Physics-Based Methods for Distinguishing Attacks from Faults

Gregory Provan, Riccardo Orizio

University College Cork

September 2017

Outline

Outline

Motivation

1. Cyber-Physical Systems (CPSs) are of great interest due to the wide application area where their model can be used.
2. System security and attacks detection can be studied through CPS models.
3. **Goal**: Detect and distinguish attacks from faults on a complex system using CPS models.

Motivation

1. Cyber-Physical Systems (CPSs) are of great interest due to the wide application area where their model can be used.
2. System security and attacks detection can be studied through CPS models.
3. **Goal**: Detect and distinguish attacks from faults on a complex system using CPS models.

Motivation

1. Cyber-Physical Systems (CPSs) are of great interest due to the wide application area where their model can be used.
2. System security and attacks detection can be studied through CPS models.
3. **Goal**: Detect and distinguish attacks from faults on a complex system using CPS models.

Contributions

1. Method for distinguishing attacks from faults on an observed-based framework.
2. Physics-based methods can be effective, but they cannot deal with every kind of attacks on the system.
3. Demonstrate approach on hydraulic benchmark system.

Contributions

1. Method for distinguishing attacks from faults on an observed-based framework.
2. Physics-based methods can be effective, but they cannot deal with every kind of attacks on the system.
3. Demonstrate approach on hydraulic benchmark system.

Contributions

1. Method for distinguishing attacks from faults on an observed-based framework.
2. Physics-based methods can be effective, but they cannot deal with every kind of attacks on the system.
3. Demonstrate approach on hydraulic benchmark system.

Outline

Preliminaries

- We assume that a CPS model is an instance of a hybrid system which can operate on different behaviours, called modes.

$$Modes : \begin{cases} y_{m_1} = g_1(x) \\ ... \\ y_{m_i} = g_i(x) \end{cases}$$

- i.e. a drone has many operating modes: take-off, landing, wandering, surface mapping, ...
- The set of modes include also faults/attacks behaviour.

Preliminaries

- We assume that a CPS model is an instance of a hybrid system which can operate on different behaviours, called modes.

$$Modes : \begin{cases} y_{m_1} = g_1(x) \\ ... \\ y_{m_i} = g_i(x) \end{cases}$$

- i.e. a drone has many operating modes: take-off, landing, wandering, surface mapping, ...

- The set of modes include also faults/attacks behaviour.

Preliminaries

- We assume that a CPS model is an instance of a hybrid system which can operate on different behaviours, called modes.

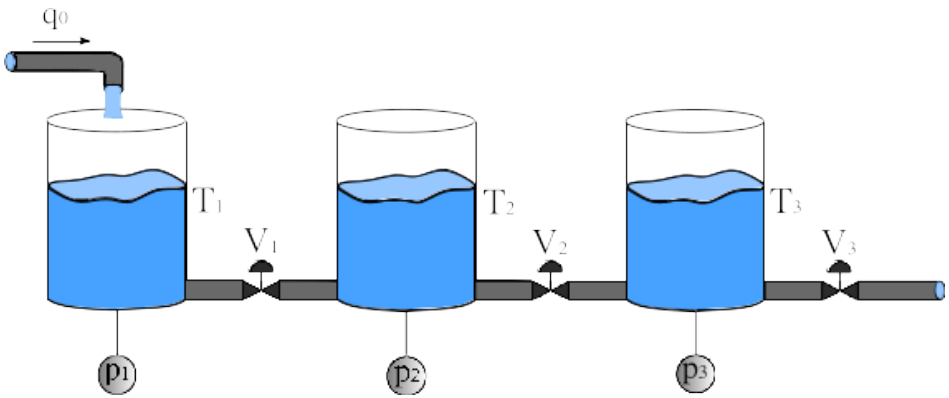
$$Modes : \begin{cases} y_{m_1} = g_1(x) \\ ... \\ y_{m_i} = g_i(x) \end{cases}$$

- i.e. a drone has many operating modes: take-off, landing, wandering, surface mapping, ...

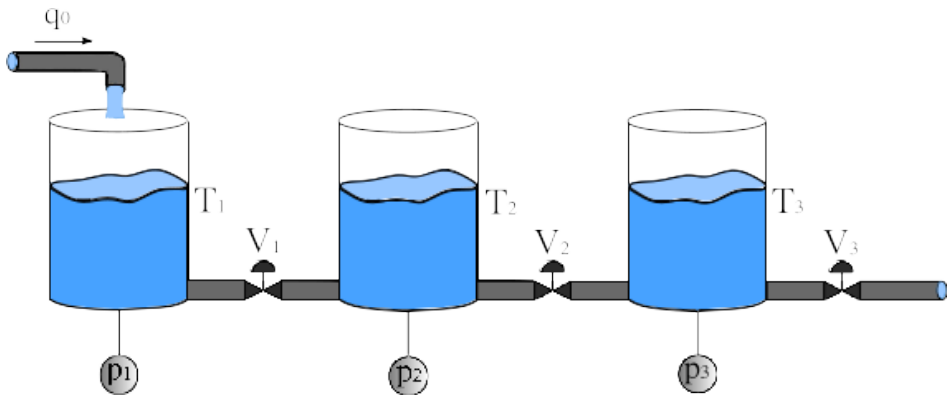
- The set of modes include also faults/attacks behaviour.

Outline

Nominal Model



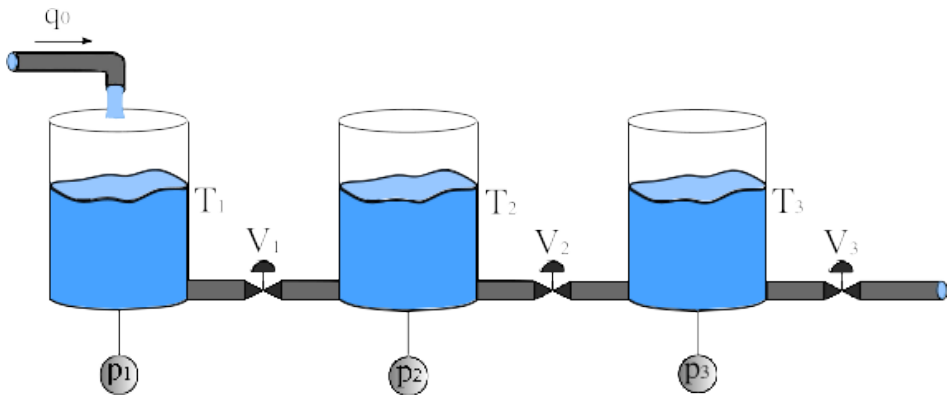
Nominal Model



$$\frac{\delta h_1}{\delta t} = q_0 - q_1 = \frac{q_0 - k_1 sign(h_1, h_2)\sqrt{|h_1 - h_2|}}{A_1}$$

$$\frac{\delta h_i}{\delta t} = \frac{k_{i-1} sign(h_{i-1}, h_i)\sqrt{|h_{i-1} - h_i|} - k_i\sqrt{h_i}}{A_1}$$

Nominal Model



$$\frac{\delta h_1}{\delta t} = q_0 - q_1 = \frac{q_0 - k_1 sign(h_1, h_2)\sqrt{|h_1 - h_2|}}{A_1}$$

$$\frac{\delta h_i}{\delta t} = \frac{k_{i-1} sign(h_{i-1}, h_i)\sqrt{|h_{i-1} - h_i|} - k_i\sqrt{h_i}}{A_1}$$

Input: $u = \{q_0, v_1, v_2, v_3\}$ Output: $y = \{p_1, p_2, p_3\}$

Faults Model

- Valve faults, leaks, sensor faults, etc..
- Additive model:

$$v_i = \begin{cases} \max\{0, v_i + \Delta_{v_i}\}, & \text{if } \Delta_{v_i} \leq 0 \\ \min\{1, v_i + \Delta_{v_i}\}, & \text{if } \Delta_{v_i} > 0 \end{cases}$$

where $\Delta_{v_i} \in [-1, 1]$

Faults Model

- Valve faults, leaks, sensor faults, etc..
- Additive model:

$$v_i = \begin{cases} \max\{0, v_i + \Delta_{v_i}\}, & \text{if } \Delta_{v_i} \leq 0 \\ \min\{1, v_i + \Delta_{v_i}\}, & \text{if } \Delta_{v_i} > 0 \end{cases}$$

where $\Delta_{v_i} \in [-1, 1]$

Attacks Model

- The attacker cannot monitor the system, only data injection.
- **Sensor:** fake sensor reading in $[0, p_i^{max}]$
- **Actuator:** fake actuator position in $[0, 1]$

Attacks Model

- The attacker cannot monitor the system, only data injection.
- **Sensor:** fake sensor reading in $[0, p_i^{max}]$
- **Actuator:** fake actuator position in $[0, 1]$

Attacks Model

- The attacker cannot monitor the system, only data injection.
- **Sensor:** fake sensor reading in $[0, p_i^{max}]$
- **Actuator:** fake actuator position in $[0, 1]$

Control Model

- Nominal system model:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + w_k \\ y_k &= C_\gamma x_k + v_k \end{aligned}$$

- Observer model:

$$\begin{aligned} \hat{x}_{k+1} &= A_\gamma \hat{x}_k + B_\gamma u_k + L_\gamma (y_k - C_\gamma \hat{x}_k) \\ \hat{y}_k &= C_\gamma \hat{x}_k + v_k \\ r_k &= y_k - C_\gamma \hat{x}_k \\ u_k &= -K_\gamma \hat{x}_k \end{aligned}$$

Control Model

- Nominal system model:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + w_k \\ y_k &= C_\gamma x_k + v_k \end{aligned}$$

- Observer model:

$$\begin{aligned} \hat{x}_{k+1} &= A_\gamma \hat{x}_k + B_\gamma u_k + L_\gamma (y_k - C_\gamma \hat{x}_k) \\ \hat{y}_k &= C_\gamma \hat{x}_k + v_k \\ r_k &= y_k - C_\gamma \hat{x}_k \\ u_k &= -K_\gamma \hat{x}_k \end{aligned}$$

External perturbation

- Faults influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_f f_k + w_k \\ y_k &= C_\gamma x_k + C_f f_k + v_k \end{aligned}$$

- Attacks influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_a a_k + w_k \\ y_k &= C_\gamma x_k + D_a a_k + v_k \end{aligned}$$

External perturbation

- Faults influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_f f_k + w_k \\ y_k &= C_\gamma x_k + C_f f_k + v_k \end{aligned}$$

- Attacks influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_a a_k + w_k \\ y_k &= C_\gamma x_k + D_a a_k + v_k \end{aligned}$$

External perturbation

- Faults influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_f f_k + w_k \\ y_k &= C_\gamma x_k + C_f f_k + v_k \end{aligned}$$

- Attacks influence:

$$\begin{aligned} x_{k+1} &= A_\gamma x_k + B_\gamma u_k + B_a a_k + w_k \\ y_k &= C_\gamma x_k + D_a a_k + v_k \end{aligned}$$

f_k and a_k are the fault and attack vector respectively.

Outline

Fault or Attack

- Our system runs over different modes, each of which have a physical model ψ_i, creating the behaviour ξ_i having measurement $\hat{y}_i$.
- **Mode estimation:** closest mode to anomalous observation $\widetilde{y}_i$

$$\psi^* = arg \min_{\psi_i \in \Psi} ||\widetilde{y}_i - \hat{y}_i|| = arg \min_{\psi_i \in \Psi} r_i$$

- **Mode identifiability:**
  - distinguishable behaviour $\xi_i \ \forall j \neq i$
  - activated residual $r_i > \delta$ if system is in mode ψ_i

Fault or Attack

- Our system runs over different modes, each of which have a physical model ψ_i, creating the behaviour ξ_i having measurement $\hat{y}_i$.
- **Mode estimation:** closest mode to anomalous observation $\widetilde{y}_i$

$$\psi^* = arg \min_{\psi_i \in \Psi} ||\widetilde{y}_i - \hat{y}_i|| = arg \min_{\psi_i \in \Psi} r_i$$

- **Mode identifiability:**
  - distinguishable behaviour $\xi_i \ \forall j \neq i$
  - activated residual $r_i > \delta$ if system is in mode ψ_i

Fault or Attack

- Our system runs over different modes, each of which have a physical model ψ_i, creating the behaviour ξ_i having measurement $\hat{y}_i$.
- **Mode estimation:** closest mode to anomalous observation $\widetilde{y}_i$

$$\psi^* = arg \min_{\psi_i \in \Psi} ||\widetilde{y}_i - \hat{y}_i|| = arg \min_{\psi_i \in \Psi} r_i$$

- **Mode identifiability:**
  - distinguishable behaviour $\xi_i \ \forall j \neq i$
  - activated residual $r_i > \delta$ if system is in mode ψ_i

Outline

Experiments

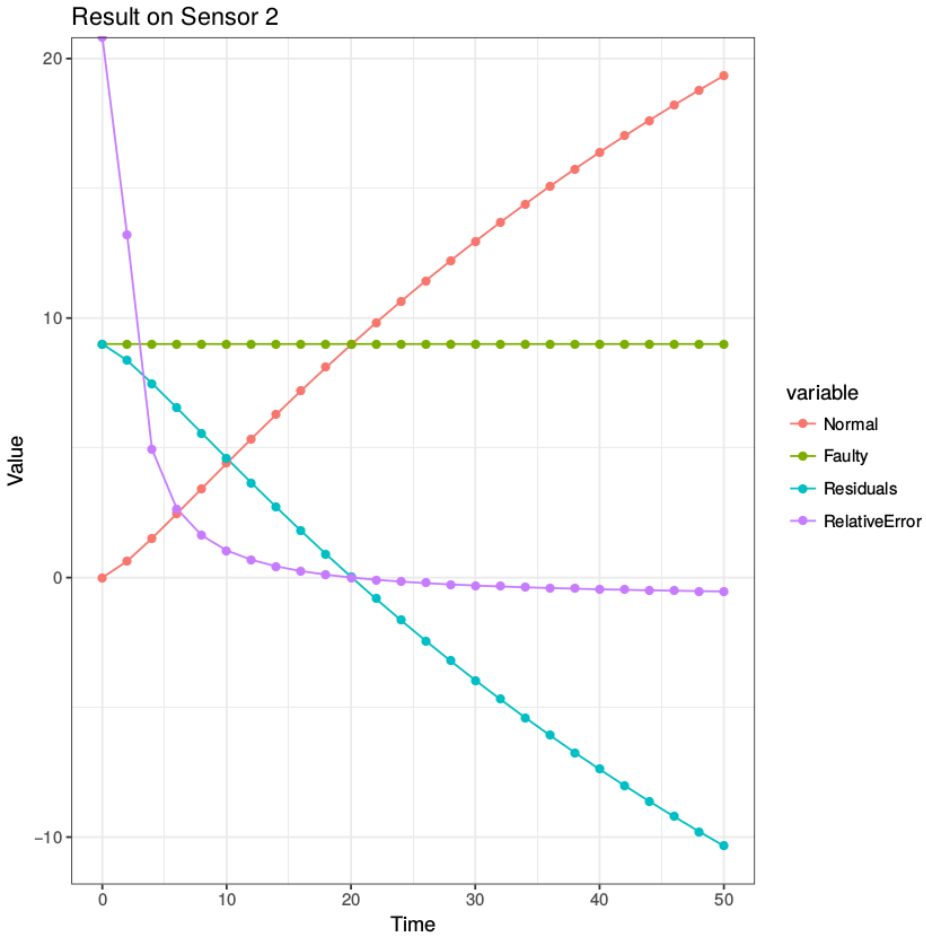
- Three types of tests:
  - Sensors attacks
  - Actuators attacks
  - Multiple components attacks
- Experiments environment:
  - Time domain: $[0, 50]$ seconds
  - Sensor data gathered every 2 seconds
  - Not faulty setting: $v_1 = v_2 = v_3 = 0.5$

Experiments

- Three types of tests:
  - Sensors attacks
  - Actuators attacks
  - Multiple components attacks
- Experiments environment:
  - Time domain: $[0, 50]$ seconds
  - Sensor data gathered every 2 seconds
  - Not faulty setting: $v_1 = v_2 = v_3 = 0.5$

Attacks on Sensors

Injected data on the second sensor of our system

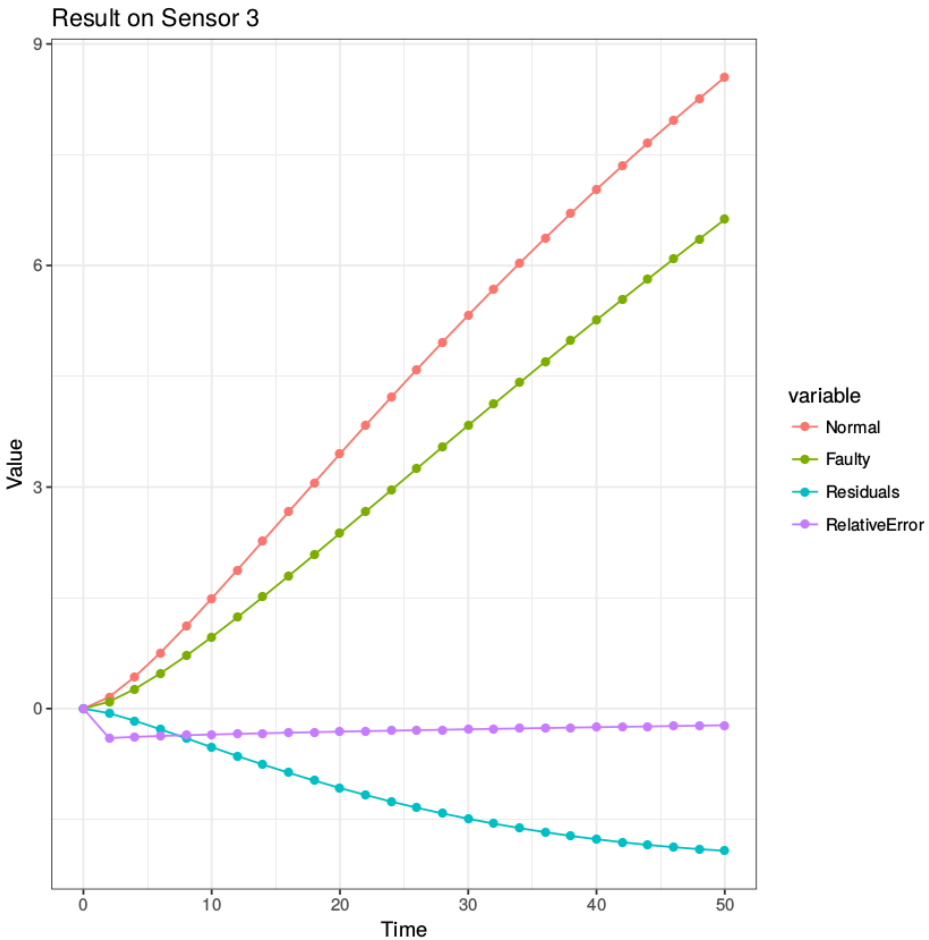


Attack identified through first derivative comparison:

$$\dot{y}_k = -\dot{r}_k$$

Attacks on Actuators

System complexity makes identifiability harder when the actuators are under attack, creating false positives.



Test	Valve 1	Valve 2	Valve 3
155	✓	X	X
355	✓	X	X
755	✓	X	X
955	✓	X	X
515	X	✓	X
535	X	✓	X
575	X	✓	X
595	X	✓	X
551			✓
553			✓
557			✓
559			✓
158	✓	X	✓
544	X	✓	✓
658	✓	X	✓
745	✓	✓	X
958	✓	X	✓
247	✓	✓	✓
638	✓	✓	✓

Multi Attacks and Results

Sensors problems correctly detected and identified.
Actuators errors detected.

Test	Valve 1	Valve 2	Valve 3	Sensor
s1_325		✓	X	1
s2_553	X		✓	2
s3_148	✓	✓		3
s12_558			✓	1-2
s23_647	✓			2-3
s31_348		✓		1-3
s123_666				1-2-3

Outline

Summary and Future Work

- Showed a security system approach based on Cyber-Physical Systems.
- Distinguishing attacks from faults is difficult when the system has a small number of sensors.
- Future work

  Deeper studies on the synergies of the system and between sensors data.

  Optimize the number of sensors in the system.

Summary and Future Work

- Showed a security system approach based on Cyber-Physical Systems.
- Distinguishing attacks from faults is difficult when the system has a small number of sensors.
- Future work

  Deeper studies on the synergies of the system and between sensors data.

  Optimize the number of sensors in the system.

Summary and Future Work

- Showed a security system approach based on Cyber-Physical Systems.
- Distinguishing attacks from faults is difficult when the system has a small number of sensors.
- Future work

  Deeper studies on the synergies of the system and between sensors data.
  Optimize the number of sensors in the system.