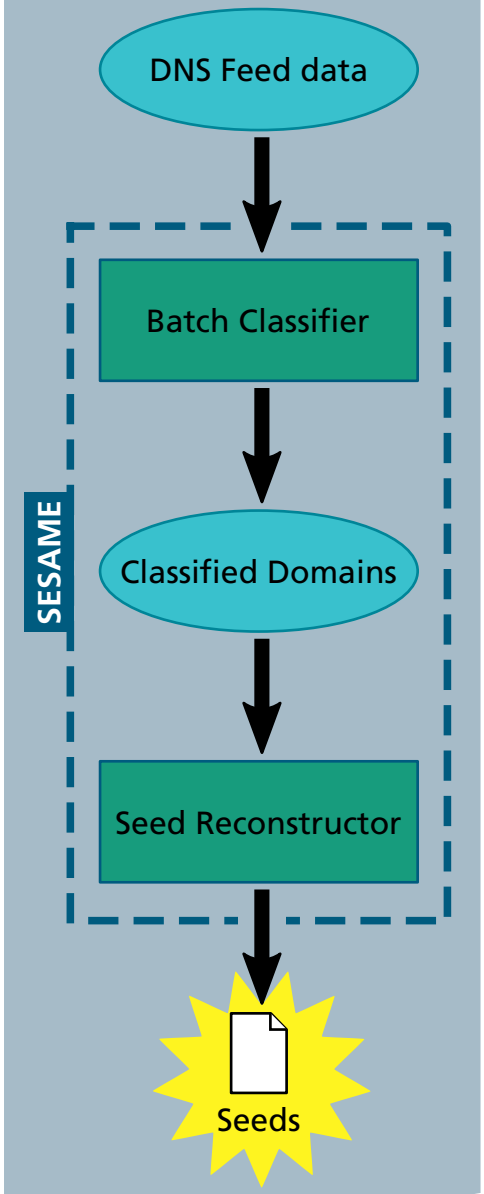
Open SESAME



Fighting Botnets with Seed Reconstructions of Domain Generation Algorithms

Motivation

Extracting seeds for Domain Generation Algorithms (DGAs) from Malware samples is time-consuming manual labor.

Most ML solutions do not identify to which DGA a generated domain belongs to.

A system is desired to:

- Automatically match domains with DGAs
- Extract previously unknown seeds
- Identify new DGAs

Methodology

Batch Classifier:

Multi-Class Residual Neural Network to match domains with DGAs and rate groups of domains according to novelty

Seed Reconstructors:

- 5 Permutators & Iterators
- 7 Bruteforcers
- 8 Smart Bruteforcers
- 6 Other Reconstructors

Evaluation Results

System applied on 21 GB of DNS data

Competetive overall accuracy of 83.89%

Automated processing:

➡ New seeds for 4 DGA families

Manual inspection of 64 highest novelty cases:

➡ 5 new DGAs

➡ 5 new DGA versions

➡ 3 new seeds

Limitations

ML Model:

- Requires groups of domains as input (likely to be generated by same DGA)

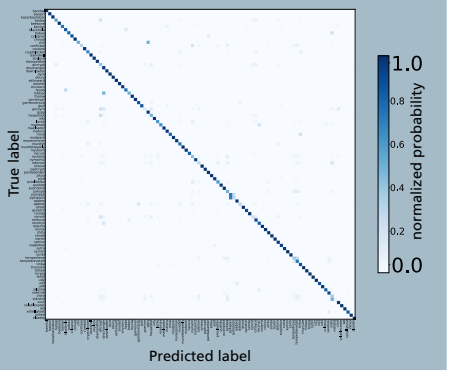
Reconstructors:

- Computational effort
- Time dependencies
- Filtering of non-DGA domains

Fraunhofer FKIE

arxiv.org/abs/2301.05048

Confusion Matrix



A perfect model would yield an identity matrix. Our confusion matrix is close to that.

N. Weissgerber, E. Padilla, T. Jenke, S. Zemanek, L. Bruckschen
Fraunhofer FKIE