# SAPPAN Innovations in DGA Detection

Arthur Drichel
RWTH Aachen University

Hugo Hromic
Hewlett Packard Enterprise

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418.

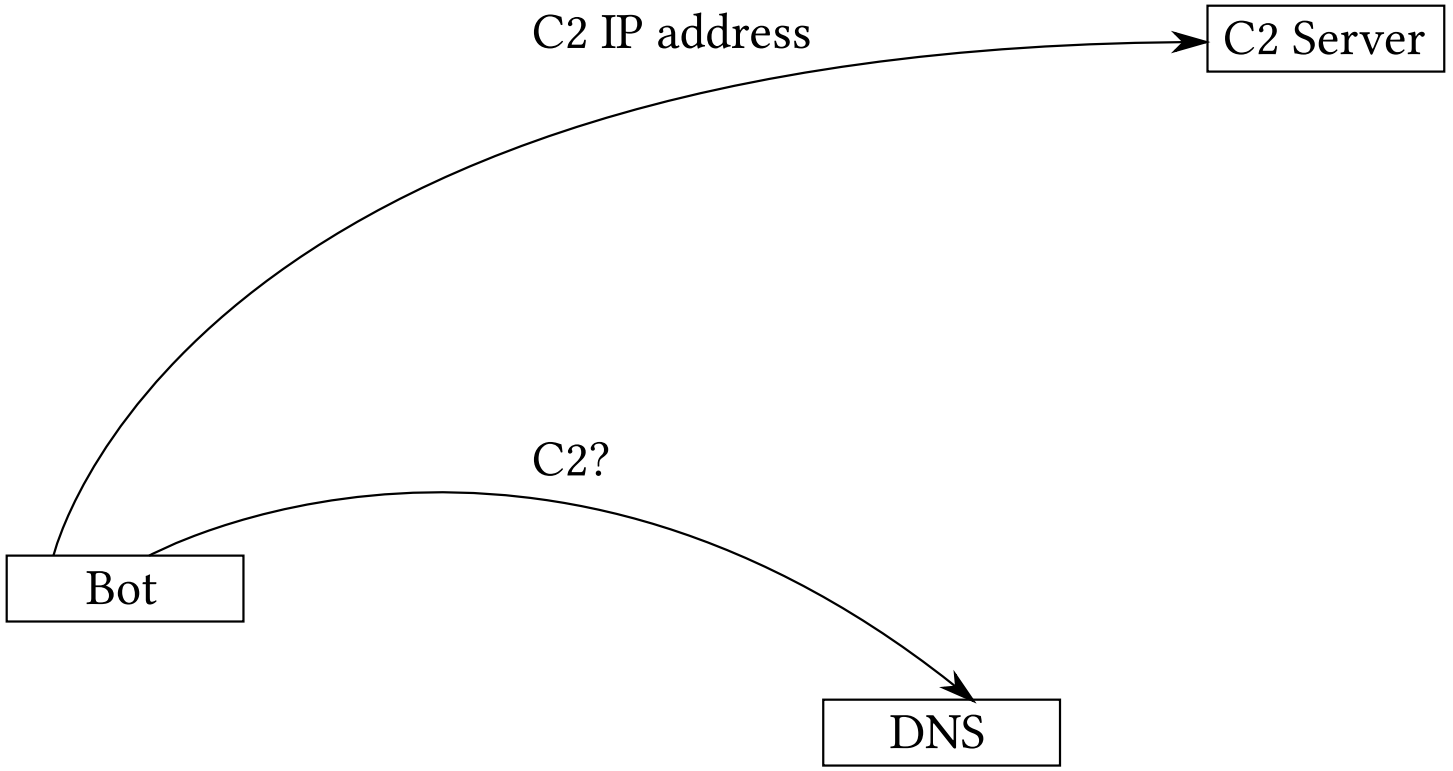
SAPPAN
*SHARING AND AUTOMATION FOR PRIVACY PRESERVING ATTACK NEUTRALIZATION*

Co-funded by the Horizon 2020 programme of the European Union

IT SEC Research Group IT-Security | RWTH AACHEN UNIVERSITY

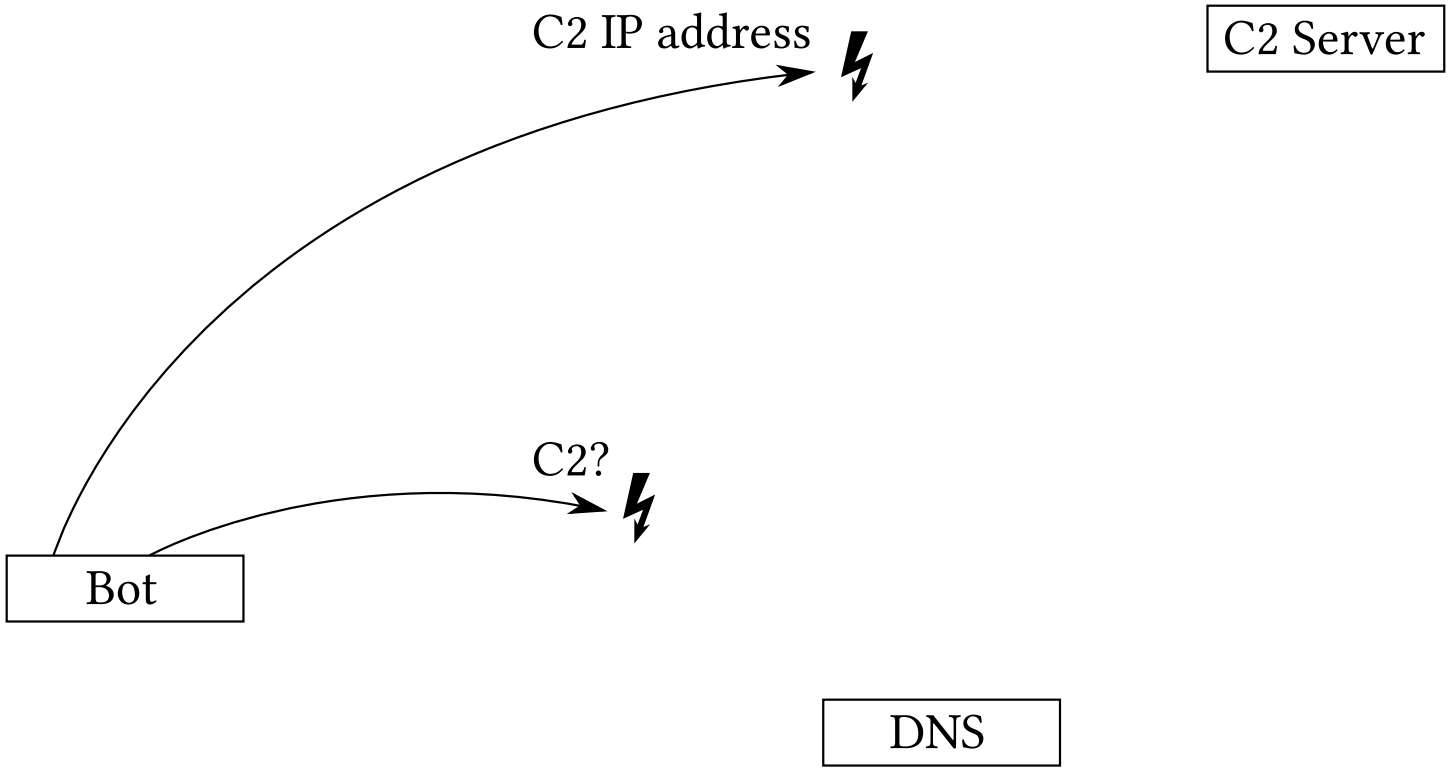
# Domain Generation Algorithms (DGAs)

## Traditional bot communication



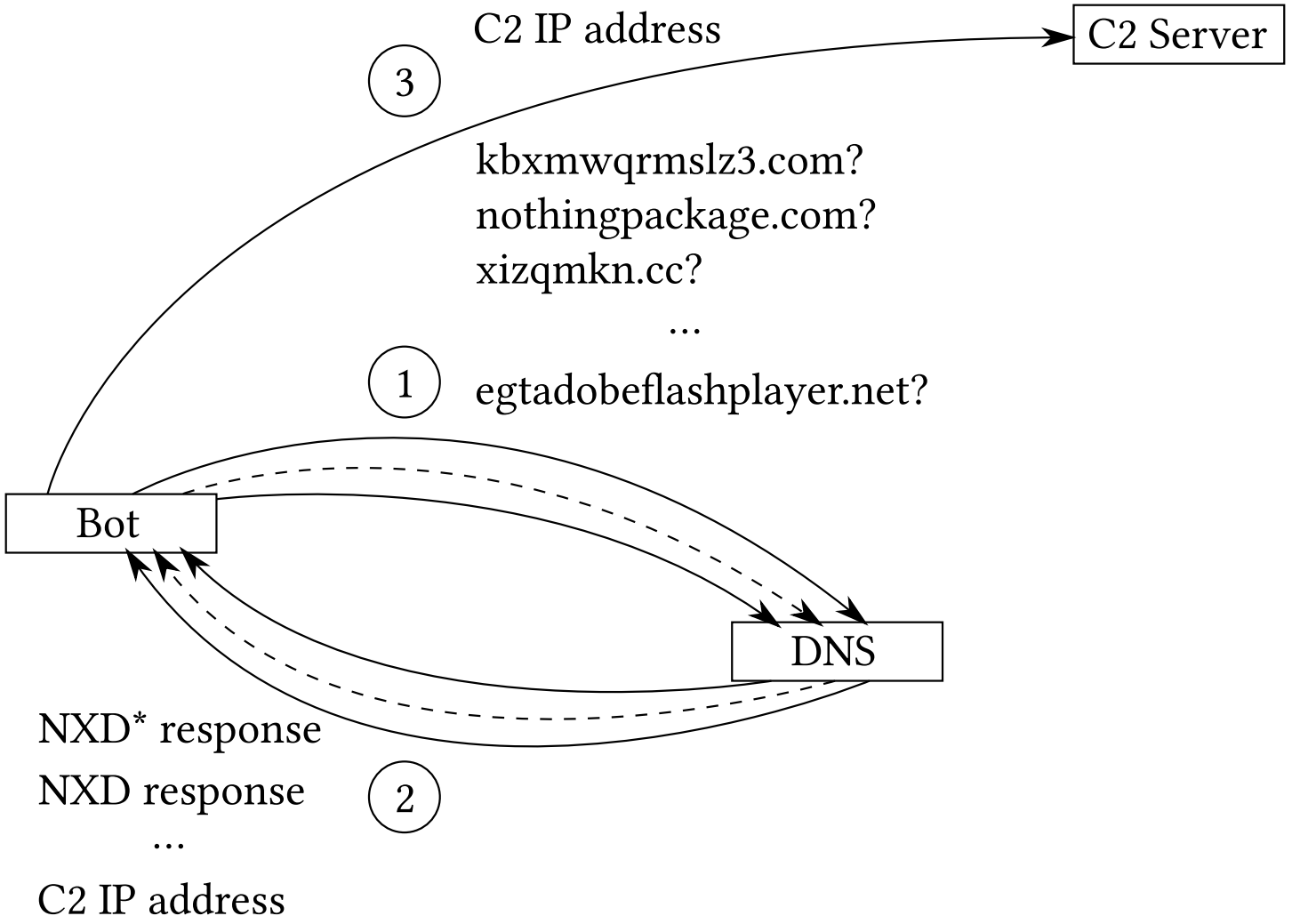
# Domain Generation Algorithms (DGAs)

## Traditional bot communication



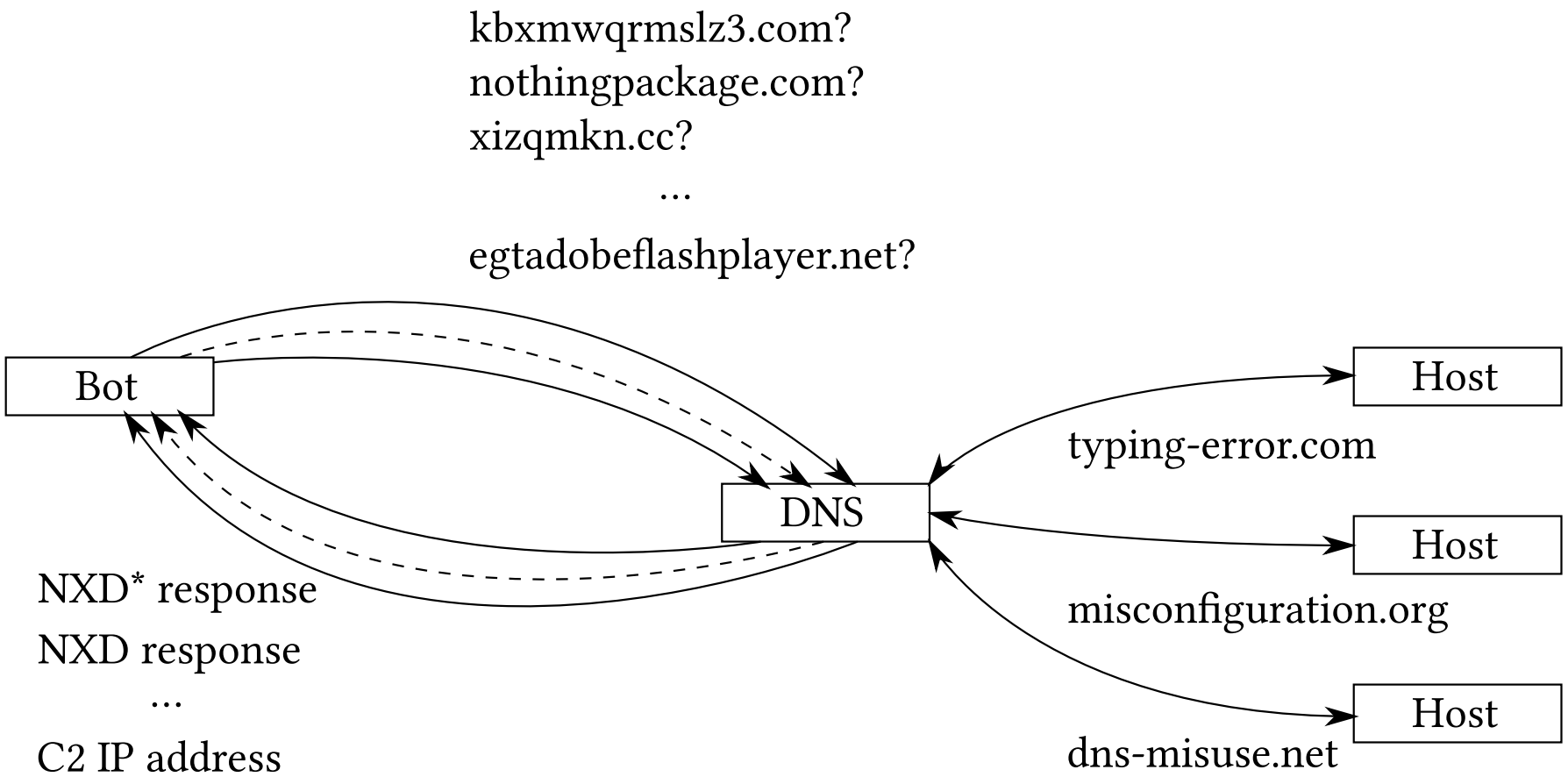

→ **Easy to block fixed domains or IP addresses**

# Domain Generation Algorithms (DGAs)




*non-existent domain (NXD)

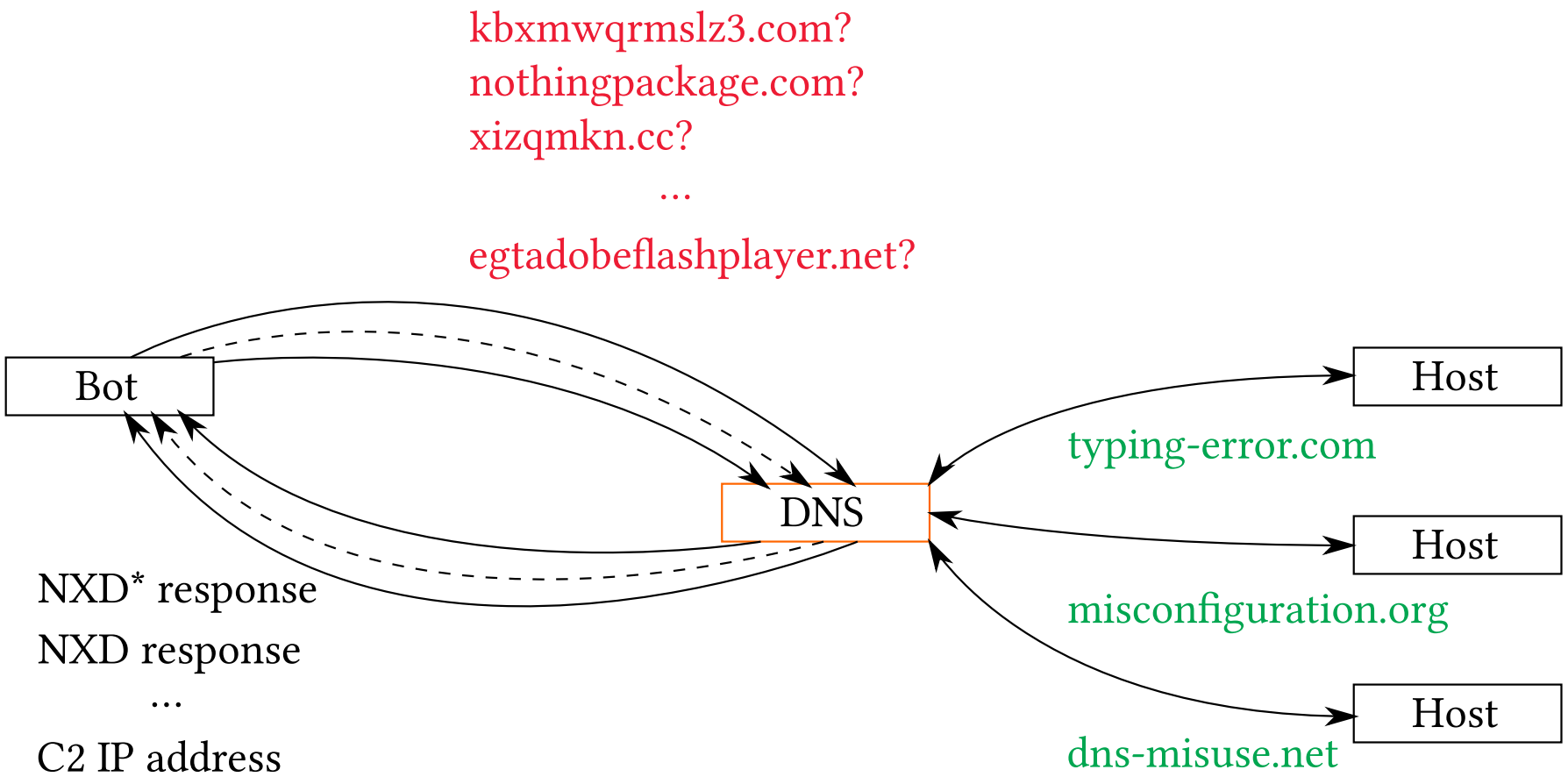
# Domain Generation Algorithms (DGAs)




*non-existent domain (NXD)

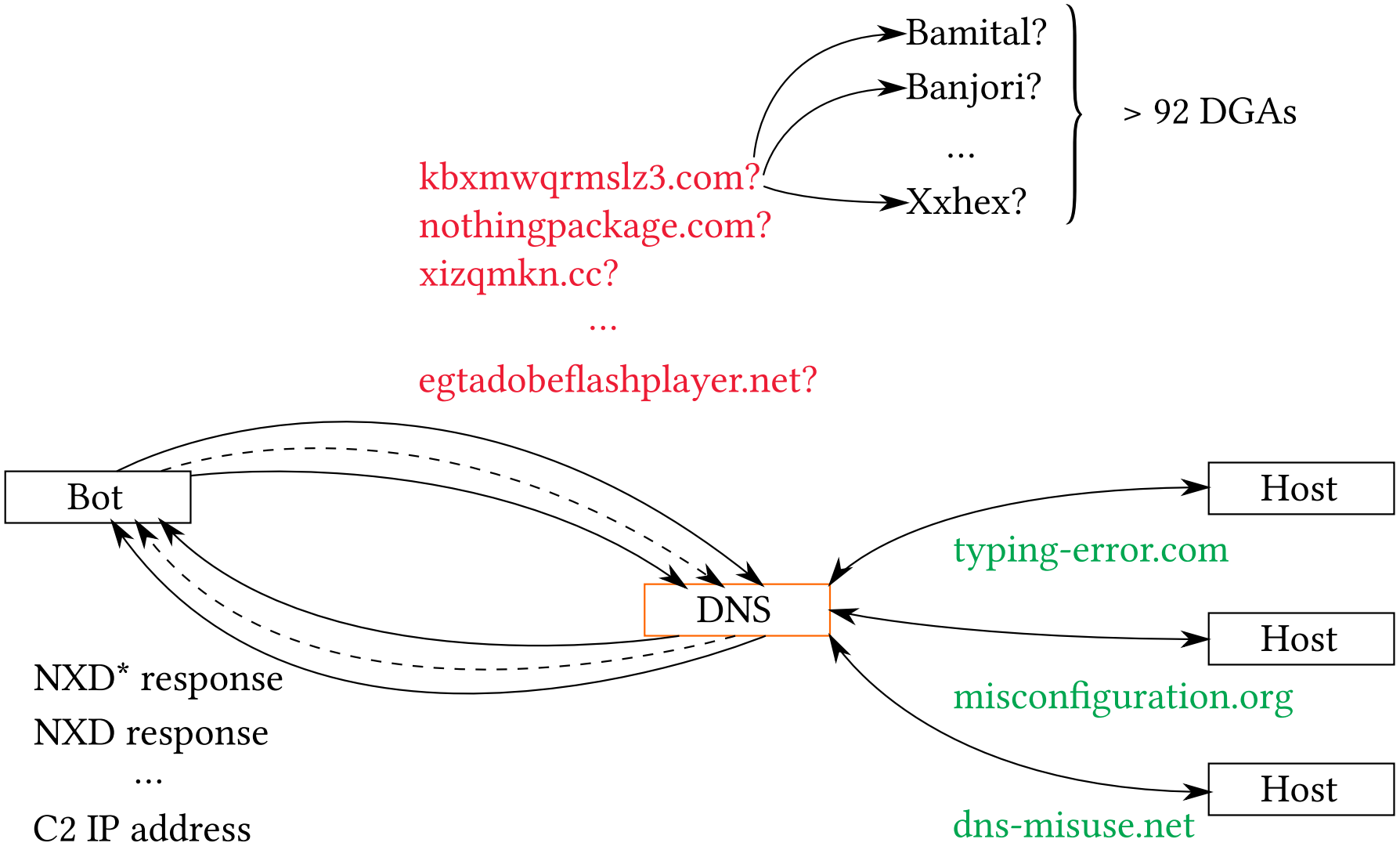
# Domain Generation Algorithms (DGAs)

## DGA binary classification




*non-existent domain (NXD)

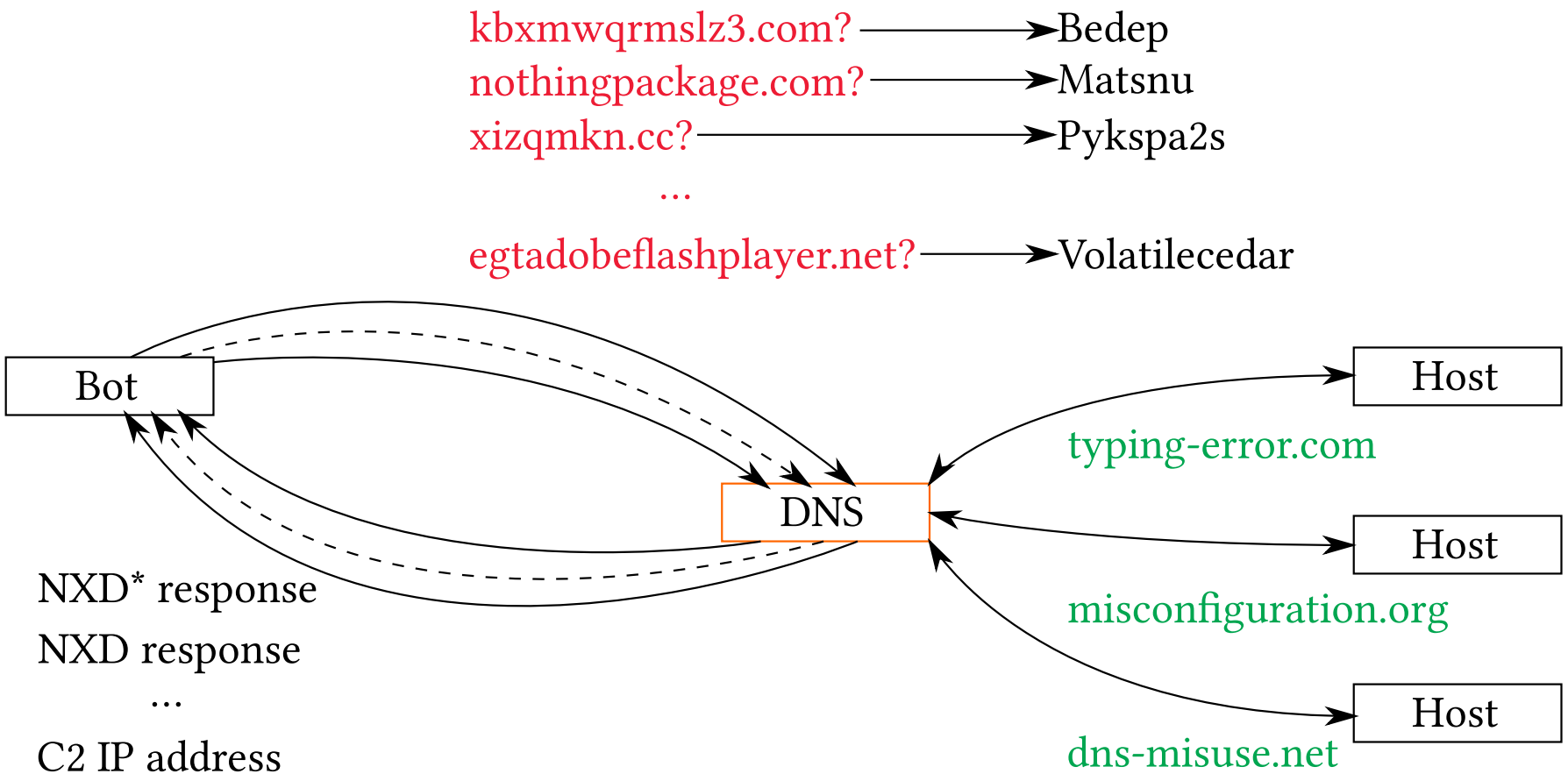
# Domain Generation Algorithms (DGAs)




*non-existent domain (NXD)

# Domain Generation Algorithms (DGAs)

## DGA multiclass classification




*non-existent domain (NXD)

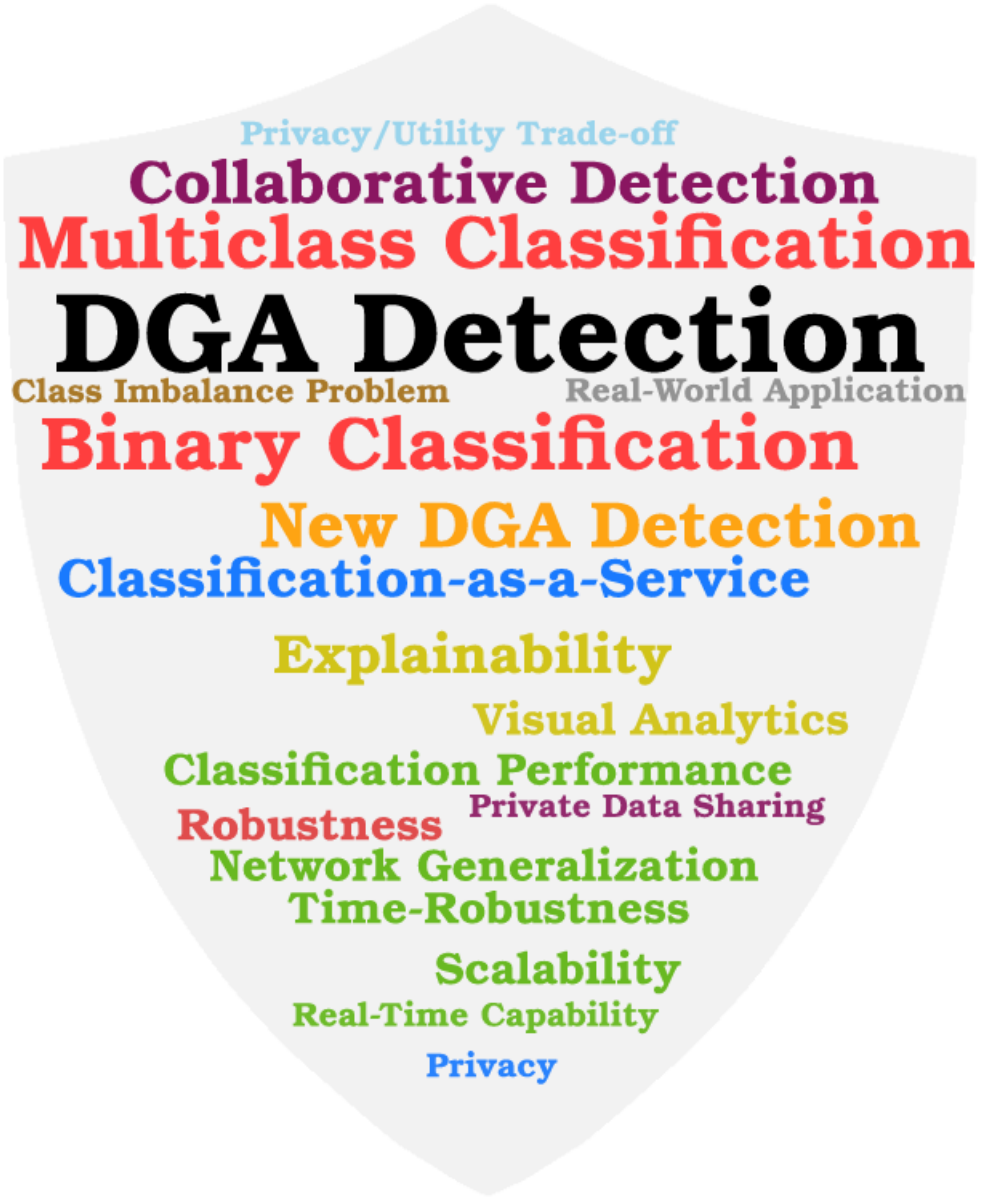
# Use-case DGA Detection in SAPPAN

Research driven approach

- 7 peer-reviewed accepted papers on DGA detection
- 1 paper currently under review

Real-world application of research results

- Classifiers are real-time capable & scalable
- Integration of research into existing Security Information and Event Management (SIEM) solutions



# DGA Binary Classification

Problem:

- How to separate benign from malicious domains?

B-ResNet: ResNet-based DGA binary classifier

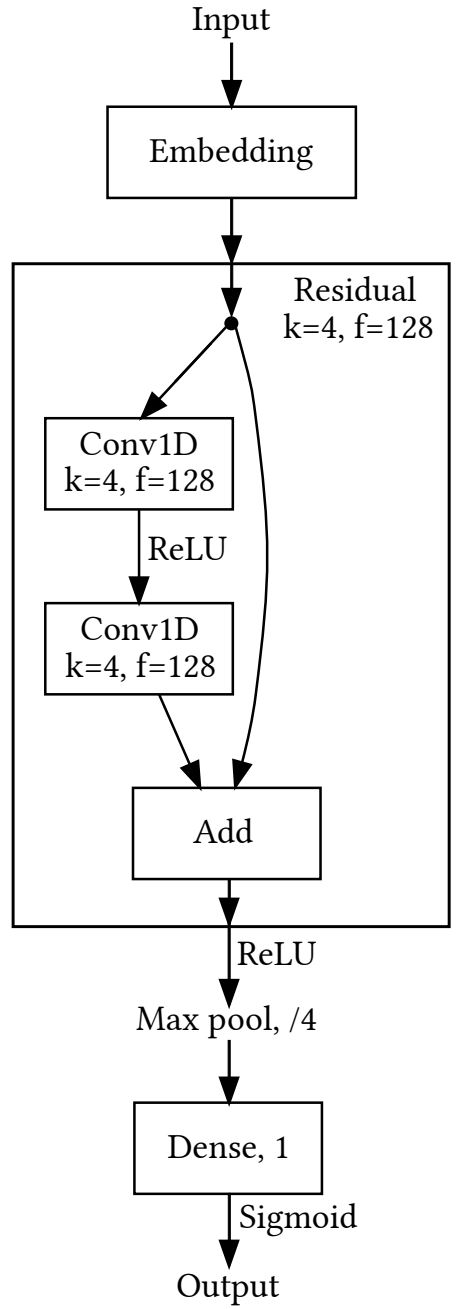
- Introduction of skip connections between convolutional layers

→ eases training and counteracts vanishing gradient problem

Comparative study with the state-of-the-art

- Reduction in false positive rate (FPR)
- B-ResNet generalizes well to different networks
  → Classification can be outsourced as a service
- B-ResNet is time-robust (even after 17 months)

→ Classifier is real-time capable



# DGA Multiclass Classification

Problem:

- How to attribute domains to either the benign class or to the DGA that generated the domain?

M-ResNet: ResNet-based DGA multiclass classifier

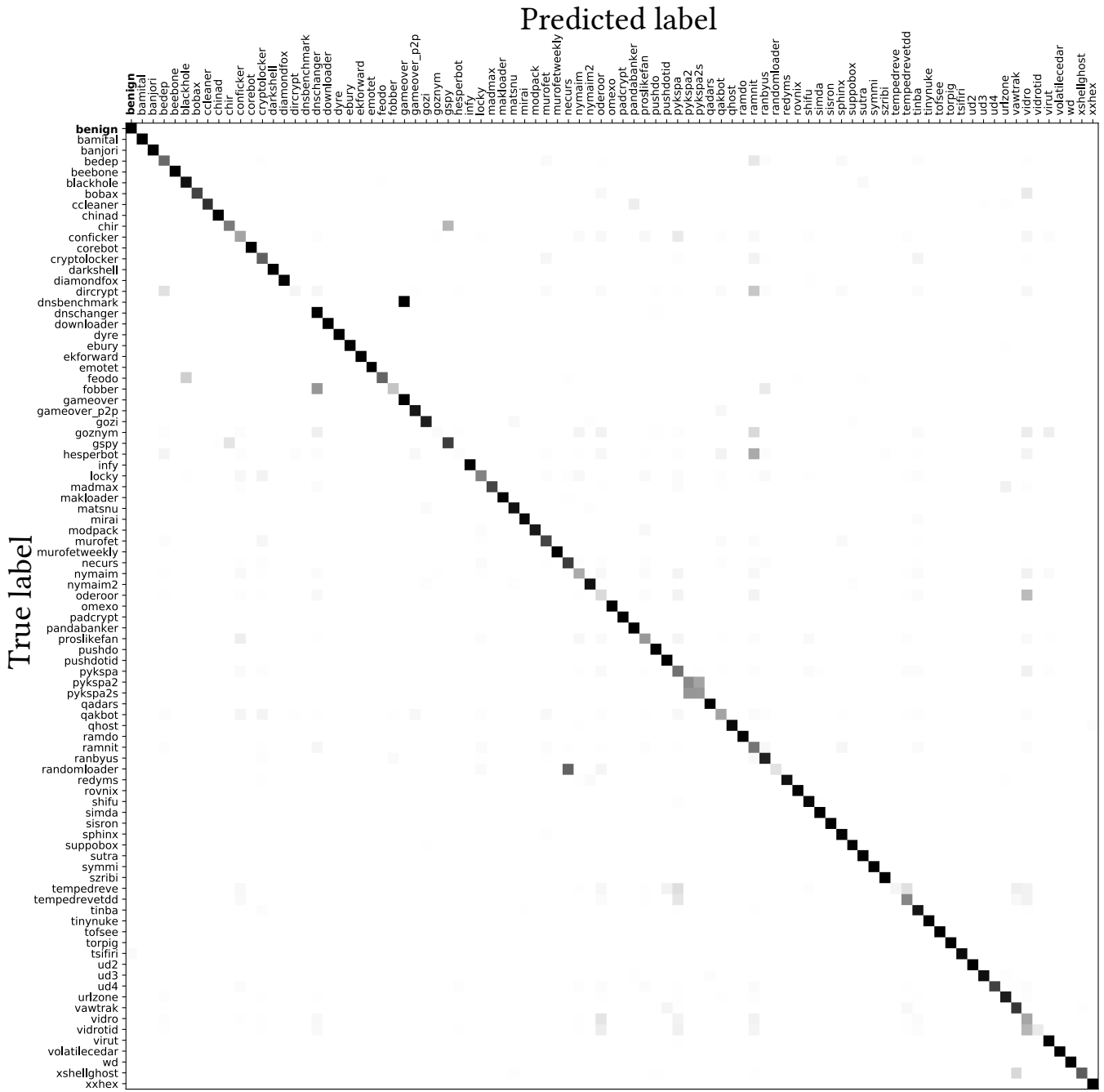
- Build up of 11 residual blocks

Comparative study with the state-of-the-art

- 30% less training time
- Improvement of over 5% in macro f1-score

→ Enables detection of several classes with high confidence

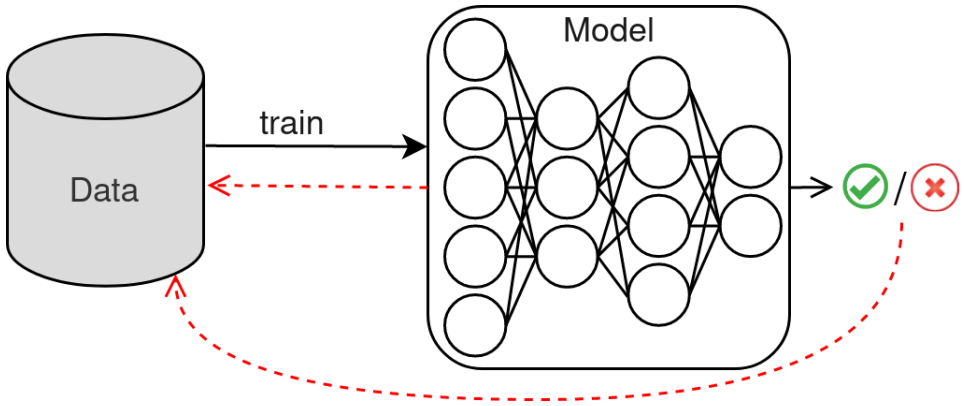
→ Classifier is real-time capable



# Class Imbalance Problem I

Problem:

- Performance of a classifier heavily depends on the used training data
- Sample distribution is heavily imbalanced
- Including underrepresented DGAs:
  - Effect on overall classification performance?
  - Ability to detect/attribute samples of underrepresented DGAs?



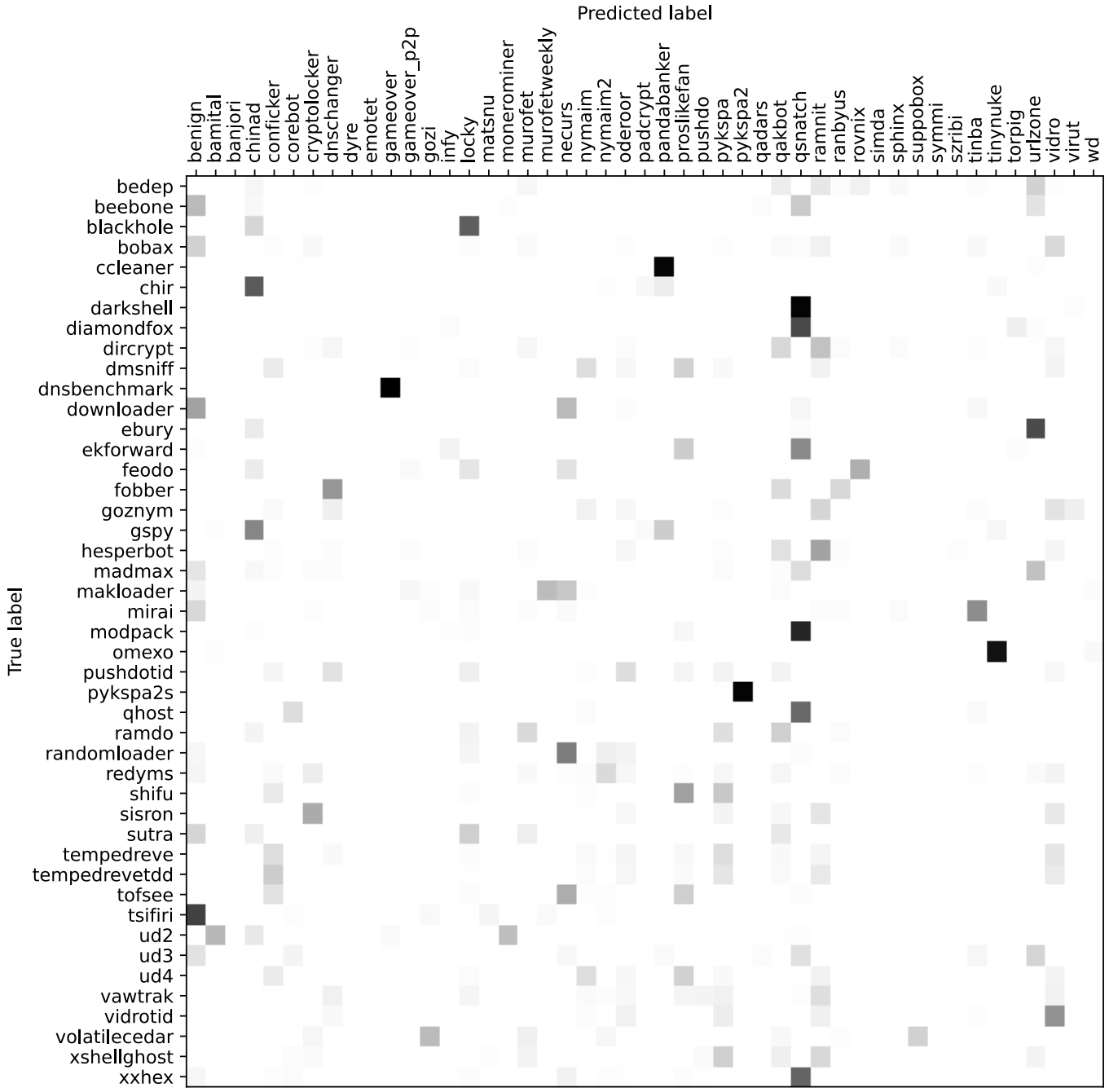

Comprehensive study

- Both classification tasks benefit from samples of weakly represented DGAs
  - Binary classification:
    - Improvement of over 10% in detection performance
  - Multiclass classification:
    - For 22/46 classes f1-score > 90%
    - For 11 classes f1-score > 99%
- No significant influence on the classification of well represented classes

# Class Imbalance Problem II - Out-Of-Distribution (OOD) Classification

Experiment:

- Train on samples of well represented group
- Classify samples of weakly represented group



# Class Imbalance Problem II - Out-Of-Distribution (OOD) Classification

Experiment:

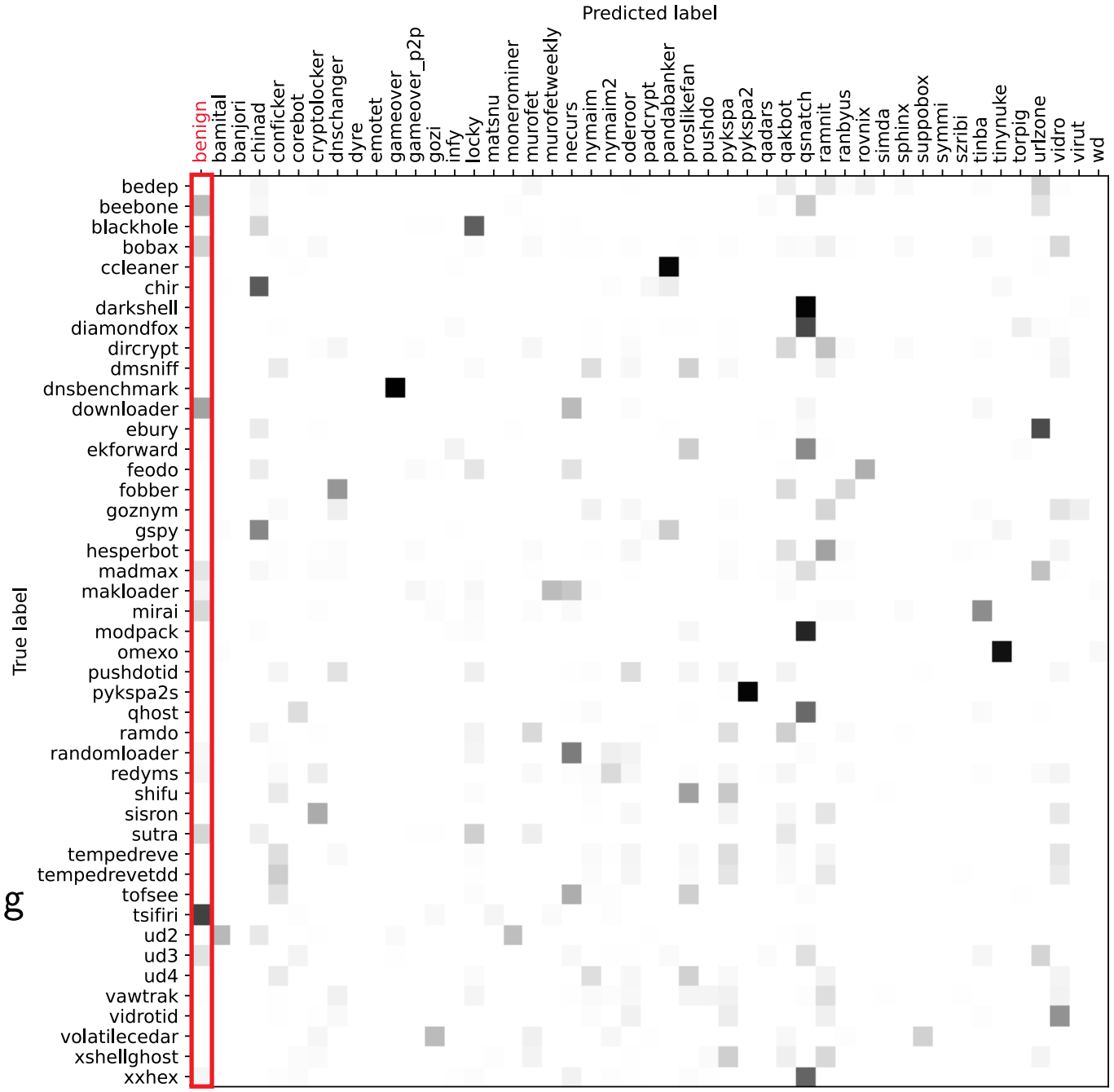
- Train on samples of well represented group
- Classify samples of weakly represented group

OOD samples are spread over malicious classes

Huge fraction is attributed to the benign class

→ These samples would be missed by an Intrusion Detection System (IDS)

→ Include weakly represented DGAs to the training



# Explainability I - Explainable AI Research

Problem:

- State-of-the-art deep learning classifiers behave like black boxes
- Difficult to evaluate their line of reasoning
  → Lack of confidence

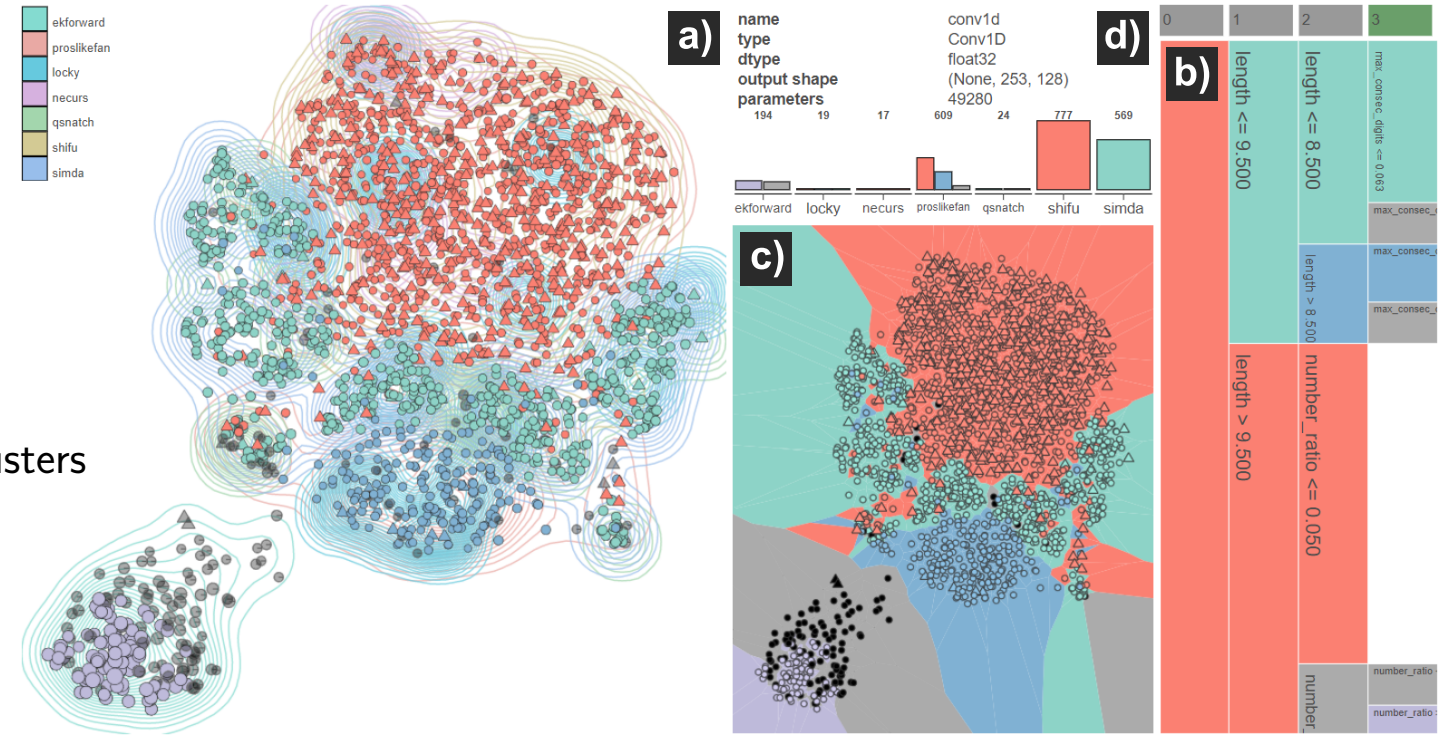
# Explainability I - Explainable AI Research

Problem:

- State-of-the-art deep learning classifiers behave like black boxes
- Difficult to evaluate their line of reasoning
  → Lack of confidence

Two approaches in SAPPAN:

1. Visual analytics system:
   - Provide understandable interpretations for predictions of deep learning classifiers
   - Cluster activations of a model's neurons
   - Leverage decision trees in order to explain clusters



# Explainability II - Feature-based DGA Detection

◤: EXPLAIN-OvR$_{Union}$ ◢: ResNet.MI

## 2. Feature-based classifier: EXPLAIN

- Random forest based classifier
- Competitive & context-less
- Explainable by design
  → Predictions are easier to interpret as they can be traced back to characteristics of used features

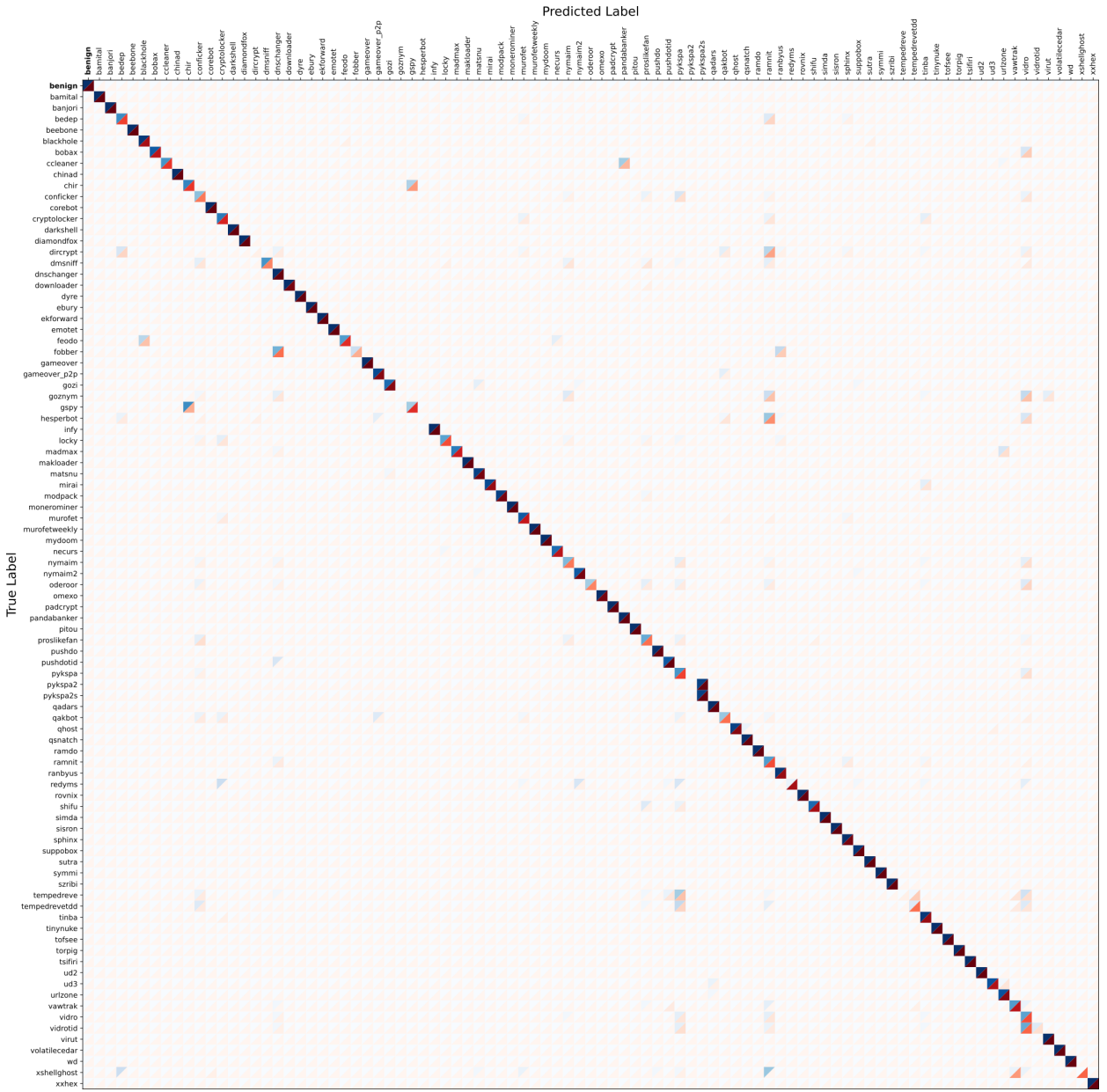
## Studied 136 context-less features

- 42 feature gathered from related work
- 94 features newly developed

## 136 Features divided in

- 64 linguistic: e.g. domain contains numbers?, vowel-ratio
- 17 structural: e.g. domain length
- 55 statistical: e.g. n-gram frequency distributions, entropy

→ Classifier is real-time capable



## Collaborative Machine Learning - Privacy-Preserving Intelligence Sharing

Problem:

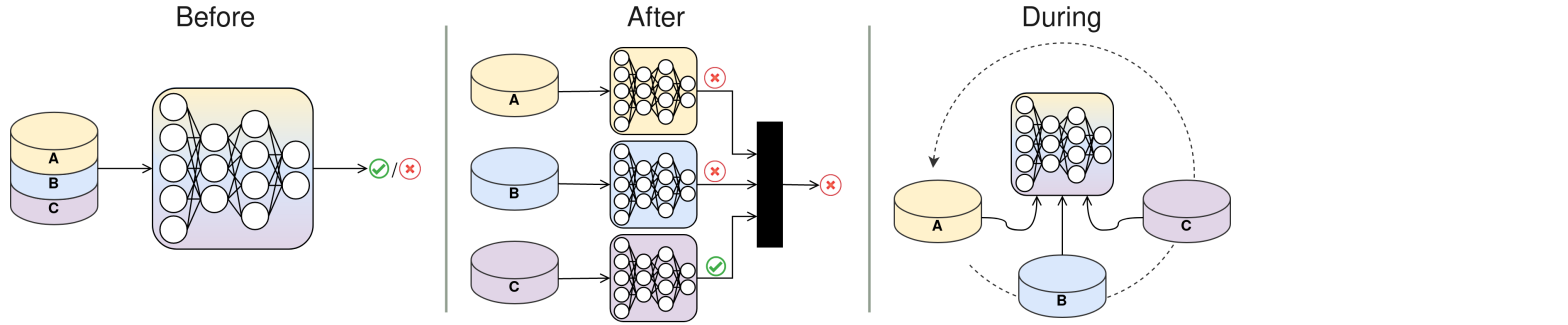
- How to improve detection by collaboration?
- Decision models are directly influenced by sensitive training data
- Models are susceptible to leak such sensitive information

# Collaborative Machine Learning - Privacy-Preserving Intelligence Sharing

Problem:

- How to improve detection by collaboration?
- Decision models are directly influenced by sensitive training data
- Models are susceptible to leak such sensitive information

Improve generalization and performance by sharing intelligence at different stages of model training:



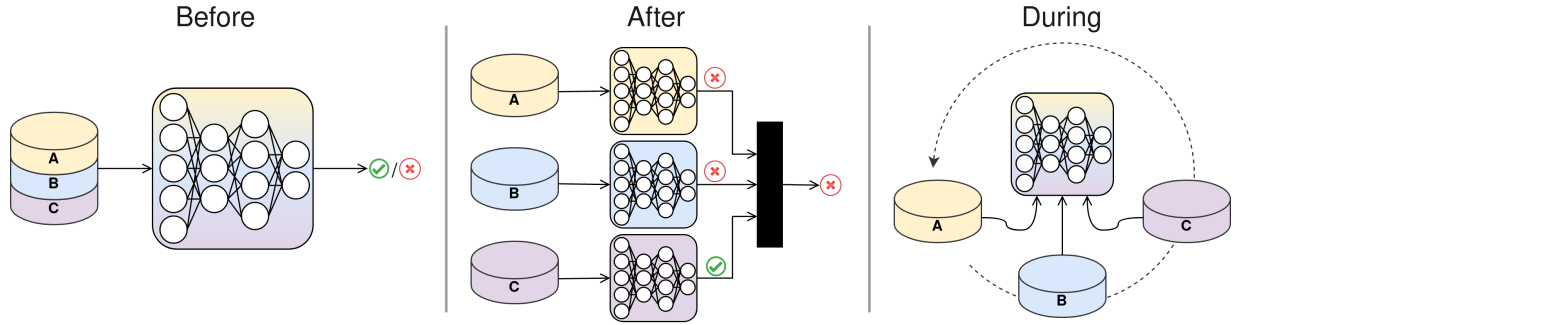

- Raw Data Sharing
- Ensemble Classification
- Feature Extractor Sharing
- Federated Learning
- Teacher-Student Learning

# Collaborative Machine Learning - Privacy-Preserving Intelligence Sharing

Problem:

- How to improve detection by collaboration?
- Decision models are directly influenced by sensitive training data
- Models are susceptible to leak such sensitive information

Improve generalization and performance by sharing intelligence at different stages of model training:




- Raw Data Sharing
- Ensemble Classification
- Feature Extractor Sharing
- Federated Learning
- Teacher-Student Learning

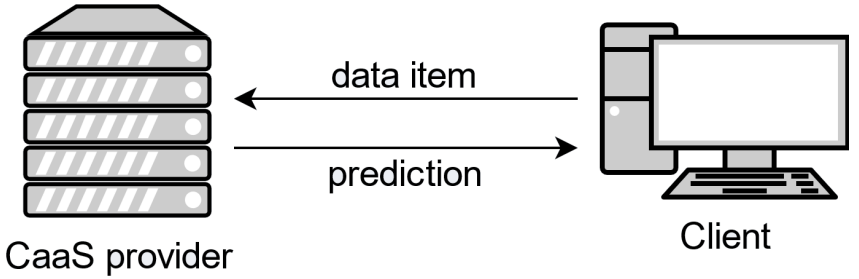
Our empirical study shows Feature Extractor Sharing and Federated Learning perform best:

- Significant reduction of false positive rate (FPR), up to 50% compared to single-party
- Reduction rate of FPR correlates with increasing number of parties
- Preliminary privacy-utility trade-off study

# Privacy-Preserving Classification as a Service (CaaS)

Problem:

- Real-world training data is mandatory for well performing classifiers
- What about resource constrained devices?
- Domain names / trained models may contain privacy-critical information



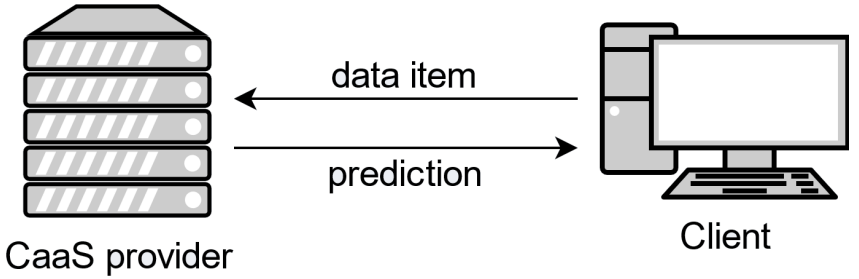

Naive application of privacy-preserving ML frameworks to existing DGA detection classifiers

→ Single inference can cost additional: 13 min inference latency, 234 GB communication

# Privacy-Preserving Classification as a Service (CaaS)

Problem:

- Real-world training data is mandatory for well performing classifiers
- What about resource constrained devices?
- Domain names / trained models may contain privacy-critical information




Naive application of privacy-preserving ML frameworks to existing DGA detection classifiers

→ Single inference can cost additional: 13 min inference latency, 234 GB communication

Comprehensive study & proposed model simplifications:

- Reduction in inference latency of up to 95%
- Reduction in communication complexity of up to 97%
- Accuracy penalty of less than 0.17%

→ Still, future work is required to make privacy-preserving CaaS feasible!

# DGA Detection - Current Research & Future Work

Robustness

- NX-classifiers more robust against adversarial attacks
- Usage of adversarial machine learning to improve robustness

# DGA Detection - Current Research & Future Work

## Robustness

- NX-classifiers more robust against adversarial attacks
- Usage of adversarial machine learning to improve robustness

## New DGA detection

- Real-world experiment: 6 unknown DGAs, 1 unknown Bamital seed
- Adaptive new DGA detection system

```
manipulation-want-date.pw
refers-spare-criticism.pp.ua
fashioned-achieve-disable.pro
```

**(a) Unknown DGA 1**

```
go2mysuite.eu
citrixgo2mypc.co.uk
gotomobileaccess.com
```

**(b) Unknown DGA 2**

```
dv4050fc.co.ir
thrsssk05.co.ir
thrl0pg13.co.ir
```

**(c) Unknown DGA 3**

```
www.c75ff6bd.com
www.94e47d25.com
www.41019163.com
```

**(d) Unknown DGA 4**

```
2b4b1d67-b38a-40c1-ba3e-af73245d7b14.com
86a94dd8-5724-4b9a-8a7a-bea8733f7e60.com
adcb3f60-d260-478a-99f2-ac24eea1de16.com
```

**(e) Unknown DGA 5**

```
egbva1b5pmgh7fb.jmrbqoa6i67zdlrwhj.com
27422j8tqot.8chcu-tza86fxaz-df70y9-t0o.com
bt-7hb7k0aqyyr-61d8o5d.dg08rz6qobme421f.com
```

**(f) Unknown DGA 6**

```
02836ae5435c57300fc95bf13e9ba7bb.info
073fcdb286615c7a6ac348f9a1ab0250.info
08211a534fad3885624a92573cc2af44.info
```

**(g) Unknown seed of *Bamital***

# DGA Detection - Current Research & Future Work

## Robustness

- NX-classifiers more robust against adversarial attacks
- Usage of adversarial machine learning to improve robustness

## New DGA detection

- Real-world experiment: 6 unknown DGAs, 1 unknown Bamital seed
- Adaptive new DGA detection system

→ **Combining all research to a single detection system**

```
manipulation-want-date.pw
refers-spare-criticism.pp.ua
fashioned-achieve-disable.pro
```

**(a) Unknown DGA 1**

```
go2mysuite.eu
citrixgo2mypc.co.uk
gotomobileaccess.com
```

**(b) Unknown DGA 2**

```
dv4050fc.co.ir
thrsssk05.co.ir
thrl0pg13.co.ir
```

**(c) Unknown DGA 3**

```
www.c75ff6bd.com
www.94e47d25.com
www.41019163.com
```

**(d) Unknown DGA 4**

```
2b4b1d67-b38a-40c1-ba3e-af73245d7b14.com
86a94dd8-5724-4b9a-8a7a-bea8733f7e60.com
adcb3f60-d260-478a-99f2-ac24eea1de16.com
```

**(e) Unknown DGA 5**

```
egbva1b5pmgh7fb.jmrbqoa6i67zdlrwhj.com
27422j8tqot.8chcu-tza86fxaz-df70y9-t0o.com
bt-7hb7k0aqyyr-61d8o5d.dg08rz6qobme421f.com
```

**(f) Unknown DGA 6**

```
02836ae5435c57300fc95bf13e9ba7bb.info
073fcdb286615c7a6ac348f9a1ab0250.info
08211a534fad3885624a92573cc2af44.info
```

**(g) Unknown seed of *Bamital***

# Impact of SAPPAN Innovations

# Impact of SAPPAN Innovations I

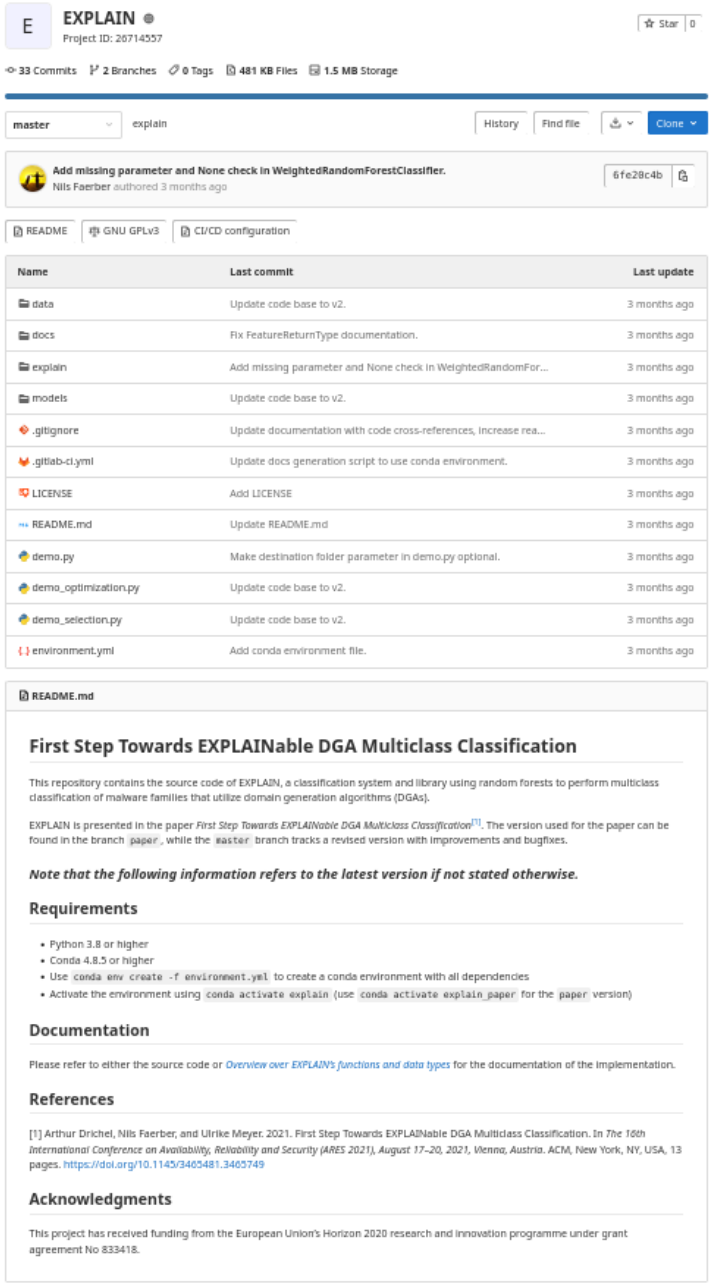
Research impact

- Improved state-of-the-art in various aspects
- 7 peer-reviewed accepted papers on DGA detection
- 1 paper currently under review

Open source software

- Binary & multiclass ResNet-based DGA models
- EXPLAIN: Feature-based multiclass classifier (https://gitlab.com/rwth-itsec/explain)

→ Classifiers are real-world applicable



E **EXPLAIN**
Project ID: 26714557
Star 0

33 Commits | 2 Branches | 0 Tags | 481 KB Files | 1.5 MB Storage

master | explain | History | Find file | Clone

Add missing parameter and None check in WeightedRandomForestClassifier.
Nils Faerber authored 3 months ago | 6fe28c4b

README | GNU GPLv3 | CI/CD configuration

| Name | Last commit | Last update |
|---|---|---|
| data | Update code base to v2. | 3 months ago |
| docs | Fix FeatureReturnType documentation. | 3 months ago |
| explain | Add missing parameter and None check in WeightedRandomFor... | 3 months ago |
| models | Update code base to v2. | 3 months ago |
| .gitignore | Update documentation with code cross-references, increase rea... | 3 months ago |
| .gitlab-ci.yml | Update docs generation script to use conda environment. | 3 months ago |
| LICENSE | Add LICENSE | 3 months ago |
| README.md | Update README.md | 3 months ago |
| demo.py | Make destination folder parameter in demo.py optional. | 3 months ago |
| demo_optimization.py | Update code base to v2. | 3 months ago |
| demo_selection.py | Update code base to v2. | 3 months ago |
| environment.yml | Add conda environment file. | 3 months ago |

README.md

**First Step Towards EXPLAINable DGA Multiclass Classification**

This repository contains the source code of EXPLAIN, a classification system and library using random forests to perform multiclass classification of malware families that utilize domain generation algorithms (DGAs).

EXPLAIN is presented in the paper *First Step Towards EXPLAINable DGA Multiclass Classification*[1]. The version used for the paper can be found in the branch `paper`, while the `master` branch tracks a revised version with improvements and bugfixes.

***Note that the following information refers to the latest version if not stated otherwise.***

**Requirements**

- Python 3.8 or higher
- Conda 4.8.5 or higher
- Use `conda env create -f environment.yml` to create a conda environment with all dependencies
- Activate the environment using `conda activate explain` (use `conda activate explain_paper` for the `paper` version)

**Documentation**

Please refer to either the source code or *Overview over EXPLAIN's functions and data types* for the documentation of the implementation.

**References**

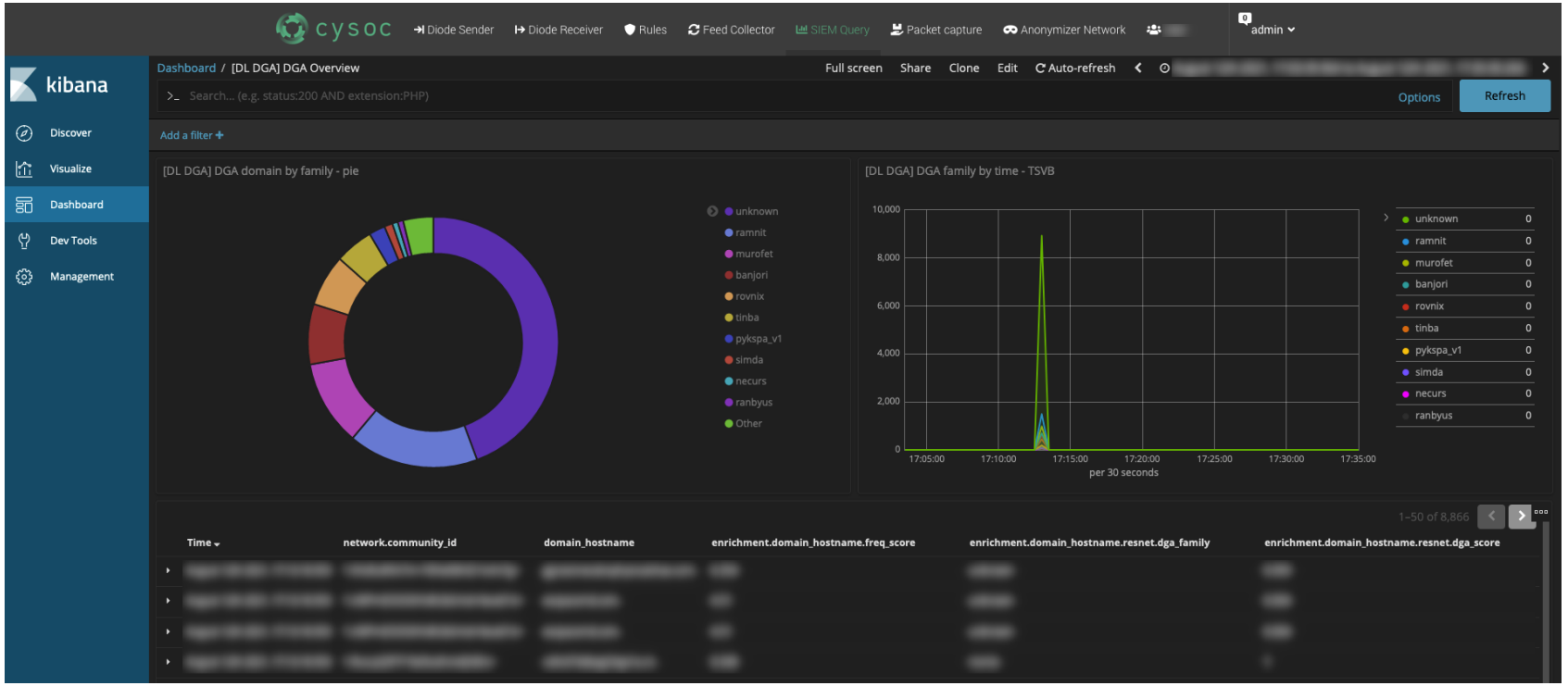
[1] Arthur Drichel, Nils Faerber, and Ulrike Meyer. 2021. First Step Towards EXPLAINable DGA Multiclass Classification. In *The 16th International Conference on Availability, Reliability and Security (ARES 2021), August 17–20, 2021, Vienna, Austria.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3465481.3465749

**Acknowledgments**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418.

# Impact of SAPPAN Innovations II

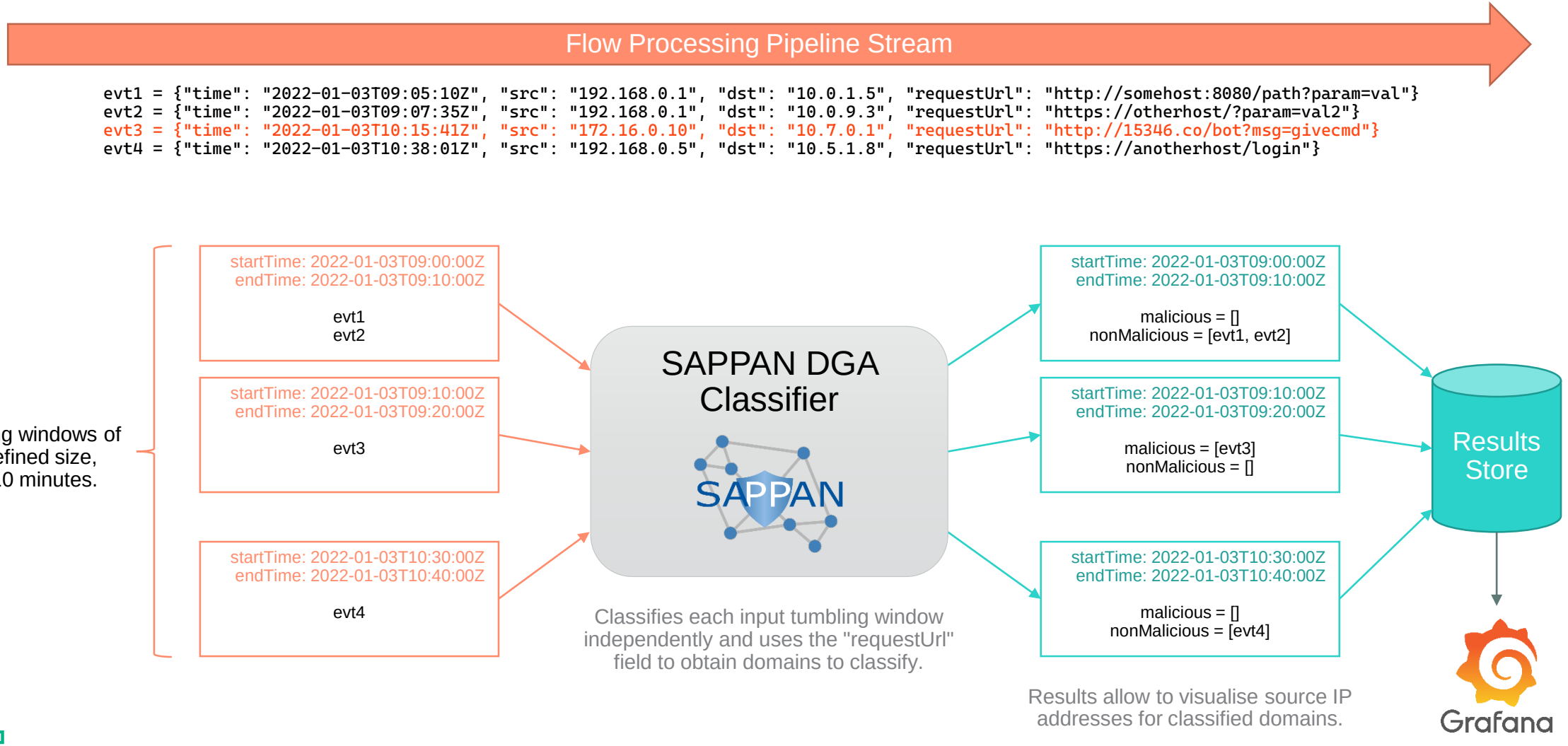
Integration of research into existing SIEM solutions:




Facilitating the work of Security Operation Center (SOC) analysts

- Improvement of detection performance
- Reduction of false positives
- Providing explanations for predictions

# Solution Design: SAPPAN DGA Classifier



# Thank You
# For Your Attention

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418.

SAPPAN

*SHARING AND AUTOMATION FOR PRIVACY PRESERVING ATTACK NEUTRALIZATION*

Co-funded by the Horizon 2020 programme of the European Union

IT SEC Research Group IT-Security | RWTH AACHEN UNIVERSITY