Cyber-Physical Resilient Systems

From Malware & Operational Security
to Feedback Truthfulness Distinguishability

Joaquin Garcia-Alfarro

Institut Mines-Télécom (Télécom SudParis)
& Université Paris-Saclay
Today’s Talk: Cyber-Physical Resilience

• Cyber-Physical Systems*
  – ICT components monitoring & controlling physical resources
  – Physical & ICT elements that interact with humans

* H. Gill, National Science Foundation, 2006.
Today’s Talk: Cyber-Physical Resilience

Subtitle was:

From Malware & Operational Security to Feedback Truthfulness Distinguishability
Malware & Operational Security

**How Stuxnet Worked**

1. **Infection**
   Stuxnet enters a system via a USB stick and proceeds to infect all machines running Microsoft Windows. By brandishing a digital certificate that seems to show that it comes from a reliable company, the worm is able to evade automated-detection systems.

2. **Search**
   Stuxnet then checks whether a given machine is part of the targeted industrial control system made by Siemens. Such systems are deployed in Iran to run high-speed centrifuges that help to enrich nuclear fuel.

3. **Update**
   If the system isn’t a target, Stuxnet does nothing; if it is, the worm attempts to access the Internet and download a more recent version of itself.

4. **Compromise**
   The worm then compromises the target system’s logic controllers, exploiting “zero day” vulnerabilities—software weaknesses that haven’t been identified by security experts.

5. **Control**
   In the beginning, Stuxnet spies on the operations of the targeted system. Then it uses the information it has gathered to take control of the centrifuges, making them spin themselves to failure.

6. **Deceive and Destroy**
   Meanwhile, it provides false feedback to outside controllers, ensuring that they won’t know what’s going wrong until it’s too late to do anything about it.
Malware & Operational Security

Ransomware takes aim at providers
Healthcare is one of 4 industries hit by 77% of all attacks

- Business/professional services, 28%
- Government, 19%
- Healthcare, 15%
- Retail, 15%
- Other, 23%

Source: NTT Security, a

*Letters represent organizations compromised.
In addition to malware ...

- Malware moving from IT Systems to Operational Systems

- Wrong configurations, lack of encryption, legacy (vulnerable) systems, intentionality...
IT & OT together ...

Asset to protect: Information

<table>
<thead>
<tr>
<th>Priority</th>
<th>IT Systems</th>
<th>MTUs to I/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Confidentiality</td>
<td>Availability</td>
</tr>
<tr>
<td>#2</td>
<td>Integrity</td>
<td>Integrity</td>
</tr>
<tr>
<td>#3</td>
<td>Availability</td>
<td>Confidentiality</td>
</tr>
</tbody>
</table>

Plus
- Reliability,
- Safety,
- Performance, ...

Dynamic Risk Assessment example

- Prevent threats (e.g., preempt exploitation of vulnerabilities)
- Use of Attack & Mission Graphs to support network administrators towards semi-automated decisions

IT Security Oriented

OT Security Oriented

http://j.mp/DRDMS
Outline

• Experience & Context

  • Cyber-Physical Systems

• Feedback Truthfulness (FT)

• Ongoing Work on FT Distinguishability

• Summary & Perspectives
The key ingredient in a CPS: Control

- **Control** means making a (dynamical) system to work as required

- **Feedback** is used to compute a corrective **control action** based on the distance between a *reference signal* and the *system output*

![Diagram of control system](image)

- Examples: dynamically follow a trajectory (robotics), regulate a temperature, regulate the sending rate of a TCP sender (TCP cong. control), controlling a pendulum in its unstable equilibrium, etc.
Networked Control System

- From a methodological standpoint, we can model a CPS using a Network Control System (NCS)

**NCS definition**

Control system whose control loops are connected through a communication network
Traditional Issues Studied in the NCS Literature

- Stabilizing a system under network delays & packet losses
- Techniques to limit data rate (e.g., from control to plant)
- Energy efficient networking for Wireless NCS
- Security?
  - Since the *stuxnet* incident, the control community seems to be heavily working as well on security issues of NCSs & CPSs
  - Control-theoretic security taxonomies?
Sample Attacks*

Replay Attack

- Step 1: Sensors output is recorded
- Step 2: Recorded sensors output is replayed and sent to the controller
- Step 3: A control signal is sent to disrupt system functionalities
Prevention & Mitigation of CPS Attacks

• A well-designed control system shall resist external disturbances (failures & attacks), to a certain degree

• Several control-theoretic techniques to prevent cyber-physical attacks have been proposed in the literature*

• Most of the techniques aim at injecting authentication to the control signal & discover anomalous measurements
  - E.g., use a noisy control authentication signal to detect integrity attacks on sensor measurements
  - In the following, we elaborate further on the aforementioned technique

Watermark Approach by Mo et al.

Idea [Mo et al., 2009, 2015]

Adding a watermark signal to the control signal which serves as an authentication signal

- Conceptually similar to a challenge-response authentication scheme
- In this case the watermark is the challenge the response is the sensor output

Main advantages:
- Only the controller has to be changed
- It does not require encryption

In a nutshell ...

- **Challenge-Response** (slight modification of normal behavior w.r.t. system dynamics)

- Control Theory & LTI models (*linear time invariant models*)

- Challenge: $u_t$ ; Response: $y_t$

- Then, statistical analysis w.r.t. $u_t$ & $y_t$:

$$g_t = \sum_{i=t-w+1}^{t} (y_i - C\hat{x}_{i|i-1})^T P^{-1} (y_i - C\hat{x}_{i|i-1})$$

- If $g_t$ exceeds the threshold $\sim$ raise alert

Initial Motivations

• Malware moving from IT Systems to Operational Systems

• Wrong configurations, lack of encryption, legacy (vulnerable) systems, third party access, ...

Proposed Methodology

• Foster new theoretical models,
• simulate/emulate case scenarios,
• validate results using training & testbeds
Preparing the Testbeds

WAGO I/O system 750-842 750-402
750-404 750-559 750-600
From China
Top-rated seller

Siemens S7 300 PLC Trainer, 8 inputs 8 outputs USB/MPI
Learn how to program Siemens PLC's, NO Software
Top-rated seller

LEGO Mindstorms EV3 Intelligent Brick # 95646c01
Brand New
Top Rated

http://j.mp/1lEAxDP
http://j.mp/1vGPIVp
http://j.mp/1qViIsG
SCADA Protocols (non exhaustive list)

• Siemens quad 4 meter
• CONITEL 2000
• CONITEL 2100
• CONITEL 3000
• CONITEL 300
• HARRIS 5000
• HARRIS 5600
• HARRIS 6000
• UCA 2.0 or MMS
• PG & E 2179
• MODBUS
• DNP3
• IEC 61850
• ...

Sample protocols

• MODBUS - Primitive with no security and not very extensible
• DNP3 – Advanced SCADA protocol
  - DNP1 and 2 are proprietary protocols
Sample Testbeds

(a) Bridge and toll testbed

(b) Industrial chain testbed

(c) Railway control testbed

(d) Autonomous industrial agents testbed

http://j.mp/TSPScada
Sample Testbed (autonomous agents testbed)

Attacks & Adversaries Implemented

- **Replay Attack**
  - Watermark Disabled
  - Watermark Enabled

- **Non-parametric Attack**
  - Stationary Watermark
  - Non-stationary Watermark

- **Parametric Attack**
  - Stationary Watermark
  - Non-stationary Watermark

- **New Parametric Attack**
  - PIETC-WD strategy

Cyber-physical industrial scenario implemented in the testbed

http://j.mp/TSPScada
Linear & Polar Representation

Normal Mode

Under Attack
Testbed Validation

- Modeled as games?
  - http://j.mp/WikiGTP

- Defender
  - Avoid collisions

- Attacker
  - Force collisions

http://j.mp/TSPScada
Outline

• Experience & Context
  • Cyber-Physical Systems
  • Feedback Truthfulness (FT)

• Ongoing Work on FT Distinguishability

• Summary & Perspectives
Feedback Truthfulness Distinguishability

- Distinguishing accidental failures and intentional manipulation
- Top-down refinement of automated runtime verification
Feedback Truthfulness Distinguishability

(1) System Dynamics

\[
\begin{align*}
\dot{x}_i &= v_i \\
\dot{v}_i &= f_i(v_i, u_i), \quad i = 1, 2.
\end{align*}
\]

(2) Threat Models

(3) Synthesis & Refinement

(4) Controllers & Artifacts

Adversary

Controllers

Network, system, sensors & actuators
Feedback Truthfulness Distinguishability

(3) Synthesis & Refinement

(4) Controllers & Artifacts

(1) High-level Abstractions

(2) Adversary Intentions

\[ \sigma \leq v_1 \leq \omega \]
\[ \sigma \leq v_2 \leq \omega \]
\[ x_2 - x_1 > \tau \]

\[ 0 \leq v_1 < \sigma \]
\[ \sigma < v_2 \leq \omega \]
\[ x_2 - x_1 \leq \tau \]
Outline

• Experience & Context
  • Cyber-Physical Systems
• Feedback Truthfulness (FT)
• Ongoing Work on FT Distinguishability
• Summary & Perspectives
Summary

• Challenging, multidisciplinary topic
  - Dynamic (networked-control) systems & data truthfulness

• Traditional ICT-based security may still be applicable
  - However, they cannot solve the problem completely
    - Fundamental differences between IT systems & CPSs

• Modeling, from a control-theoretic perspective, shall
  - Pay attention to adversary strategies from the attacker’s angle
  - Assume attackers with knowledge about information systems & physical systems at the same time

• Perspectives
  - Automated techniques for the verification of feedback truthfulness distinguishability is a must
References


• Krotofil & Larsen. Hacking Chemical Plants for Competition and Extortion, DefCon23, 2015


