DETECTING RESOLVERS AT .NZ

Jing Qiao, Sebastian Castro
DNS-OARC 29
Amsterdam, October 2018

BACKGROUND

DNS TRAFFIC IS NOISY

Despite general belief, not all the sources at auth nameserver are resolvers

Non-resolvers' traffic could skew our Domain Popularity ranking result

If we can tell whether a source is a resolver?

OBJECTIVE

A classifier to predict the probability of a source being a resolver

It's not a simple task:

1. Trail blazing
2. Uncertainty about resolvers' pattern

"In the search of resolvers", OARC 25

"Understanding traffic sources seen at .nz", OARC 27

USING DNS KNOWLEDGE TO REDUCE THE SCOPE

REMOVE NOISE BY CRITERIA

4 weeks (28/08/2017 – 24/09/2017)

27.8% of sources only queried for 1 domain
45.5% of sources only queried for 1 query type
25% of sources only queried 1 of 7 .nz servers
65.8% of sources sent no more than 10 queries per day

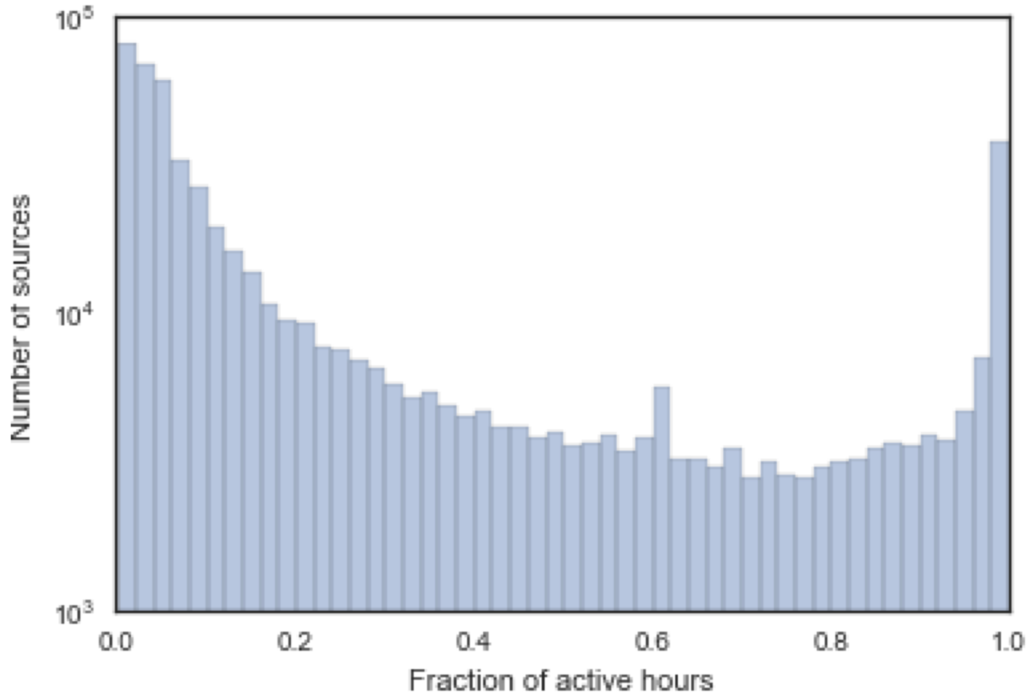
The address list is reduced from 2M to 550k

FOCUS ON ACTIVE SOURCES

Active at least 75% of total hours

Active at least 5/7 days per week

Address list is further reduced to 82k



ASSUMPTION: THE REST IS EITHER RESOLVER OR MONITOR.

THE KEY IS TO FIND DISCRIMINATING FEATURES.

DATASET

.nz DNS traffic across 4 weeks

82k unique sources

KNOWN SAMPLES

2515 Resolvers

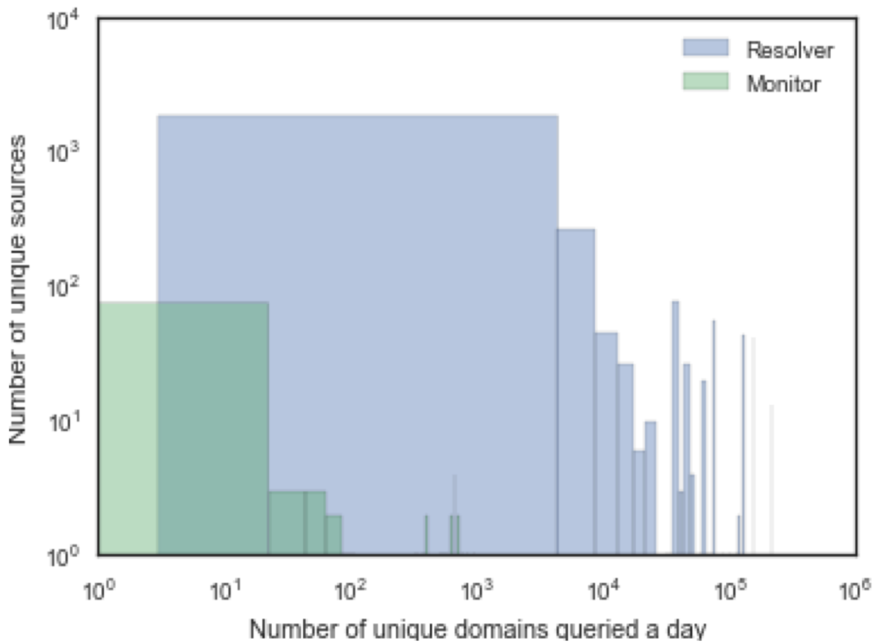
- ISP, Google DNS, OpenDNS, Education & Research, addresses collected from RIPE Atlas probes

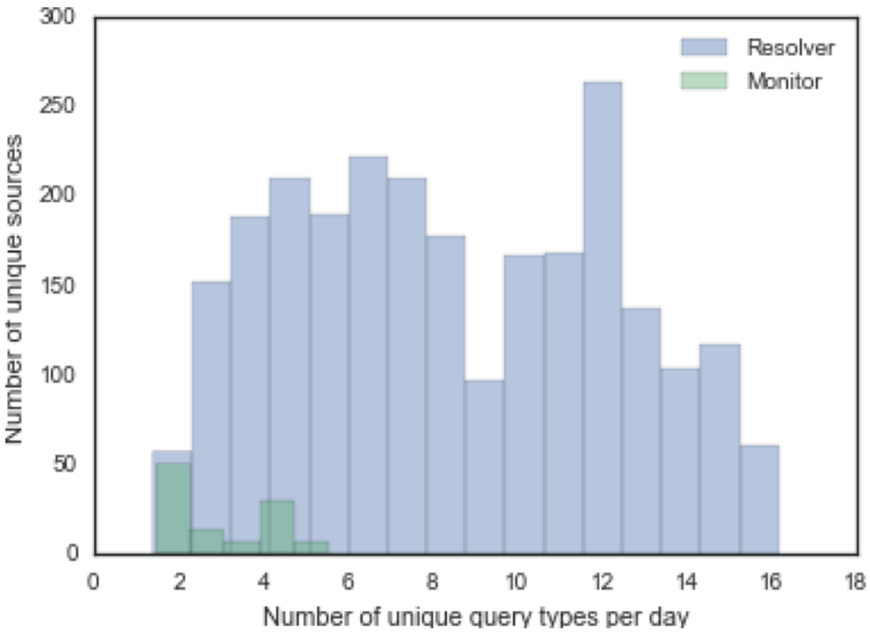
106 Monitors

- ICANN, Pingdom, ThousandEyes, RIPE Atlas probes, RIPE Atlas anchors

NO CLEAR SPLIT OVER SINGLE FEATURE

Resolvers and monitors are distributed across overlapped ranges





TEMPORAL FEATURES

Query features: query type, query name, query rate, DNS flag, response code, …

- Each feature can be described as a time series
- mean, standard deviation, percentiles (10, 90)

ENTROPY FEATURES

Queries from a resolver should be more random than those from a monitor

Timing entropy:

- Time lag between successive queries
- Time lag between successive queries of same query name and query type

Query name entropy:

- Similarity of the query names (Jaro-Winkler string distance) between successive queries

VARIABILITY

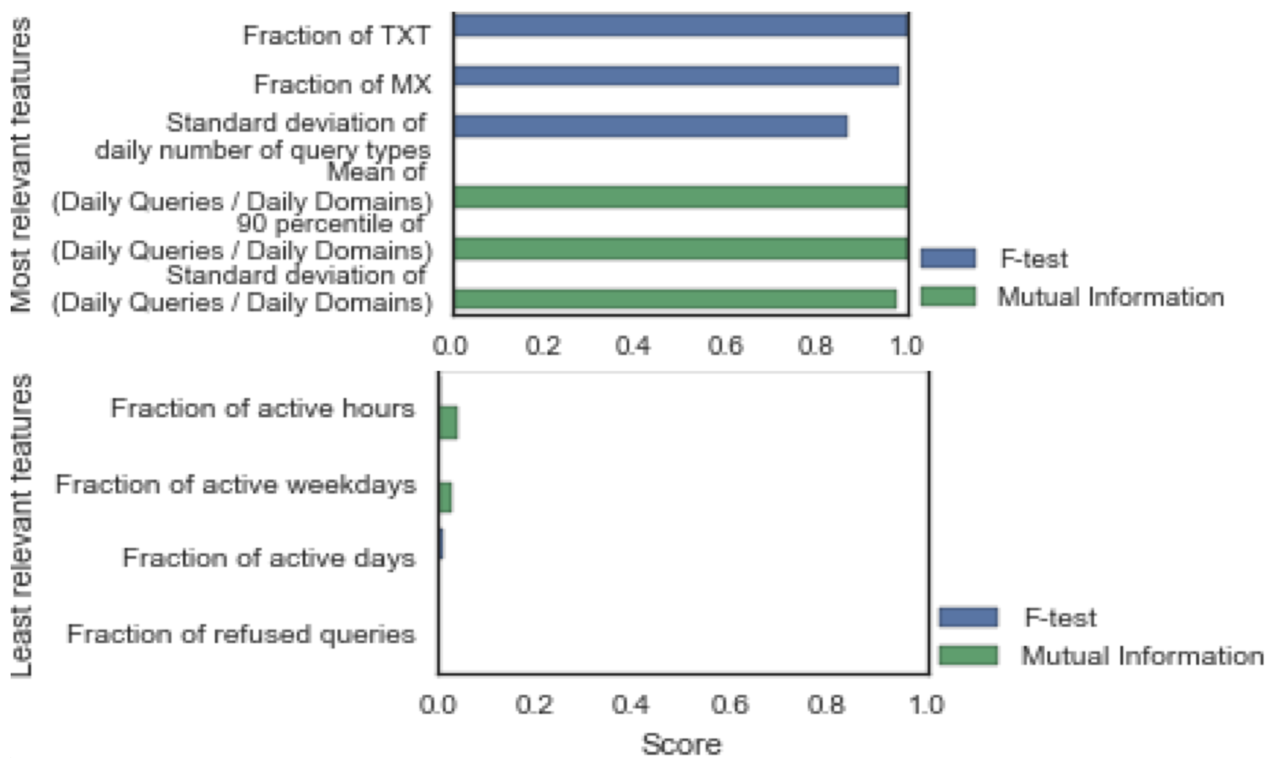
A monitor's query flow is likely to be less variable compared to a resolver

- Finer aggregation across hour tiles
- Variance metrics
  - Interquartile Range
  - Quartile Coefficient of Dispersion
  - Mean Absolute Difference
  - Median Absolute Deviation
  - Coefficient of Variation

FEATURE SELECTION

Remove redundant features with above 0.95 correlation

Select most relevant features by statistical tests, such as F-test and Mutual Information (MI)



FINAL FEATURE SET

50 features of different scales

Normalize to comparable scales

- Standardization: zero mean and unit variance
- Quantile Transformation: robust to outliers

VERIFY THE FEATURE SET BY CLUSTERING

Algorithms:

- K-Means
- Gaussian Mixture Model
- MeanShift
- DBSCAN
- Agglomerative Clustering

Evaluation metrics:

- Adjust Rand Index
- Homogeneity Score
- Completeness Score

CLUSTERING RESULT

The model with best performance on the known samples:

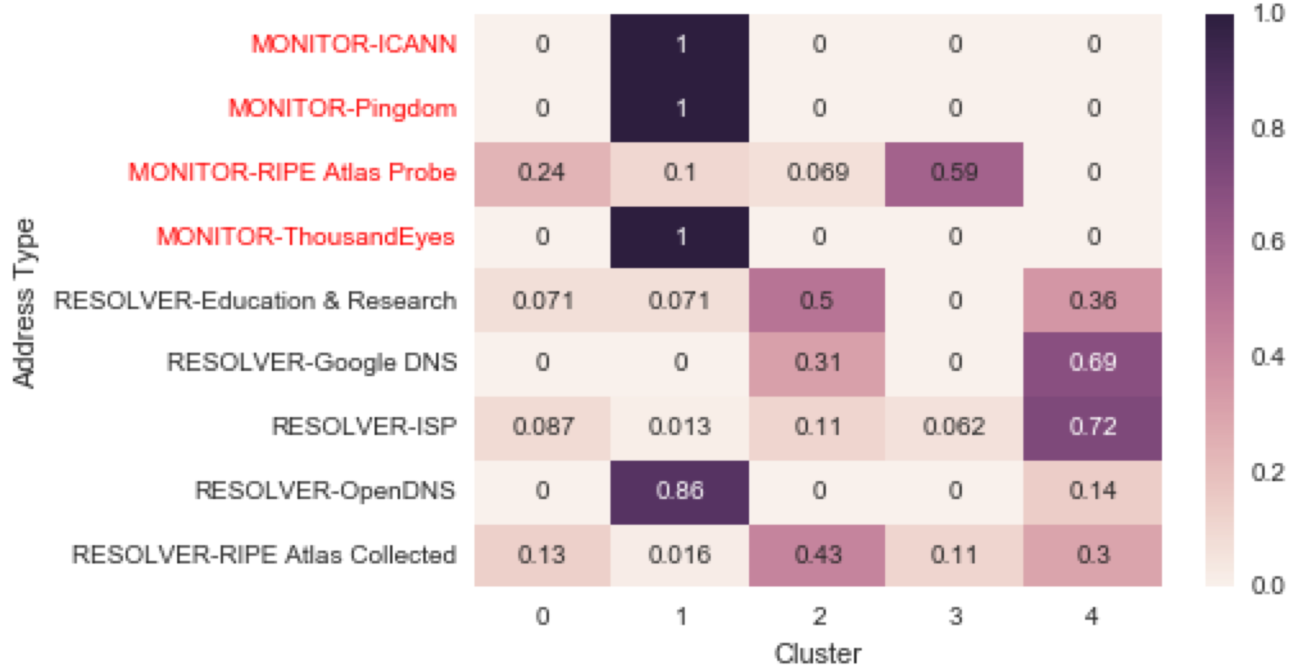
Gaussian Mixture with 5 clusters

- Adjust Rand Index: 0.849789 (good)
- Homogeneity score: 0.960086 (good)
- Completeness score: 0.671609 (average but acceptable as resolvers can be separated into a set of clusters different from monitors)

VERIFY WITH GROUND THRUTH

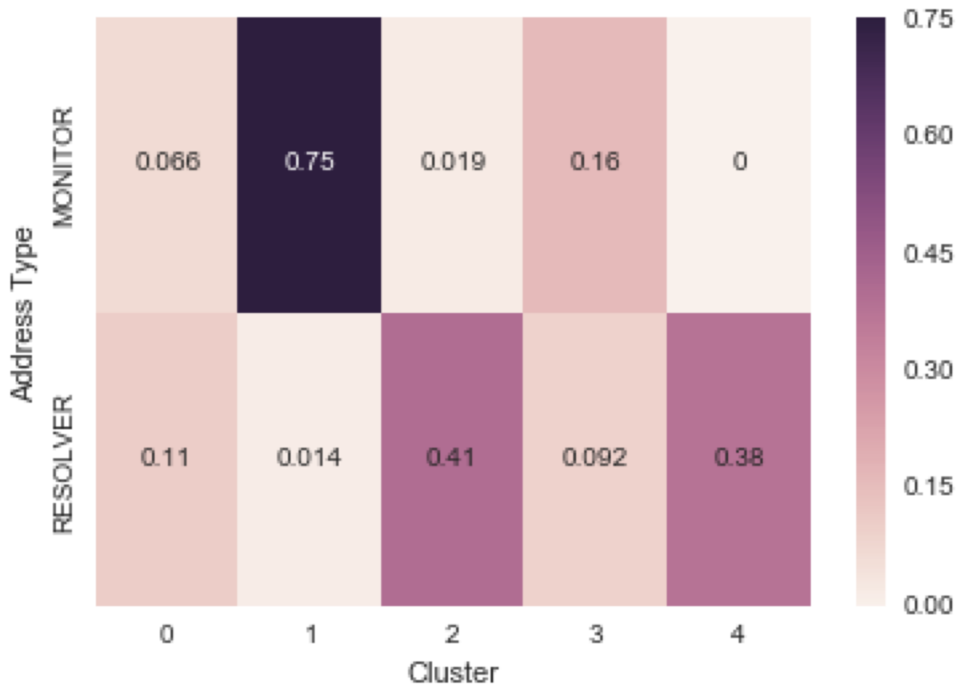
Most resolvers are separated from monitors OpenDNS is a special case

- Only 7 samples
- Specific behavior



CONCLUSION

Exclude samples from OpenDNS
90% resolvers (Cluster 0, 2, 4) are separated from 91% monitors (Cluster 1, 3) using our feature set



SUPERVISED CLASSIFIER

Training data

- 2515 positive (resolver)
- 106 negative (monitor)

50 features

4 week period

AUTOML

Challenges of training a classifier:

- Searching algorithms and hyperparameters
- Performance and efficiency

We use auto-sklearn to solve the challenges

- Off-the-shelf toolkit that automates the process of model selection and hyperparameter tuning
- Improve performance and efficiency using Bayesian optimization, meta-learning, ensemble construction
- Efficient and Robust Automated Machine Learning, Feurer et al., Advances in Neural Information Processing Systems 28 (NIPS 2015).

PRELIMINARY RESULT

An ensemble of 28 models:

- Running time: 10 minutes
- Accuracy score: 0.991
- Precision score: 0.991
- Recall score: 1.000
- F1 score: 0.995

CURRENT STATUS

- Train the classifier on a sliding window regularly
  - Takes a few hours, mostly on feature generation
- Predict type of source on a given threshold (probability of being like a resolver)
  - We picked probabilities higher than 0.7
- Integrate with popularity ranking algorithm
  - Improved ranking for known domains

CURRENT STATUS

- Out of 100k source address classified, 73k were detected as resolvers

- The model identified 96% of Illimunati probes as resolvers, from a request from SIDN

POTENTIAL USES & FUTURE WORK

With adjustment of feature set and training data, source classification can be used to measure:

- validating resolvers
- adoption of QNAME minimization in the wild
- etc

from authoritative data, e.g. the root servers

ACKNOWLEDGEMENTS

Daniel Griggs for his input to reduce noise.
Sebastian Castro for the variability metrics.

REFERENCE

https://blog.nzrs.net.nz/source-address-clustering-feature-engineering/

https://blog.nzrs.net.nz/source-address-classification-clustering/

THANK YOU!

jing@internetnz.net.nz

sebastian@internetnz.net.nz

InternetNZ .nz