INTERNET-SCALE PROBING OF CYBER-PHYSICAL SYSTEMS
INFEERENCE, CHARACTERIZATION AND ORCHESTRATION ANALYSIS

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

- Control
- Software
- Systems

Transportation
- Interactive traffic control

Healthcare
- Wearable sensors

Defense
- 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

- Water: 19.0%
- Critical Manufacturing: 18.0%
- Energy-Electricity: 12.0%
- Energy-Petroleum: 8.0%
- Communications: 11.0%
- Energy-Natural Gas: 4.0%
- Unknown: 7.0%
- Governmental Facilities: 7.0%
- Energy-Misc: 3.0%
- Health Care: 2.0%
- Transportation: 3.0%
- Chemical: 2.0%
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

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

Models describing the underlying physical phenomena enables the prediction of future behavior and, more importantly, unforeseen deviations from it.
## Related Work: Cyber Security Approaches

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

- Probing analysis
  - Inference
  - Analysis
  - Measurements

- Network Telescope: Measurements & Analysis

- CPS Traffic Analysis
Passive Measurements
Darknet Preprocessing Model

\[
P_{misc}(d_i) = \frac{n_s(d_i)}{\sum_{d_j \in D} n_s(d_j)}
\]

\[
P_{mat}(d_i) = \frac{1}{|D|}
\]

How unusual the access to a darknet IP \( d \) is
Darknet Preprocessing Model

\[ P_{\text{misc}}(d_i) = \frac{n_s(d_i)}{\sum_{d_j \in D} n_s(d_j)} \]

\[ P_{\text{mal}}(d_i) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Darknet Preprocessing Model

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

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

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

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

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

A source accessing a predefined \( n \) darknet destinations

A source accessing a number of darknet destinations
Darknet Preprocessing Model

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

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

\[ L_{misc}(D_i) = -\ln P_{misc}(D_i) \]
\[ L_{mal}(D_i) = -\ln P_{mal}(D_i) \]
\[ L_{mal}(D_i) - L_{misc}(D_i) > 0 \]

**Algorithm 1** Inferring misconfiguration flows using the probabilistic model

1: **Input:** Darknet Flows, DarkFlows
2: **Output:** Flag, MiscFlag, indicating that the DarkFlow is originating from a misconfigured source
3: 
4: **for** DarkFlows **do**
5: \hspace{1em} MiscFlag \leftarrow 0
6: \hspace{1em} \textit{i} \leftarrow \text{DarkFlows.getUniqueSources()} \hspace{1em} \text{Amalgamate DarkFlows}_i \text{ originating from a specific source } s_i \hspace{1em} \text{Update } s_i(D_i) \hspace{1em} \text{Compute } P_{misc}(D_i), P_{mal}(D_i) \hspace{1em} \text{if } P_{misc}(D_i) > P_{mal}(D_i) \text{ then} \hspace{1em} \text{MiscFlag} \leftarrow 1 \hspace{1em} \text{end if} \hspace{1em} \text{end for}
**Algorithm 2 CPS Scanning Inference Algorithm**

1. **Input:** A set \( F \) of unique darknet flows \( f \),
2. Each flow \( f \) contains packet count \( \text{pkt}_c \) and rate \( \text{rate} \).
3. **SP:** CPS Service Port
   - \( Tw \): Time window
   - \( Pth \): Packet threshold
   - \( Rth \): Rate threshold
   - \( Tn \): Time of packet number \( n \) in a flow
4. **pkt:** Packet
5. **Output:** CPS flag, \( CPS\_flag \)
6. for Each \( f \) in \( F \) do
7.   while \( \text{pkt} \) in \( f \) do
8.     if \( \text{pkt.contains()} \neq \text{SP} \) then
9.       \( CPS\_flag() \leftarrow 0 \)
10.    end if
11.   if \( \text{pkt.contains()} = \text{SP} \) then
12.      \( CPS\_flag() \leftarrow 1 \)
13.   end if
14. end while
15. \( \text{pkt}_c \leftarrow 0 \)
16. \( TI \leftarrow \text{pkt.gettime()} \)
17. \( Tf \leftarrow TI + Tw \)
18. while \( \text{pkt} \) in \( f \) do
19.   \( Tr = \text{pkt.gettime()} \)
20.   if \( Tr < Tf \) then
21.     \( \text{pkt}_c \leftarrow \text{pkt}_c + 1 \)
22. end if
23. end while
24. \( \text{rate} \leftarrow \frac{\text{pkt}_c}{Tw} \)
25. if \( \text{pkt}_c < Pth \) \( \text{or} \) \( \text{rate} < Rth \) then
26.   \( CPS\_flag() \leftarrow 0 \)
27. end if
28. end for

---

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

<table>
<thead>
<tr>
<th></th>
<th>April Week 3</th>
<th>April Week 4</th>
<th>May Week 1</th>
<th>May Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Scanners</td>
<td>7954</td>
<td>8871</td>
<td>8731</td>
<td>8341</td>
</tr>
<tr>
<td>Total Uniq Scanners</td>
<td>3007</td>
<td>3727</td>
<td>3950</td>
<td>3731</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th></th>
<th>April Week 3</th>
<th>April Week 4</th>
<th>May Week 1</th>
<th>May Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port</td>
<td>6000</td>
<td>53 (535)</td>
<td>1048 (785)</td>
<td>6000 (426)</td>
</tr>
<tr>
<td></td>
<td>53933</td>
<td>43490 (356)</td>
<td>42880 (576)</td>
<td>60000 (330)</td>
</tr>
<tr>
<td></td>
<td>53 (315)</td>
<td>6000 (235)</td>
<td>53 (334)</td>
<td>53 (314)</td>
</tr>
<tr>
<td></td>
<td>43490 (267)</td>
<td>22 (214)</td>
<td>59651 (223)</td>
<td>63030 (156)</td>
</tr>
<tr>
<td></td>
<td>59531 (244)</td>
<td>1048 (146)</td>
<td>58017 (221)</td>
<td>50449 (128)</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>April Week 3</th>
<th>April Week 4</th>
<th>May Week 1</th>
<th>May Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xd431 (13060)</td>
<td>0xd431 (12632)</td>
<td>0xd431 (11640)</td>
<td>0xd431 (12849)</td>
</tr>
<tr>
<td>0x0100 (820)</td>
<td>0x0100 (343)</td>
<td>0x0100 (566)</td>
<td>0x0100 (530)</td>
</tr>
<tr>
<td>0x0049 (11)</td>
<td>0x0b1c (10)</td>
<td>0x843d (9)</td>
<td>0x0438 (13)</td>
</tr>
<tr>
<td>0x9625 (9)</td>
<td>0x052a (10)</td>
<td>0x591e (9)</td>
<td>0xb530 (9)</td>
</tr>
<tr>
<td>0x0ae7 (9)</td>
<td>0x058d (9)</td>
<td>0x01da (9)</td>
<td>0x8faf (9)</td>
</tr>
</tbody>
</table>
Empirical Findings: Sources of Probes
Empirical Findings: Top Targeted CPS Services

- Modbus (502)
- ICCP (102)
- Niagara_Fox (1911/4911)
- Ethernet (2222/44818)
- BACnet (47808)
- DNP/DNP3 (19999/20000)
- Foundation_Fieldbus_HSE (Multiple)
- ROC_Plus (4000)
- Red_Lion (789)
- Telvent_OASys_DNA (Multiple)
- IEC-104 (2404)
- SNC_GEnE (Multiple)
- Iconic_Genesis32_GenBroker (18000)
- OPC_UA_Discovery_Server (4840)
- SCADA_Node_Ports (4592/14592)
- OSIsoft_PI_Server (5450)
- Foxboro_DCS_FoxAPI (55555)
- ABB_Ranger_2003 (Multiple)
- Siemens_Spectrum_Power_TG (Multiple)

Probe Packet Count

- April-W3
- April-W4
- May-W1
- May-W2
Empirical Findings: Co-occurrence Patterns

The diagram illustrates the co-occurrence patterns of different CPS services over various weeks. Each service is represented by a different color, and the bars indicate the percentage occurrence over different periods.

- **BACnet-udp-47808**: Represented in blue, with values ranging from 57% to 69%.
- **DNP3-tcp-20000**: Represented in orange, with values ranging from 57% to 75%.
- **Modbus-tcp-502**: Represented in green, with values ranging from 27% to 82%.
- **PROFINET-tcp-34962**: Represented in dark blue, with values ranging from 69% to 86%.
- **Ethernet(multiple)**: Represented in light blue, with values ranging from 61% to 75%.
- **Foundation(multiple)**: Represented in brown, with values ranging from 66% to 86%.

The data is divided into weeks:
- **April-Week3**: Values vary for each service, showing the percentage occurrence.
- **April-Week4**: Similar to April-Week3, showing percentage occurrence.
- **May-Week1**: Values vary for each service, showing the percentage occurrence.
- **May-Week2**: Similar to May-Week1, showing percentage occurrence.
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

<table>
<thead>
<tr>
<th>Reference Source</th>
<th>Source Domain</th>
<th>Number of Distinct IP Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>*.edu</td>
<td>64</td>
</tr>
<tr>
<td>B</td>
<td>*.io</td>
<td>136</td>
</tr>
<tr>
<td>C</td>
<td>*.com, *.de</td>
<td>188</td>
</tr>
<tr>
<td>D</td>
<td>*.cn</td>
<td>116</td>
</tr>
<tr>
<td>E</td>
<td>*.ru</td>
<td>54</td>
</tr>
</tbody>
</table>

- **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

- 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)

<table>
<thead>
<tr>
<th>Reference Source</th>
<th>Source Domain</th>
<th>Number of Distinct IP Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>*.edu</td>
<td>64</td>
</tr>
<tr>
<td>B</td>
<td>*.io</td>
<td>136</td>
</tr>
<tr>
<td>C</td>
<td>*.com</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>*.de</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>*.cn</td>
<td>116</td>
</tr>
<tr>
<td>E</td>
<td>*.ru</td>
<td>54</td>
</tr>
</tbody>
</table>
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

Thank you