INL/CON-22-68728-Revision-1

Micro Baselines for Operational Technology Environments

September 2022

Gabriel Arthur Weaver, Samuel Patrick Farnan, Dan Gunter

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

INL/CON-22-68728-Revision-1

Micro Baselines for Operational Technology Environments

Gabriel Arthur Weaver, Samuel Patrick Farnan, Dan Gunter

September 2022

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

U.S. DEPARTMENT OF ENERGY | Office of Cybersecurity, Energy Security, and Emergency Response

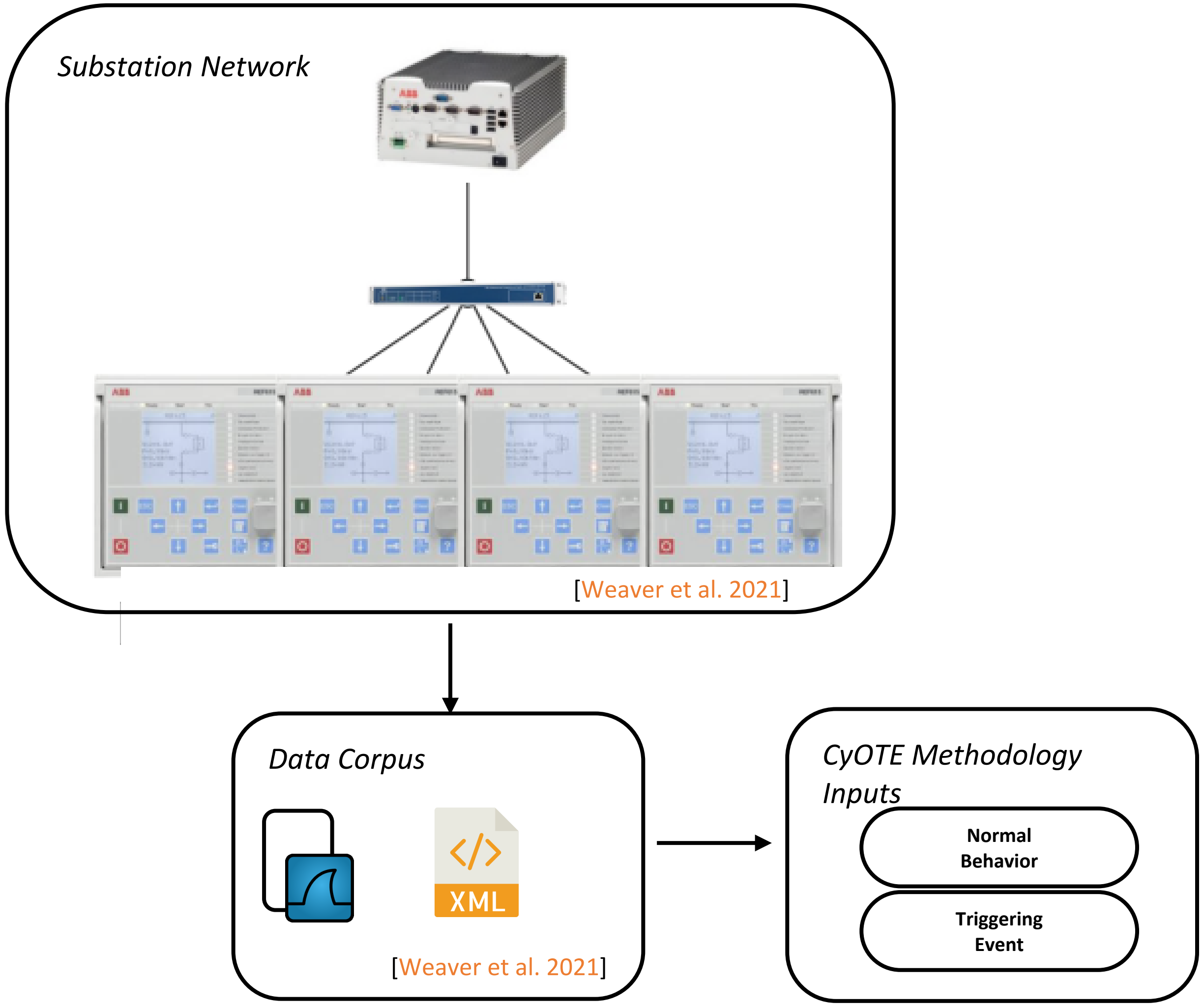
Micro-Baselines for Operational Technology Environments

Gabriel Weaver, Dan Gunter, Sam Chanoski, and Sam Farnan

2022 MORS Emerging Techniques Forum

What is Network Baselining?

- Within a computer network, we want to understand ‘normal’ behavior.
- Useful to monitor
  - configuration changes,
  - malicious communications,
  - unexpected behavior
- Cybersecurity for the Operational Technology Environment (CyOTE) requires us to understand:
  - Expected behavior due to operational events
  - A specific notion of context



Substation Network

[Weaver et al. 2021]

Data Corpus

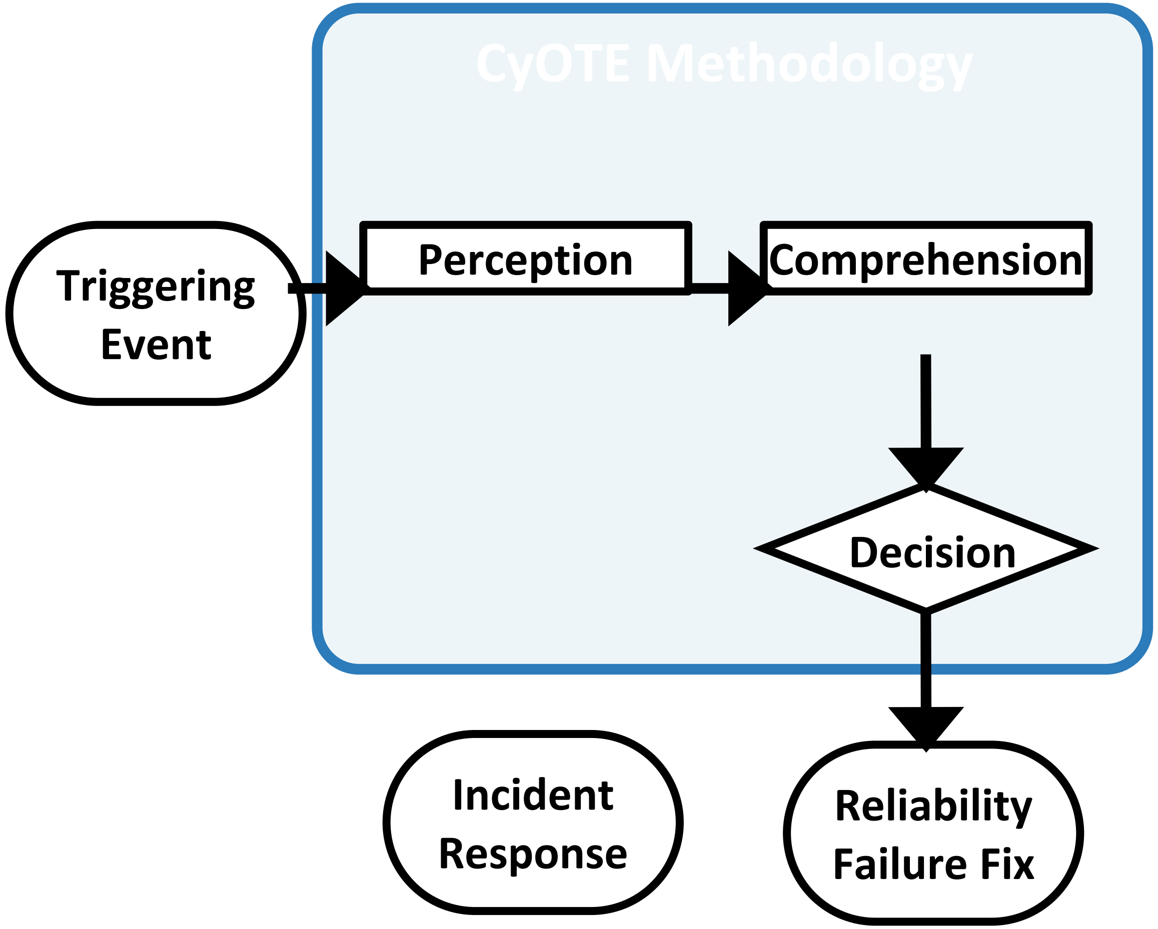
[Weaver et al. 2021]

CyOTE Methodology Inputs

Normal Behavior

Triggering Event

CyOTE Methodology Overview



- **How to understand the information you have, not get more data**
- **Applies concepts of perception and comprehension to a world of Knowns and Unknowns**
- **Endpoint is making a risk-informed decision to conduct incident response or to treat as a reliability failure**
- Over time, detect fainter signals sooner

Challenges of Network Baselining

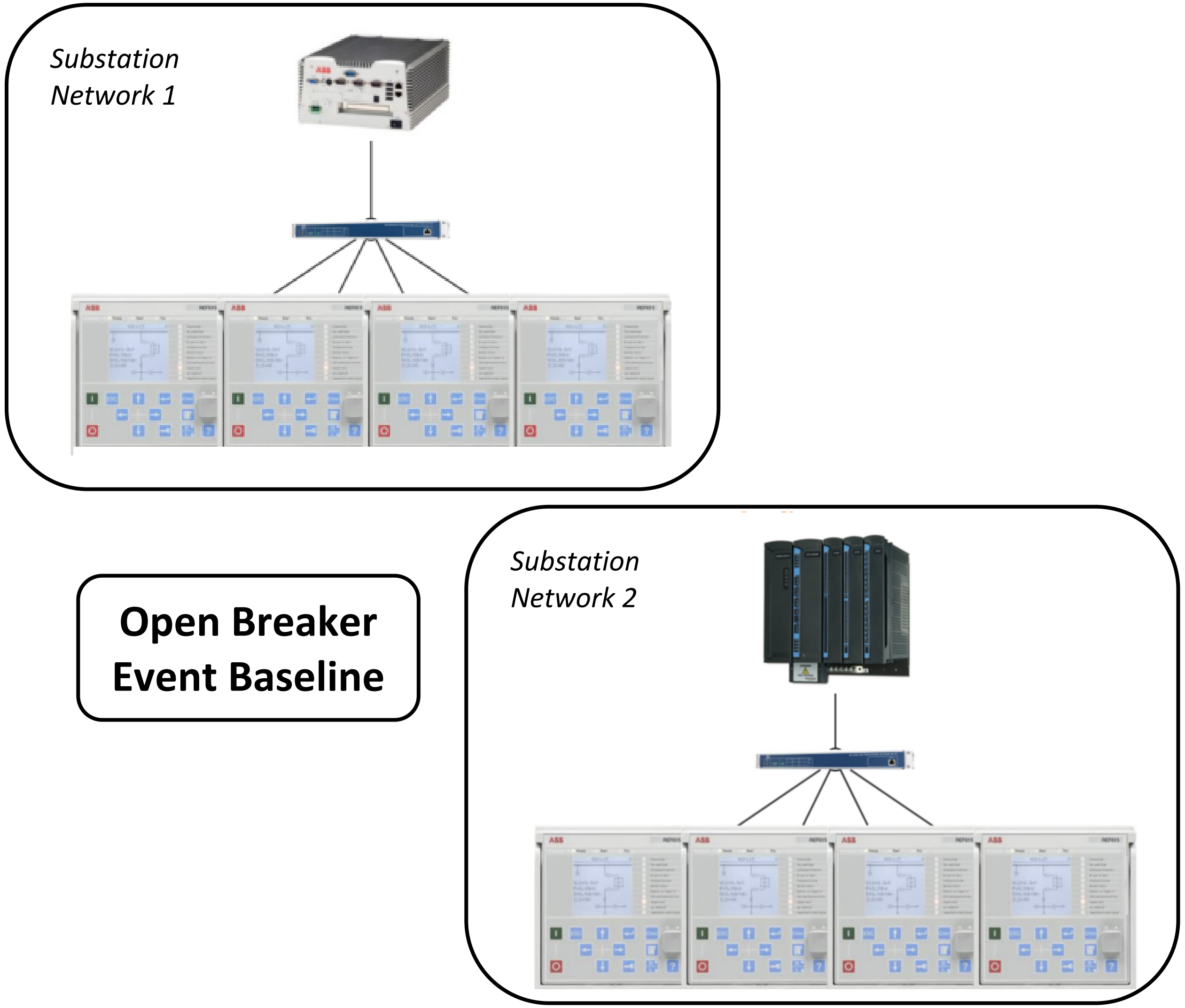
- **Top-down approaches to network baselining rely on <u>generally-available observables</u>.**
- **But these observables lack properties upon which traditional statistical tools depend [Schulz et al. 2019]**
  - **Stationarity**
  - **Memorylessness**
- **Behavior may vary depending upon location and time.**
  - **Power system in summer versus winter**

Data link, network, and transport layer from Network Packet	Ethernet addresses, IP addresses, TCP/UDP Ports	• Any context of what the communication is about • OT data
Session, Presentation, Application Layer from Network Packet	IT & OT Protocols: • Header metadata • Packet content OT Protocols: • Function Codes • Exception Codes • Device State	• Proprietary/unknown packet data • Data outside monitoring point

Our Proposed Contributions

Therefore, we propose <u>micro-baselines</u>: event-specific signatures within operational networks

1. **Construct behavioral baselines for specific operational events.**
   - Breaker open/close
   - Specific maintenance activity
   - Specific configuration change
2. **Evaluate the ability to repurpose baselines across different operational contexts:**
   - Geographic location
   - Time
   - Devices



Outline

1. Introduction/Motivation
2. Data Collection and Curation
3. Stateful Micro Baselines
4. Future Work
5. Conclusion

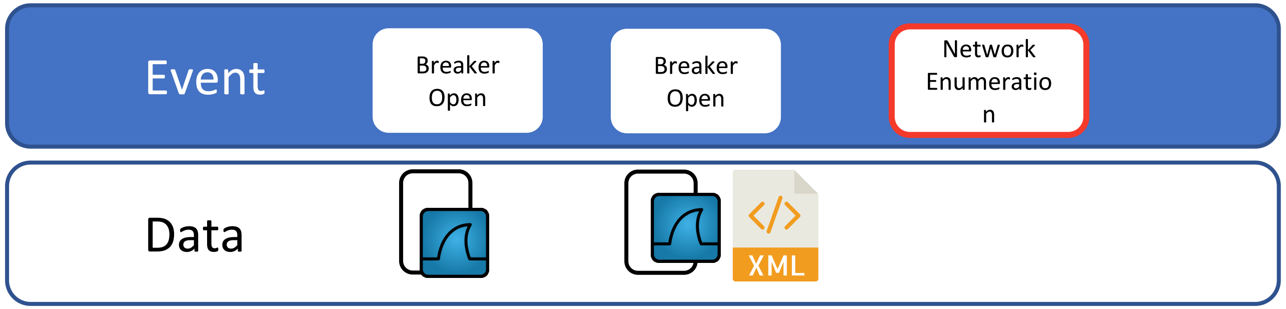
Idaho National Laboratory

Insane FORENSICS

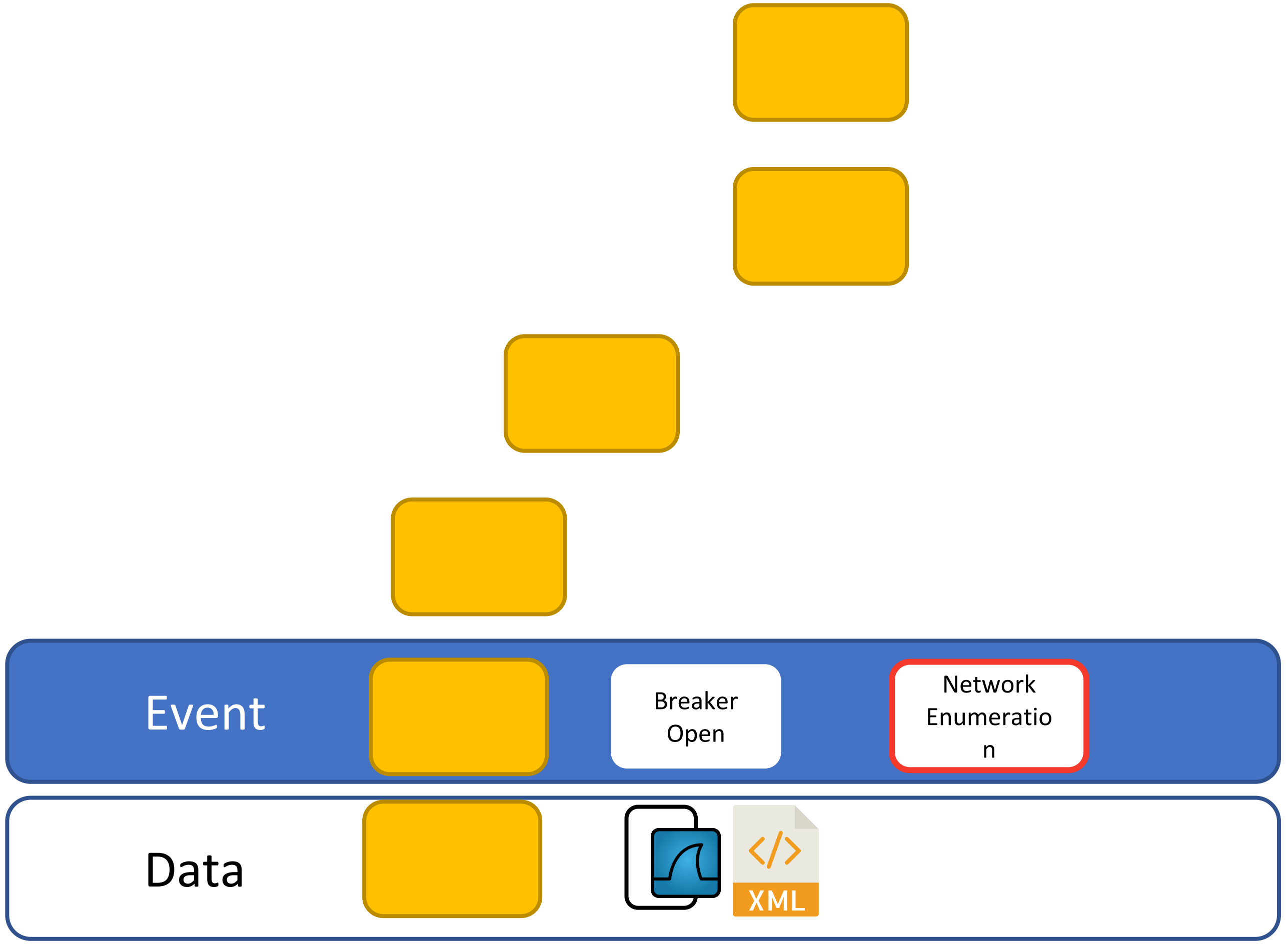
Micro-Baseline: Data Collection and Curation

Objective: Compare the same operational/adversarial events across different operational contexts.

- Collect network and host-level data for a given set of events.
  - Breaker open/close
  - Breaker trip
- Augment the collected data with attributes for *context metadata*.
  - Location (lat/lon, facility, device)
  - Time
- Construct event baselines relative to different data contexts.



Data Curation and Citation for Micro-Baselining



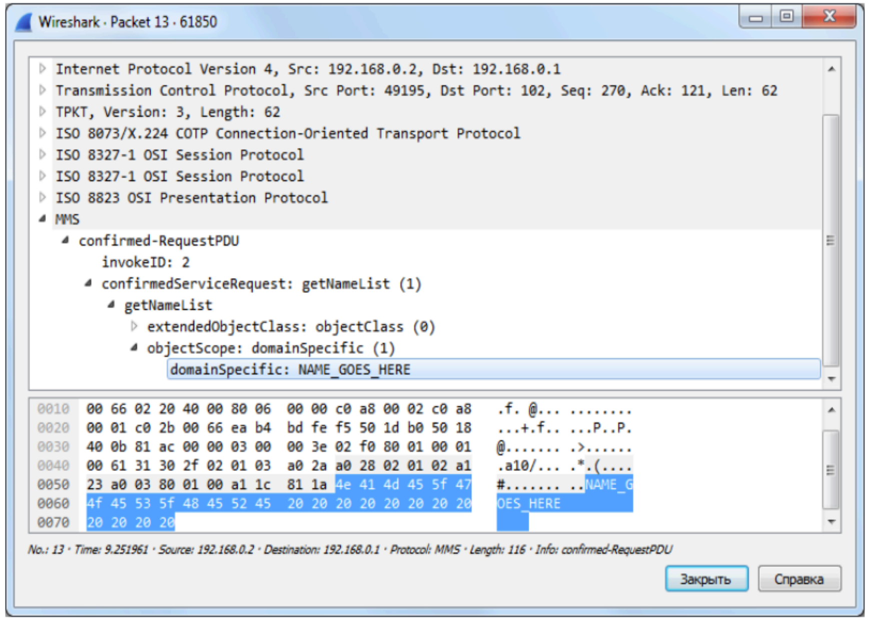
Event

Breaker Open

Network Enumeratio n

Data

Illinois.ComEd.Substation1.January.BreakerOpen.pcap1



```
▷ Internet Protocol Version 4, Src: 192.168.0.2, Dst: 192.168.0.1
▷ Transmission Control Protocol, Src Port: 49195, Dst Port: 102, Seq: 270, Ack: 121, Len: 62
▷ TPKT, Version: 3, Length: 62
▷ ISO 8073/X.224 COTP Connection-Oriented Transport Protocol
▷ ISO 8327-1 OSI Session Protocol
▷ ISO 8327-1 OSI Session Protocol
▷ ISO 8823 OSI Presentation Protocol
◢ MMS
  ◢ confirmed-RequestPDU
      invokeID: 2
    ◢ confirmedServiceRequest: getNameList (1)
      ◢ getNameList
        ▷ extendedObjectClass: objectClass (0)
        ◢ objectScope: domainSpecific (1)
            domainSpecific: NAME_GOES_HERE
```

```
0010  00 66 02 20 40 00 80 06  00 00 c0 a8 00 02 c0 a8   .f. @... ........
0020  00 01 c0 2b 00 66 ea b4  bd fe f5 50 1d b0 50 18   ...+.f.. ...P..P.
0030  40 0b 81 ac 00 00 03 00  00 3e 02 f0 80 01 00 01   @....... .>......
0040  00 61 31 30 2f 02 01 03  a0 2a a0 28 02 01 02 a1   .a10/... .*.(....
0050  23 a0 03 80 01 00 a1 1c  81 1a 4e 41 4d 45 5f 47   #....... ..NAME_G
0060  4f 45 53 5f 48 45 52 45  20 20 20 20 20 20 20 20   OES_HERE
0070  20 20 20 20
```

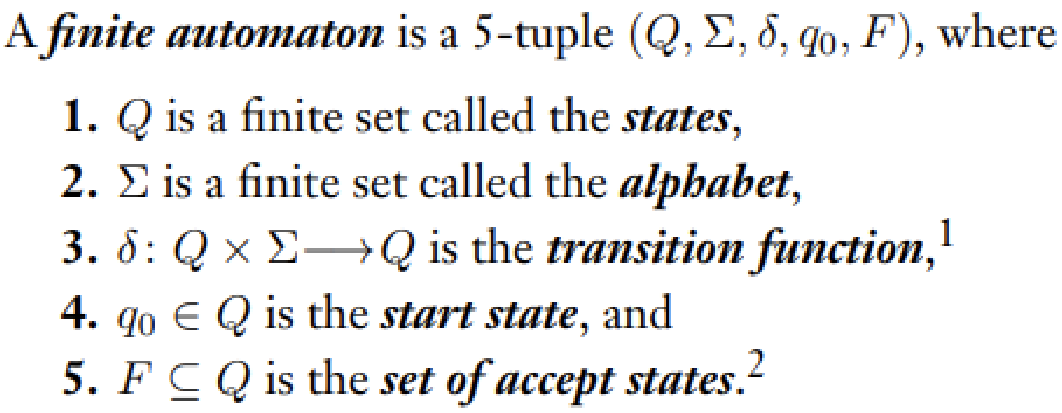
No.: 13 · Time: 9.251961 · Source: 192.168.0.2 · Destination: 192.168.0.1 · Protocol: MMS · Length: 116 · Info: confirmed-RequestPDU

Network Behavior and Finite State Machines

Objective: Represent the behavior of ICS protocols relative to specific operational events.

- Some ICS protocols already have a state machine defined (e.g. DNP3)

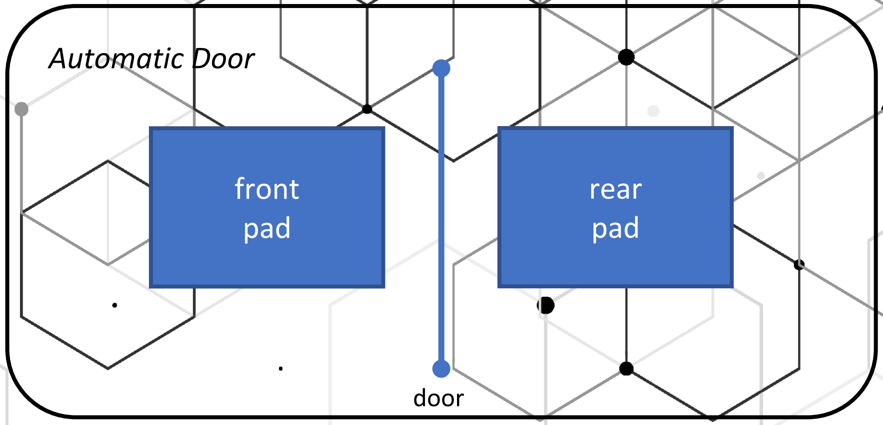
Approach: Use finite state machines to represent behavior.

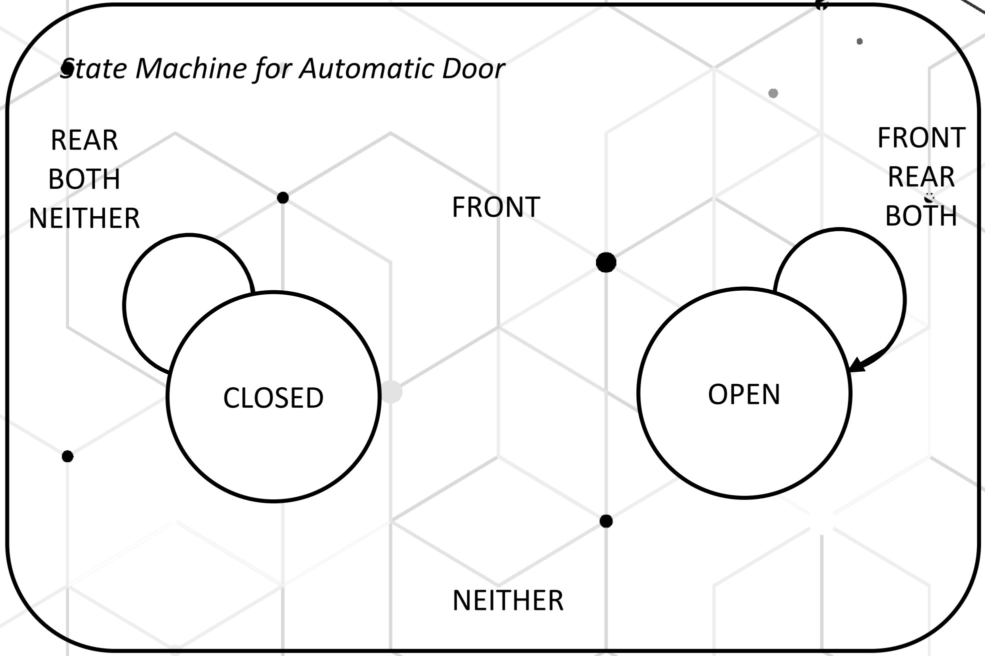


A *finite automaton* is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$, where

1. Q is a finite set called the ***states***,
2. Σ is a finite set called the ***alphabet***,
3. $\delta\colon Q \times \Sigma \longrightarrow Q$ is the ***transition function***,[1]
4. $q_0 \in Q$ is the ***start state***, and
5. $F \subseteq Q$ is the ***set of accept states***.[2]

[Sipser 2013]

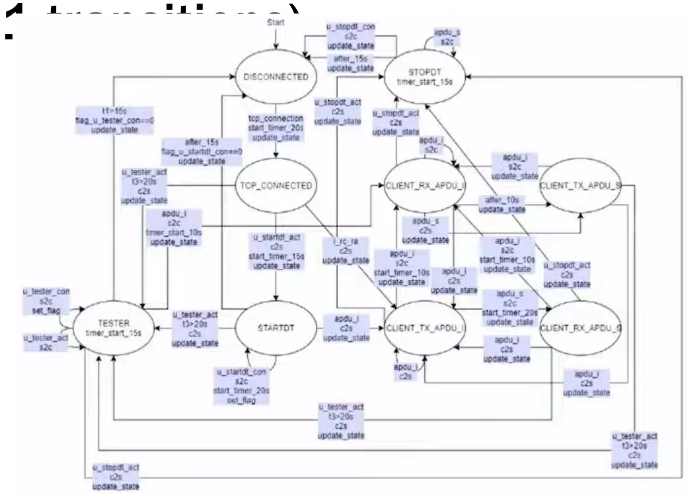




Previous Work: Baselining and Finite State Machines

Stateful Protocol Hunting

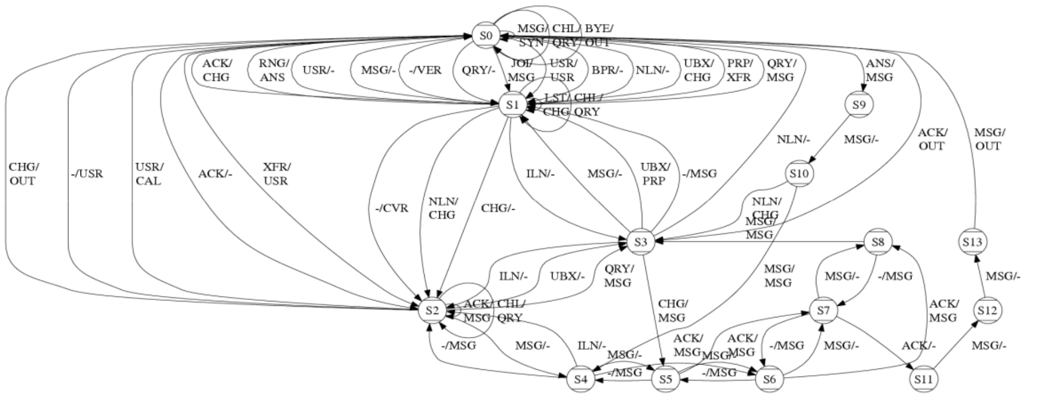
- Analyze malware packet captures based on state sequences/misses [Gunter et al. 2019]
- Noted that real-world attacks (CRASHOVERRIDE, Stuxnet) inject packets out of state.
    - Applied to IEC 104 Protocol Specification
    - Manual build of state machine (9 states, 31 transitions)



State Machine Inference

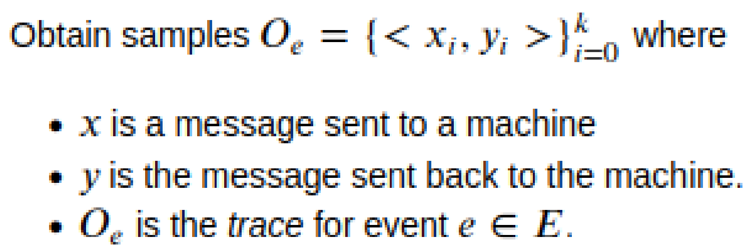
Synthesize a protocol from I/O behavior.

- Passive Synthesis
    - Given a set of network traces, infer a finite state machine.
    - Constructing a minimized FSM that contains a set of traces and only those traces is NP-hard. There are polynomial time algorithms to construct a reduced (not minimized) FSM that contains a set of traces and only those traces. [Hsu et al. 2008]
    - Demonstrate approach on MSN Instant Messaging Protocol, 14 states, 48 transitions



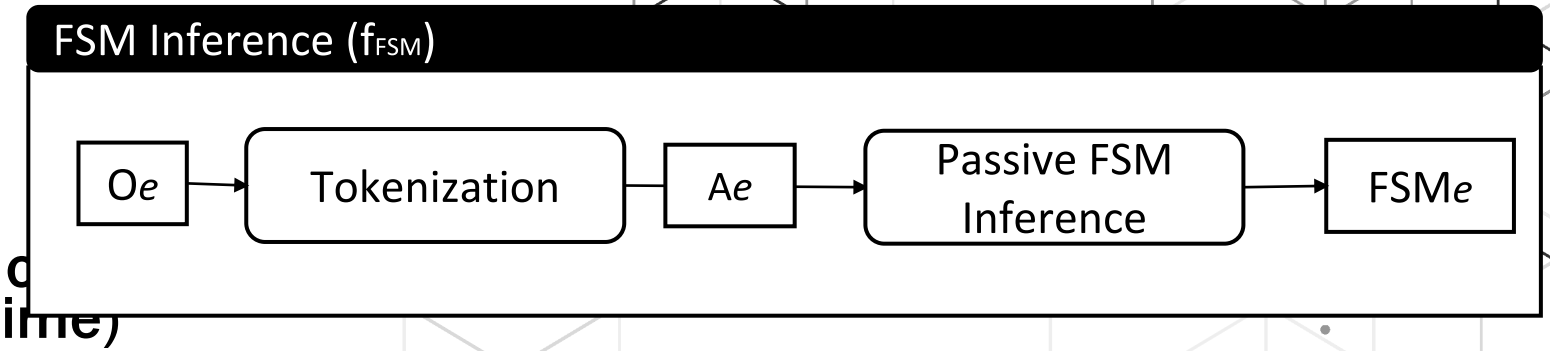
Proposed Work: Micro Baselines

- Use passive synthesis to infer a finite state machine for a specific operational event (*e*).

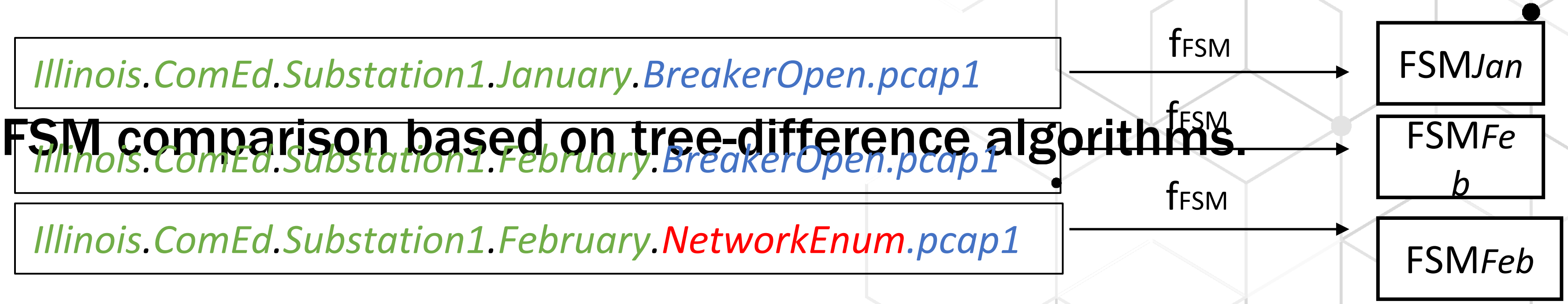


Obtain samples $O_e = \{< x_i, y_i >\}_{i=0}^{k}$ where

- x is a message sent to a machine
- y is the message sent back to the machine.
- O_e is the *trace* for event $e \in E$.



- (c
contexts (e.g. substations over time)



Illinois.ComEd.Substation1.January.BreakerOpen.pcap1
Illinois.ComEd.Substation1.February.BreakerOpen.pcap1
Illinois.ComEd.Substation1.February.NetworkEnum.pcap1
f_{FSM}
f_{FSM}
f_{FSM}
FSM*Jan*
FSM*Feb*
FSM*Feb*

- Potential FSM comparison based on tree-difference algorithms.

Potential Data Sources for Network Baselining

Sector	Type	Layer	Source	Context	Data Format
Communications	Network	Physical	Utility Device Inventory	Utility/Facility	PCAP, SCL
		Data Link	Device Interfaces and Links	Facility	PCAP, SCL
		Network	Device IP Addresses and Routes	Facility	PCAP, SCL
		Transport	Service Ports	Facility	PCAP, nmap
	Flows	Data Link	Layer 2 Frames Interface Statistics	Facility/Device	PCAP, bmon
		Network	Layer 3 Packets	Facility	PCAP
		Transport	TCP Connections	Facility/Device	PCAP, netstat
		Application	Application semantics	Facility/Device	PCAP

[Weaver et al. 2021]

Next Steps

- Explore data sources to evaluate micro-baselines in context.
    - Software Defined Networks (SDN)
    - Hardware Testbeds
    - Digital Twins
    - Datasets from previous exercises

- Construct an initial use case, based on DNP3, to prototype FSM inference engine.

- Evaluate data drift across different contexts to understand ability to apply baselines to other facilities.

- Explore the ability to use the FSM to generate samples of adversarial behavior as part of imbalanced multiclass identification problem.

[via https://inl.gov/ics-celr/]

For questions contact:

Gabriel Weaver, CyOTE Precursor Analysis Program Analyst, Gabriel.Weaver@inl.gov

@DOE_CESER

linkedin.com/company/office-of-cybersecurity-energy-security-and-emergency-response

energy.gov/CESER

This presentation was prepared by Idaho National Laboratory (INL) under an agreement with and funded by the U.S. Department of Energy. INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.