Formal Impact Metric for Cyber-Physical Attacks

Massimo Merro

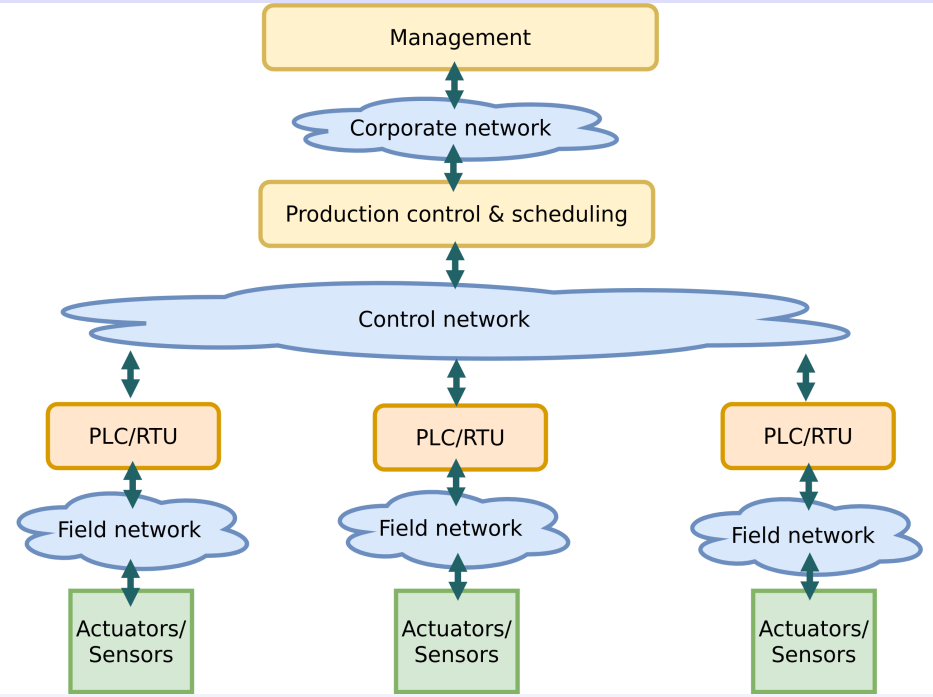
(joint work with Ruggero Lanotte, Andrei Munteanu and Simone Tini)

Department of Computer Science
University of Verona - Italy

34th IEEE CSF - June 23 2021, Virtual Conference

Cyber-Physical Systems

In Cyber-Physical Systems (CPSs) embedded computers and networks monitor and control physical processes, usually with feedback loops where physical processes affect computation and vice versa



Attacking cyber-physical systems

1. Stuxnet: centrifuges of a nuclear plant in Iran
2. BlackEnergy/CRASHOVERRIDE: power outages in Ukraine
3. TRITON/TRITIS: petrochemical plants in Saudi Arabia
4. . . . the list of cyber-physical attacks is quite long

Attacking cyber-physical systems

1. Stuxnet: centrifuges of a nuclear plant in Iran
2. BlackEnergy/CRASHOVERRIDE: power outages in Ukraine
3. TRITON/TRITIS: petrochemical plants in Saudi Arabia
4. . . . the list of cyber-physical attacks is quite long

A cyber-physical attack:

Attacking cyber-physical systems

1. Stuxnet: centrifuges of a nuclear plant in Iran
2. BlackEnergy/CRASHOVERRIDE: power outages in Ukraine
3. TRITON/TRITIS: petrochemical plants in Saudi Arabia
4. . . . the list of cyber-physical attacks is quite long

A cyber-physical attack:

- disrupts the **plant** evolution via **cyber attacks** to **control systems**

Attacking cyber-physical systems

1. Stuxnet: centrifuges of a nuclear plant in Iran
2. BlackEnergy/CRASHOVERRIDE: power outages in Ukraine
3. TRITON/TRITIS: petrochemical plants in Saudi Arabia
4. . . . the list of cyber-physical attacks is quite long

A cyber-physical attack:

- disrupts the **plant** evolution via **cyber attacks** to **control systems**
- brings the plant into a unsafe/incorrect state

Attacking cyber-physical systems

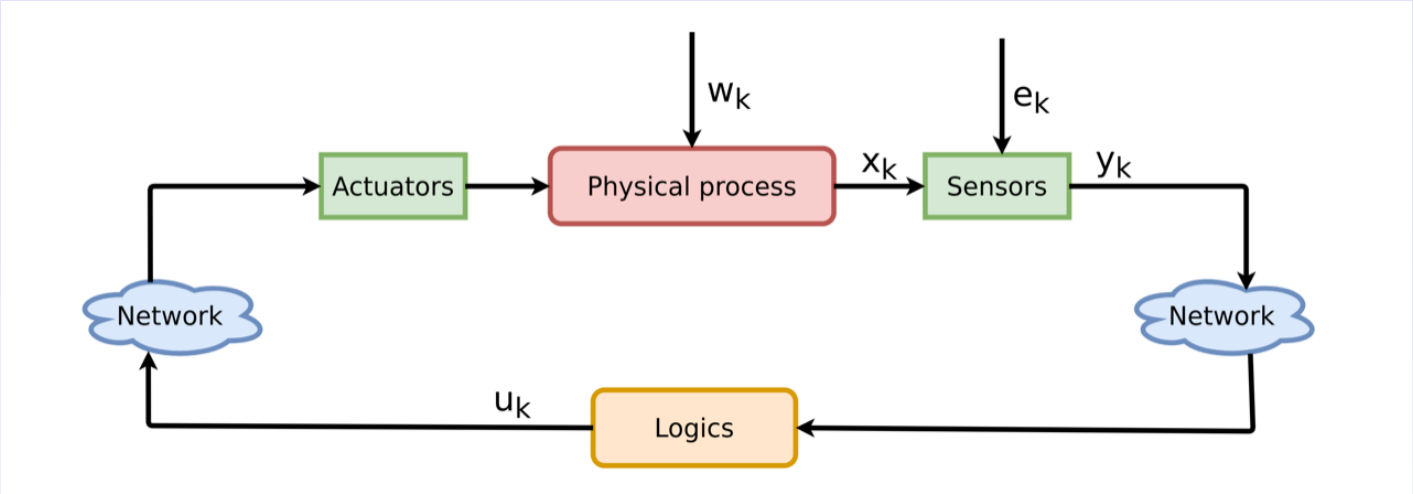
1. Stuxnet: centrifuges of a nuclear plant in Iran
2. BlackEnergy/CRASHOVERRIDE: power outages in Ukraine
3. TRITON/TRITIS: petrochemical plants in Saudi Arabia
4. . . . the list of cyber-physical attacks is quite long

A cyber-physical attack:

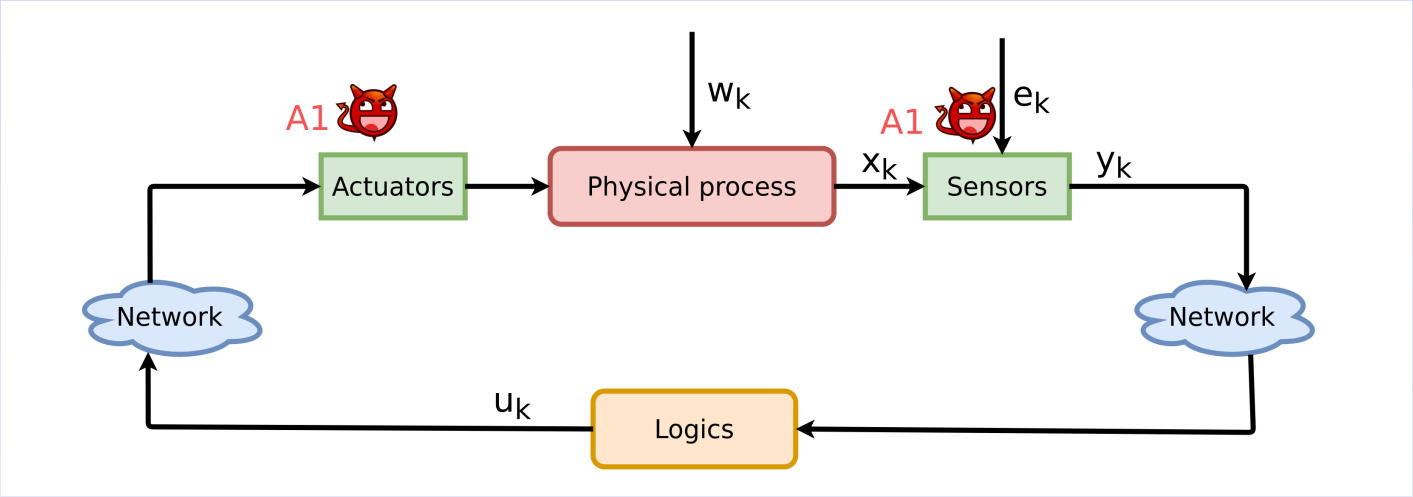
- disrupts the **plant** evolution via **cyber attacks** to **control systems**
- brings the plant into a unsafe/incorrect state
- requires expert **knowledge** in the physical domain

CPSs in a nutshell (focus on the field network)

- Physical process: often represented as *discrete-time state-space model*
- Logics: controllers, IDSs, supervisors, etc. (cyber-components)
- Network: to connect plant and logics, forming feedback loops

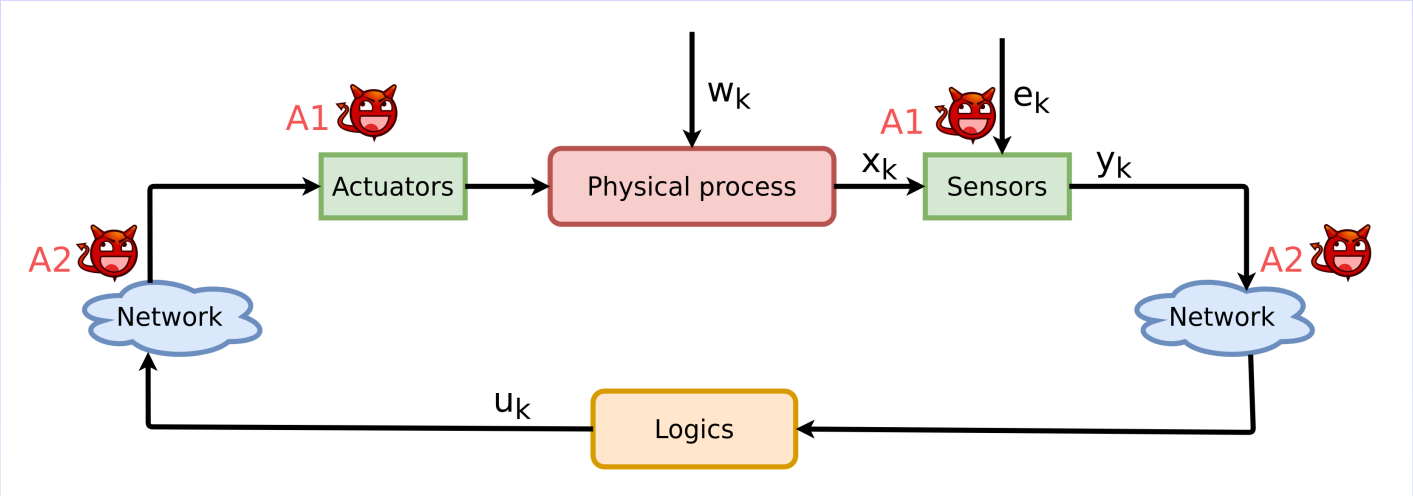


A Threat model for Physics-based attacks



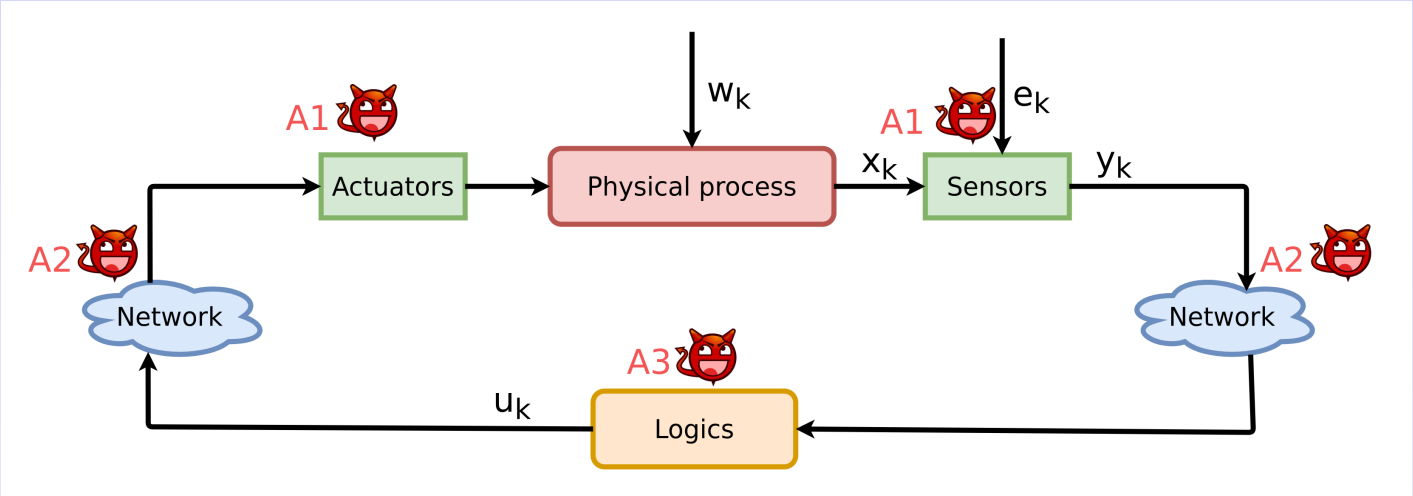
- A1 - actuators/sensors (forge/drop measurements and/or commands)

A Threat model for Physics-based attacks



- A1 - actuators/sensors (forge/drop measurements and/or commands)
- A2 - network (DoS, integrity, replay attacks ...)

A Threat model for Physics-based attacks



- A1 - actuators/sensors (forge/drop measurements and/or commands)
- A2 - network (DoS, integrity, replay attacks ...)
- A3 - logics (malware tampering with control sw or register's data)

Claim: physical aspects put CPS security apart from IT security!

Three crucial aspects: Timing, Duration and Impact

Three crucial aspects: Timing, Duration and Impact

- Timing: physical states might be more vulnerable to attacks at certain points in time, for instance, when they reach a local maximum/minimum

Three crucial aspects: Timing, Duration and Impact

- Timing: physical states might be more vulnerable to attacks at certain points in time, for instance, when they reach a local maximum/minimum
- Duration: it may take minutes for a chemical reactor to rupture, hours to heat a tank of water or burn out a motor, and days to destroy centrifuges

Three crucial aspects: Timing, Duration and Impact

- Timing: physical states might be more vulnerable to attacks at certain points in time, for instance, when they reach a local maximum/minimum
- Duration: it may take minutes for a chemical reactor to rupture, hours to heat a tank of water or burn out a motor, and days to destroy centrifuges
- Impact: understanding both attacker's goals and the severity of the damages inflicted when such goals are achieved is fundamental to conduct a risk assessment

Three crucial aspects: Timing, Duration and Impact

- Timing: physical states might be more vulnerable to attacks at certain points in time, for instance, when they reach a local maximum/minimum
- Duration: it may take minutes for a chemical reactor to rupture, hours to heat a tank of water or burn out a motor, and days to destroy centrifuges
- Impact: understanding both attacker's goals and the severity of the damages inflicted when such goals are achieved is fundamental to conduct a risk assessment

Impact metrics should quantify the perturbation introduced by an attack in both the physical and the logical behaviour of the target system

Three crucial aspects: Timing, Duration and Impact

- Timing: physical states might be more vulnerable to attacks at certain points in time, for instance, when they reach a local maximum/minimum
- Duration: it may take minutes for a chemical reactor to rupture, hours to heat a tank of water or burn out a motor, and days to destroy centrifuges
- Impact: understanding both attacker's goals and the severity of the damages inflicted when such goals are achieved is fundamental to conduct a risk assessment

Impact metrics should quantify the perturbation introduced by an attack in both the physical and the logical behaviour of the target system

Question: Can we use formal methodologies for defining impact metrics?

Our proposal

Our proposal

Given a CPS equipped with an IDS, and exposed to the malicious activities on an attacker, define *two metrics* relying on three concepts:

Our proposal

Given a CPS equipped with an IDS, and exposed to the malicious activities on an attacker, define *two metrics* relying on three concepts:

- a **timed probabilistic LTS** of the system under attack

Our proposal

Given a CPS equipped with an IDS, and exposed to the malicious activities on an attacker, define *two metrics* relying on three concepts:

- a **timed probabilistic LTS** of the system under attack
- a set $\mathcal{I} = \{(i_1, w_1), \ldots, (i_k, w_k)\}$ of **weighted attacker's goal indicators** denoting how close is the attacker in reaching those goals

Our proposal

Given a CPS equipped with an IDS, and exposed to the malicious activities on an attacker, define *two metrics* relying on three concepts:

- a **timed probabilistic LTS** of the system under attack
- a set $\mathcal{I} = \{(i_1, w_1), \ldots, (i_k, w_k)\}$ of **weighted attacker's goal indicators** denoting how close is the attacker in reaching those goals
- a **detection policy** $\mathcal{P}$ that given an alert signal and a goal indicator returns a statically-determined estimate of the attacker's progresses wrt that goal

For goals achieved via stealthy attacks, $\mathcal{P}$ gives little/vague estimates

Our proposal

Given a CPS equipped with an IDS, and exposed to the malicious activities on an attacker, define *two metrics* relying on three concepts:

- a **timed probabilistic LTS** of the system under attack
- a set $\mathcal{I} = \{(i_1, w_1), \ldots, (i_k, w_k)\}$ of **weighted attacker's goal indicators** denoting how close is the attacker in reaching those goals
- a **detection policy** $\mathcal{P}$ that given an alert signal and a goal indicator returns a statically-determined estimate of the attacker's progresses wrt that goal

For goals achieved via stealthy attacks, $\mathcal{P}$ gives little/vague estimates

FN$^n_{\mathcal{I},\mathcal{P}}$: average false negatives in the system under attack in n time units

FP$^n_{\mathcal{I},\mathcal{P}}$: average false positives in the system under attack in n time units

What the metric $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

What the metric $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average effectiveness of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the actual damage wrt goals in $\mathcal{I}$ and the estimated attacker's progresses (via $\mathcal{P}$)

What the metric $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average effectiveness of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the actual damage wrt goals in $\mathcal{I}$ and the estimated attacker's progresses (via $\mathcal{P}$)
- $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j

What the metric $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ actually measures

- The average effectiveness of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the actual damage wrt goals in $\mathcal{I}$ and the estimated attacker's progresses (via $\mathcal{P}$)
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is highly effective in detecting malicious activities of the attacker aiming at reaching goals in $\mathcal{I}$

What the metric $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ actually measures

- The average effectiveness of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the actual damage wrt goals in $\mathcal{I}$ and the estimated attacker's progresses (via $\mathcal{P}$)
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is highly effective in detecting malicious activities of the attacker aiming at reaching goals in $\mathcal{I}$
- $0 < \mathbf{FN}^n_{\mathcal{I},\mathcal{P}} \leq 1$: the detection mechanism underestimates the progresses of malicious activities of the attacker to reach goals in $\mathcal{I}$ (many false negatives)

What the metric $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ actually measures

- The average effectiveness of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the actual damage wrt goals in $\mathcal{I}$ and the estimated attacker's progresses (via $\mathcal{P}$)
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is highly effective in detecting malicious activities of the attacker aiming at reaching goals in $\mathcal{I}$
- $0 < \mathbf{FN}^n_{\mathcal{I},\mathcal{P}} \leq 1$: the detection mechanism underestimates the progresses of malicious activities of the attacker to reach goals in $\mathcal{I}$ (many false negatives)

An aggressive detection mechanism may be very effective in detecting all significant malicious activities but .... scarcely precise!

What the metric $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ actually measures

What the metric $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average precision of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the estimated attacker's progresses (via $\mathcal{P}$) and the actual damage wrt goals in $\mathcal{I}$

What the metric $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average precision of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the estimated attacker's progresses (via $\mathcal{P}$) and the actual damage wrt goals in $\mathcal{I}$
- Again, $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j

What the metric $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average precision of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the estimated attacker's progresses (via $\mathcal{P}$) and the actual damage wrt goals in $\mathcal{I}$
- Again, $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is very precise in recognising malicious activities of the attacker reaching goals in $\mathcal{I}$

What the metric $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average precision of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the estimated attacker's progresses (via $\mathcal{P}$) and the actual damage wrt goals in $\mathcal{I}$
- Again, $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is very precise in recognising malicious activities of the attacker reaching goals in $\mathcal{I}$
- $0 < \mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}} \leq 1$: the detection mechanism overestimates the progresses of malicious activities of the attacker to reach goals in $\mathcal{I}$ (many false positives)

What the metric $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ actually measures

- The average precision of the detection mechanism (IDS + detection policy $\mathcal{P}$) in terms of the deviation between the estimated attacker's progresses (via $\mathcal{P}$) and the actual damage wrt goals in $\mathcal{I}$
- Again, $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ is weighted considering both the severities w_j of the goals i_j and the probability that the attacker achieves the goals i_j
- $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}} = 0$: the detection mechanism is very precise in recognising malicious activities of the attacker reaching goals in $\mathcal{I}$
- $0 < \mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}} \leq 1$: the detection mechanism overestimates the progresses of malicious activities of the attacker to reach goals in $\mathcal{I}$ (many false positives)

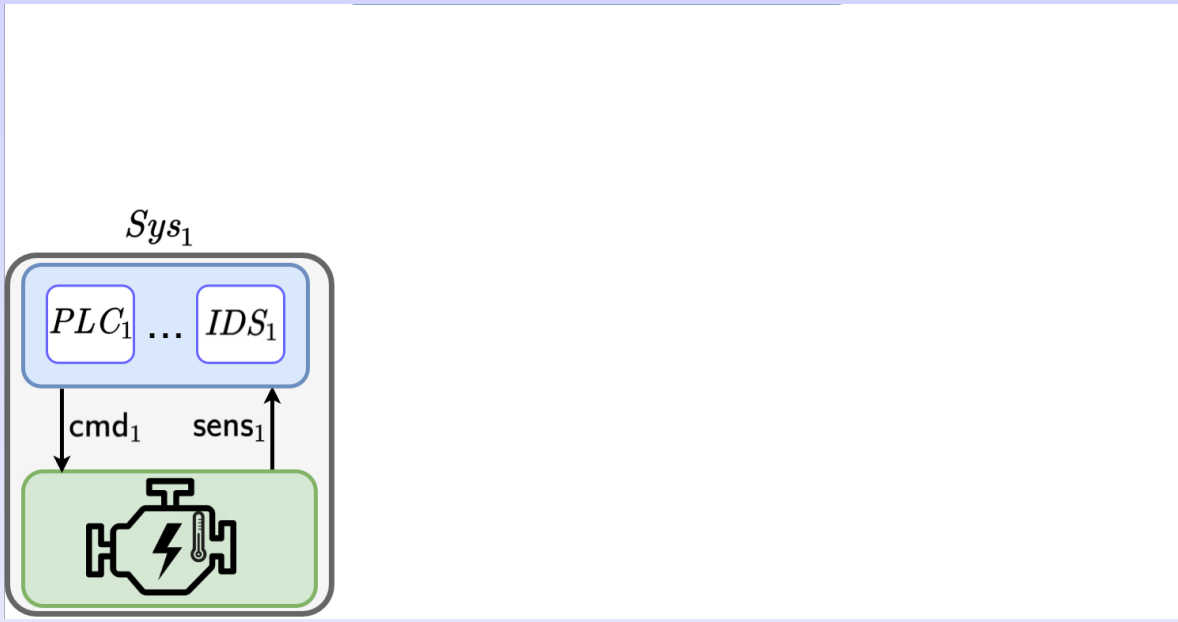
A loose detection mechanism may be very precise when it raises an alert but ... scarcely effective!

$\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ can be computed in a compositional way

What kind of compositionality

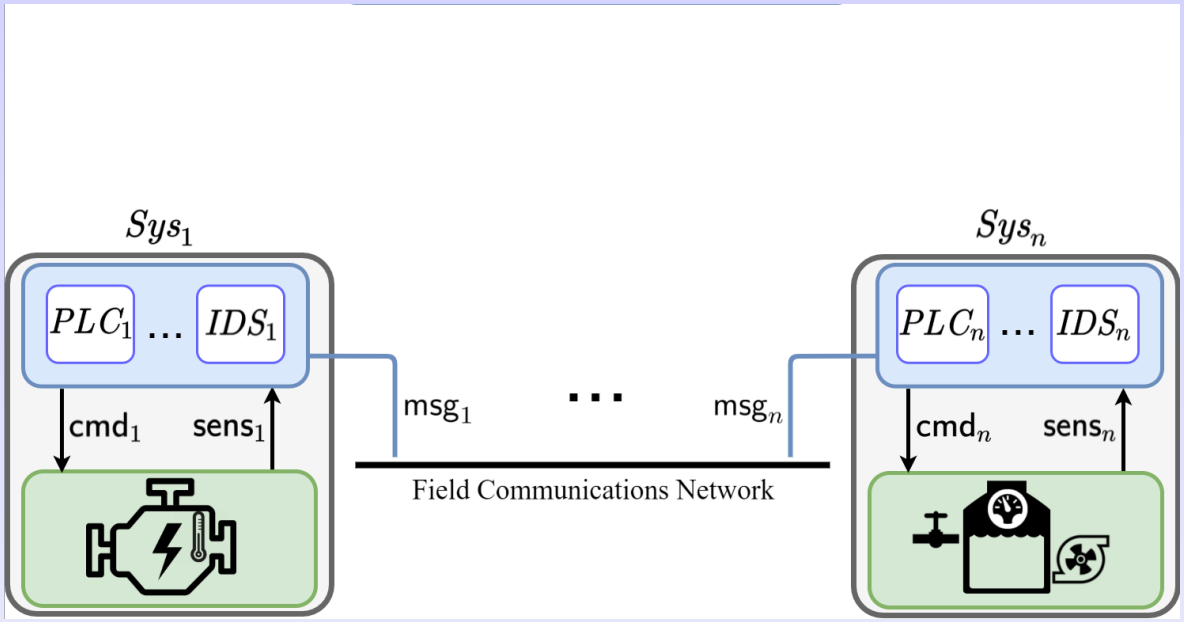
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ can be computed in a compositional way on CPSs whose sub-systems may physically interact only if the associated logical components agree on when and how that interaction should happen
- The sub-systems Sys_i are said physically-disjoint
- Formally, $((Sys_1 \uplus \ldots \uplus Sys_n) \parallel SCADA \parallel HMI \parallel Hist) \setminus \{state_i\}$

What kind of compositionality



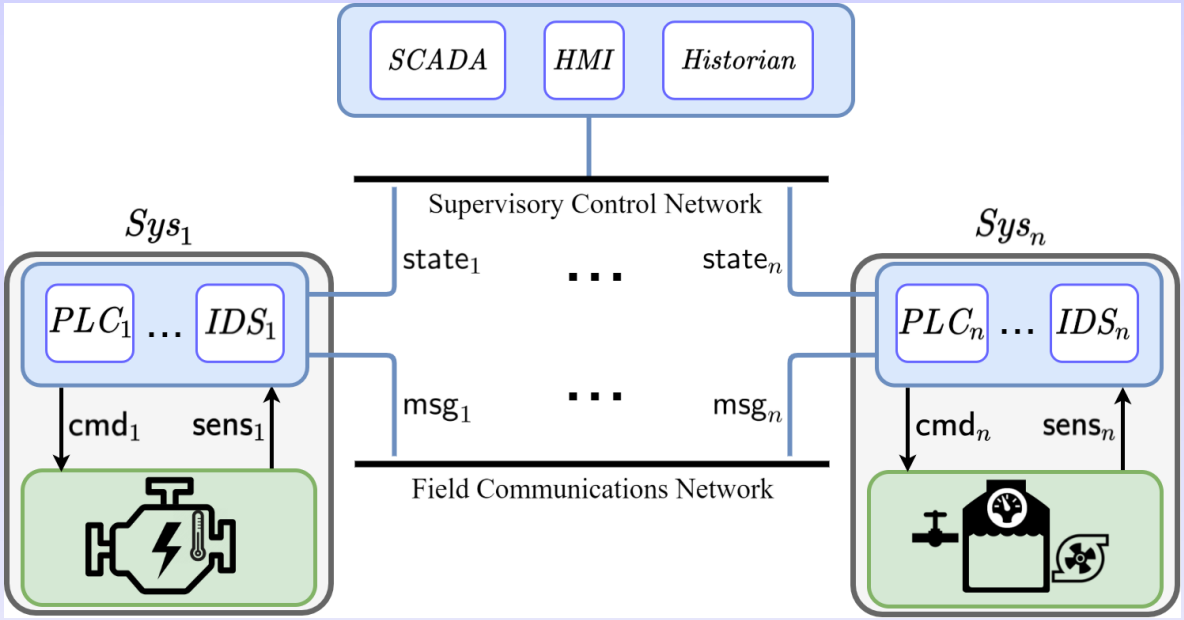
- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ can be computed in a compositional way on CPSs whose sub-systems may physically interact only if the associated logical components agree on when and how that interaction should happen
- The sub-systems Sys_i are said physically-disjoint
- Formally, $((Sys_1 \uplus \ldots \uplus Sys_n) \parallel SCADA \parallel HMI \parallel Hist) \setminus \{state_i\}$

What kind of compositionality



- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ can be computed in a compositional way on CPSs whose sub-systems may physically interact only if the associated logical components agree on when and how that interaction should happen
- The sub-systems Sys_i are said physically-disjoint
- Formally, $((Sys_1 \uplus \ldots \uplus Sys_n) \parallel SCADA \parallel HMI \parallel Hist) \setminus \{state_i\}$

What kind of compositionality



- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ can be computed in a compositional way on CPSs whose sub-systems may physically interact only if the associated logical components agree on when and how that interaction should happen
- The sub-systems Sys_i are said physically-disjoint
- Formally, $((Sys_1 \uplus \ldots \uplus Sys_n) \parallel SCADA \parallel HMI \parallel Hist) \setminus \{state_i\}$

Compositionality results

Compositionality results

Suppose a composite system $M_1 \uplus \ldots \uplus M_k$ whose sub-systems M_i have different security clearances (e.g., engine control vs. air conditioning)

Compositionality results

Suppose a composite system $M_1 \uplus \ldots \uplus M_k$ whose sub-systems M_i have different security clearances (e.g., engine control vs. air conditioning)

Theorem (Compositionality w.r.t. $\uplus$)

The metrics $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ *and* $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ *of a system* $M_1 \uplus \ldots \uplus M_k$ *are given by the weighted sum of the metrics on the single components* M_i

Compositionality results

Suppose a composite system $M_1 \uplus \ldots \uplus M_k$ whose sub-systems M_i have different security clearances (e.g., engine control vs. air conditioning)

Theorem (Compositionality w.r.t. $\uplus$)

The metrics $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}}$ *and* $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}}$ *of a system* $M_1 \uplus \ldots \uplus M_k$ *are given by the weighted sum of the metrics on the single components* M_i

Theorem (Compositionality on supervised systems)

Let M *be a CPS supervised by a process Sup which amplifies* the estimates of the attacker's progresses made by* M*... omissis... Then,*

- $\mathbf{FN}^n_{\mathcal{I},\mathcal{P}'}((M \parallel Sup) \setminus \mathcal{A}) \leq \mathbf{FN}^n_{\mathcal{I},\mathcal{P}}(M)$
- $\mathbf{FP}^n_{\mathcal{I},\mathcal{P}'}((M \parallel Sup) \setminus \mathcal{A}) \geq \mathbf{FP}^n_{\mathcal{I},\mathcal{P}}(M)$

The increased aggressiveness of the detection mechanism results in smaller number of false negatives and in greater number of false positives.

A non-trivial case study

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds
- two IDSs checking whether the cooling systems are active when the temperature is above the given threshold. In that case:

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds
- two IDSs checking whether the cooling systems are active when the temperature is above the given threshold. In that case:
  1. sends a warning to the supervisory level

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds
- two IDSs checking whether the cooling systems are active when the temperature is above the given threshold. In that case:
  1. sends a warning to the supervisory level
  2. asks the controller to slow down the engine (mitigation)

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds
- two IDSs checking whether the cooling systems are active when the temperature is above the given threshold. In that case:
  1. sends a warning to the supervisory level
  2. asks the controller to slow down the engine (mitigation)
  3. demand full power from the other engine (compensation)

A non-trivial case study

- A system with two supervised and coordinated refrigerated engines
- temperature maintained within a safety region by cooling systems
- engine speeds: $\{\text{slow}, \text{half}, \text{full}\}$; normal pace is half power
- a goal indicator *stress* recording persistent violations of the thresholds
- two IDSs checking whether the cooling systems are active when the temperature is above the given threshold. In that case:
  1. sends a warning to the supervisory level
  2. asks the controller to slow down the engine (mitigation)
  3. demand full power from the other engine (compensation)

  when the overheating is resolved engines' paces return to half power

Examples of attacks 1/3

Examples of attacks 1/3

Consider an attack targeting one of the two engines

Examples of attacks 1/3

Consider an attack targeting one of the two engines

Attack1: integrity on the actuator of the cooling system

Forge fake actuator commands to turn off the cooling system when the temperature is close to the threshold

Stealthy attack: system remains close to threshold and sends no warnings

Examples of attacks 1/3

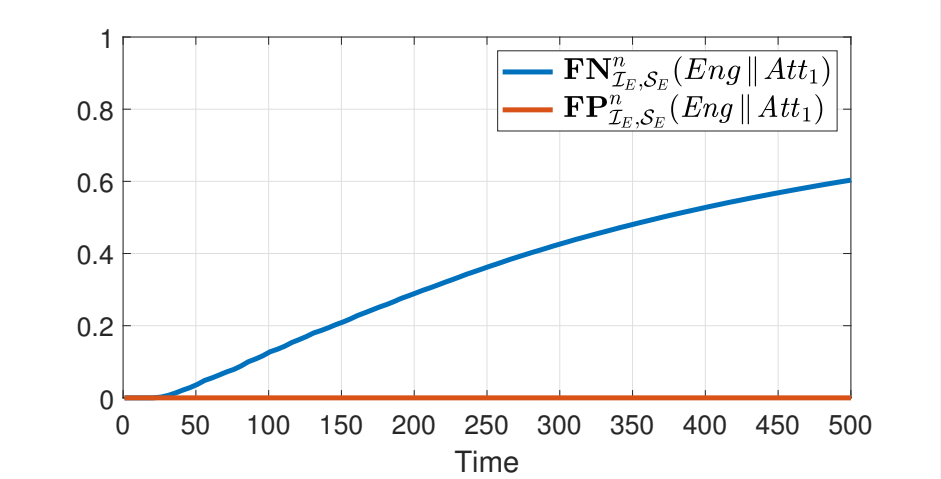
Consider an attack targeting one of the two engines

Attack1: integrity on the actuator of the cooling system

Forge fake actuator commands to turn off the cooling system when the temperature is close to the threshold

Stealthy attack: system remains close to threshold and sends no warnings

- False negatives grow with the size of the attack window (*stress*)
- No false positives



Examples of attacks 2/3

Examples of attacks 2/3

Consider another attack targeting one of the two engines

Examples of attacks 2/3

Consider another attack targeting one of the two engines

Attack2: integrity on the sensor of the temperature

Add a negative offset to temperature detected by the sensor of one engine

This is not a stealthy attack as the IDS sends warnings

Examples of attacks 2/3

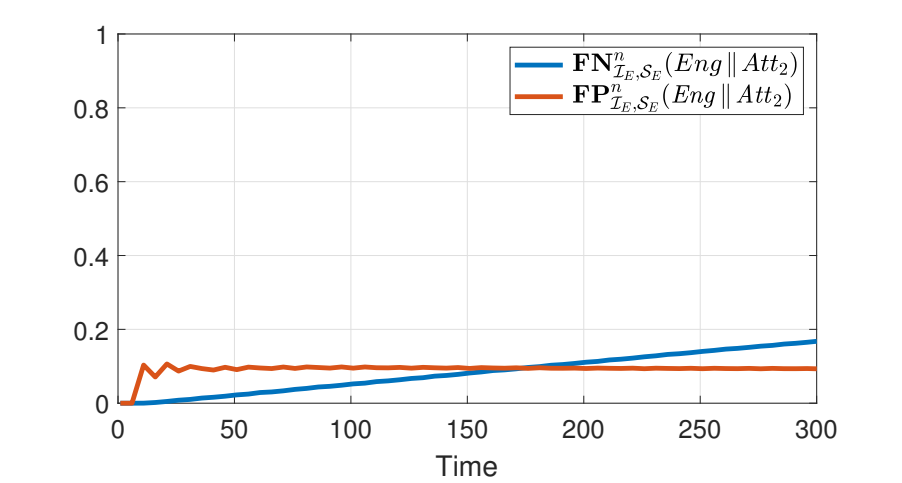
Consider another attack targeting one of the two engines

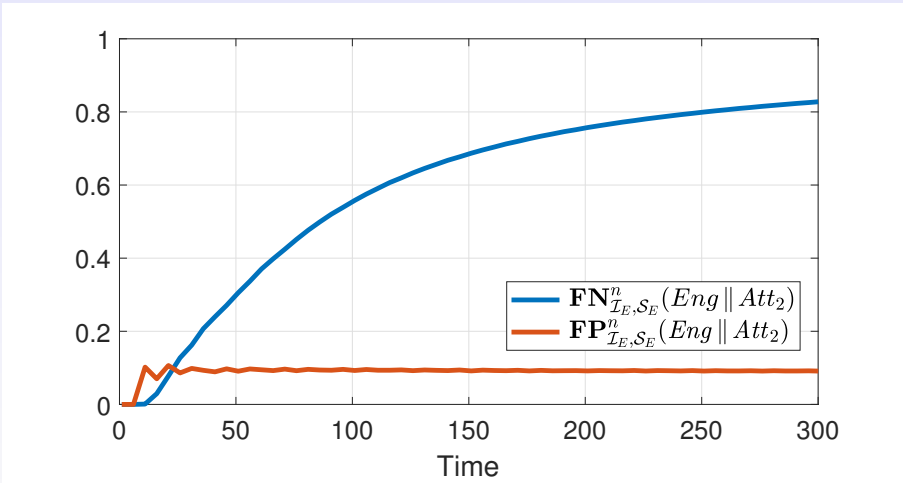
Attack2: integrity on the sensor of the temperature

Add a negative offset to temperature detected by the sensor of one engine

This is not a stealthy attack as the IDS sends warnings
Impact with and without mitigation (slowing down overheating engines)

- False negatives grow with the size of the attack window (*stress*)
- False positives stabilises to 0.1: 10 false positives in 100 time instants





Examples of attacks 3/3

Examples of attacks 3/3

Consider a combined attack targeting both engines

Examples of attacks 3/3

Consider a combined attack targeting both engines

Attack3:

Attack1 tampers with one engine and Attack2 with the other

Examples of attacks 3/3

Consider a combined attack targeting both engines

Attack3:

Attack1 tampers with one engine and Attack2 with the other

- This is not a stealthy attack as the IDS of the engine exposed to Attack2 sends warnings
- The metrics $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$ of the whole system exposed to both attacks are given by the weighted sums of the metrics obtained for the single attacks (by compositionality results)

Conclusions and future work

Summary:

- Two probabilistic impact metrics $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$
- Our metrics rely on three general concepts
  - a *timed probabilistic LTS* of the system under attack
  - a set $\mathcal{I}$ of *weighted attacker's goal indicators* denoting how close is the attacker in reaching those goals
  - a *detection policy* $\mathcal{P}$ returning a *statically-determined estimate* of the attacker's progresses wrt that goal
- our metrics are compositional under specific conditions

Conclusions and future work

Summary:

- Two probabilistic impact metrics $\mathbf{FN}^{n}_{\mathcal{I},\mathcal{P}}$ and $\mathbf{FP}^{n}_{\mathcal{I},\mathcal{P}}$
- Our metrics rely on three general concepts
  - a *timed probabilistic LTS* of the system under attack
  - a set $\mathcal{I}$ of *weighted attacker's goal indicators* denoting how close is the attacker in reaching those goals
  - a *detection policy* $\mathcal{P}$ returning a *statically-determined estimate* of the attacker's progresses wrt that goal
- our metrics are compositional under specific conditions

Future work:

- to improve computational efficiency select traces via Monte Carlo simulation methods, working up to a certain accuracy
- move to statistical impact metrics to be used on existing datasets (e.g., the SWaT system)

Thank you!


obrigado
Dank U
Merci
mahalo
Köszi
спасибо
Grazie
Thank you
maururuuru
Takk
Gracias
Dziękuję
Děkuju
danke
Kiitos