May 25, 2023

Gabriel A. Weaver
Senior Critical Infrastructure Analyst

Dan Gunter
CEO & Founder of Insane Forensics

Language-Theoretic Data Collection to Support ICS

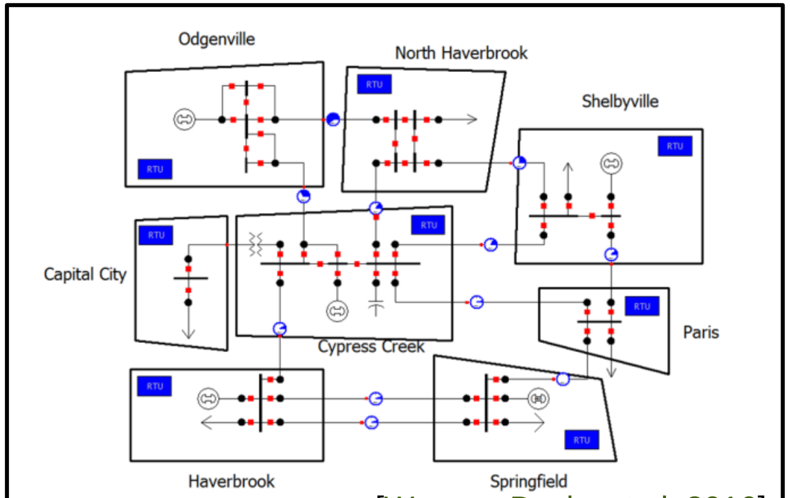
Introduction

An **Industrial Control System** (ICS) is an information system used to control industrial processes.

Characteristics of ICS networks that **suggest alignment with LangSec** include:

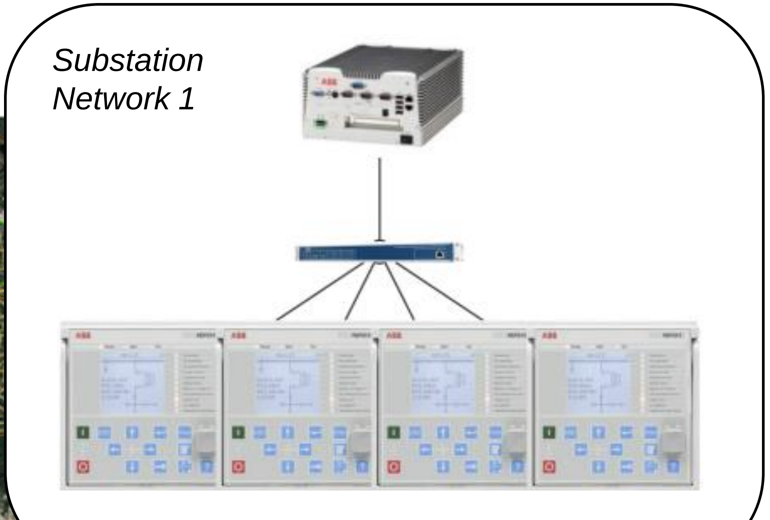
- Distributed across broad geographic regions (*secure communication boundaries)*
- Diversity of devices and protocol implementations *(mutually-intelligible dialects*)

We argue that **ICS provide a potentially rich application domain for Language-Theoretic Security** (LangSec)



[Weaver, Davis, et al. 2016]

[Fletcher et al. 2020]



[Weaver, Yardley, Emmerich 2021]

Approach

ICS provide a potentially rich application domain for Language-Theoretic Security LangSec Applications.

Hypothesis: ICS **processes generate artifacts** expressed across heterogeneous data sources; these **artifacts form a language** in the language-theoretic sense.

Three real-world problems based on engagement to-date with industry:

1. *Fusion of Network and Device Data* with Grammars
2. ICS *Device Fingerprinting* via Language Dialects
3. *System Baselining* with Security Automata

Intent: Foster discussion of these approaches/problems benefits the LangSec community.

Grammars for Data Fusion

Problem The need to fuse network and device data is an acknowledged industry gap.

- Integration of OSISoft PI Historian [Johnson-Barbier and Gunter 2019]
- Acknowledged need to analyze ICS protocols relative to high-level primitives [McFail 2022 (via Tsamis)]

Approach

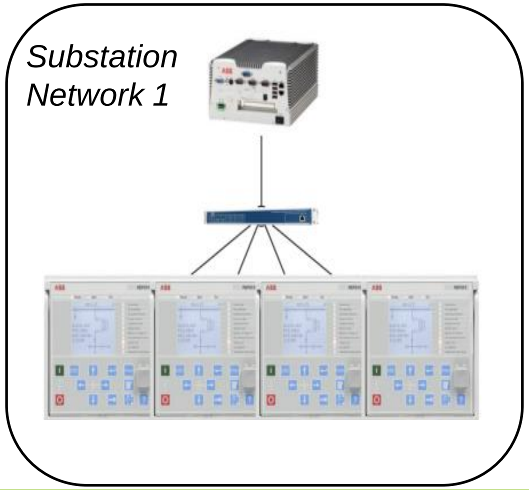
- Can we use deterministic, context-free grammars as a tool to analyze security artifacts?
- Use non-terminals to group different expressions of the same procedure (within and across multiple protocols).

Impact

- Formal **specification of** data sources that provide **features to enable** system-wide **fingerprinting and baselining**
- Multiple sources of data to **mitigate** informational **single points of failure**

Example Breaker Open

- DNP3:
  `{DIRECT OPERATE}`;
  `{SELECT, OPERATE}`
- Modbus:
  Read/write to registers



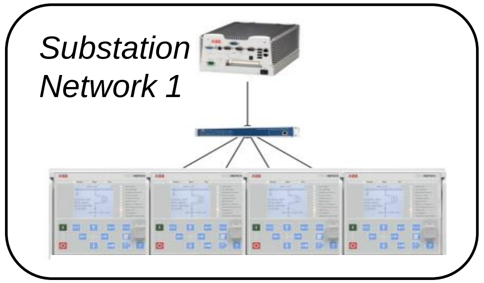
Language Dialects for Device Fingerprinting

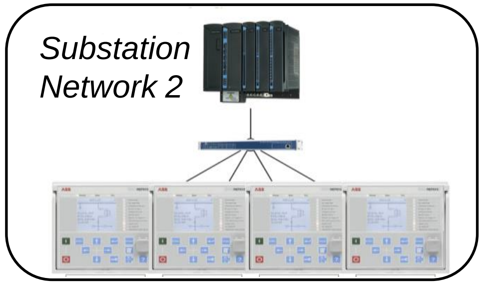
Problem

- ICS protocols are **notorious for being poorly implemented** relative to their standard
- Differences in the IEC 61850 protocol communications can identify specific vendor product lines [Brizinov 2022 via SANS]
- Parse tree differential analysis may enhance fingerprinting based attacks [Sassaman et al. 2013]

Approach

- Look at differences across various implementations of ICS protocols and use that to fingerprint systems.





Impact

- Although a **diversity** of vendors and devices can **help** asset owners **fingerprint** devices and **avoid a monoculture**…
- This diversity **may** also **introduce vulnerabilities** due to differences among mutually-intelligible protocol implementations.

Example

Adversaries may implement industrial protocols differently (Industroyer/CrashOverride) [Gunter and Michaud-Soucy 2019]

Security Automata and System Baselining

Problem

Current approaches to network baselining rely on generally-available observables.

- srcIP, dstIP, etc.
- Basic Asset ID Obsessed
- IT Connection metadata centric

Low-level observables:

- no semantics on business process or environmental context
- lack properties upon which traditional statistical tools depend [Schulz et al. 2019]

Approach

Are there practical approaches to construct security automata? (e.g. [Schneider 2000])

Finite State Machine Inference

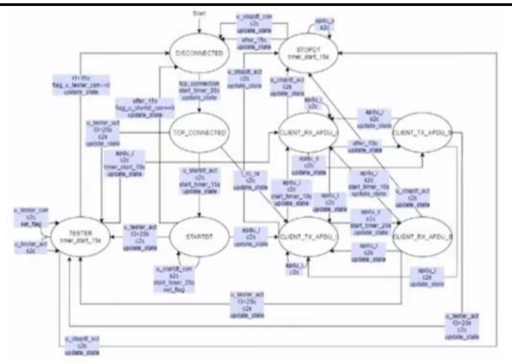
- Passive FSM inference: Given a set of traces, produce a state machine.
- In general, NP hard, but applied successfully to real-world problems:
  - botnet c2 protocols
  - modeling microservices in Kubernetes

Impact

- Researchers have noted while discussing the **problems with IDS in general**, there is **promise in approaches such as security automata** for application-specific security policies [Sassaman et al. 2013].

Example

- Analyze malware packet captures based on state sequences/misses
- Device modes as states



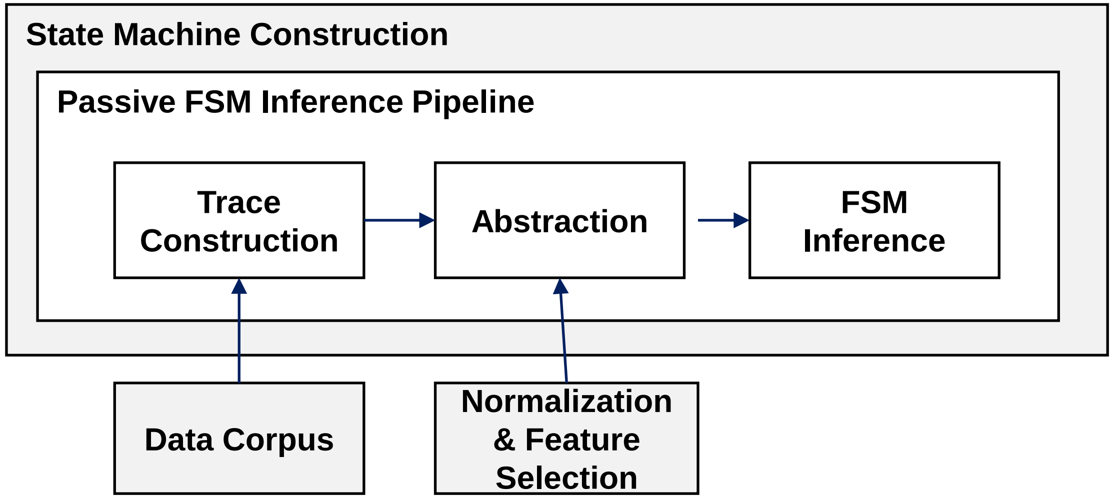
Conclusion

ICS provide a potentially rich application domain for Language-Theoretic Security LangSec Applications.

Hypothesis: ICS **processes generate artifacts** expressed across heterogeneous data sources; these **artifacts form a language** in the language-theoretic sense.

Three real-world problems based on engagement to-date with industry:

1. *Grammars for Data Fusion*
2. Language Dialects for Device Fingerprinting
3. *Security Automata for System Baselining*



Idaho National Laboratory