INTERNET-SCALE PROBING OF CYBER-PHYSICAL SYSTEMS

INFERENCE, CHARACTERIZATION AND ORCHESTRATION ANALYSIS

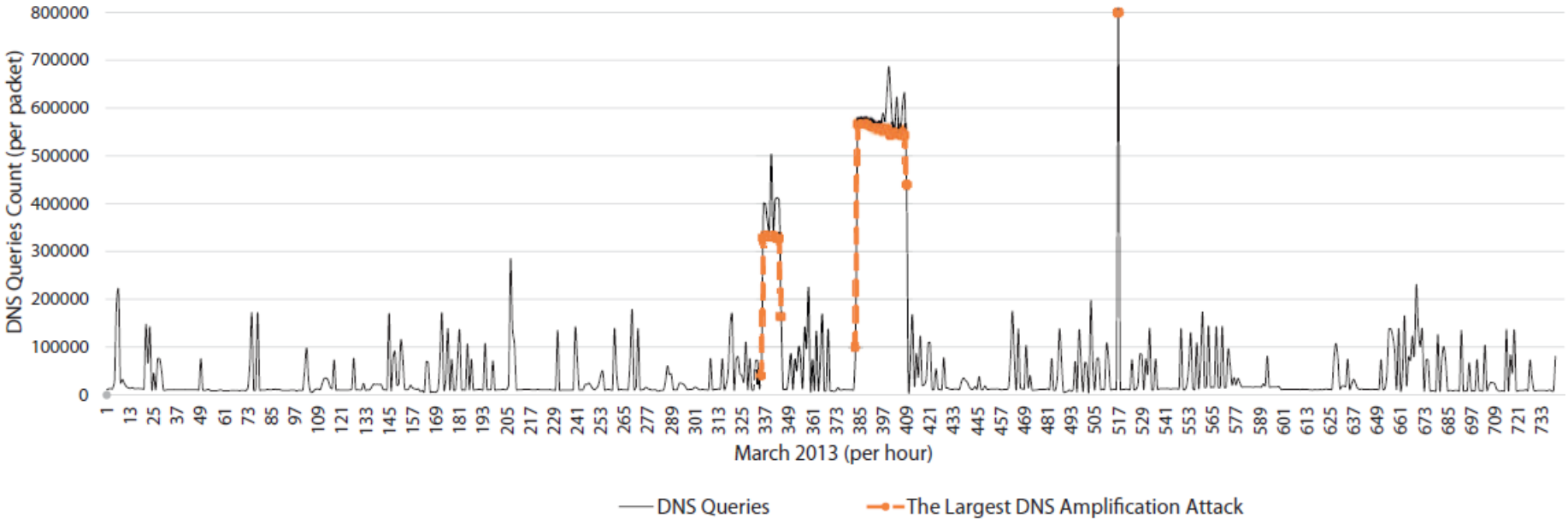
Claude Fachkha, Elias Bou-Harb, Anastasis Keliris, Nasir Memon, Mustaque Ahamad

1st of March, 2017

The Network and Distributed System Security Symposium (NDSS) 2017

Some of our previous works

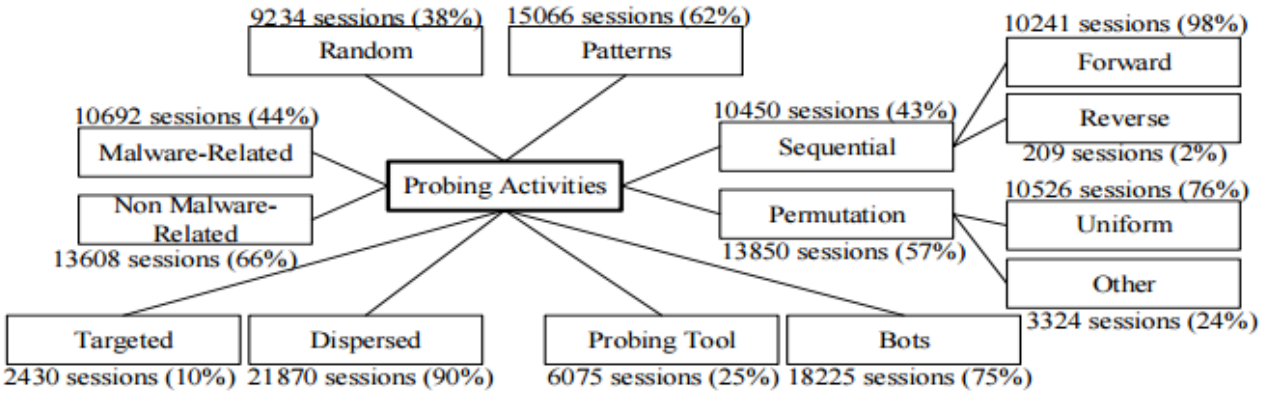
- An operational capability to *passively* identify DDoS amplification (reflection) attempts

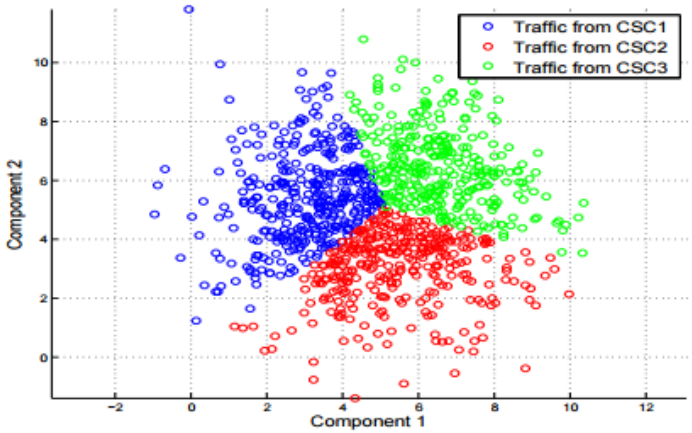


Traces from the largest (300 Gbps) DNS amplification attack in 2013 against Spamhaus

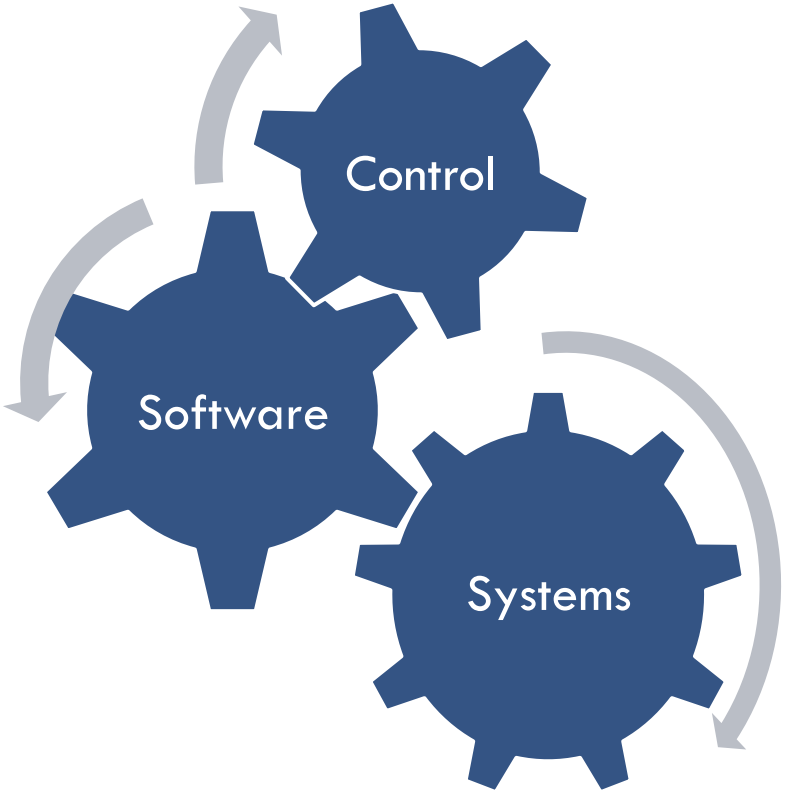
Some of our previous works

- An operational capability to *passively* identify large-scale orchestrated probing campaigns





Cyber-Physical Systems



Transportation	Healthcare	Defense
• Interactive traffic control	• Wearable sensors	• Smart Unmanned Vehicles

In the news…

Alert (ICS-ALERT-14-281-01E)
Ongoing Sophisticated Malware Campaign Compromising ICS

European renewable power grid rocked by cyber-attack

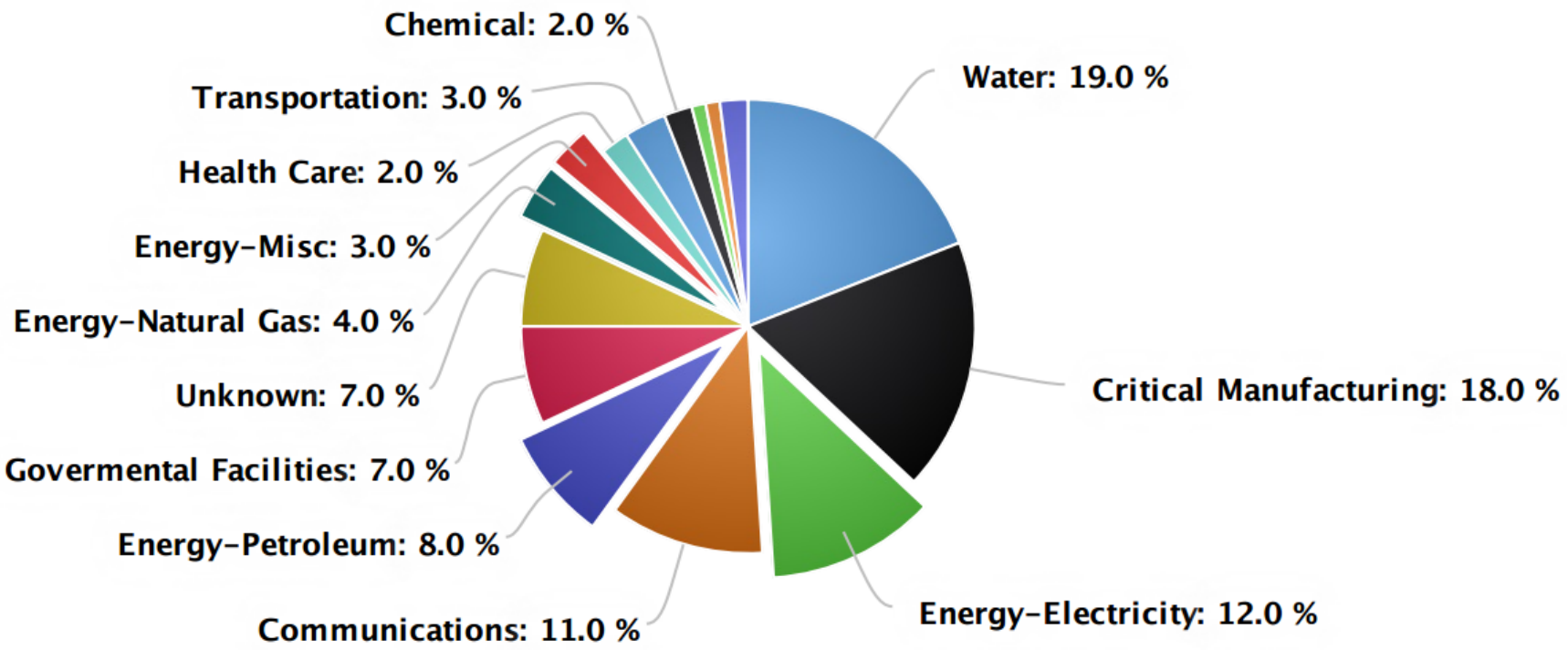
Water Treatment Plant Hit by Cyber-attack

Cyber Risks On The Rise For Transportation

The four amigos: Stuxnet, Flame, Gauss & DuQu

German Nuclear Power Plant Shut Down due to Malware On Chernobyl's 30th Anniversary

DHS reported CPS threats



Motivation

- Properly comprehending and accurately characterizing malicious attackers' capabilities, intents and aims, remains challenging

- Lack of real malicious empirical data that can be captured, inferred, and analyzed from within the boundaries of operational CPS realms
  - lack of complete maturity related to CPS
  - the significant diversity of such types of systems
  - logistic and privacy constraints

Contributions

- Automated approaches that aim at disclosing real CPS attackers' strategies, by *passively* inferring, characterizing, and correlating CPS probing events
  - Proposing a formal preprocessing probabilistic model that aims at filtering noise (i.e., misconfiguration traffic)
  - Executing multidimensional investigation of probing activities targeting more than 25 communication and control CPS services distributed over 120 ports
  - Validating the proposed models, methods and approaches by experimenting with 50 GB of darknet data

Related Work: Control-Theoretic Approaches

Type of system	Noise	Attack model	Defense mechanisms
Control system	Noisy	Faults	Filters, hypothesis testing, X^2 detector
Static power grid	Noisy	False-data injection (sensor attack)	Residue detector
Wireless control network	none	Malicious nodes with arbitrary state attacks	Intrusion detector, output estimation
Distributed network	none	Malicious nodes with arbitrary state attacks	Combinatorial estimator
Consensus network	none	Malicious or faulty nodes	Detection and identification filters
Sensor network	Noisy	Dynamic false-data injection (sensor attack)	Residue detector

Models describing the underlying physical phenomena enables the prediction of future behavior and, more importantly, unforeseen **deviations** from it

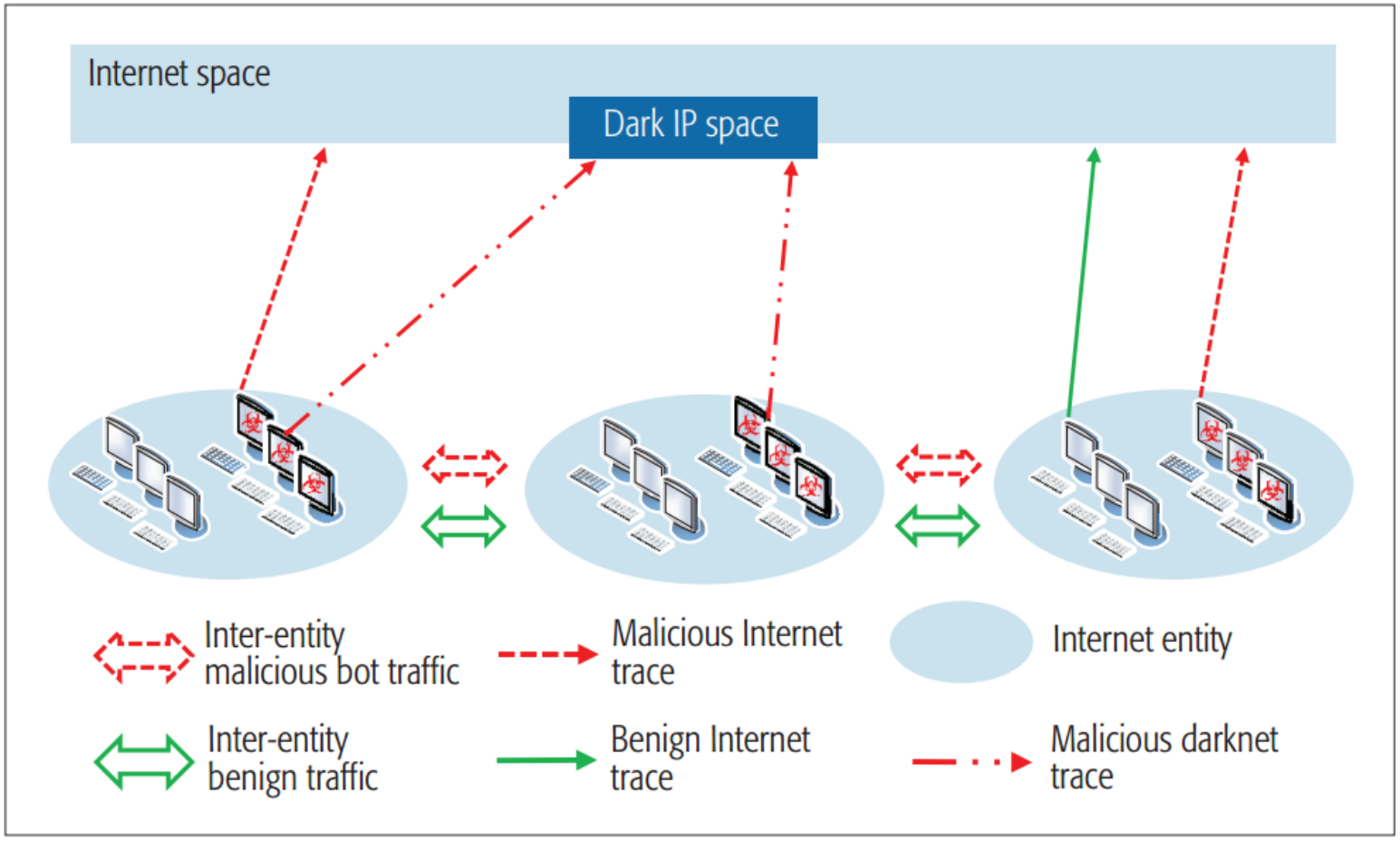
Related Work: Cyber Security Approaches

Level	Impact		Attack description
1	Data integrity		Corrupt integrity by adding data to the packet.
2	IT System	Reconnaissance	Analyse functionality a PLC implements.
		Integrity	Exploit lack of specification compliance.
			Perform unauthorized use of an administrative command.
		Denial of service	Perform MITM to enforce system delay.
			Perform unauthorized use of administrative command.
3	Process	Reconnaissance	Analyse structure of memory map.
		Direct control	Perform change on process variable.
		Indirect control	Tamper with process values.

Related Work

- Probing analysis
  - Inference
  - Analysis
  - Measurements
- Network Telescope: Measurements & Analysis
- CPS Traffic Analysis

Passive Measurements

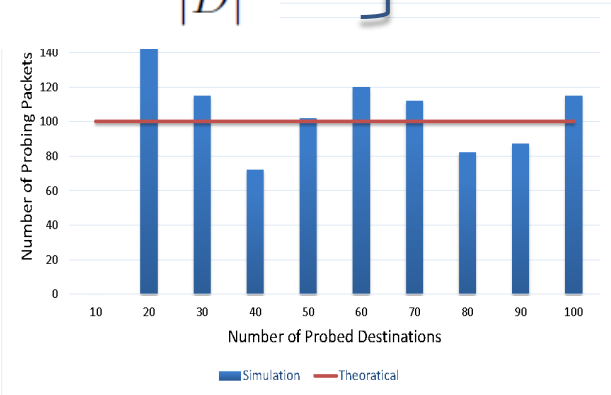


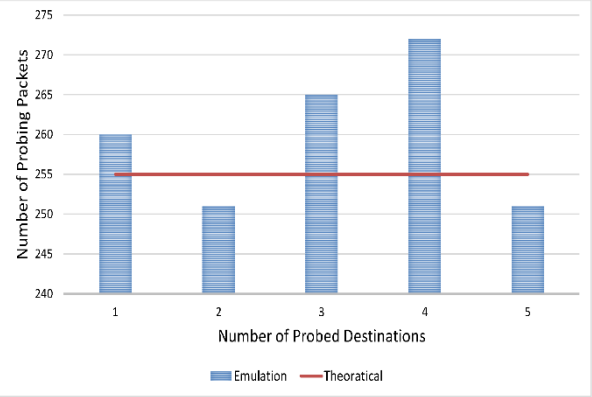
Darknet Preprocessing Model

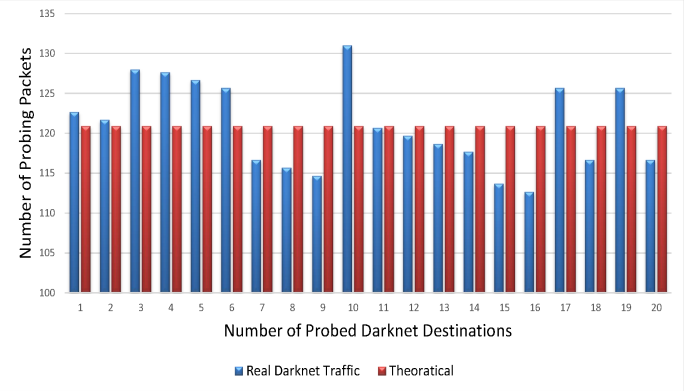
$$P_{misc}(d_i) = \frac{n_s(d_i)}{\sum_{\forall d_j \in D} n_s(d_j)}$$

$$P_{mal}(d_i) = \frac{1}{|D|}$$

How unusual the access to a darknet IP *d* is



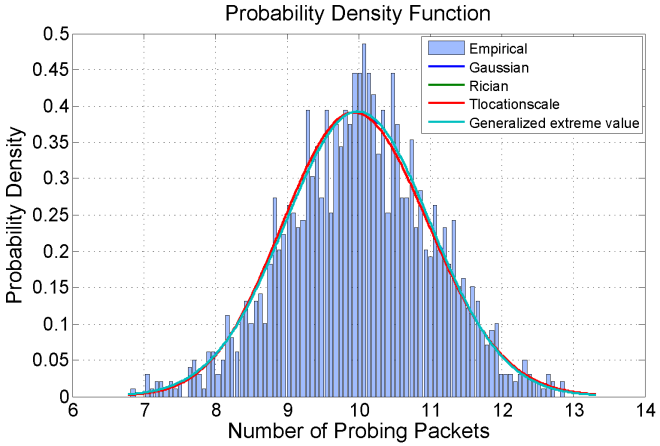


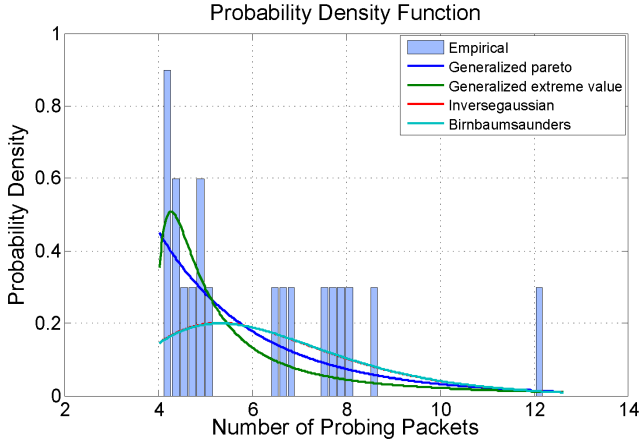


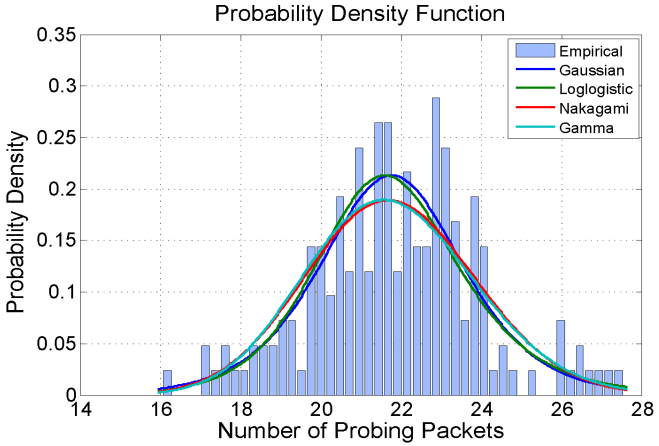
Darknet Preprocessing Model

$$P_{misc}(d_i) = \frac{n_s(d_i)}{\sum_{\forall d_j \in D} n_s(d_j)}$$

$$P_{mal}(d_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$







Darknet Preprocessing Model

$$P(D_i) = P(D_i = \{d_{i1}, d_{i2}, \cdots, d_{in}\} \mid |D_i| = n) \times P(|D_i| = n)$$

$$P_{misc}(D_i = \{d_{i1}, d_{i2}, \cdots\} \mid |D_i|) = \frac{1}{K} \prod_{\forall d_j \in D_i} P_{misc}(d_i)$$

$$P_{mal}(D_i = \{d_{i1}, d_{i2}, \cdots\} \mid |D_i|) = \frac{1}{K} \prod_{\forall d_j \in D_i} P_{mal}(d_i)$$

A source accessing a predefined *n* darknet destinations

$$P_{misc}(|D_i|) = \frac{1}{(e-1)|D_i|!}$$

$$P_{mal}(|D_i|) = \frac{1}{|D|}$$

A source accessing *a* number of darknet destinations

Darknet Preprocessing Model

$$P_{misc}(D_i) = \frac{1}{K(e-1)|D_i|!} \prod_{\forall d_j \in D_i} P_{misc}(d_i)$$

$$P_{mal}(D_i) = \frac{1}{K|D|} \prod_{\forall d_j \in D_i} P_{mal}(d_i)$$

$$L_{misc}(D_i) = -lnP_{misc}(D_i)$$
$$L_{mal}(D_i) = -lnP_{mal}(D_i)$$

$$L_{mal}(D_i) - L_{misc}(D_i) > 0$$

Algorithm 1 Inferring misconfiguration flows using the probabilistic model

Input: Darknet Flows, *DarkFlows*
Output: Flag, *MiscFlag*, indicating that the DarkFlow is originating from a misconfigured source

for $DarkFlows$ **do**
    $MiscFlag \leftarrow 0$
    $i \leftarrow$ DarkFlows.*getUniqueSources()*
    Amalgamate $DarkFlows_i$ originating from a specific source s_i
    Update $s_i(D_i)$
    Compute $P_{misc}(D_i), P_{mal}(D_i)$
    if $P_{misc}(D_i) > P_{mal}(D_i)$ **then**
        $MiscFlag \leftarrow 1$
    end if
end for

CPS Probing Inference

Algorithm 2 CPS Scanning Inference Algorithm

Input: A set (F) of unique darknet flows (f),
Each flow f contains packet count (pkt_cnt) and rate ($rate$)
SP: CPS Service Port
Tw: Time window
Pth: Packet threshold
Rth: Rate threshold,
Tn: Time of packet number n in a flow
pkt: Packet
Output: CPS flag, CPS_flag

for Each f in F **do**
  while pkt in f **do**
    if $pkt.contains() != SP$ **then**
      $CPS\_flag() \leftarrow 0$
    end if
    if $pkt.contains() = SP$ **then**
      $CPS\_flag() \leftarrow 1$
    end if
  end while

  $pkt\_cnt \leftarrow 0$
  $T1 \leftarrow$ pkt_gettime()
  $Tf \leftarrow T1 + Tw$
  while pkt in f **do**
    $Tn =$ pkt_gettime()
    if $Tn < Tf$ **then**
      pkt_cnt $\leftarrow$ pkt_cnt + 1
    end if
  end while
  $rate \leftarrow \frac{pkt\_cnt}{Tw}$
  if $pkt\_cnt < Pth \;||\; rate < Rth$ **then**
    $CPS\_flag() \leftarrow 0$
  end if
end for

CPS Communication & Control Protocols	Port Number	Type
ABB Ranger 2003	10307/10311/10364, etc.	Registered
BACnet/IP	47808	Registered
DNP/DNP3	19999/20000	Registered
Emerson/Fisher ROC Plus	4000	Registered
EtherCAT	34980	Registered
EtherNet/IP	2222/44818	Registered
FL-net Reception/Transmission	55000-55003	Dynamic/Private
Foundation Fieldbus HSE	1089/1090/1091	Registered
Foxboor/Invensys Foxboro DCS	55550	Dynamic/Private
Iconic Genesis32 GenBroker	18000	Registered
ICCP	102	Well-known
IEC-104	2404	Registered
Johnson Controls Metasys N1	11001	Registered
Modbus	502	Well-known
MQ Telemetry Transport	1883	Registered
Niagara Fox	1911/4911	Registered
OPC UA Discovery Server	3480	Registered
OSIsoft PI Server	5450	Registered
PROFINET	34962/24963/34964	Registered
Project/SCADA Node Primary Port	4592	Registered
Red Lion	789	Well-known
ROC Plus	4000	Registered
SCADA Node Ports	4592/14592	Registered
Siemens Spectrum Power TG	50001/50018/50020, etc.	Dynamic/Private
SNC GENe	62900/62911/62924, etc.	Dynamic/Private
Telvent OASyS DNA	5050/5052/5065, etc.	Registered

CPS Characterization and Co-occurrence

- Amalgamated Statistics
- Significance and Prevalence
- Distribution of different types of scans

- Jaccard similarity to infer co-occurrence patterns

CPS Probing Orchestration Fingerprinting

- Large-scale probing events
  - the population of the participating bots is several orders of magnitude larger
  - the target scope is generally the entire IP address space
  - the sources adopt well-orchestrated, often botmaster-coordinated, stealth scan strategies that maximize targets' coverage while minimizing redundancy and overlap

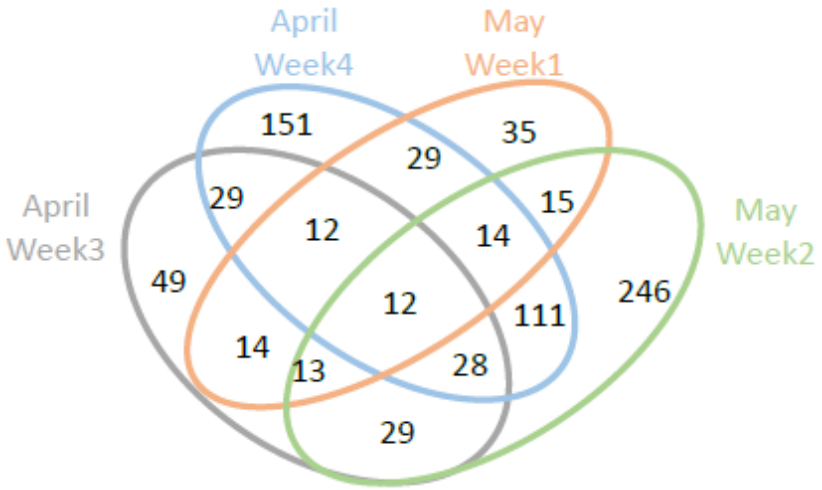
CPS Probing Orchestration Fingerprinting

- Inferring CPS large-scale probing events
  - Time series analysis
    - Infer temporal similarities
    - Dynamic Time Warping (DTW) technique
  - Netflow analysis
    - Infer netflow characteristics
    - Context triggered piecewise hashing (CTPH)
- Select and cluster CPS probing sessions that minimize the DTW similarity metric while maximizing the CTPH measure

Empirical Findings: Characterization

	April Week 3	April Week 4	May Week 1	May Week 2
Total Scanners	7954	8871	8731	8341
Total Uniq Scanners	3007	3727	3950	3731

Top Scanners



Consistency and overlap targeting Modbus

Validation:

- AbuseIPDB and Cymon: 4.37% of scanners were involved in various malicious reported activities (hacking (41.25%), portscan (31.46%), FTP/SSH, brute force (13.28%), and DDoS (6.29%)).
- Dshield: 88.1% found.
- Remaining: never reported

Empirical Findings: Characterization

Top five used/abused src-port

April Week 3	April Week 4	May Week 1	May Week 2
6000 (609)	53 (535)	1048 (785)	6000 (426)
53933 (348)	43490 (356)	42880 (576)	60000 (330)
53 (315)	6000 (235)	53 (334)	53 (314)
43490 (267)	22 (214)	59651 (223)	63030 (156)
59531 (244)	1048 (146)	58017 (221)	50449 (128)

Common used ports:

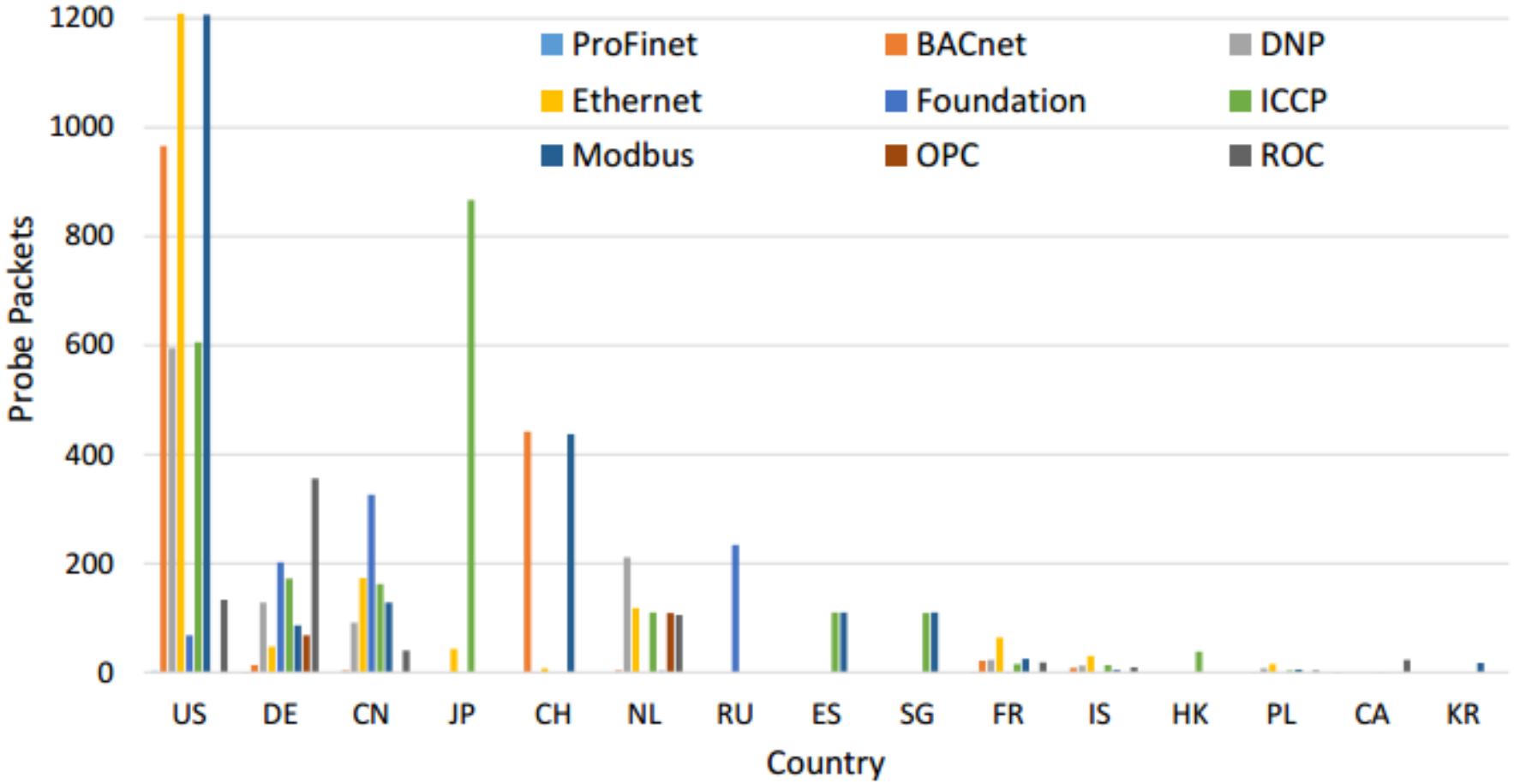
- Port 6000 (often reported to be used by trojans)
- 40k and 60k range
- For Modbus communication, 30% of its traffic originated from source port 6706

Empirical Findings: Characterization

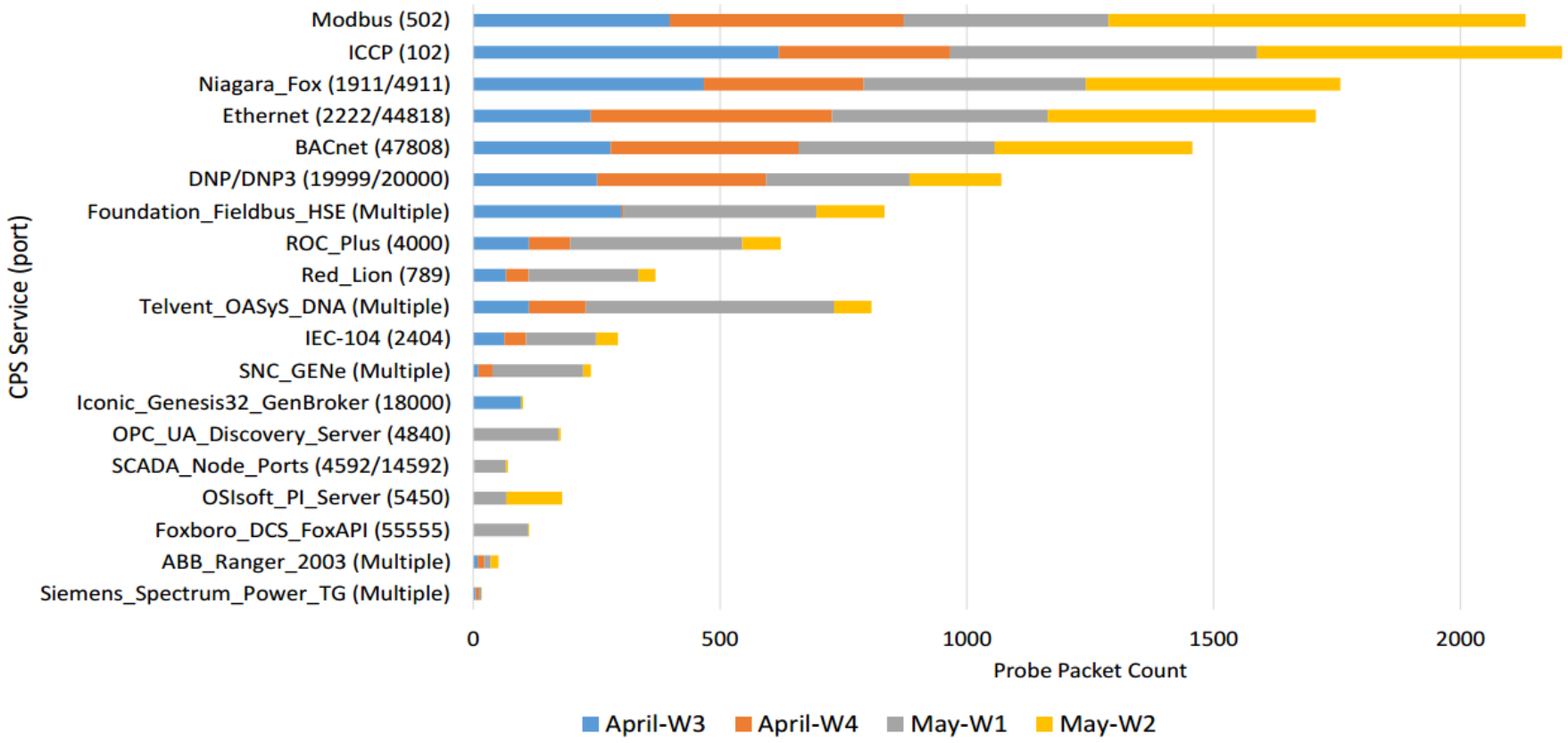
Top five IP-ID values
(Probe packet count)

April Week 3	April Week 4	May Week 1	May Week 2
0xd431 (13060)	0xd431 (12632)	0xd431 (11640)	0xd431 (12849)
0x0100 (820)	0x0100 (343)	0x0100 (566)	0x0100 (530)
0x0049 (11)	0x0b1c (10)	0x843d (9)	0x0438 (13)
0x9625 (9)	0x052a (10)	0x591e (9)	0xb530 (9)
0x0ae7 (9)	0x058d (9)	0x01da (9)	0x8faf (9)

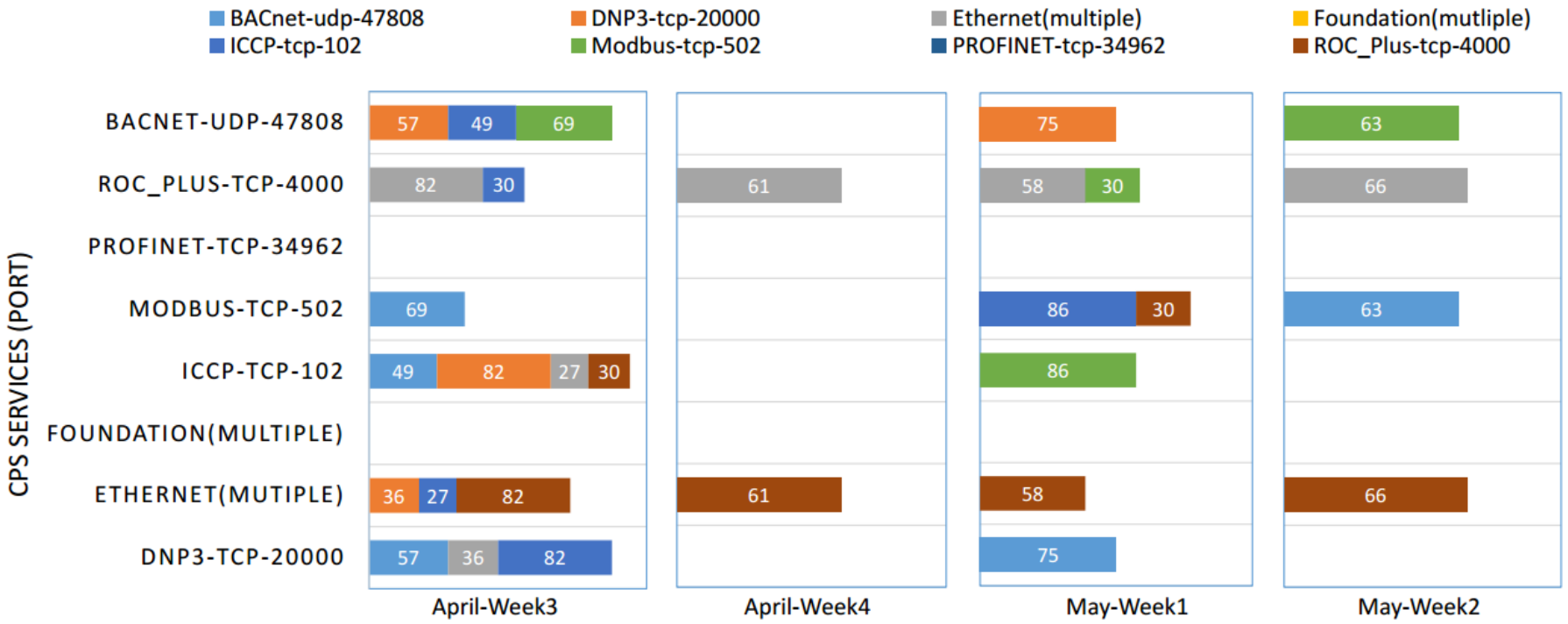
Empirical Findings: Sources of Probes



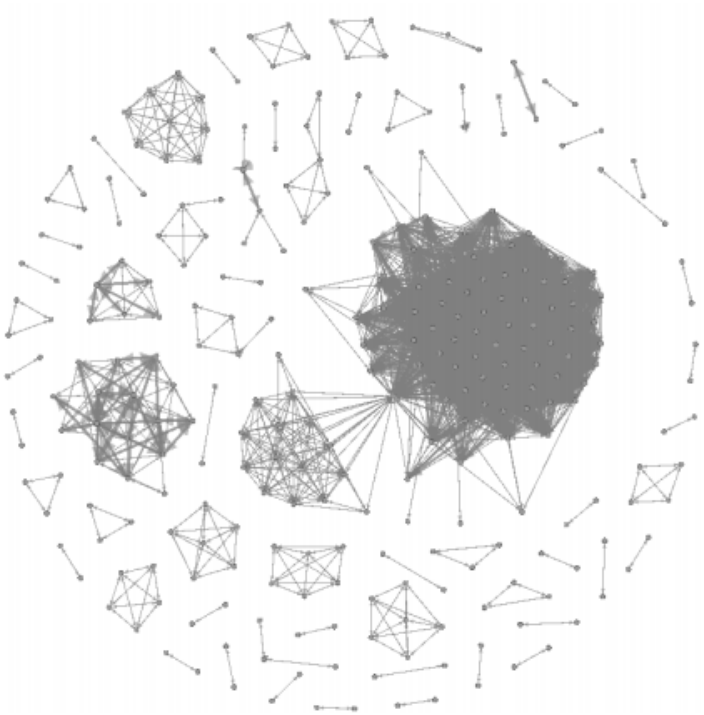
Empirical Findings: Top Targeted CPS Services



Empirical Findings: Co-occurrence Patterns



Empirical Findings: Orchestrated Campaigns



- 58 inferred campaigns
- Some employ very low probing rate
- 5 large-scale coordinated events (more than 50 hosts)

Empirical Findings: Orchestrated Campaigns

Reference Source	Source Domain	Number of Distinct IP Addresses
A	*.edu	64
B	*.io	136
C	*.com *.de	188
D	*.cn	116
E	*.ru	54

- Focused (A)
  - Modbus on TCP port 502, Niagara Fox on TCP port 1911 and BACnet on TCP port 47808 (CPS-specific)
  - Employed unique hosts
- Distributed (B)
  - Probed 191 services, including, Modbus and BACnet
  - Recycled 13 hosts per week

Empirical Findings: Orchestrated Campaigns

Reference Source	Source Domain	Number of Distinct IP Addresses
A	*.edu	64
B	*.io	136
C	*.com *.de	188
D	*.cn	116
E	*.ru	54

- C, D, E: Possibly malicious campaigns
- C, D:
  - Sources from US, Germany and China
  - Large-scale stealthy probing
  - Dedicated for brute force attacks (HMI exploitations)
- E:
  - Attributed to Russia
  - Probed almost all the darknet IP space
  - Focused on coordinated scanning towards Foundation Fieldbus systems (factory automation)

Discussion

- Challenges
- Attackers' IP Address Selection
  - Particularly or randomly targeted?
- Incomplete view of the CPS abuse
- Defense against scanning
  - Blacklisting
- Research Trends
  - Collaborative approach for CPS security

Concluding Remarks

- Attempt to generate unsolicited empirical data related to CPS activities
- 33 thousand probes towards ample of CPS protocols
- 74% of CPS probes that were persistent throughout the entire analyzed period
- Thousands of large-scale, stealthy, previously undocumented orchestrated probing events
- CPS targets in rarely investigated CPS realms such as manufacturing and building automation systems

Future Work

- Fuse the obtained data with CPS honeypot data to build broader notions of CPS maliciousness
- Identify attack models for CPS in the health and cargo terminal (ports) sectors
- Empirical measurements in the IoT paradigm for inference and resiliency

Acknowledgment

- Prof. Nasir Memon (**NYU**)
- Prof. Mustaque Ahamad (**Georgia Tech**)
- Our team (Elias and Tasos)
- NYU AD: CISO and the networking team

Questions

Claude Fachkha
Assistant Professor
claude.fachkha@gmail.com

Elias Bou-Harb
Assistant Professor
ebouharb@fau.edu