DNS Big Data Analytics

NCSC One

April 14th 2015

Maarten Wullink, SIDN

Analyzing malicious activity

with

DNS data analytics

SIDN labs Internet Research & Innovation SIDN

SIDN

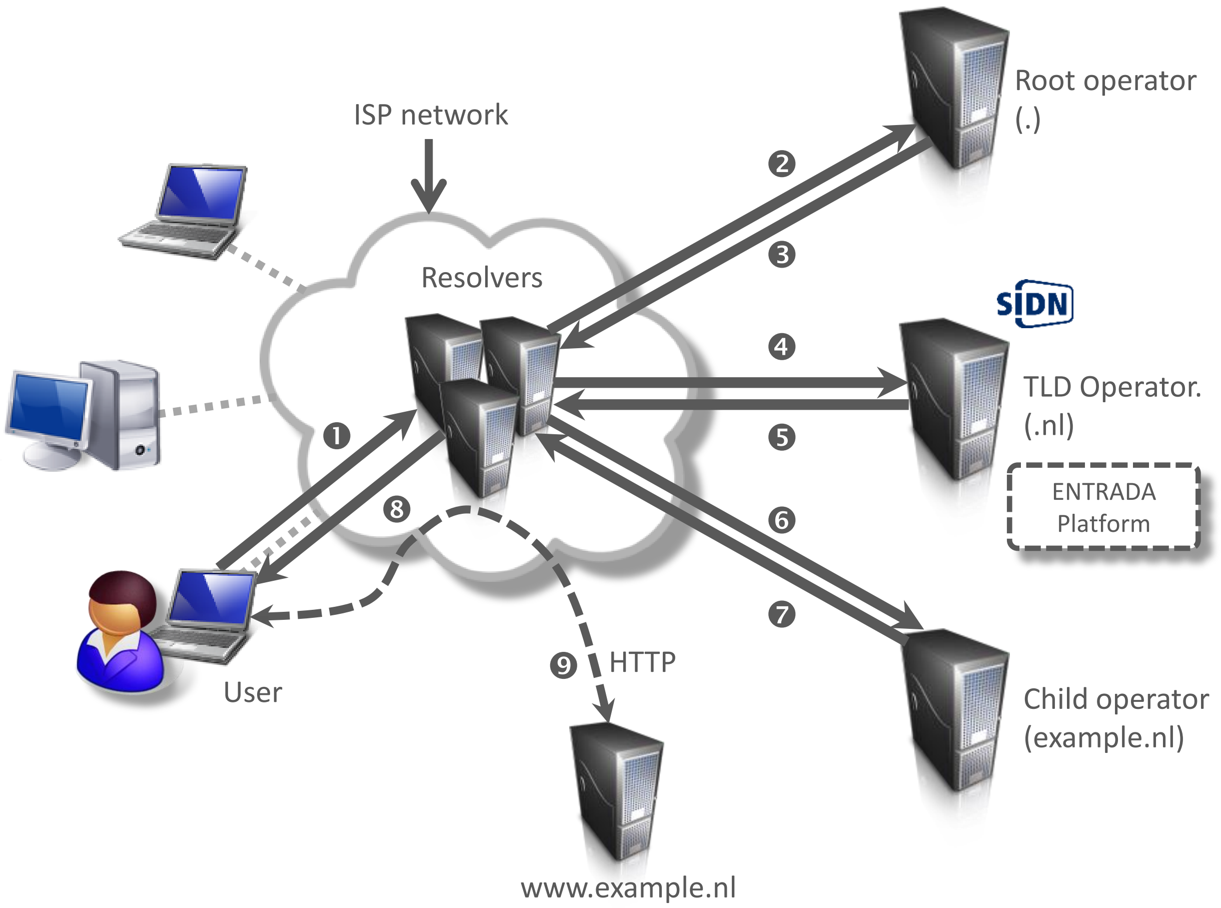
- Domain name registry for .nl ccTLD

- > 5,5 million domain names

- 2,4 million domain names secured with DNSSEC

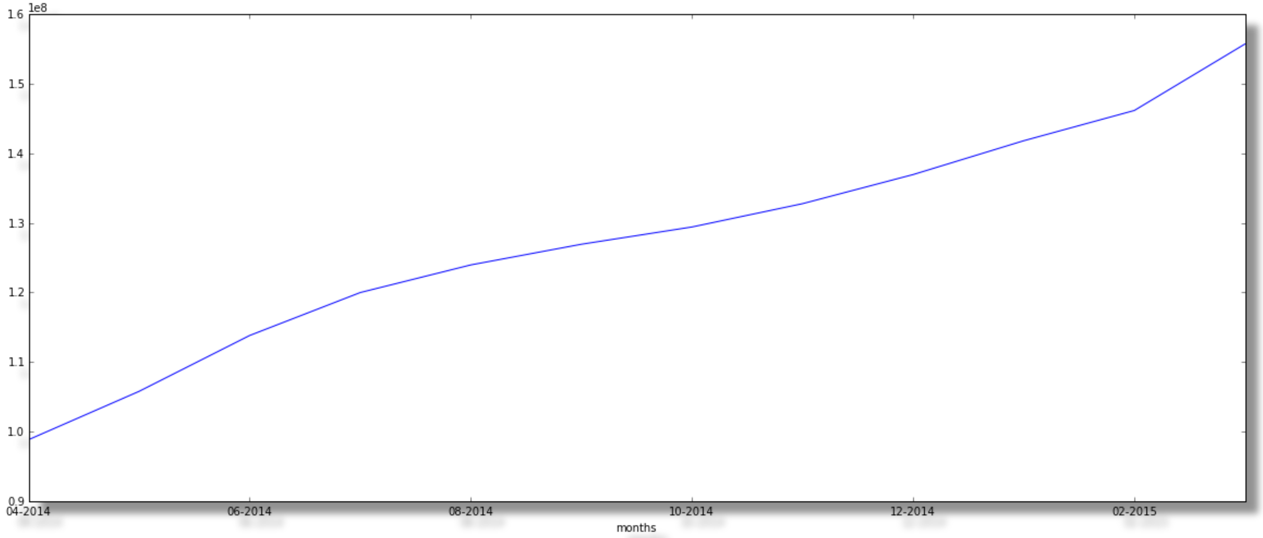
- SIDN Labs is the R&D team of SIDN

Domain Name System (mini)tutorial



DNS (big) data @SIDN

- > 3.100.000 distinct resolvers
- > 1.300.000.000 query's daily
- > 300 GB of raw data daily
- ~10% of this data is captured and stored



Traffic increase over the last 10 months

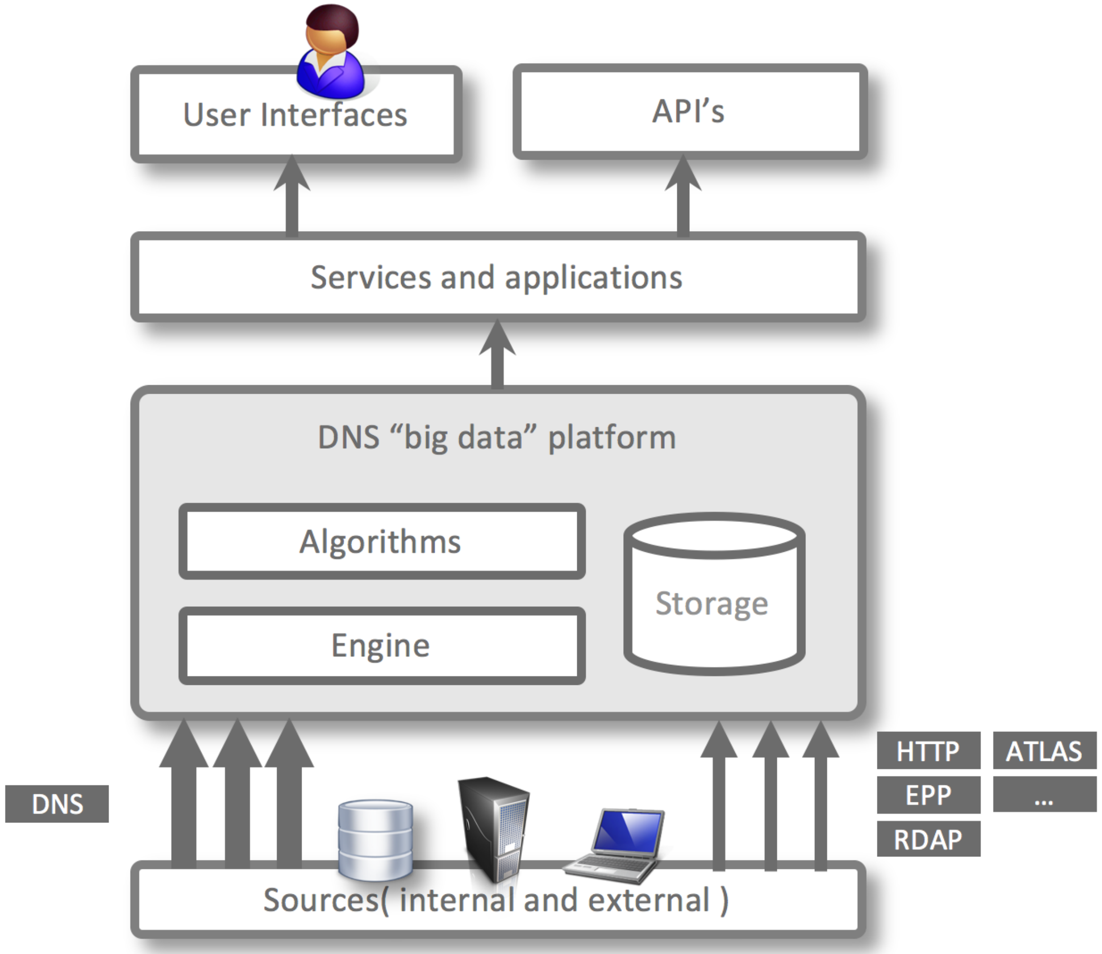
ENTRADA

- **E**Nhanced **T**op-level domain **R**esilience through **A**dvanced **D**ata **A**nalysis

- Big data platform used for research by SIDN Labs & partners

- Goal: Increase the security and stability of the .nl zone and the internet as a whole

ENTRADA highlevel architecture

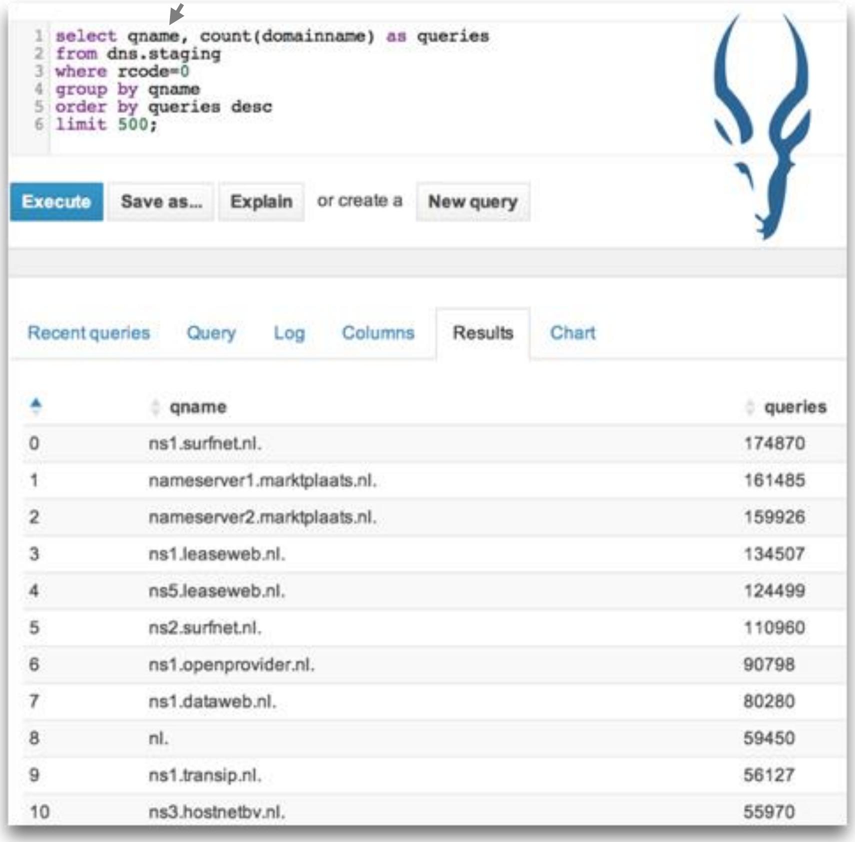


ENTRADA technology

- **Requirements**:
  - Performance
  - Scalability
  - Fault tolerance /redundancy
  - SQL compatibility

- **Choices:**
  - Apache Parquet as storage format
  - Apache Hadoop HDFS as storage
  - Cloudera Impala as Query engine

SQL

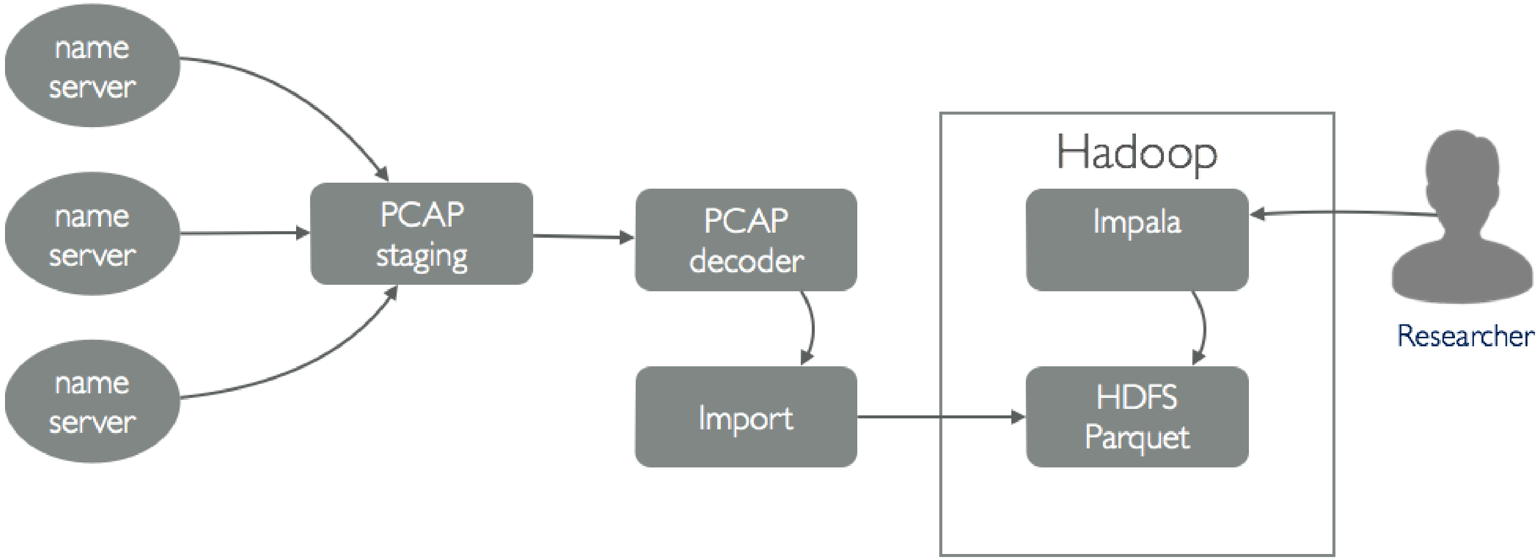


```
select qname, count(domainname) as queries
from dns.staging
where rcode=0
group by qname
order by queries desc
limit 500;
```

	qname	queries
0	ns1.surfnet.nl.	174870
1	nameserver1.marktplaats.nl.	161485
2	nameserver2.marktplaats.nl.	159926
3	ns1.leaseweb.nl.	134507
4	ns5.leaseweb.nl.	124499
5	ns2.surfnet.nl.	110960
6	ns1.openprovider.nl.	90798
7	ns1.dataweb.nl.	80280
8	nl.	59450
9	ns1.transip.nl.	56127
10	ns3.hostnetbv.nl.	55970

Interactive query interface

ENTRADA workflow



Query data available for analysis in under 15 minutes

ENTRADA status

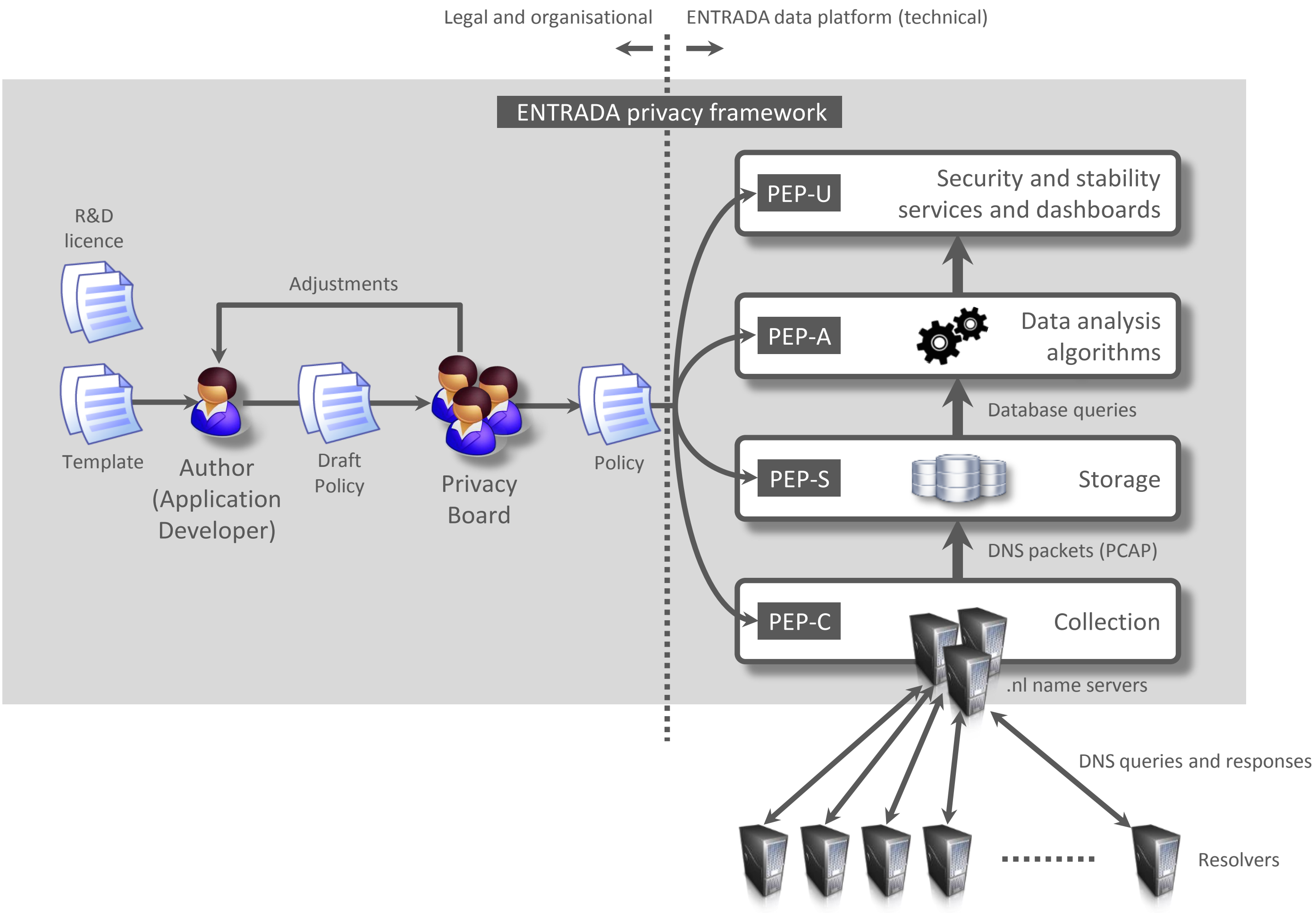
- Stored 1 year of data from a single .nl name server

- > 48.000.000.000 (48 B) rows

- > 7 TB data

- Cluster capacity ~150 B rows (easy to expand to trillions of rows)

Privacy framework



Download paper:
http://goo.gl/GvsfzQ

Policy elements:

- Purpose
- Data that is used
- Filters on the data
- Retention period
- Access to the data
- Type of application (Research vs. Production)

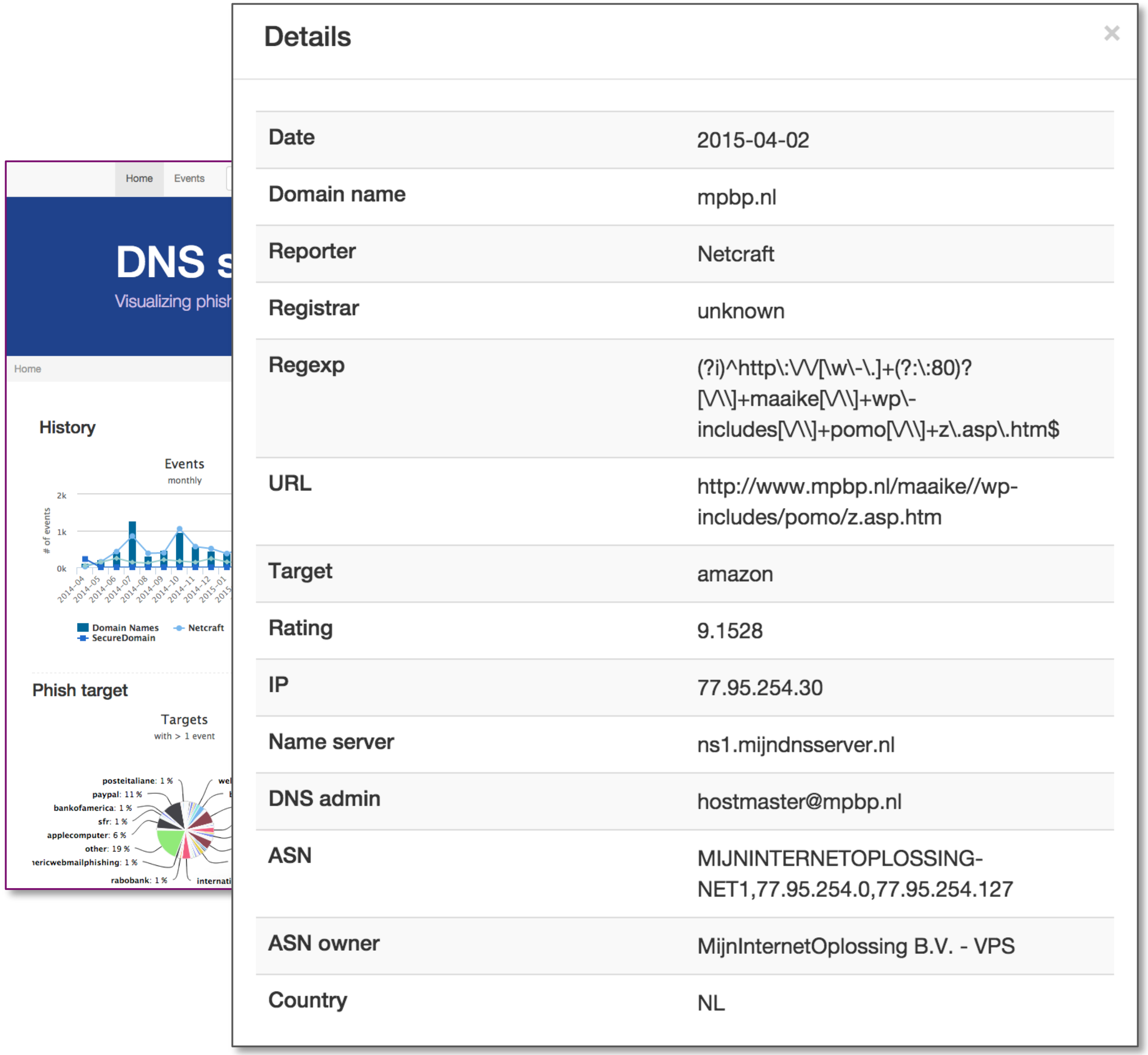
Example applications

- DNS security scoreboard
- Resolver reputation

DNS security scoreboard

Goal: Visualize DNS patterns for malicious activity

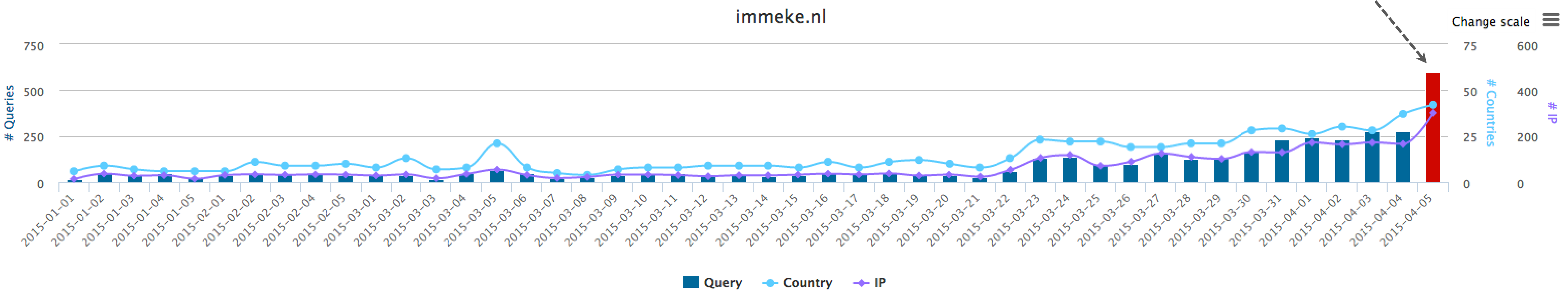
How: Combining external phising feeds with passive DNS data



Details

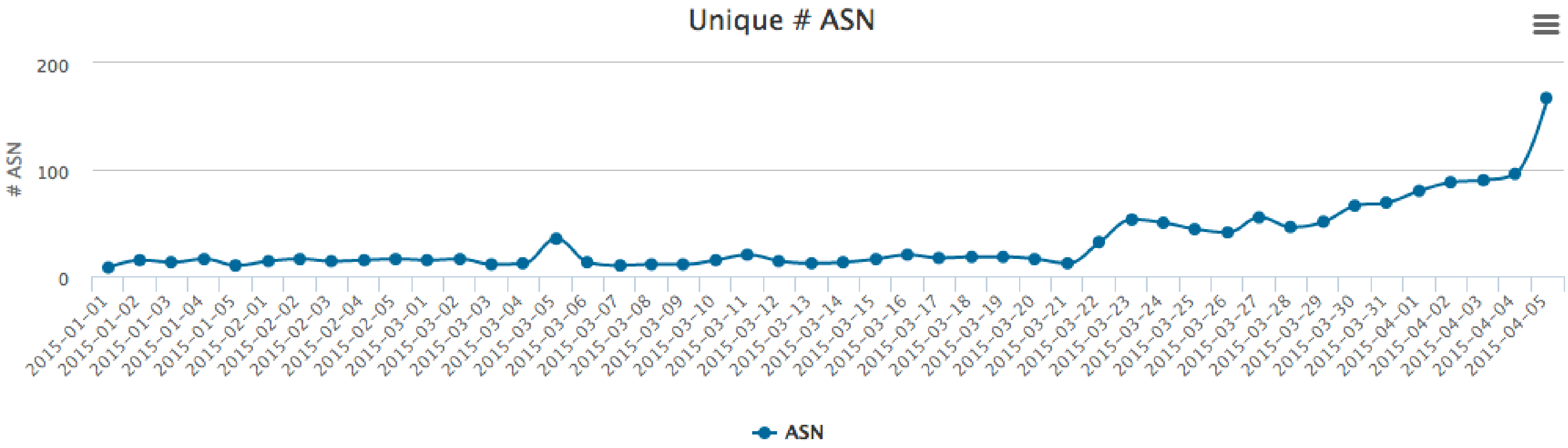
Date	2015-04-02
Domain name	mpbp.nl
Reporter	Netcraft
Registrar	unknown
Regexp	(?i)^http\:\/\/[\w\-\.]+(?:\:80)?[\/\\]+maaike[\/\\]+wp\-includes[\/\\]+pomo[\/\\]+z\.asp\.htm$
URL	http://www.mpbp.nl/maaike//wp-includes/pomo/z.asp.htm
Target	amazon
Rating	9.1528
IP	77.95.254.30
Name server	ns1.mijndnsserver.nl
DNS admin	hostmaster@mpbp.nl
ASN	MIJNINTERNETOPLOSSING-NET1,77.95.254.0,77.95.254.127
ASN owner	MijnInternetOplossing B.V. - VPS
Country	NL

Volumetric pattern

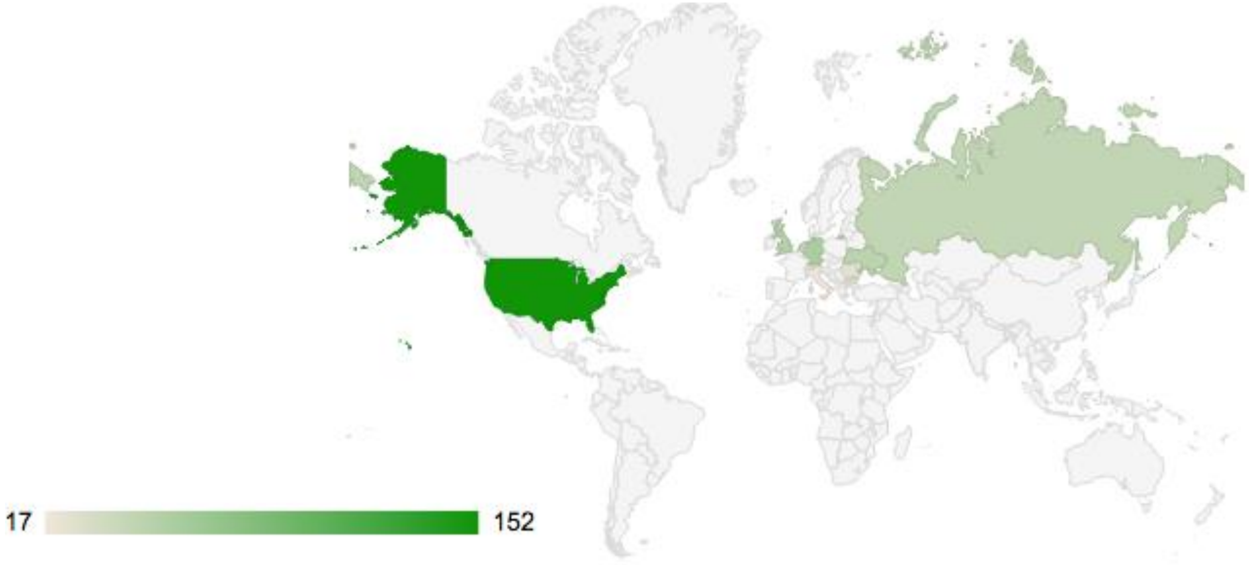


Phish feed alert

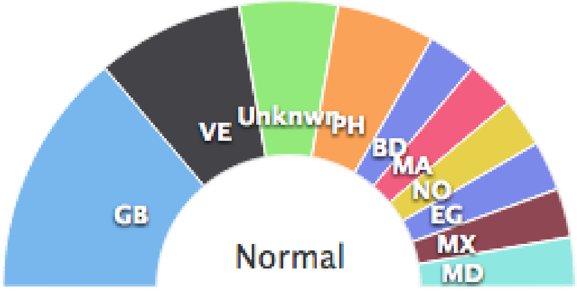
Network pattern



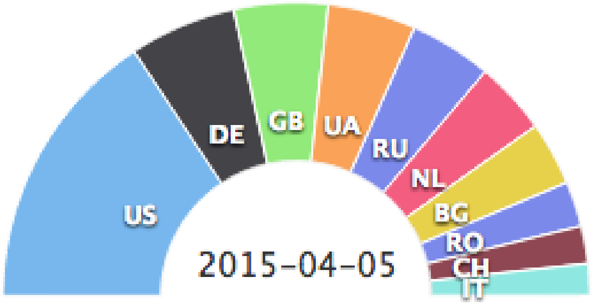
GEO pattern



17
152



GB
VE
Unknwn
PH
BD
MA
NO
EG
MX
MD
Normal



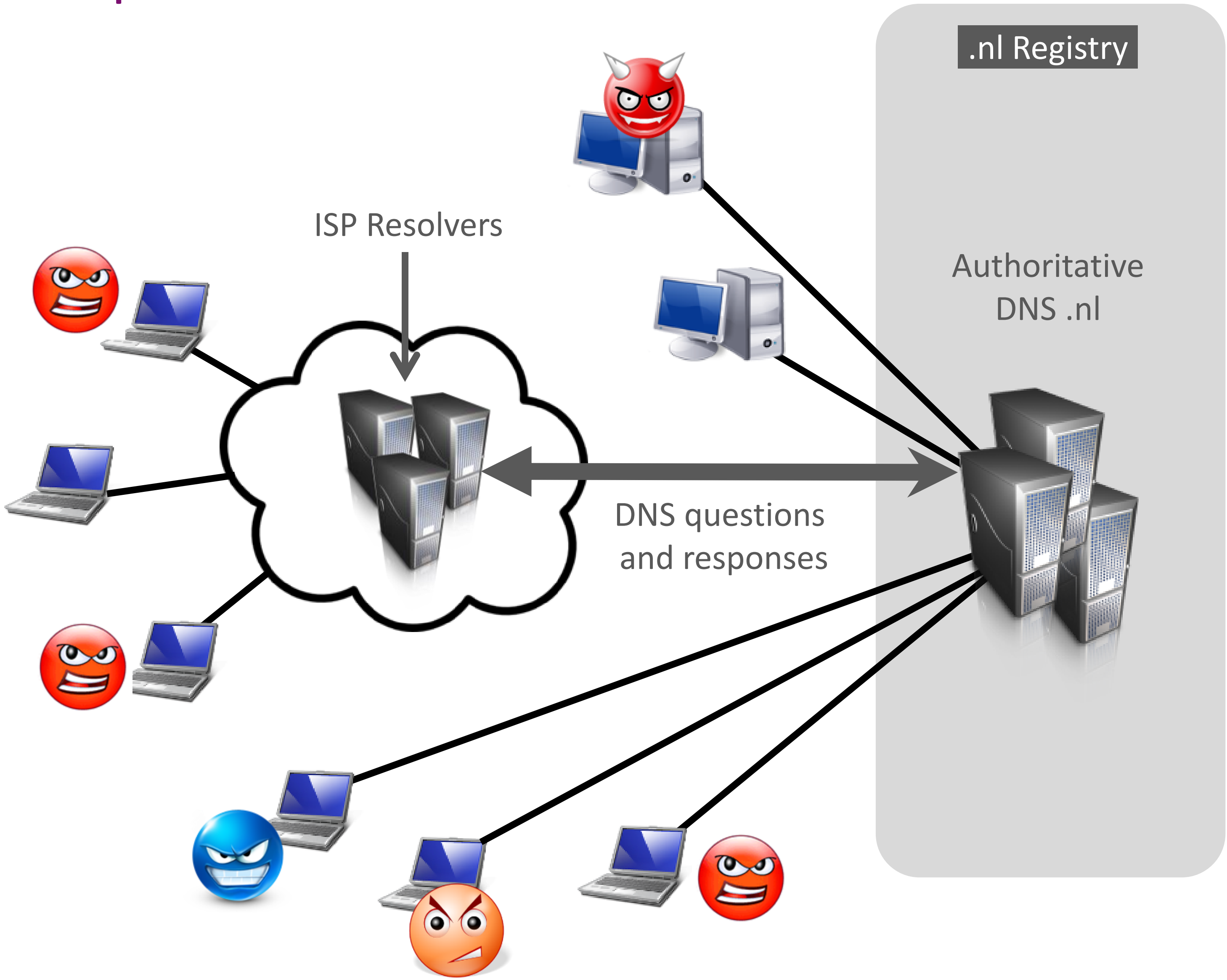
US
DE
GB
UA
RU
NL
BG
RO
CH
IT
2015-04-05

Resolver reputation

Goal: Find out if malicious activity can be mittigated by assigning reputation scores to resolvers

How: “fingerprinting” resolver behaviour

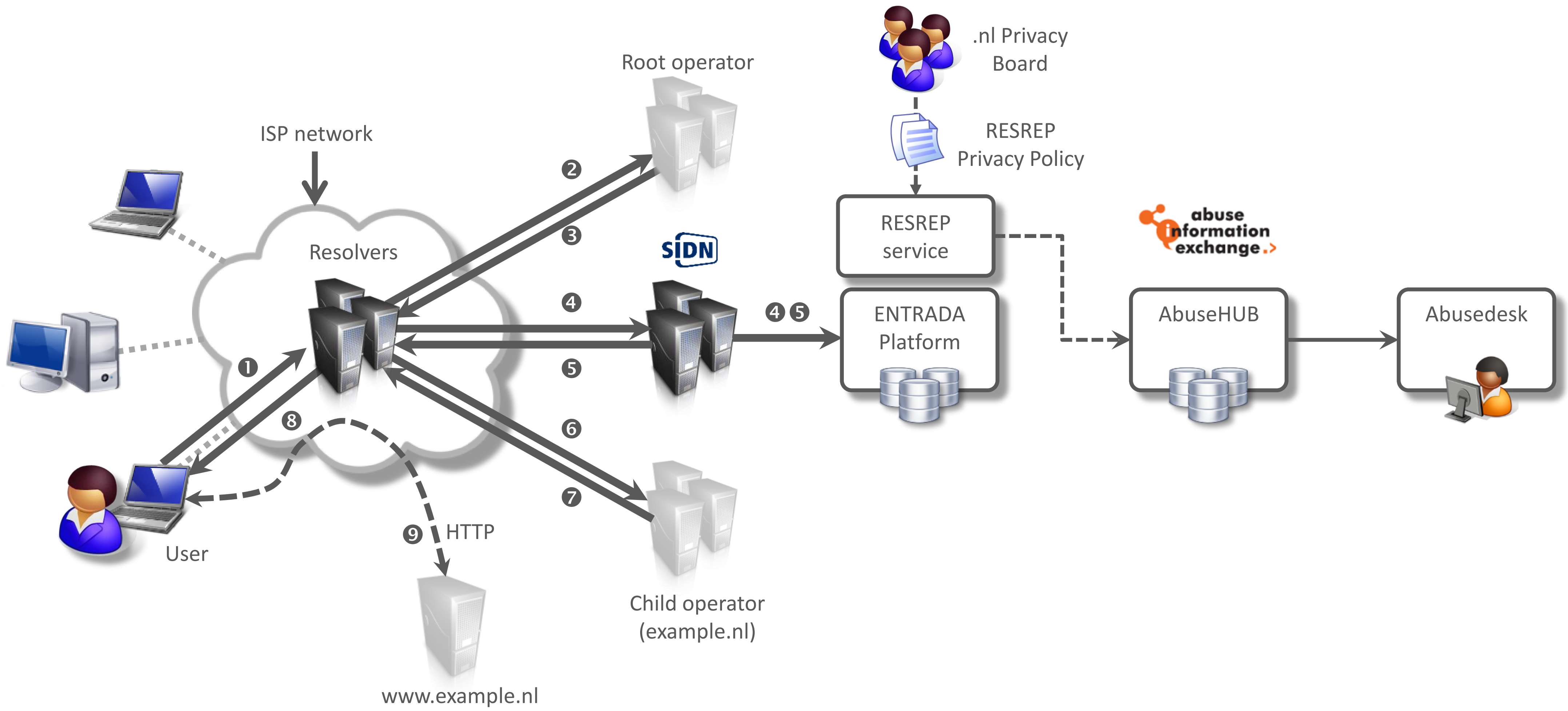
Concept



Malicious activity:

- Spam-runs
- Botnets like Cutwail
- DNS-amplification attacks

Architecture



Conclusion

Technical:

- Hadoop HDFS / Parquet / Impala is a winning combination!

Contributions:

- Data used for research by SIDN Labs and universities
- Identified malicious domain names and botnets
- Data feed to Abuse Information Exchange (soon)
- Insights into the DNS query data

Questions

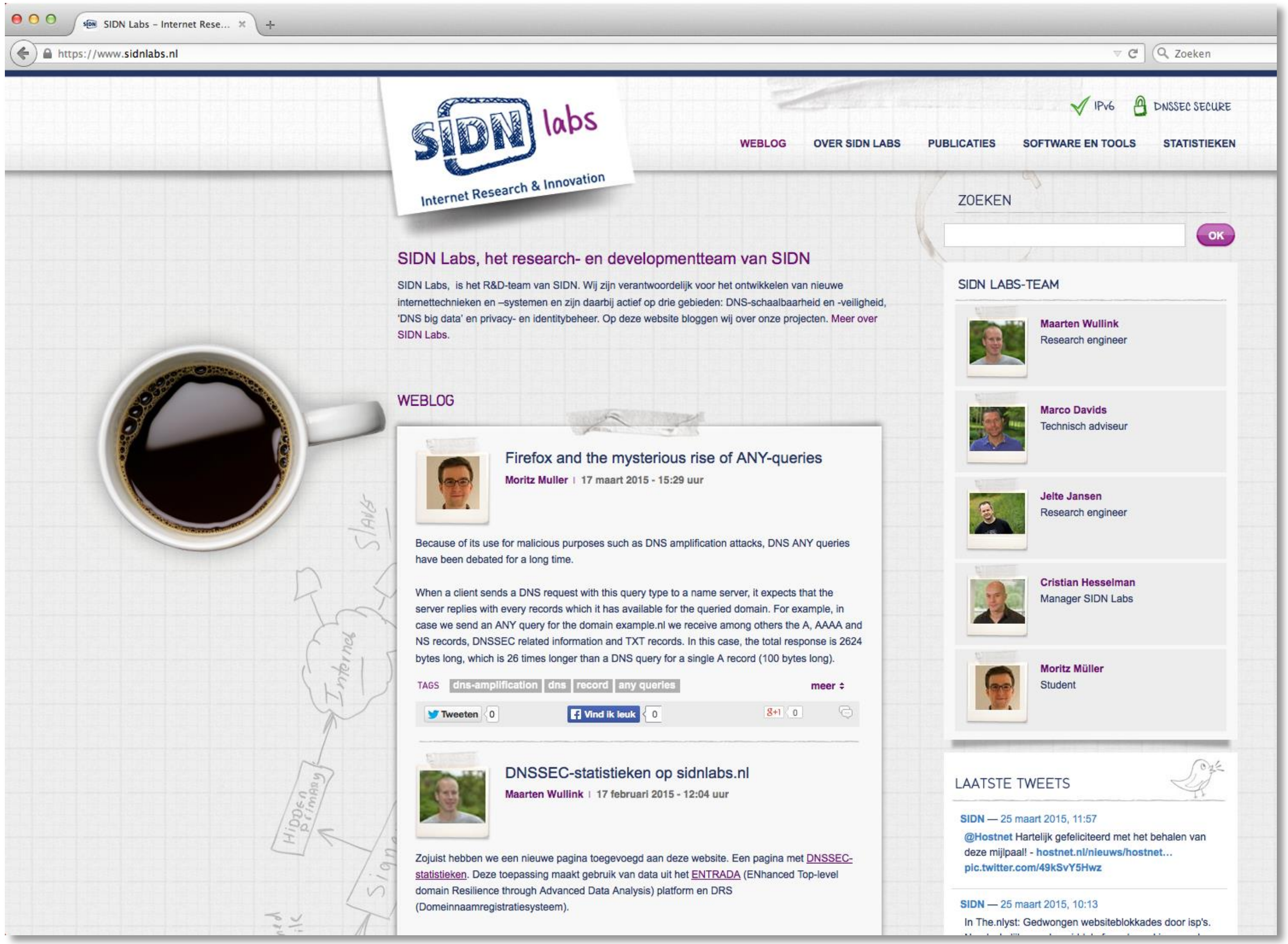
Maarten Wullink

Research Engineer

maarten.wullink@sidn.nl

@wulliak

www.sidnlabs.nl



SIDN Labs, het research- en developmentteam van SIDN

SIDN Labs, is het R&D-team van SIDN. Wij zijn verantwoordelijk voor het ontwikkelen van nieuwe internettechnieken en –systemen en zijn daarbij actief op drie gebieden: DNS-schaalbaarheid en -veiligheid, 'DNS big data' en privacy- en identitybeheer. Op deze website bloggen wij over onze projecten. Meer over SIDN Labs.

WEBLOG

Firefox and the mysterious rise of ANY-queries

Moritz Muller | 17 maart 2015 - 15:29 uur

Because of its use for malicious purposes such as DNS amplification attacks, DNS ANY queries have been debated for a long time.

When a client sends a DNS request with this query type to a name server, it expects that the server replies with every records which it has available for the queried domain. For example, in case we send an ANY query for the domain example.nl we receive among others the A, AAAA and NS records, DNSSEC related information and TXT records. In this case, the total response is 2624 bytes long, which is 26 times longer than a DNS query for a single A record (100 bytes long).

DNSSEC-statistieken op sidnlabs.nl

Maarten Wullink | 17 februari 2015 - 12:04 uur

Zojuist hebben we een nieuwe pagina toegevoegd aan deze website. Een pagina met DNSSEC-statistieken. Deze toepassing maakt gebruik van data uit het ENTRADA (ENhanced Top-level domain Resilience through Advanced Data Analysis) platform en DRS (Domeinnaamregistratiesysteem).

SIDN LABS-TEAM

Maarten Wullink
Research engineer

Marco Davids
Technisch adviseur

Jelte Jansen
Research engineer

Cristian Hesselman
Manager SIDN Labs

Moritz Müller
Student

LAATSTE TWEETS

SIDN — 25 maart 2015, 11:57
@Hostnet Hartelijk gefeliciteerd met het behalen van deze mijlpaal! - hostnet.nl/nieuws/hostnet... pic.twitter.com/49kSvY5Hwz

SIDN — 25 maart 2015, 10:13
In The.nlyst: Gedwongen websiteblokkades door isp's.