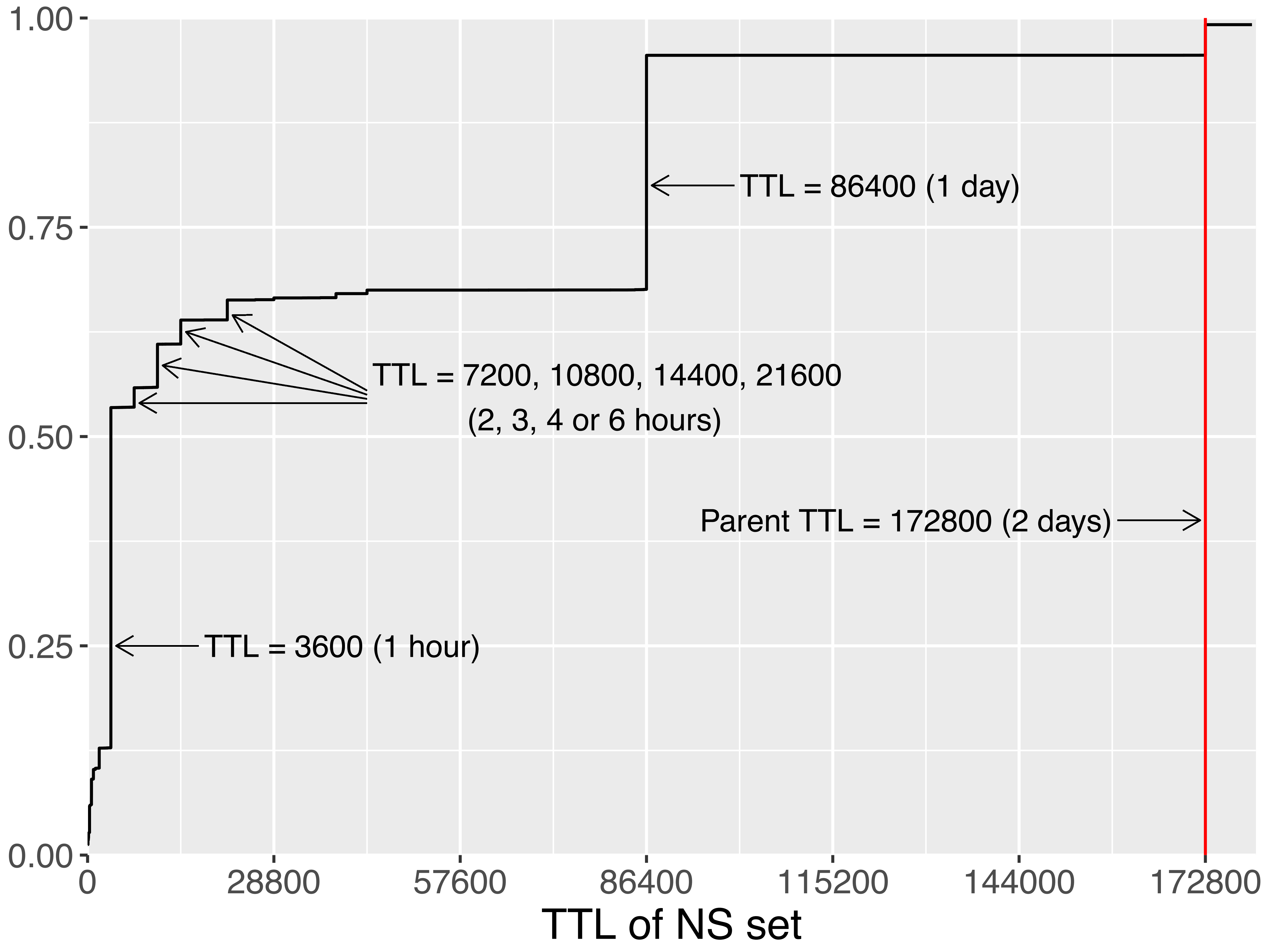
- Makes extensive use of OpenINTEL to map points of failure, e.g.:

- *Parent/child delegation mismatches*
- *Parent/child delegation TTL mismatches*
- *Shared infrastructure*
- *Topological bottlenecks*

Example 2: DNS resilience

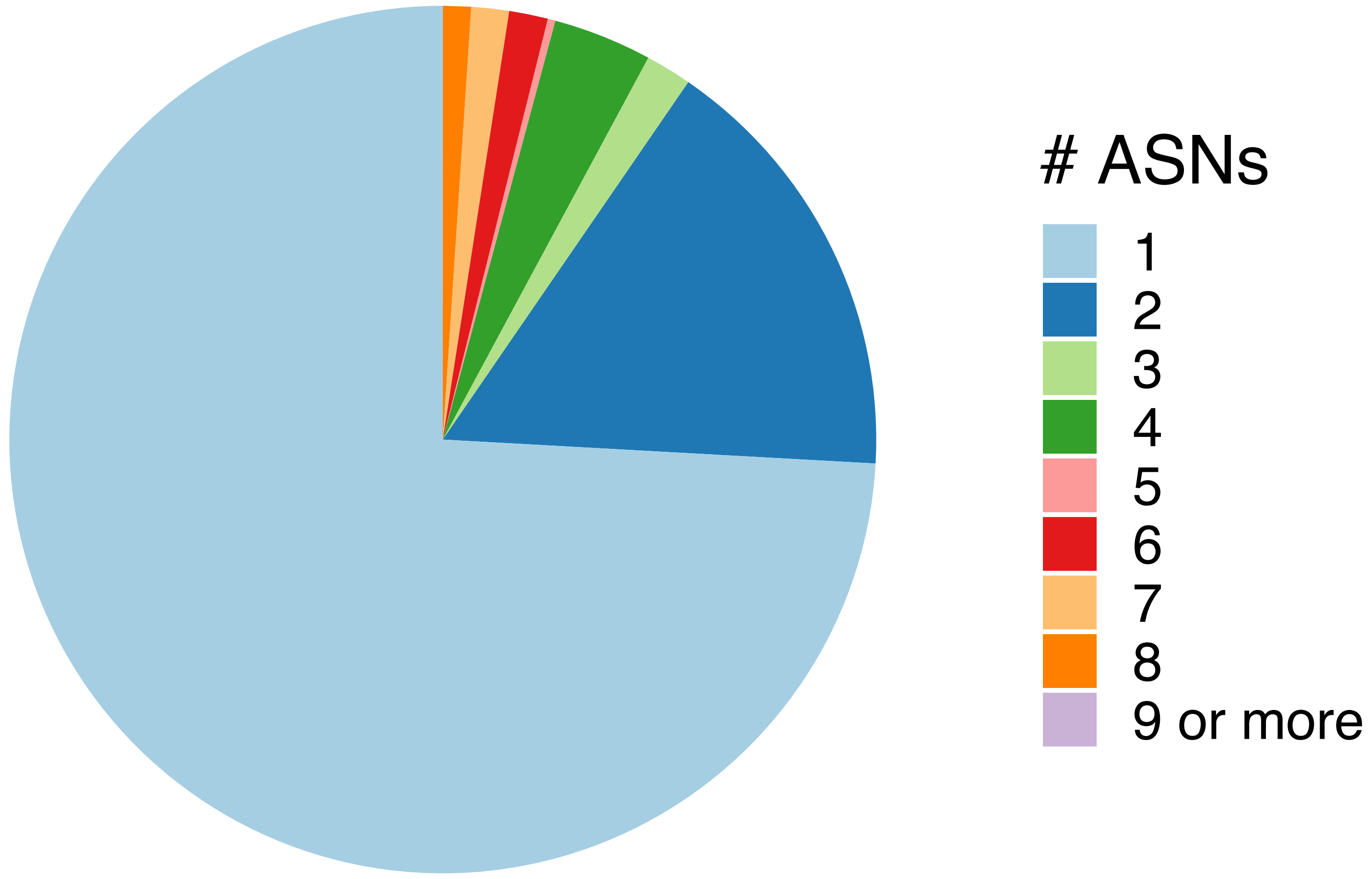
- We are currently **studying parent/child** delegation **TTL mismatches**

- These **impact resilience under DDoS** (time to change) and how long a **DNS hijack lingers**

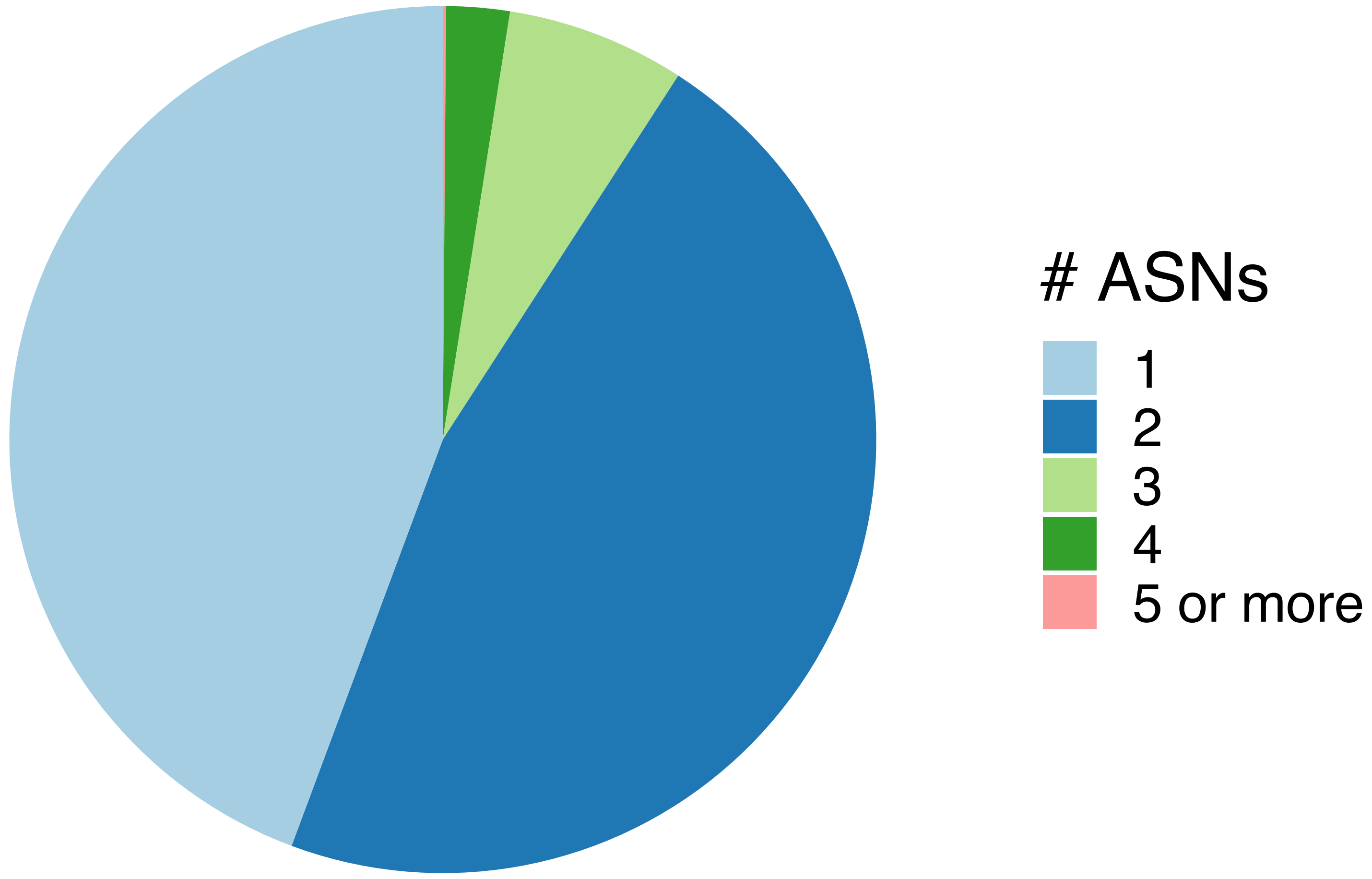


Example 2: DNS resilience

- **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**



.com

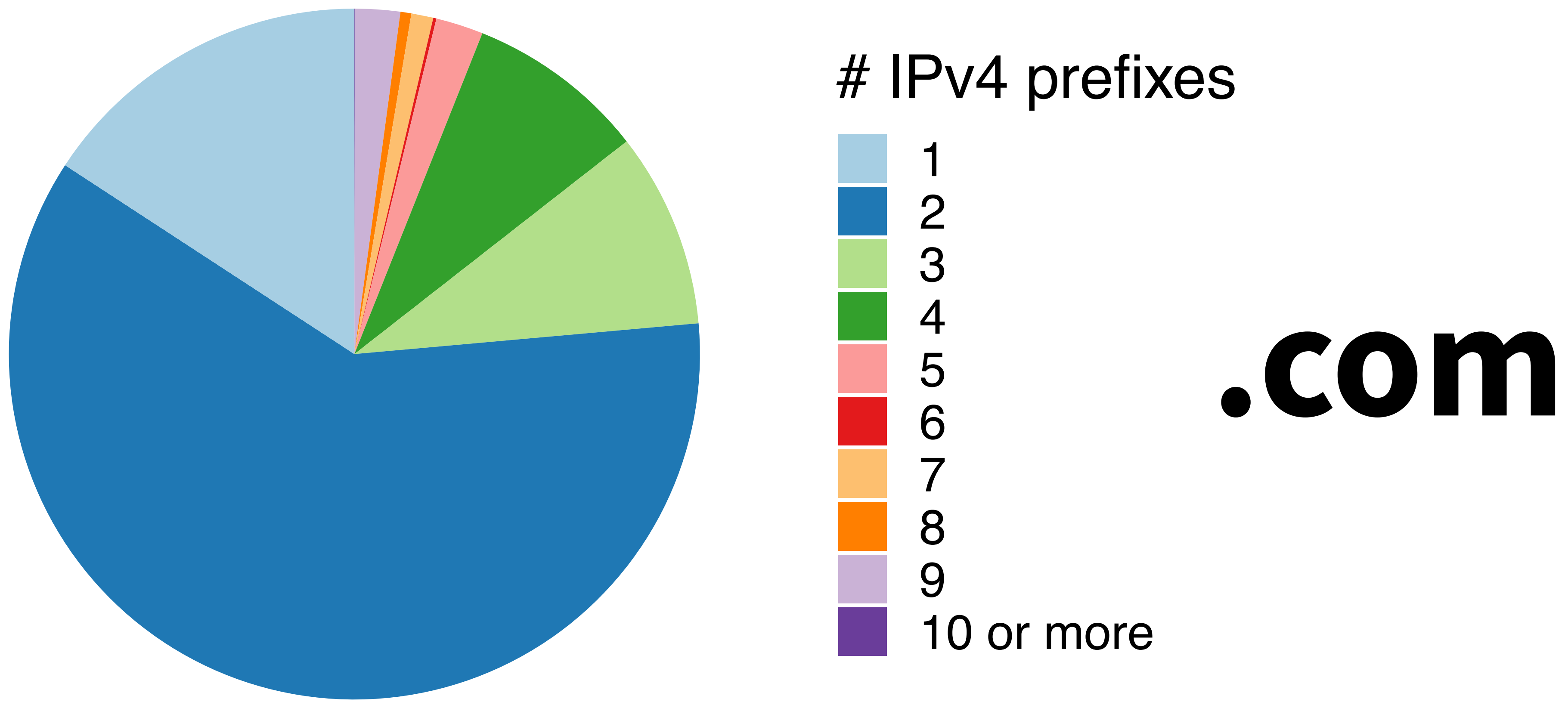


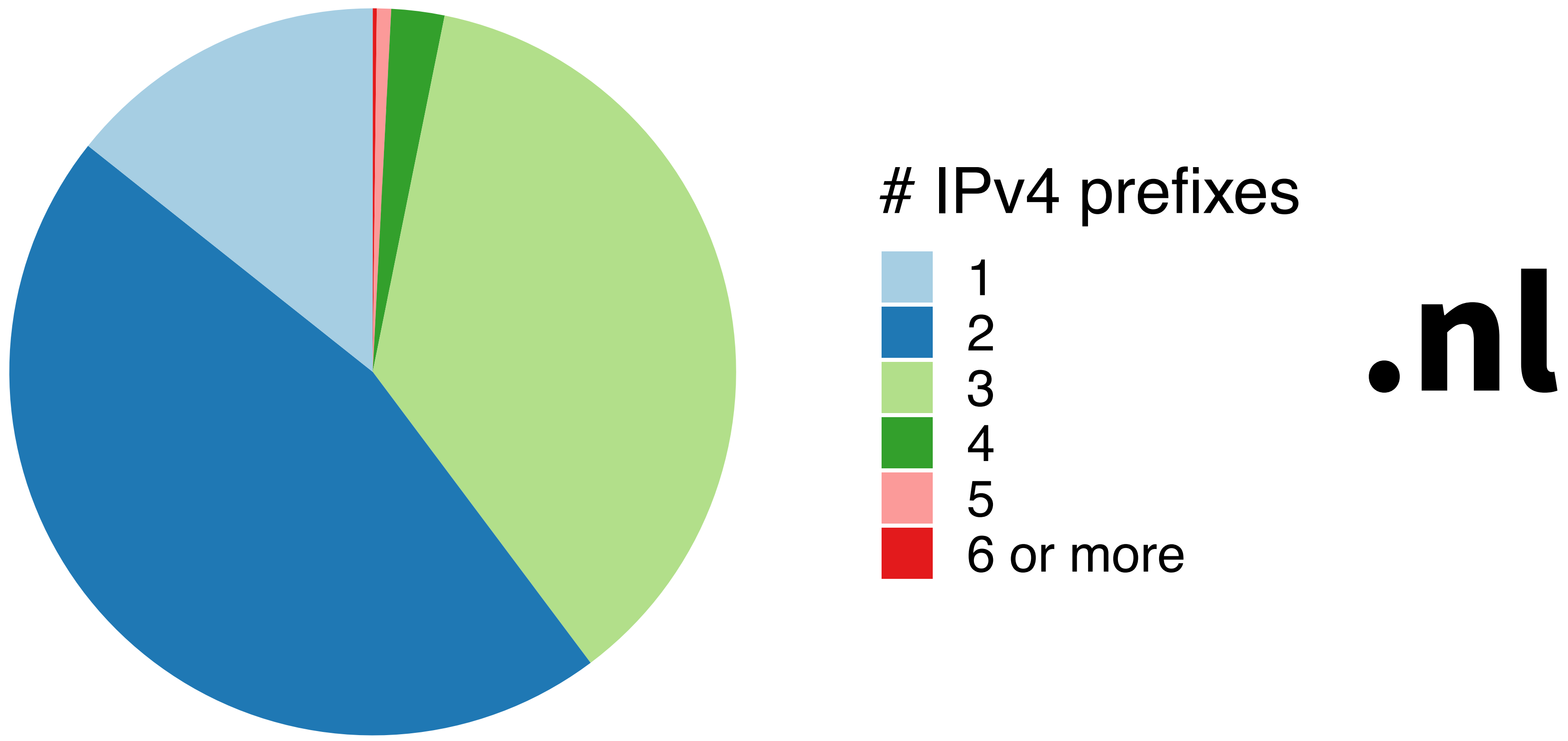
.nl

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”

Drum roll...

And the winner is…

```
-----BEGIN RSA PRIVATE KEY-----
MIICXwIBAAKBgQC36kRNc5OwG3uDlRy0OxU+9X5LYlhdj0D+ax6BiC27W7iweVwf
wupxsMvLBhhgegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoRthy07SSLsFAC
koXP++JxZ7bIakqdj5wAyIJ53zSJu7wKImH1Eha7+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-----
MIICXwIBAAKBgQC36kRNc5OwG3uDlRy0OxU+9X5LYlhdj0D+ax6BiC27W7iweVwf
wupxsMvLBhhgegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoRthy07SSLsFAC
koXP++JxZ7bIakqdj5wAyIJ53zSJu7wKImH1Eha7+Myip9LG8HPfsZtY3wIDAQAB
... <– I left this part out...
-----END RSA PRIVATE KEY-----
```

MATCH!!!

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

```
<redacteddomain.tld> IN TXT "v=DKIM1; k=rsa;
p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQC36kRNc5OwG3uDlRy0OxU+9X5LYlhdj
0D+ax6BiC27W7iweVwfwupxsMvLBhhgegptc5tqb1puXPkCxA6aHwhToFtKSEy4fIWTjWoR
thy07SSLsFACkoXP+JxZ7bIakqdj5wAyIJ53zSJu7wKImH1Eha7+Myip9LG8HPfsZtY3wID
AQAB"
```

Future of the project

- **Short term** challenges:
  - Ensure **robust data archival**
  - **Expand** the number of **ccTLD**s 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?

Thank you for your attention!

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