DNS resolver reputation

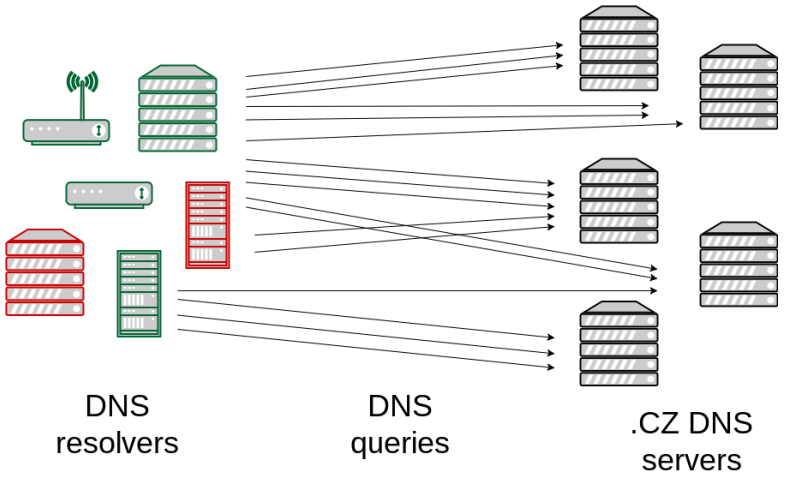
An insight into DNS traffic

Maciej Andziński • maciej.andzinski@nic.cz • 15.11.2018

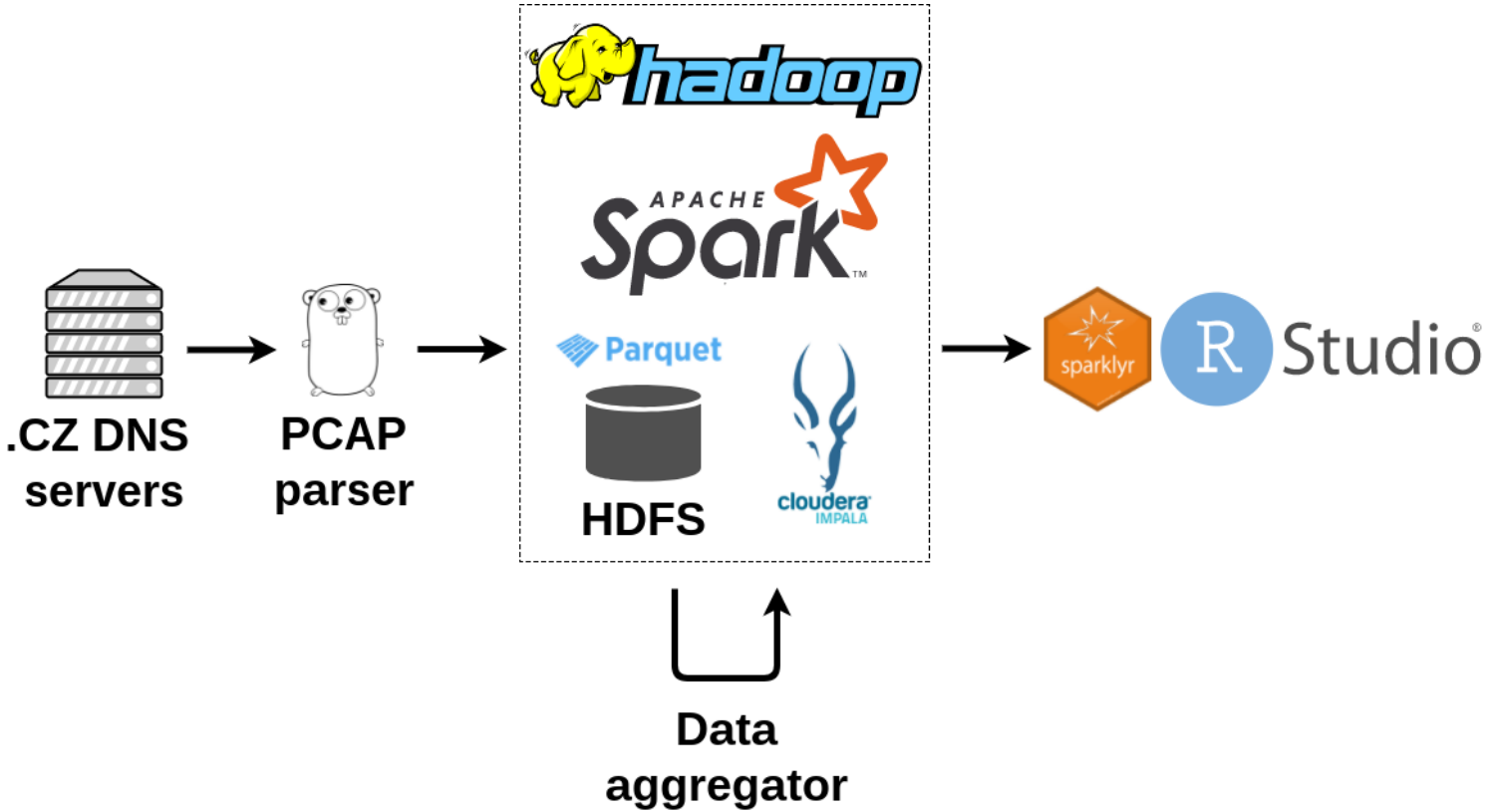
CZ DOMAIN REGISTRY

DNS resolver reputation?

- Identify anomalous sources of DNS queries
- What is an anomaly?
  - Hard to define
    - Scanners?
    - Monitors?
    - Misconfigured resolvers?
  - **Unusual behaviour**

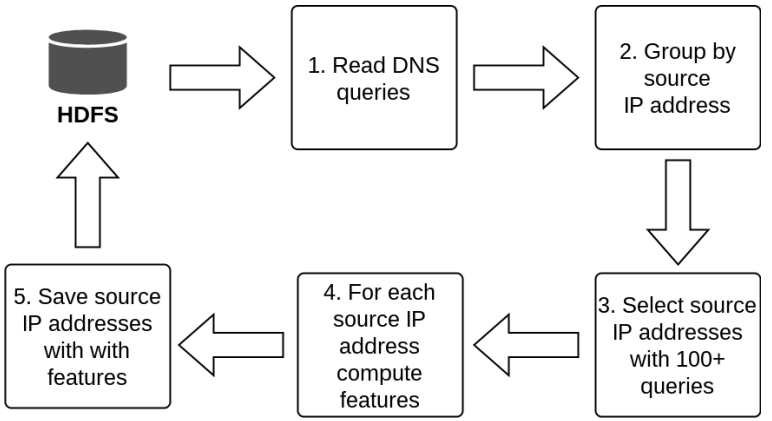


Data flow


.CZ DNS servers
PCAP parser
hadoop
APACHE Spark
Parquet
HDFS
cloudera IMPALA
sparklyr
R Studio
Data aggregator

Data aggregation

1) Group DNS queries by source IP address
2) For each source IP address compute statistics (features)
   - Take only IP addresses which send min. 100 queries daily
   - Time window = 1 day



Features #1

- **Entropy (normalised Shannon Index)**
  - Source port
  - Transaction ID
- **Coefficient of variation ($C_V=\sigma/\mu$)**
  - Idletime
  - Packet length

Features #2

- Amplification factor
- Mean domain name length
- Domain name diversity

Features #3

- **Observed DNS QTYPEs**
  - A + AAAA
  - NS
  - DNSSEC RRs
  - Popular RRs
  - Weird RRs

Features #4

- **Observed DNS RCODEs**
  - NOERROR
  - NXDOMAIN
- **Observed DNS FLAGs**
  - RD
  - EDNS0 DO

Features #5

- **Observed DNS QCLASSes**
  - IN
- **Observed DNS OPCODEs**
  - QUERY

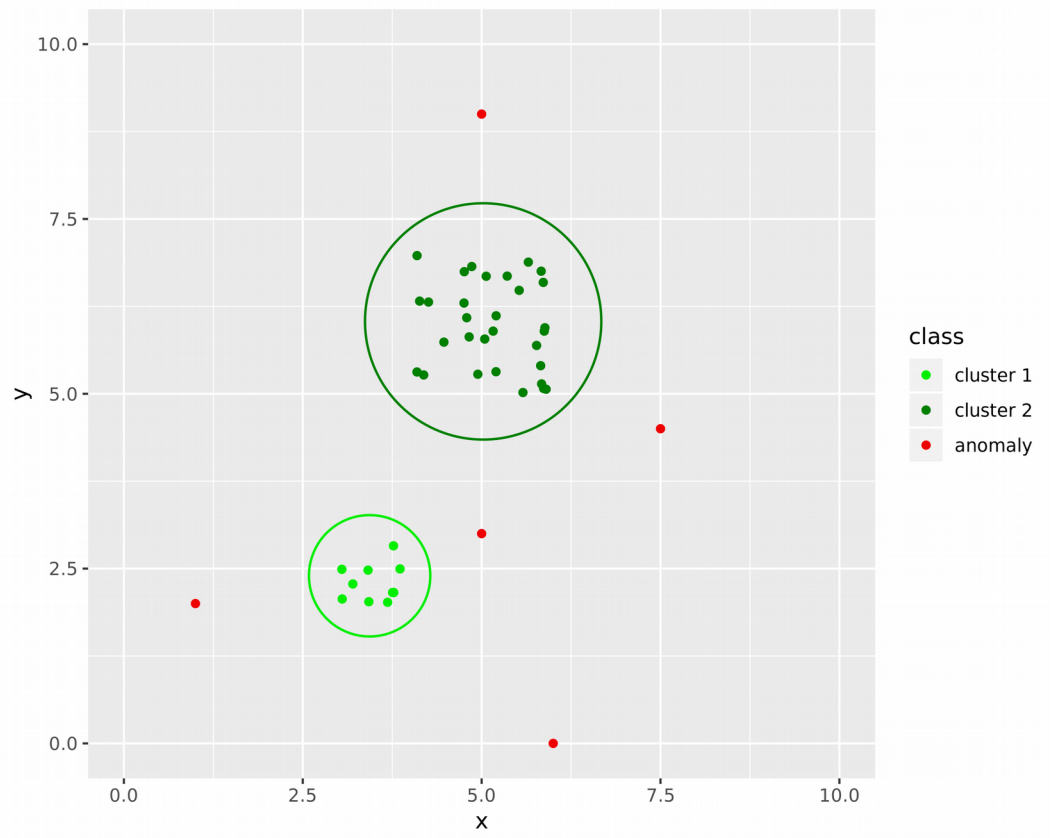
Features – an example

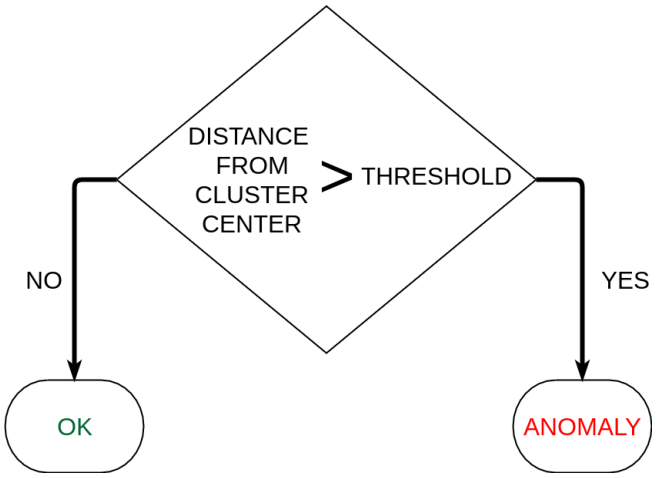
- 217.31.204.130 on 23 October 2018

  (CZ.NIC open DNS resolver)

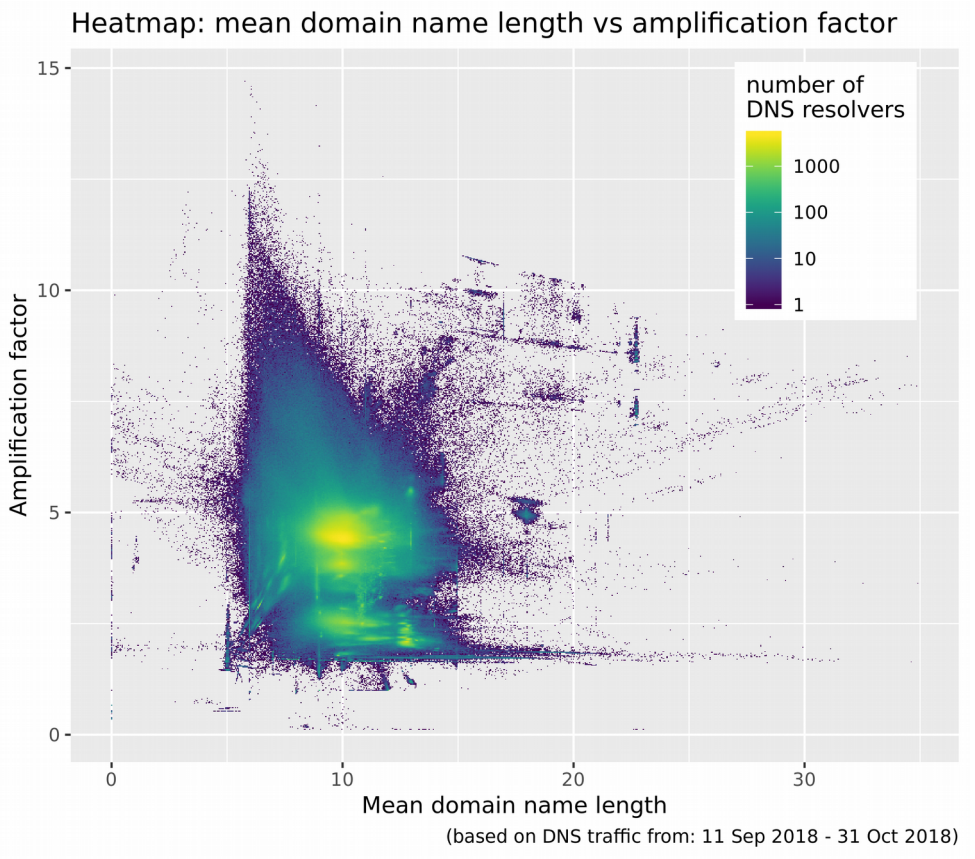
srcp_sh_ix_n	0.9876179
id_sh_ix_n	0.9879639
idletime_cv	0.709647
dn_len_mean	11.57786
dn_perc	0.292901
rcode_noerror_perc	0.9828929
rcode_nxdomain_perc	0.01710712
qtype_common_perc	0.975649
qtype_weird_perc	0.001121778
qtype_dnssec_perc	0.1577407
qtype_ns_perc	0.01025492
qtype_addr_perc	0.9247106
qclass_in_perc	1
edns_do_perc	1
flag_rd_perc	0
ampl_factor	4.743633
len_cv	0.1132715
opcode_query_perc	1

Anomaly detection concept

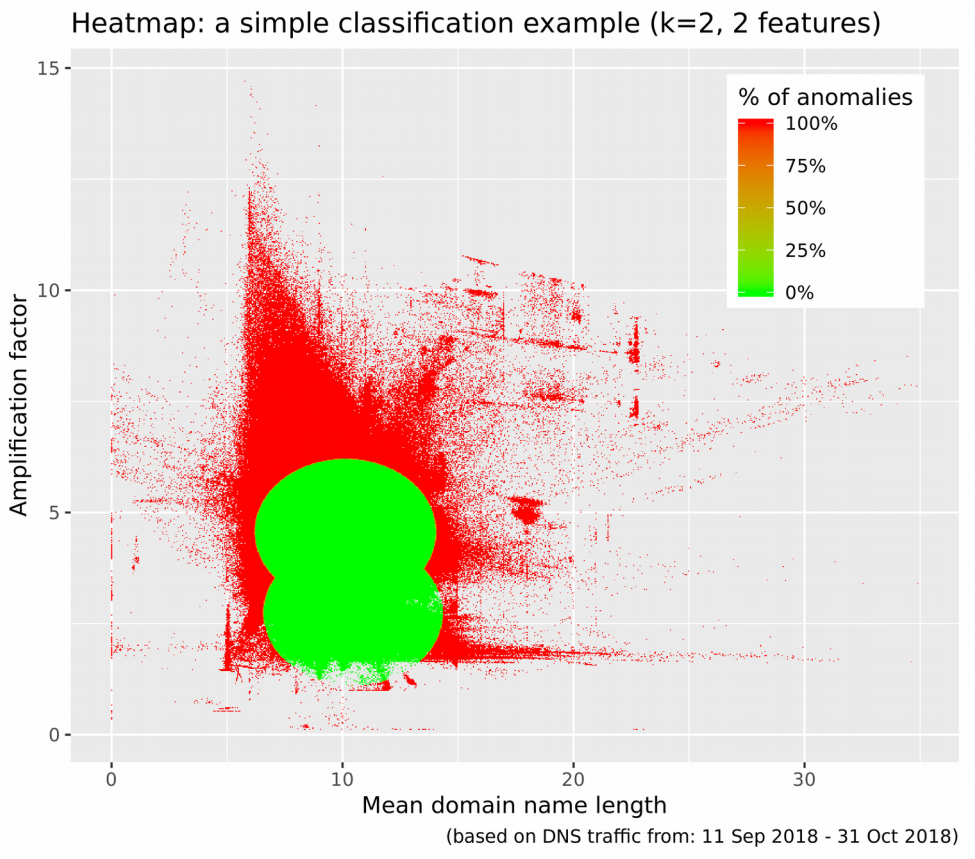




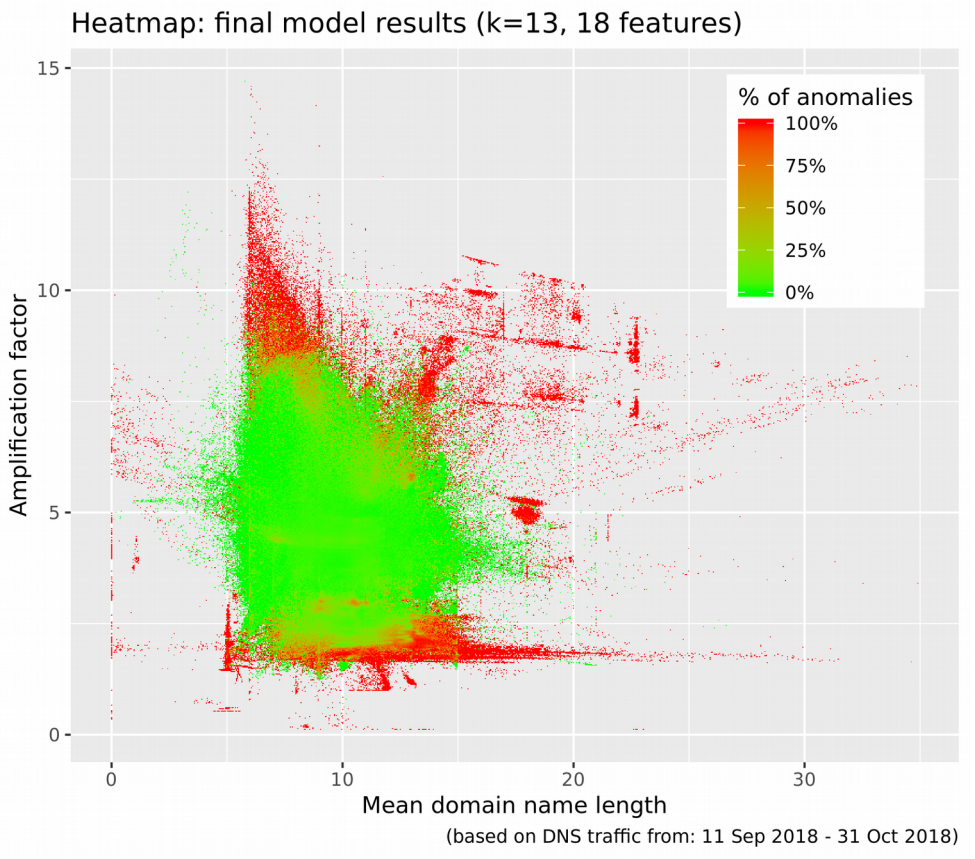
Anomaly detection concept



Anomaly detection concept



Anomaly detection concept



Model

- Spark MLlib
- K-means clustering
  - UDF to compute distance from cluster center
- MinMaxScaler
  - Entire dataset used for scaling (some features in training set were meaningful but had “near zero” variance)

Model

- Training set
  - Real DNS resolvers (each RIPE Atlas probe was employed to query its local DNS resolver for whoami.akamai.net)
    - Gathered 3 430 unique IP addresses
    - 51 days = 137 701 observations
    - Filtered out weird observations

Model

- Test/Validation set
  - Difficult to measure anomaly detection performance
  - Needed for grid search to select best model parameters (best F-score)

Model

- Test/Validation set #1
  - Real DNS resolvers
    - DNS resolvers of RIPE Atlas probes
    - Google Public DNS
    - Cloudflare
    - Quad9
    - OpenDNS (Cisco)
    - Dyn
    - Level3
    - Yandex
    - CZ.NIC

Model

- Test/Validation set #2
  - Known anomalies
    - DNSMON
    - Domain name scanners
    - Misconfigured DNS resolvers

Model

- Model parameters
  - k = 13
  - Threshold (maximal distance from cluster center) = 3 * Q3 (third quartile)

Model performance

F-score: **0.9894033**

- Real DNS resolvers

dataset	total	anomaly	%anomaly
atlas_resolvers	3193	48	1.5 %
google	1250	0	0.0 %
quad9	224	0	0.0 %
opendns	107	2	1.9 %
dyn	107	3	2.8 %
level3	160	4	2.5 %
cloudflare	180	2	1.1 %
yandex	82	2	2.4 %
cznic	2	0	0.0 %

Model performance

F-score: **0.9894033**

- Known anomalies

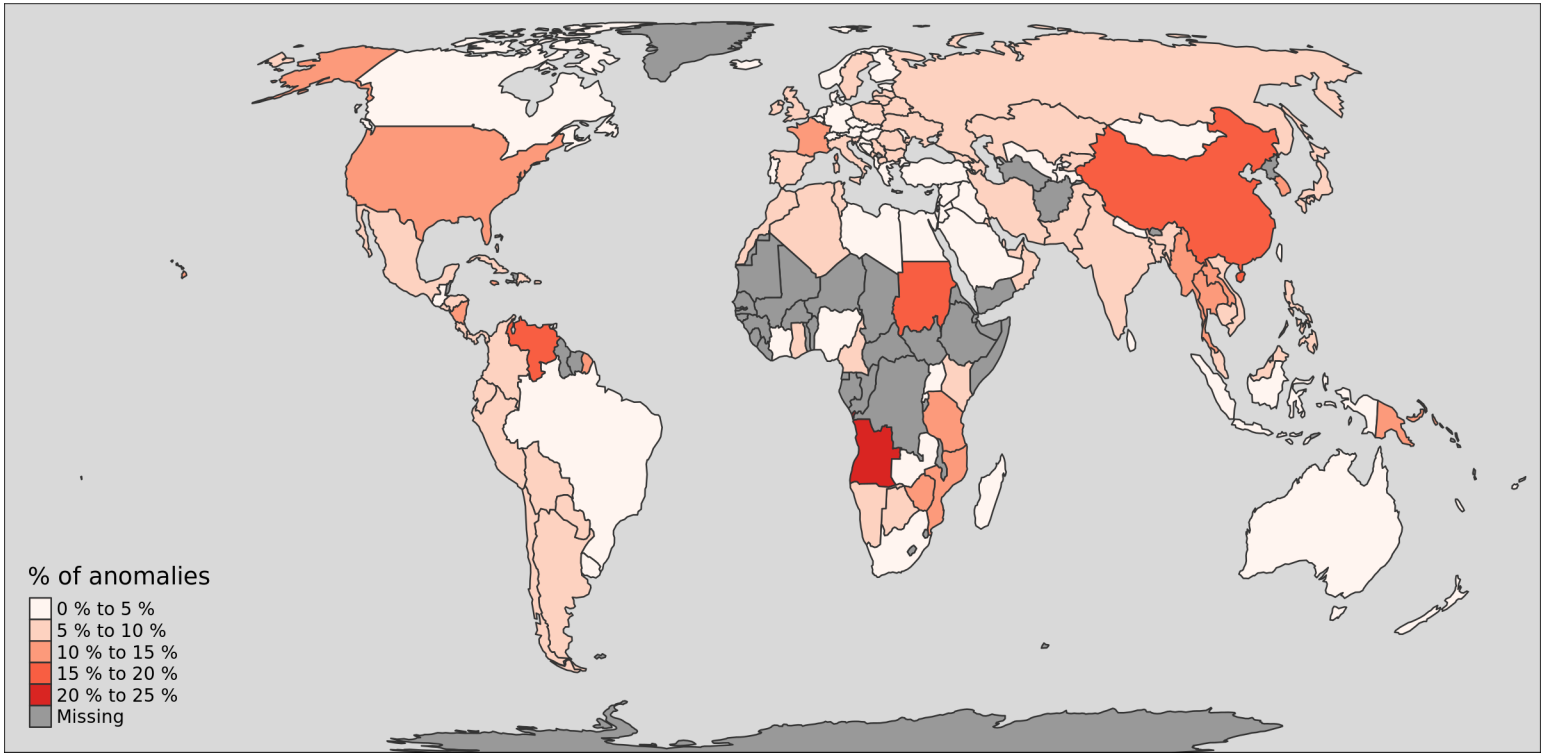
dataset	total	anomaly	%anomaly
dnsmon	38	38	100.0 %
scanners	25	25	100.0 %
scanners2	100	100	100.0 %
misconfigured	99	99	100.0 %
dnsviz	1	0	0.0 %

Results

- DNS traffic from 11 Sept 2018 - 31 Oct 2018
  - 737 729 out of 9 918 267 observations **(7.4%)** were classified as **anomaly**
  - 8 649 294 465 out of 40 073 507 471 queries **(17.7%)** were originated in **anomalous source**

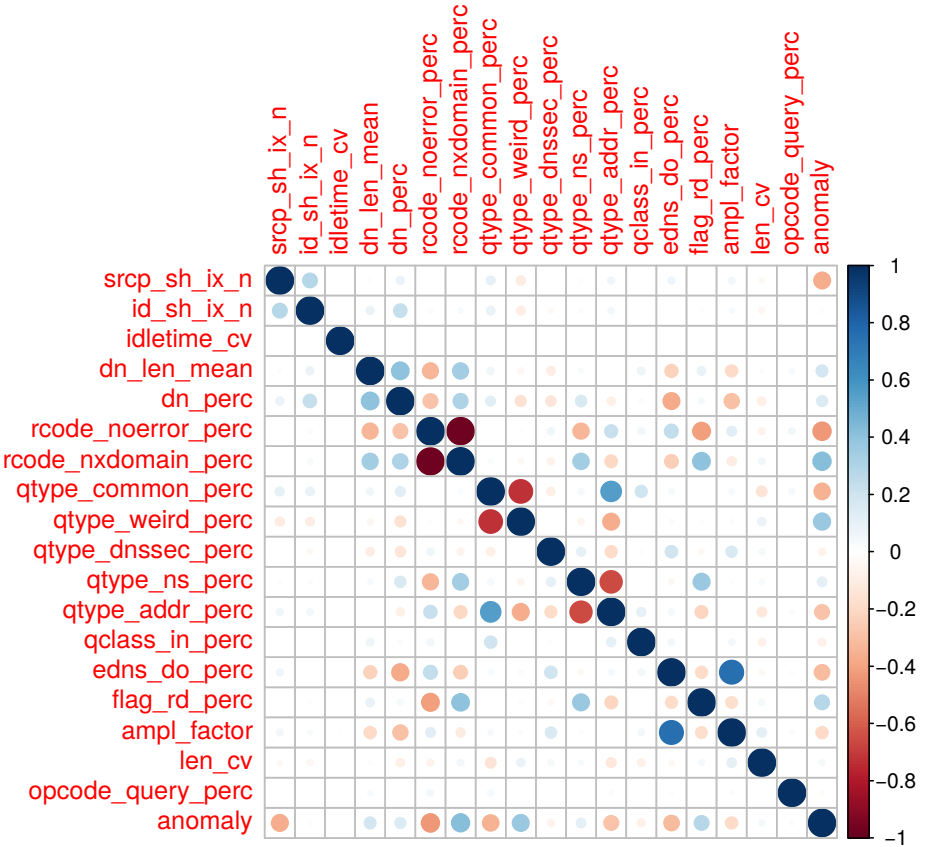
Results

Anomalous sources by country (only countries with >1000 observations)



(based on DNS traffic from 11 Sept 2018 – 31 Oct 2018)

Results



Feature correlation matrix (Pearson)

Findings

- A security issue with one of the DNS operators

  (details to be disclosed later)

Findings

- AS25192 (CZ.NIC, z.s.p.o.)
  - 5th biggest in terms of query number
  - 2 496 observations (**128 unique IP addresses**)
    - 525 (21 %) classified as anomaly (22 unique IP addresses)
  - 1 731 755 782 queries
    - 87 432 646 (5 %) from anomalous sources

Findings in AS25192 (CZ.NIC, z.s.p.o.)

- 32 out of 128 IP addresses were observed every day
  - 19 were **never anomalous** (0%)
  - 5 were **almost never anomalous** (<5%)
  - 7 were **always anomalous** (100%)
  - 1 was **almost always anomalous** (>90%)

Findings in AS25192 (CZ.NIC, z.s.p.o.)

- **Always classified as anomaly (100%)**
  - Incigna monitoring system (IPv4+IPv6)
  - Domain name crawler
  - RIPE Atlas anchor (IPv4 + IPv6)
  - A monitoring system without name
  - DNS resolver for Hadoop cluster (IPv6)
- **Almost always classified as anomaly (>90%)**
  - DNS resolver for Hadoop cluster (IPv4)

Findings in AS25192 (CZ.NIC, z.s.p.o.)

- **Never classified as anomaly (0%)**
  - Real DNS resolvers
- **Ocasionally classified as anomaly (<5%)**
  - DNS resolver for mail server
    - A configuration issue was discovered
  - NAT gateways

Future work

- Add more classes
  - Scanner, monitor, misconfigured, under attack, etc.
- Extend / modify feature set
- Try different algorithms
- Collect better ground truth
- Visualise results

Thank You

Maciej Andziński • maciej.andzinski@nic.cz

CZ DOMAIN REGISTRY