DGA-based Malware Detection and Classification using P4-Programmable Data Plane Networks

Ali AlSabeh, Jorge Crichigno

Integrated Information Technology Department, University of South Carolina, Columbia, South Carolina

Abstract

- Domain Generation Algorithms (DGAs) are the de facto dynamic Command and Control (C2) communication method used by malware, including botnets, ransomware, and many others
- Approaches that analyze the behavior of the network traffic (context-aware) fingerprint DGAs are slow since they typically analyze batches of traffic offline
- Alternatively, approaches that analyze the domain name (context-less) achieve high accuracy with ML models but might create a bottleneck as they use a general-purpose CPU/GPU
- This project proposes a novel scheme that uses the high-speed P4-Programmable Data Plane (PDP) switches to detect and classify DGAs
- The switch (Intel Tofino ASIC) performs Deep Packet Inspection (DPI) to analyze domain names and detect DGAs entirely within the data plane
- The switch sends the collected features to an intelligent control plane for further classification of the DGA family
- Evaluation results on 50 DGAs families and benign traffic from a campus network show a high detection and classification accuracy within a few packets

Related Work

- DGA binary detection (DGA or benign):
  - [1, 2] use NetFlow and an SDN controller to collect context-aware features
  - [3] uses ML models on context-aware and context-less features on batches of DNS traffic
  - [4-7] use machine learning trained on features of the domain name (statistical, structural, linguistic, etc.)
- DGA multiclass classification (ransomware, bot, Trojan, etc.):
  - [8] and [9] extract numerous features from a domain name to classify DGAs

Approach	DGA multiclass.	Context-less	Context-aware	F.E. latency
[1]			✓	$minutes$ ●
[2]			✓	$seconds$ ●
EXPOSURE [3]		✓	✓	$minutes$ ●
FANCI [4]		✓		ms ●
ANCS [5]		✓		ms ●
[6]		✓		ms ●
[7]		✓		ms ●
EXPLAIN [8]	✓	✓		100's μs ●
[9]	✓	✓		ms ●
Our approach	✓	✓	✓	**2-3** μs ⋆

⋆ : ASIC processing ● : CPU/GPU processing

[1] M. Grill, et al., "Detecting DGA Malware using NetFlow," in 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 1304–1309, IEEE, 2015.

[2] Y. Iuchi, et al., "Detection and Blocking of DGA-Based Bot Infected Computers by Monitoring NXDOMAIN Responses," in 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), pp. 82–87, IEEE, 2020.

[3] L. Bilge, et al., "Exposure: A passive DNS Analysis Service to Detect and Report Malicious Domains," ACM Transactions on Information and System Security (TISSEC), vol. 16, no. 4, pp. 1–28, 2014.

[4] S. Schuppen, et al., "FANCI: Feature-based Automated NXDomain Classification and Intelligence," in 27th USENIX Security Symposium (USENIX Security 18), pp. 1165– 1181, 2018.

[5] L. Fang, X. et al., "ANCS: Automatic NXDomain Classification System Based on Incremental Fuzzy Rough Sets Machine Learning," IEEE Transactions on Fuzzy Systems, vol. 29, no. 4, pp. 742–756, 2020.

[6] K. Highnam, et al., "Real-time Detection of Dictionary DGA Network Traffic Using Deep Learning," SN Computer Science, vol. 2, no. 2, pp. 1–17, 2021.

[7] K. Highnam, et al., "Real-time Detection of Dictionary DGA Network Traffic Using Deep Learning," SN Computer Science, vol. 2, no. 2, pp. 1–17, 2021.

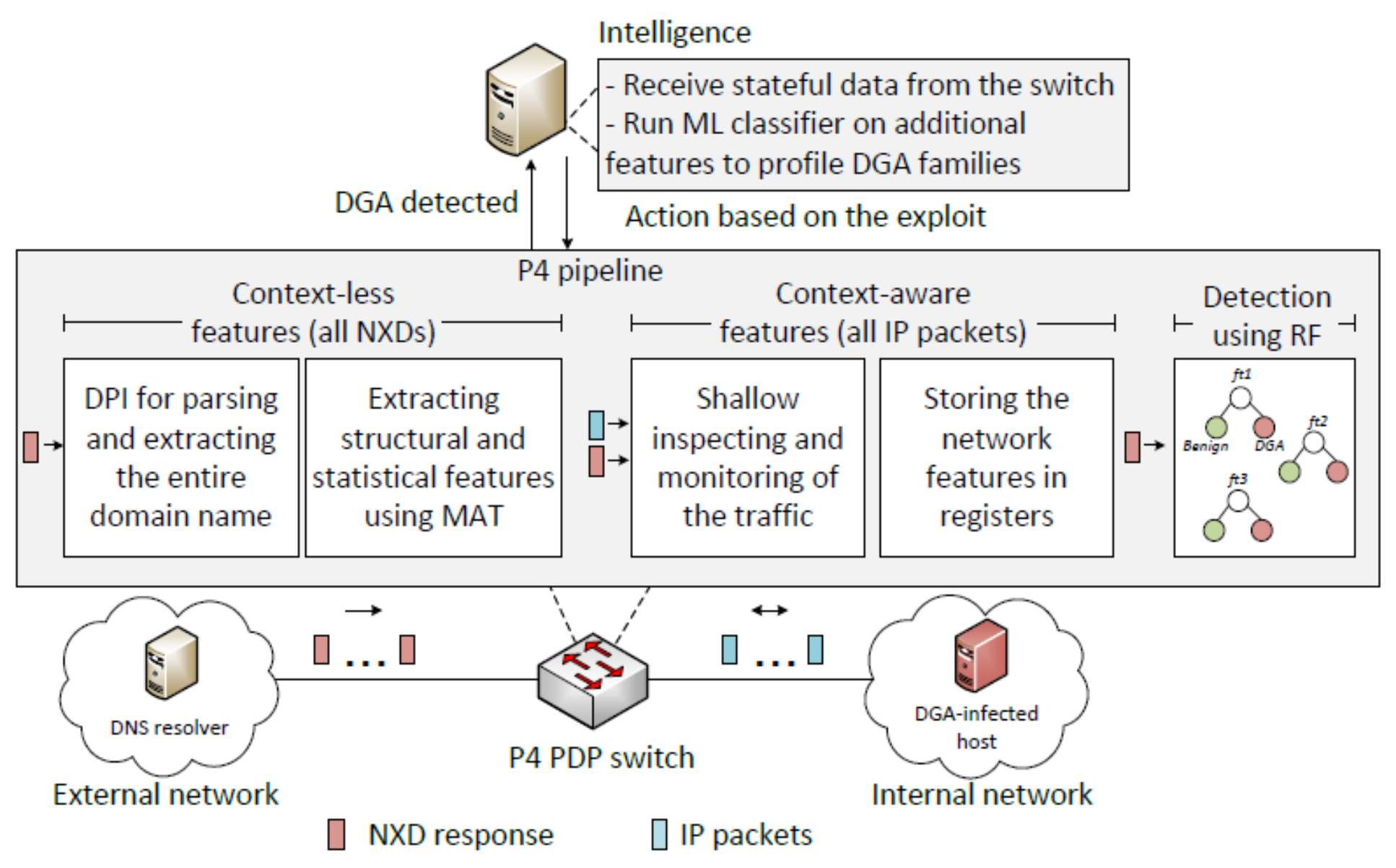
[8] A. Drichel, et al., "First Step Towards Explainable DGA Multiclass Classification," in The 16th International Conference on Availability, Reliability and Security, pp. 1–13, 2021.

[9] T. A. Tuan, et al., "On Detecting and Classifying DGA Botnets and their Families," Computers & Security, vol. 113, p. 102549, 2022.

Methodology

The high-level architecture of the proposed approach is shown in the figure below and can be summarized in the following steps:

- The P4 PDP switch collects and stores the context-aware features of the hosts
- When an NXD response is received, the switch performs DPI on the domain name to extract its context-less features
- The P4 PDP switch implements a Random Forest (RF) classifier to detect DGAs based on few context-less and context-aware features
- If a DGA is detected with a high confidence, the P4 PDP switch sends the collected features to the control plane
- The control plane runs the intelligence to classify the DGA family and initiate the appropriate incident response
  - The control plane implements a more accurate classifier that operates on all the features collected by the switch


Intelligence
- Receive stateful data from the switch
- Run ML classifier on additional features to profile DGA families
DGA detected
Action based on the exploit
P4 pipeline
Context-less features (all NXDs)
Context-aware features (all IP packets)
Detection using RF
DPI for parsing and extracting the entire domain name
Extracting structural and statistical features using MAT
Shallow inspecting and monitoring of the traffic
Storing the network features in registers
ft1
Benign
DGA
ft2
ft3
DNS resolver
External network
P4 PDP switch
DGA-infected host
Internal network
NXD response
IP packets

Context-aware features:

- For each host in the network, the following features are collected by the P4 PDP switch
  - Number of unique IP addresses contacted
  - Number of DNS requests made
  - Inter-arrival Time (IAT) between subsequent NXD responses
- Such features are collected in the data plane without involving the control plane (until a DGA is detected)

Context-less features:

- It computes the bigram of the domain name; a bigram model may suffice to predict whether a domain name is a legitimate human readable domain
- Other domain name attributes include length of the domain name and number of subdomains
- For each NXD response received, the data plane extracts the following features from the domain name

$$score\ (d) = \sum_{\forall\ subdomain\ s\ \in\ d} \left(\sum_{\forall\ bigram\ b\ \in\ s} f_s^b \right)$$

Where f_s^b is the frequency of the bigram b in the subdomain *s*

- **Example**: bigrams of "google" are: "$g", "go", "oo", "og", "gl", "le", "e$"

Acknowledgement

- This work was supported by the Office of Naval Research (ONR), Grant N00014-23-1-2245.

Evaluation Results

Dataset:

- Hundreds of GB of malware samples from cyber security websites were crawled
- Each sample was instrumented in an isolated environment to capture its network traffic behavior
- The resulting dataset includes 1,311 samples containing 50 DGA families
  - We publish the dataset online: https://github.com/aalsabeh/P4-DGA-Multiclass
- CTU-13 dataset was used for benign samples to evaluate the DGA detection module
  - Link to the dataset: https://www.stratosphereips.org/datasets-overview

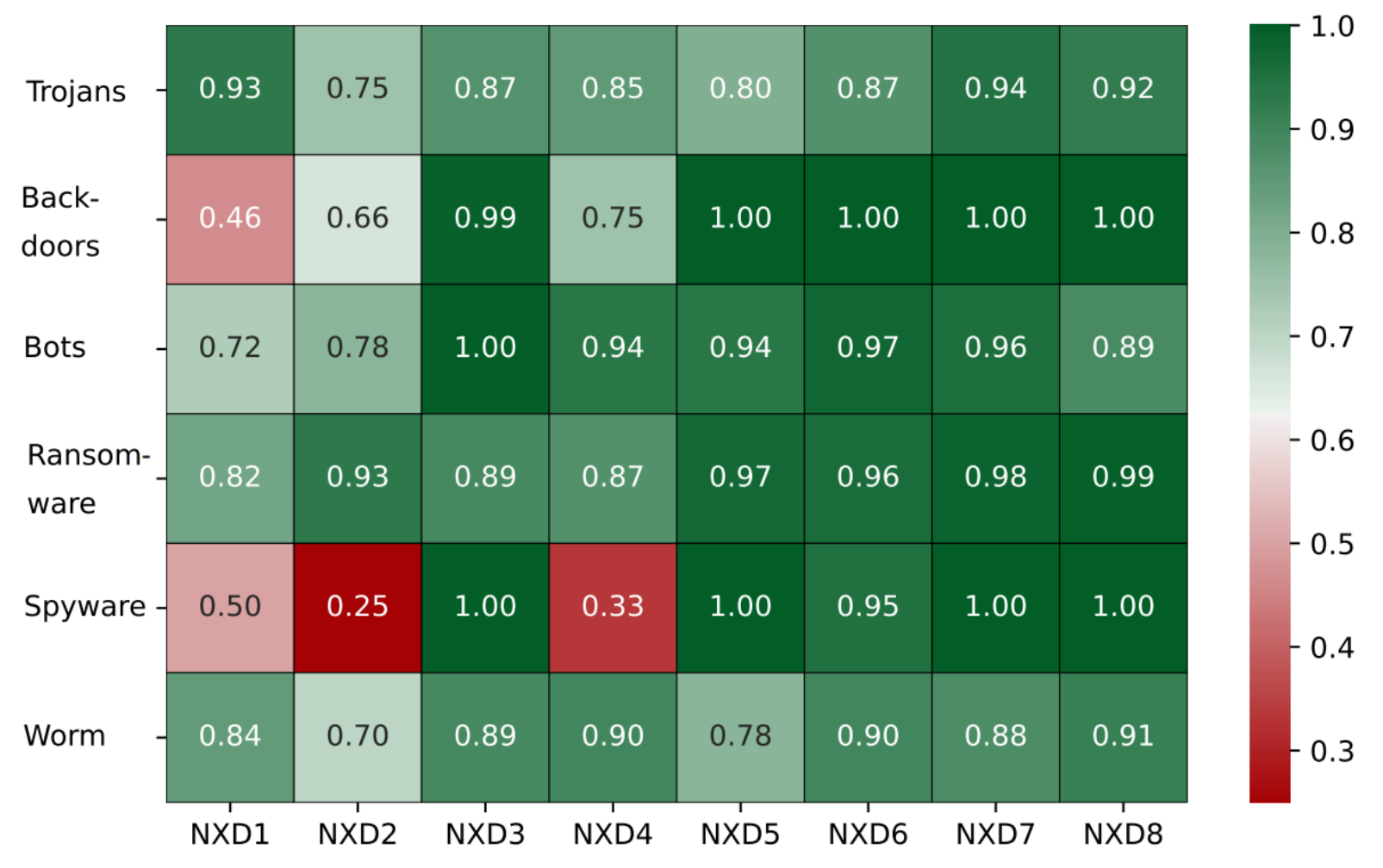
Evaluation results

- The DGA detection module in the data plane uses the features: unique IP addresses contacted, number of DNS requests made, and the bigram frequency value of the domain for each NXD received
- The RF classifier can detect DGAs with an accuracy ranging between 97-98% starting from the second NXD domain
- The DGA classification is a more challenging task, thus, the classifier is more complex and had to be implemented in the control plane. The classification results are shown below

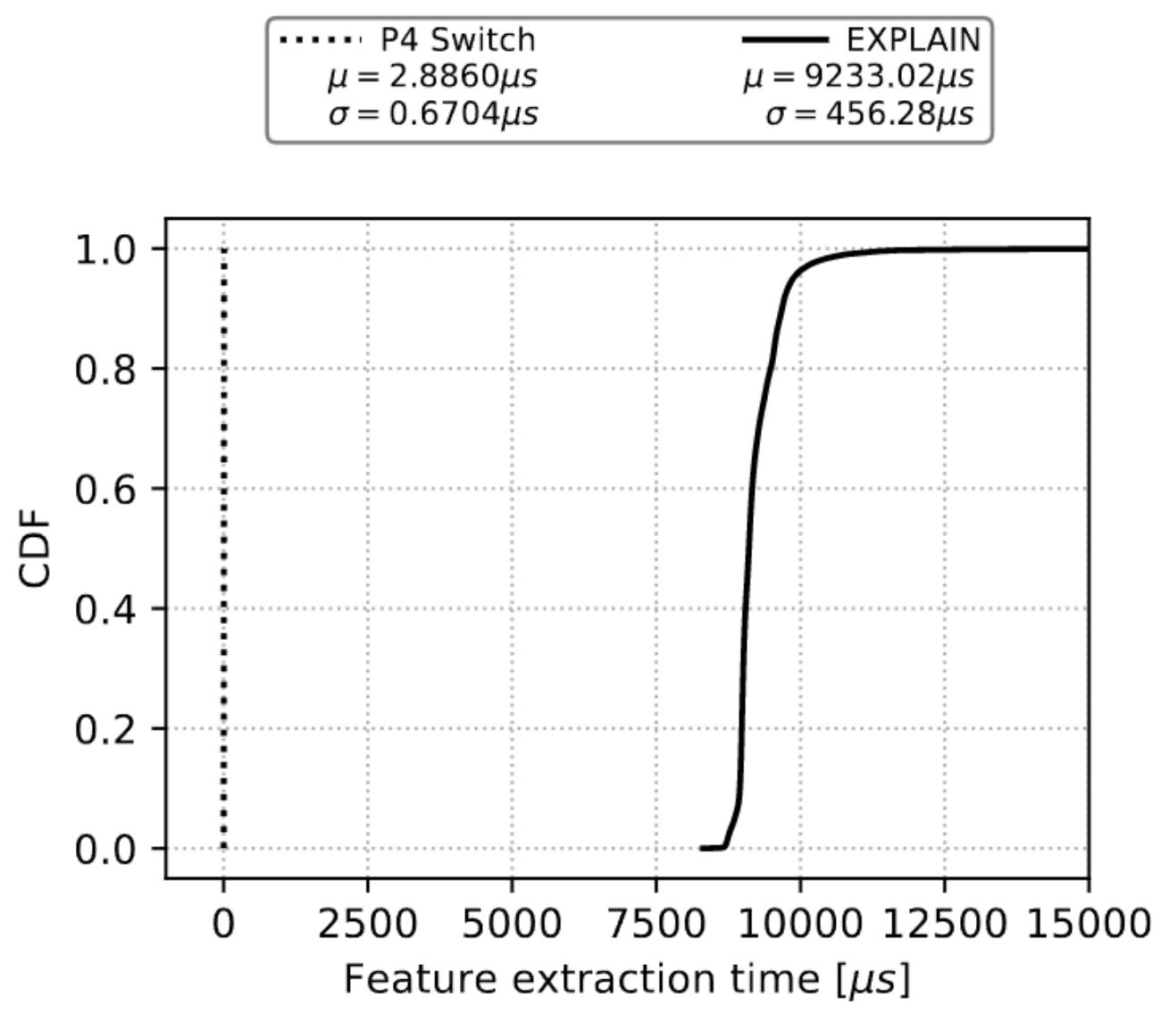
The RF model performed best with an Accuracy (Acc) starting at 92% from the first NXD response received and reaching 95% by the 8th NXD response

NXD count	RF			SVM			MLP			LR			GNB		
	Acc	F1	Prec	Acc	F1	Prec	Acc	F1	Prec	Acc	F1	Prec	Acc	F1	Prec
NXD 1	0.923	0.907	0.902	0.872	0.856	0.847	0.87	0.843	0.829	0.716	0.679	0.667	0.726	0.688	0.688
NXD 2	0.951	0.943	0.943	0.899	0.893	0.893	0.904	0.897	0.9	0.76	0.741	0.747	0.727	0.701	0.707
NXD 3	0.964	0.958	0.964	0.918	0.913	0.914	0.924	0.914	0.912	0.767	0.74	0.743	0.649	0.668	0.732
NXD 4	0.966	0.961	0.963	0.906	0.905	0.912	0.916	0.909	0.915	0.79	0.765	0.758	0.633	0.635	0.692
NXD 5	0.97	0.966	0.967	0.915	0.91	0.911	0.919	0.91	0.907	0.77	0.735	0.746	0.604	0.615	0.689
NXD 6	0.975	0.972	0.973	0.914	0.911	0.912	0.922	0.915	0.918	0.794	0.767	0.783	0.617	0.627	0.716
NXD 7	0.977	0.976	0.979	0.92	0.915	0.915	0.929	0.924	0.93	0.799	0.771	0.78	0.61	0.613	0.714
NXD 8	0.98	0.979	0.981	0.917	0.912	0.914	0.93	0.923	0.921	0.764	0.73	0.735	0.631	0.618	0.65

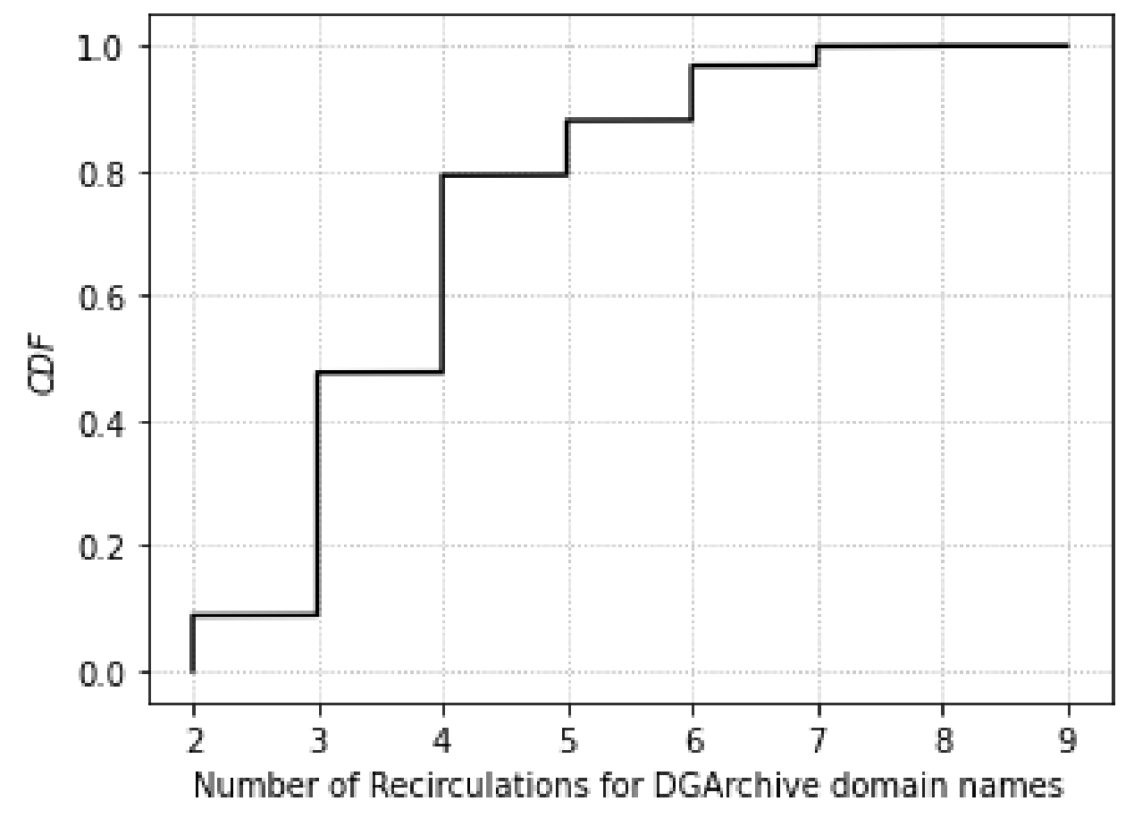
Performance of the proposed approach amid varying NXD responses on a subset of samples grouped by their attack category



Feature extraction time of our work and EXPLAIN [8]

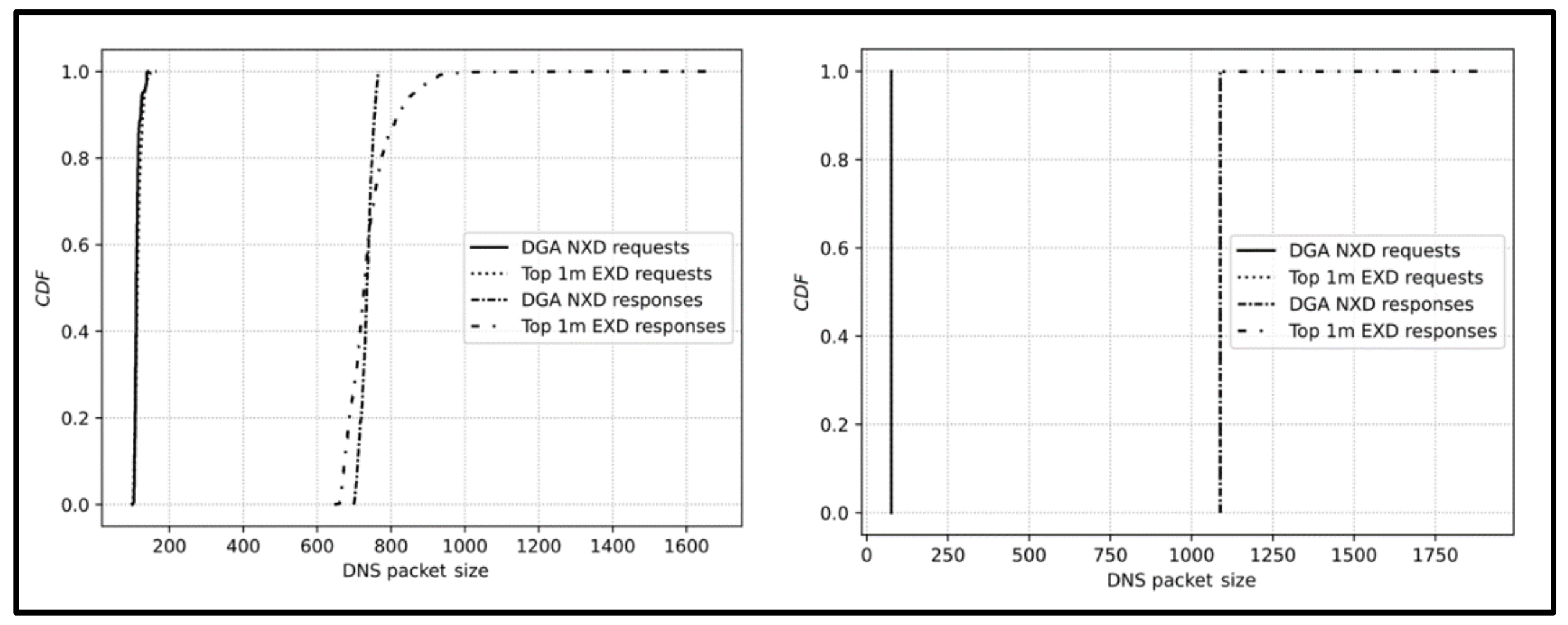


Number of recirculation in the P4 program for domain names in DGArchive dataset (https://tinyurl.com/5c3vp54k)

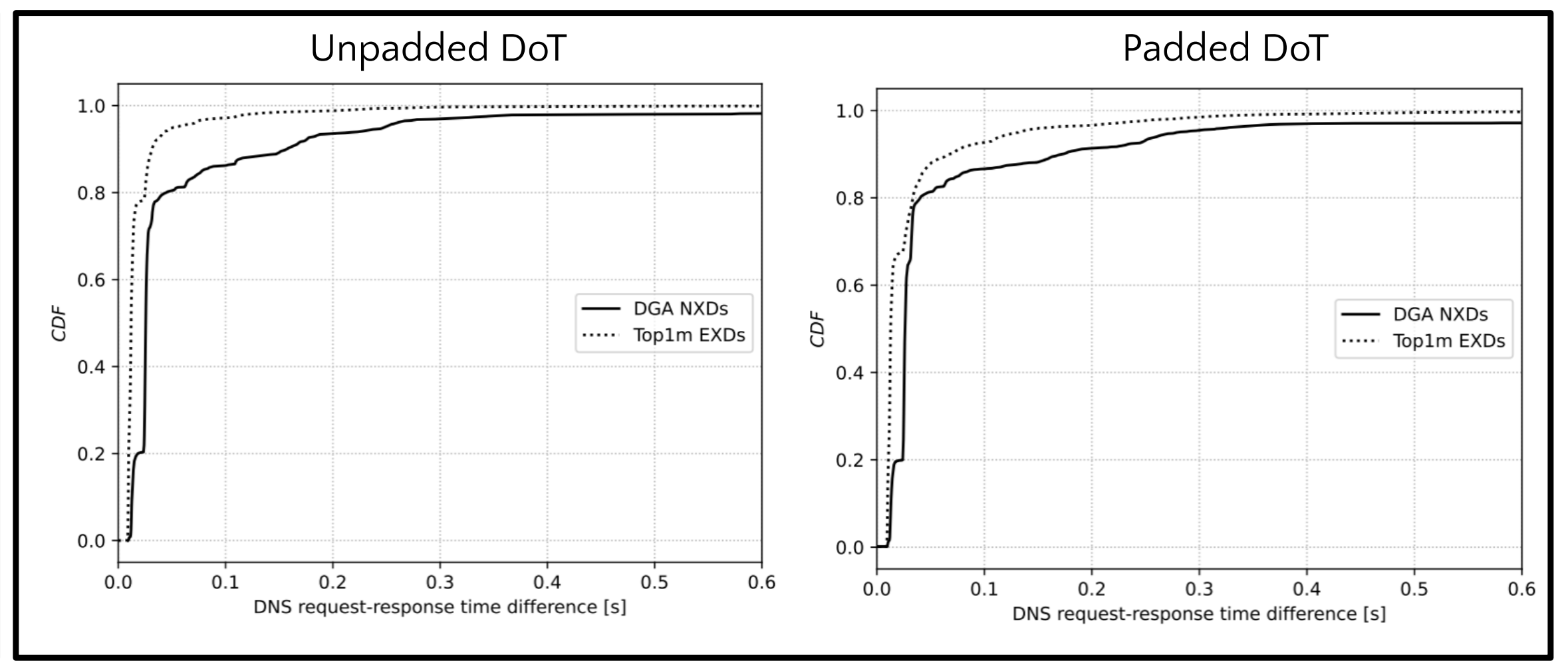


Ongoing Work on Encrypted DNS Traffic

CDFs of the DNS packet size belonging to benign and DGA domain names using unpadded DNS over TLS (DoT) (left figure) and padded DoT (right figure)



CDFs of the DNS request-response resolving time of NXDs from DGAs, and Existent Domains (EXDs) from the Top 1 million domain names dataset


Unpadded DoT
Padded DoT
CDF
DGA NXDs
Top1m EXDs
DNS request-response time difference [s]