Real time data analytics for cyber-physical grid security

Anna Scaglione
(Anna.Scaglione@asu.edu)
Motivation

• In July this year DHS warned:

“.. hackers accessed utility networks by compromising vendor companies who do business with the utilities, officials said. The attacks began in 2016, continued through last year and are ongoing.”


• big complex infrastructures are porous...
The interaction of *open interoperable systems* with the physical world defines the cyber-security threat.
Common Threats

- Denial of Service attacks
- Unauthorized access
- Man-in-the-middle attacks
- Rogue device installation
- Malicious software patch
- Data injection
- Spoofing

Pillars of CPS Security

- Availability
  - Ultra-reliable communications, delay
  - Secure timing and location
- Physical Security
  - Managing access to control physical devices
- Confidentiality (privacy)
  - Managing access to sensors
- Situation awareness
  - Intrusion analysis
Preventive Actions

- **Recommended Practices:**
  - Centralized user authentication
  - End-to-End Secure Delivery of Messages/Encryption
  - Scalable Key Management
  - Secure Software Patching
  - Tamper-resistant Credential Protection

- There have been efforts to add security to legacy systems by default (e.g. IEC 62351, NERC CIP)
  - authentication of data transfer through digital signatures
  - ensuring only authenticated access
  - prevention of eavesdropping
If systems are breached?
Intrusion Detection Systems (IDSs)

Traditional network IDSs:
- data analysis tools monitoring the computer data provenance and network traffic in search for suspicious activities

CPS network IDS
- analyzes also the behavior of the embedded physical components and physical environment
  - under attack they display unexpected physical properties

Examples of IDS software used in industrial control:
- Tofino Security Appliance
- Digital Bond Quickdraw
- Radiflow Secure Gateway
- Bro Security Framework
CPS Intrusion Detection components

- Detection technique
  - Knowledge based
  - Behavior based
    - Behavior specification based

- Audit material
  - Host based
  - Network based

Defines data analytics that detect misbehavior of a physical component and its cause.

Defines "how" the IDS collects data for the data analysis.
Knowledge-based Detection

- Looks for runtime features that *match a specific pattern of misbehavior*

- **Advantage:**
  - Low FPR (false positive rates)

- **Disadvantage:**
  - Need for an up-to-date dictionary that specify each attack vector
  - Unfortunately many attacks to CPS are “zero-day”
Behavior-based Detection

• Look for runtime features that are out of ordinary

• Defining ordinary

• Advantage:
  ▪ No need for attack dictionary

• Disadvantages:
  ▪ Susceptibility to false positive
  ▪ Need for training/profiling phase

• Further Classification

  Unsupervised: based on the history of the test signal and the system physics

  Semi-supervised: based on a collection of real data

  Statistics-based approaches

  Non-parametric methods (e.g., data clustering, support vector machines)
Behavior-specification-based Detection

• **Possible unsupervised approach:** detect intrusion when the system departs from its physical and its normal operational model

• **Advantages:**
  ▪ Low False Negative Rates
  ▪ No need for training phase
  ▪ *Exploits physics and known reliability limits*

• **Disadvantage:**
  ▪ generating formal specifications can be difficult, particularly for networked systems
What data? SCADA vs PMU data

• SCADA updated at best every 3-5 seconds: some events missed
• Do not read the state ➔ Data injection attacks at the device level can mislead physics-aware IDS

A possible solution:
• Combining high resolution synchrophasor data that reflects the physical state of the grid as input for IDS
• Each PMU reading can be utilized in a stand alone fashion --> it is a reading of the state
### Audit Material

<table>
<thead>
<tr>
<th>Host-Based</th>
<th>Network-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Analyze logs maintained by a node</td>
<td>• Analyze network activity fusing all information</td>
</tr>
<tr>
<td>• Advantages:</td>
<td>• Advantage:</td>
</tr>
<tr>
<td>▪ Distributed control</td>
<td>▪ Individual nodes are free of the requirement to maintain/analyze their logs</td>
</tr>
<tr>
<td>▪ Ease of specifying/detecting host-level misbehavior</td>
<td>• Disadvantage:</td>
</tr>
<tr>
<td>• Disadvantages:</td>
<td>▪ Slower response</td>
</tr>
<tr>
<td>▪ Additional node workload</td>
<td>▪ Slower, hard to configure and audit sensors to get complete activity of the network</td>
</tr>
<tr>
<td>▪ Can be mislead by manipulated logs at the device level</td>
<td>▪ Additional cost in CPS IDS</td>
</tr>
<tr>
<td>▪ Can be OS or application specific</td>
<td></td>
</tr>
</tbody>
</table>
Combined benefits: Fog computing

Cloud computing
- a host for co-analysis of data from multiple sensors
- event localization and categorization (natural vs malicious anomaly)

Fog computing
- near real-time analysis of data (1 sec has achieved),
- prioritizing communication of eventful segments,
- security analysis with partial information,
- blind to network topology
- scalability

Figure: Intrusion Detection Architecture.
Low measurement regime

- Local rules for Fog computing have one or very few data sources to detect events
- Example: Rule for detecting a transient
  - The low rank properties of PMU data hold only in a quasi-steady state condition

IEEE-39 bus New England test case, with a fault introduced at line 16-19, lasting for .1. Anomaly recorded locally (top) and with 60% of PMUs
Localization

- Fault localization through few PMUs → loss in resolution

The data are an "under-sampled graph signal"

IEEE 34 buses test case
Blockchain: distributed and secure intelligence

- Trust no-one: Every node needs to agree with a common version of the facts

**Difference with other blockchains?**

1. Adversarial model
2. Non-binary validity
3. Validity enforcement

*The P2P distributed intelligence validates the data through non binary consensus*

*Validated data are recorded in the distributed ledger*

*Data are broadcast to the P2P network*

*Blockchain data inform control and decision functionalities*

*Stabilize  Control & Protect  Optimize*
Summary

• Real time data analytics are essential to provide situation awareness
  ▪ This is still missing in recommended best practices
• Behavior based detection can leverage basic reliability and physical knowledge of the system
• How we audit is important for real time awareness
  ▪ Fog computing – rules that rely on few measurements and local processing
• This may lead to secure network data exchanges: “fog” rules used to chain information in a distributed ledger (a new blockchain)