

**Integrated Project**  
**Cyber security of energy systems for the digital-energy transition**

***Power Consumption Anomaly Detection Using Threat-Driven Data Injection***



- Power monitoring is conducted in different areas of the grid, and on different lines
- To optimize power flow throughout the grid
  - Very important to avoid overloading
- To compute bills for customers
- ...

**But what if consumption is anomalous due to attacks?**



- Typically, a specific case study is targeted and analysed, then either
  - A prototype is made, monitoring its behavior (power consumption) under normal operating conditions and when specific faults or attacks are simulated
  - A Digital Twin is deployed, and simulations are conducted in this virtual environment that is as close as possible to the final product

However, both approaches require experimentations and ways to simulate attacks within the system

- What if I cannot do that, or if I cannot have a good “twin” of a system?

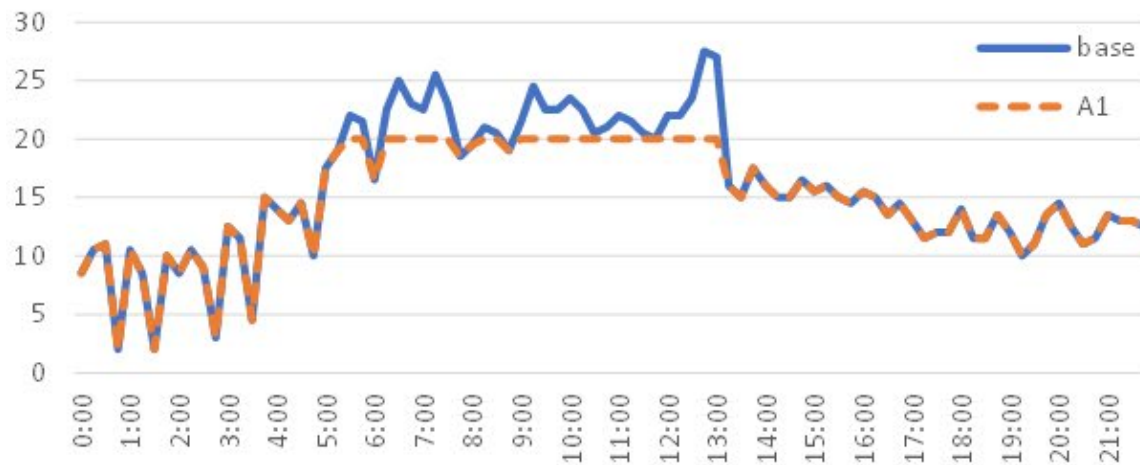
**In other words: I just have normal data! Can I make good use of it?**

- There exist some guidelines for conducting threat and risk assessments in smart grid scenarios
- E.g., ENISA Smart Grid Threat Landscape
- A baseline to customize at will

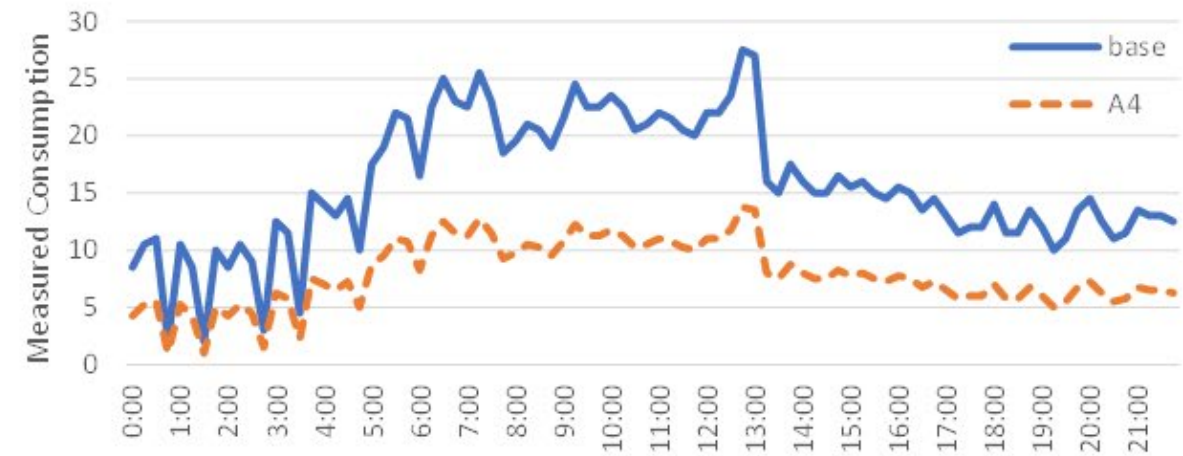


Threat Group	Threat	Threat details	Threat Agent	Trend <sup>44</sup>	Comments
Physical attack (deliberate/intentional)					
	<i>Bomb attack / threat</i>				
	<i>Fraud</i>				
		Fraud by employees	Employees	Increasing	
	<i>Sabotage</i>		Corporations Cybercriminals Employees Hacktivists Nation States Terrorists <sup>45</sup>		
	<i>Vandalism</i>		Employees Terrorists Rioter		

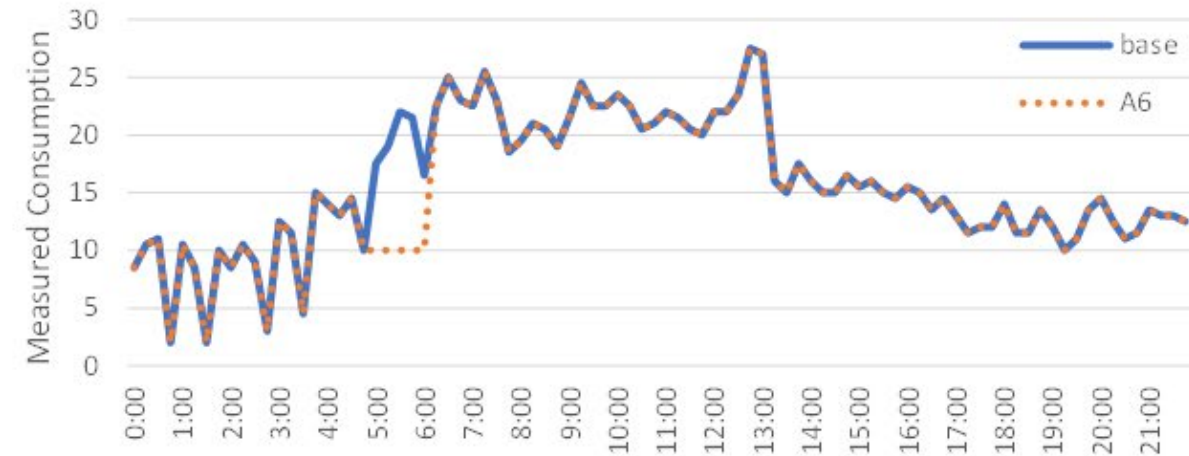
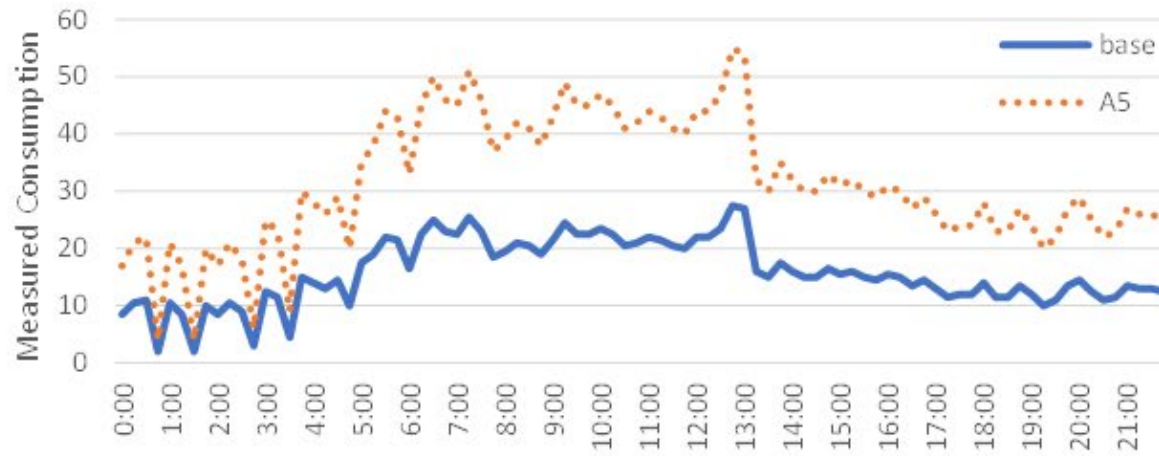
- Within the old PTR22-24, we figured
  - a characterization of anomalies



- Within the old PTR22-24, we figured
  - a characterization of anomalies



- Within the old PTR22-24, we figured
  - a characterization of anomalies





- Within the old PTR22-24, we figured
  - a characterization of anomalies
  - And their mapping to ENISA threats (i.e., what threat could generate what anomaly in power consumption)

Threat Group	Threat ID	Threat	Effects
Physical attack (deliberate intentional)	PA.1	Bomb attack / threat	A3
	PA.2	Fraud / Sabotage / Vandalism	A3
	PA.3	Theft (of devices, storage media, documentt)	A3
	PA.4	Information leakage/sharing	-
	PA.5	Unauthorized physical access / entry to premises	A1, A2, A3, A4, A5, A6
	PA.6	Coercion, extortion or corruption	A3



- Then, we inject them into normal data traces
  - Using false data injection
  - Obtaining normal/anomalous labelled datasets

<b>kWh</b>	1.19	1.17	1.20	1.26	1.22	0.75	0.51	0.03	2.15	2.33	1.65
<b>Label</b>	normal	normal	normal	normal	normal	A2	A2	A2	A2	A2	A2



- Then, we inject them into normal data traces
  - Using false data injection
  - Obtaining normal/anomalous labelled datasets

<b>kWh</b>	1.19	1.17	1.20	1.26	1.22	0.75	0.51	0.03	2.15	2.33	1.65
<b>Label</b>	normal	normal	normal	normal	normal	A2	A2	A2	A2	A2	A2

- For consumption anomaly detection!
- However...
  - No matter how hard we tried
    - Supervised, unsupervised, ensembles, meta-learning
  - We did not get any “doable” result



## Why? Too little information (no context)!



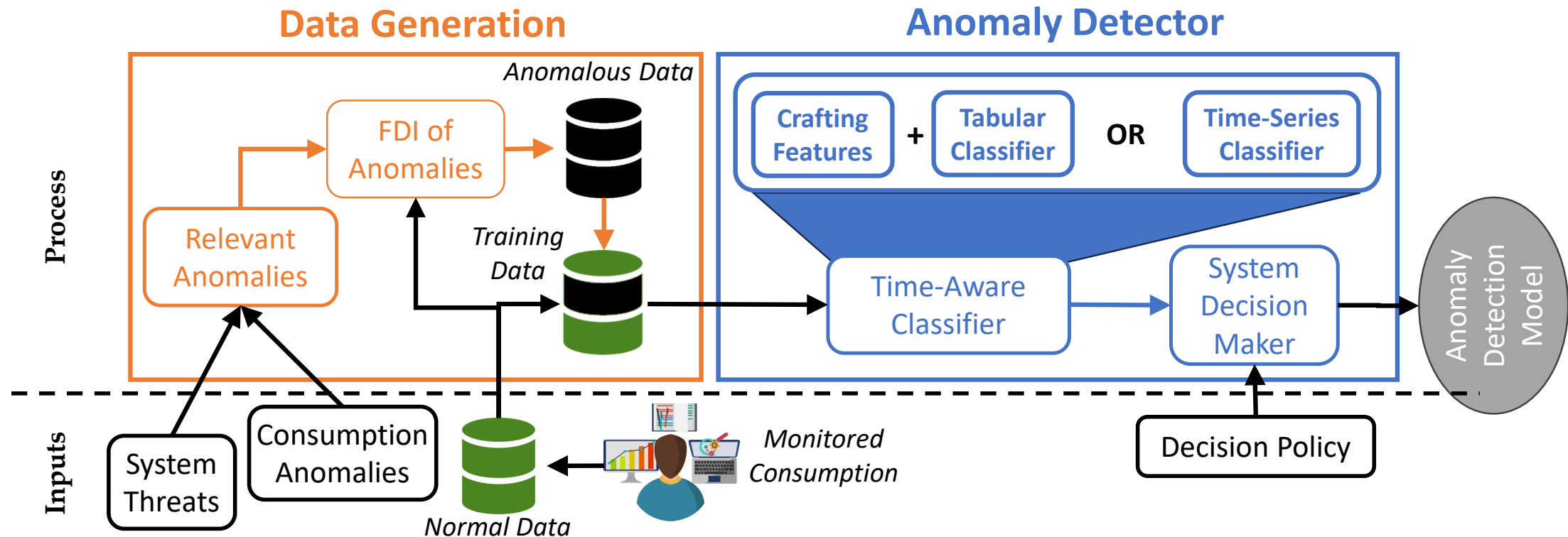
<b>kWh</b>	1.19	1.17	1.20	1.26	1.22	0.75	0.51	0.03	2.15	2.33	1.65
<b>Label</b>	normal	normal	normal	normal	normal	A2	A2	A2	A2	A2	A2

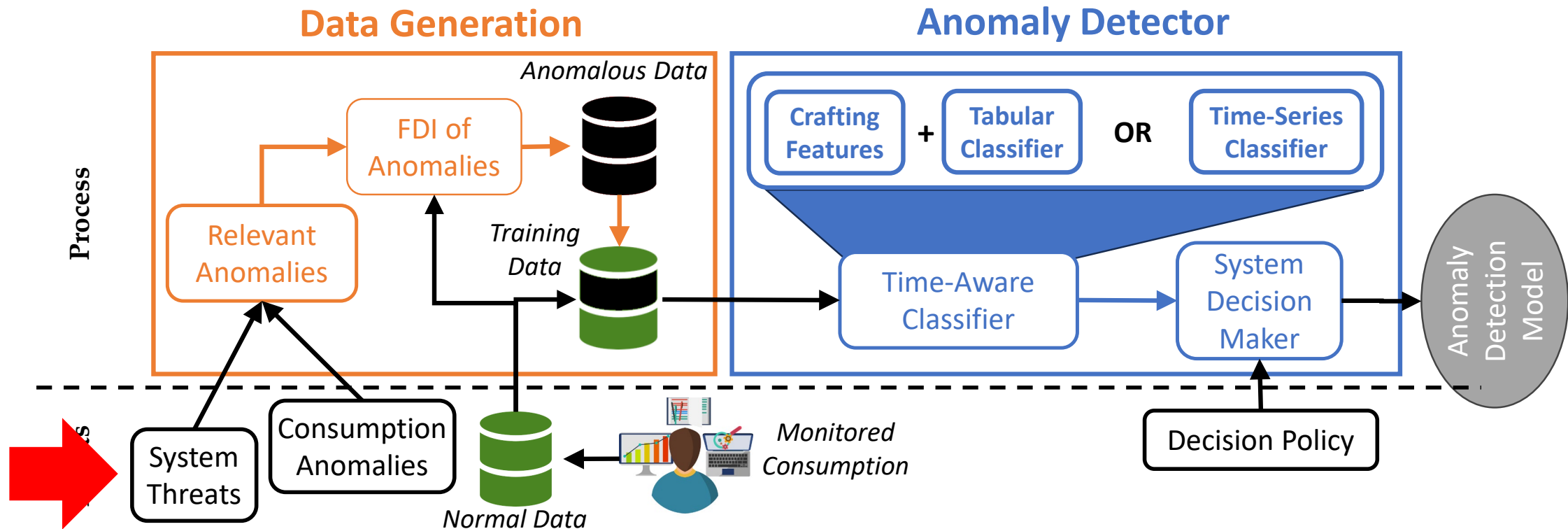
<b>kWh</b>	1.19	1.17	1.20	1.26	1.22	0.75	0.51	0.03	2.15	2.33	1.65
<b>Diff</b>	0.00	0.02	-0.03	-0.07	0.05	0.47	0.24	0.48	-2.12	-0.18	0.68
<b>Diff %</b>	0.00	0.02	-0.02	-0.05	0.04	0.63	0.47	14.67	-0.98	-0.08	0.41
<b>D. MA</b>	0.00	0.00	0.01	0.08	-0.01	-0.49	-0.48	-0.60	1.88	1.24	-0.59
<b>Label</b>	normal	normal	normal	normal	normal	A2	A2	A2	A2	A2	A2

- Context is important for time-series detection
  - i.e., the current state w.r.t. previous states
  - Which, in this case, gets to almost perfect coverage and low false alarms using ensembles of decision trees
    - XGBoost, Random Forests

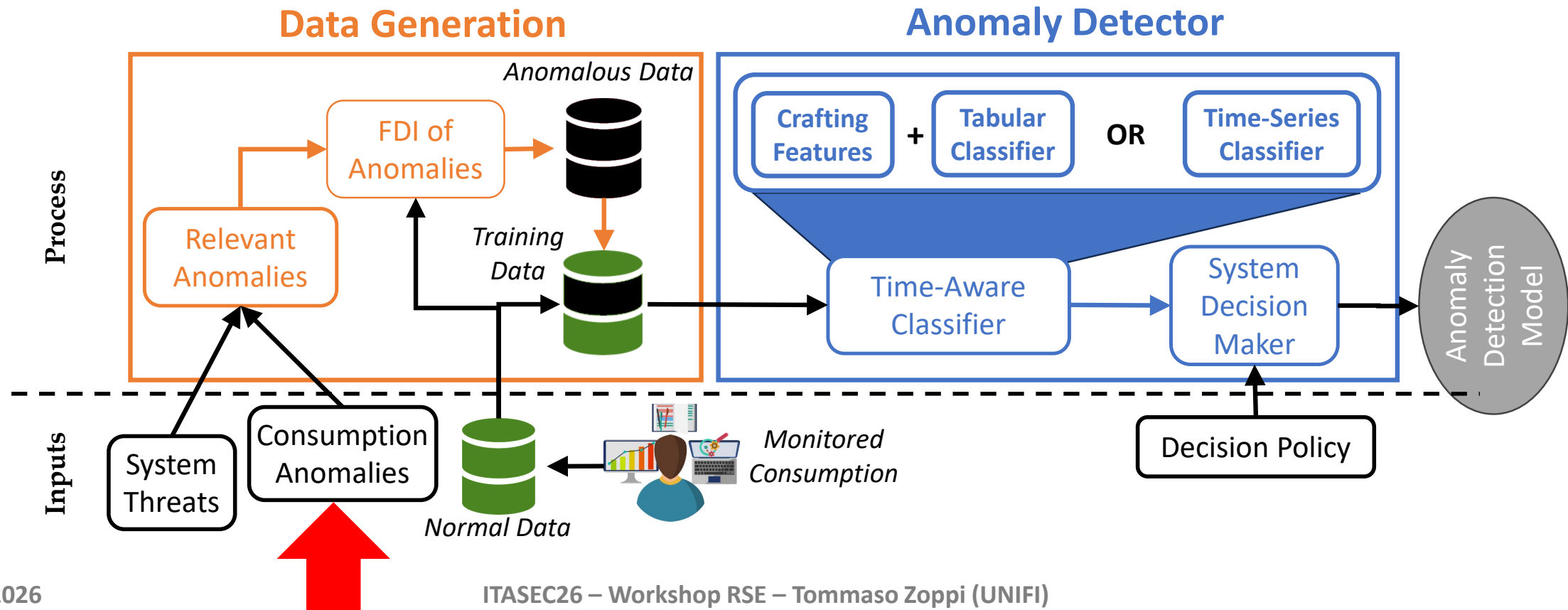
- For any case study, we apply the following:



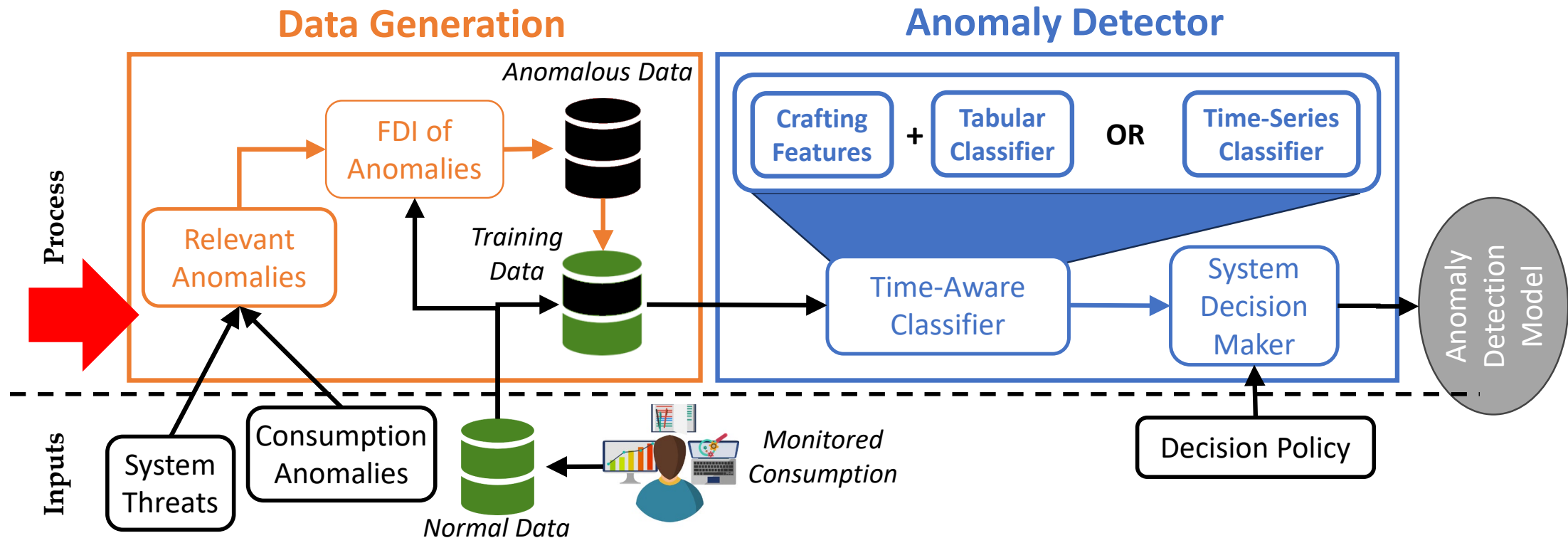
- For any case study, we apply the following:
  - The stakeholder or system owner conducts a threat analysis to devise relevant threats for a given scenario



- For any case study, we apply the following:
  - Threats are automatically mapped to anomalies

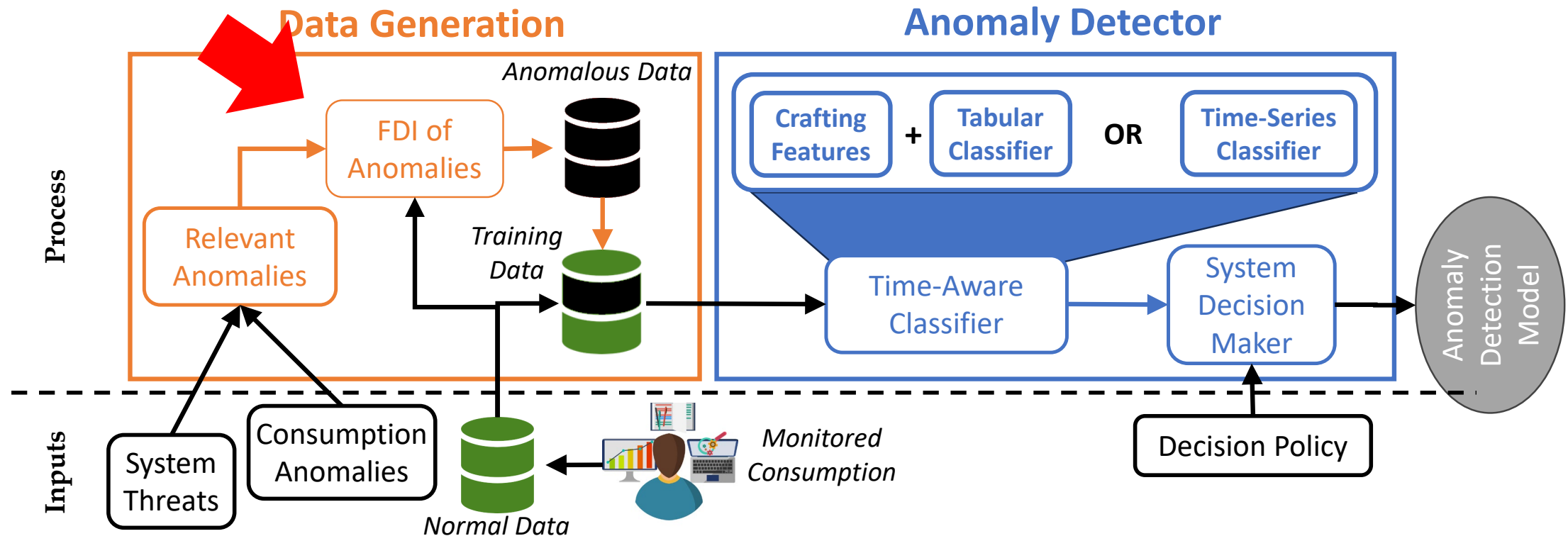


- For any case study, we apply the following:
  - Anomalies that are linked to 1+ relevant threats are selected for experiments

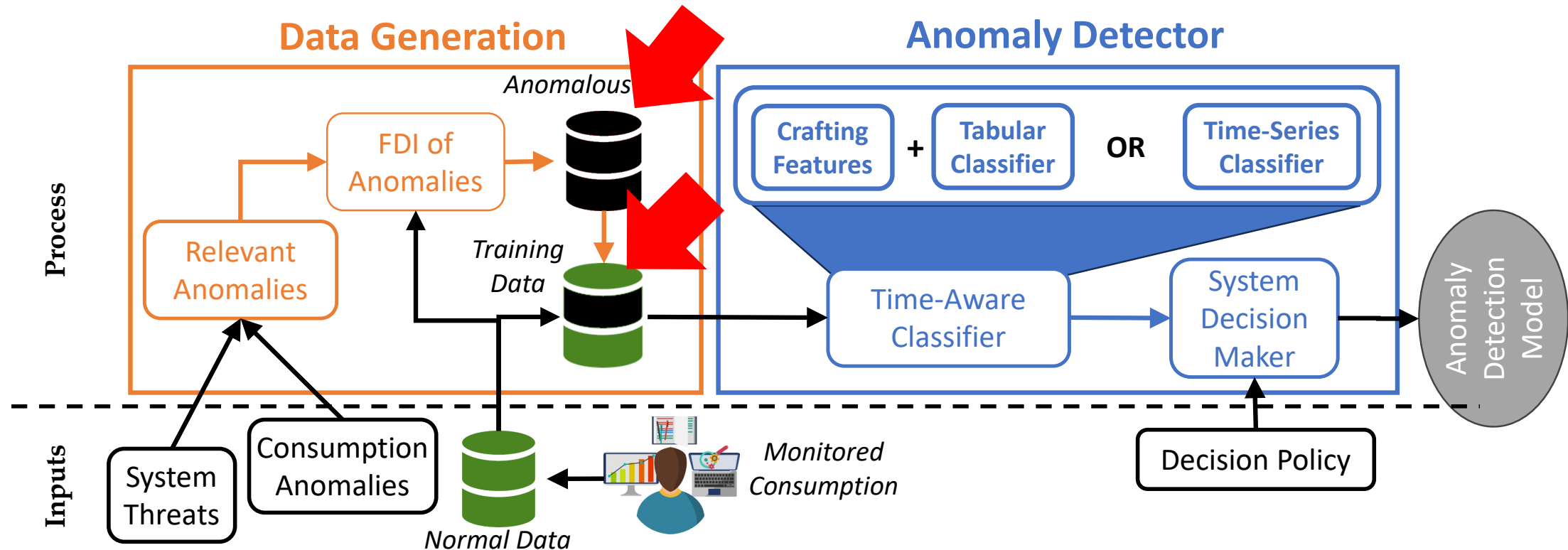




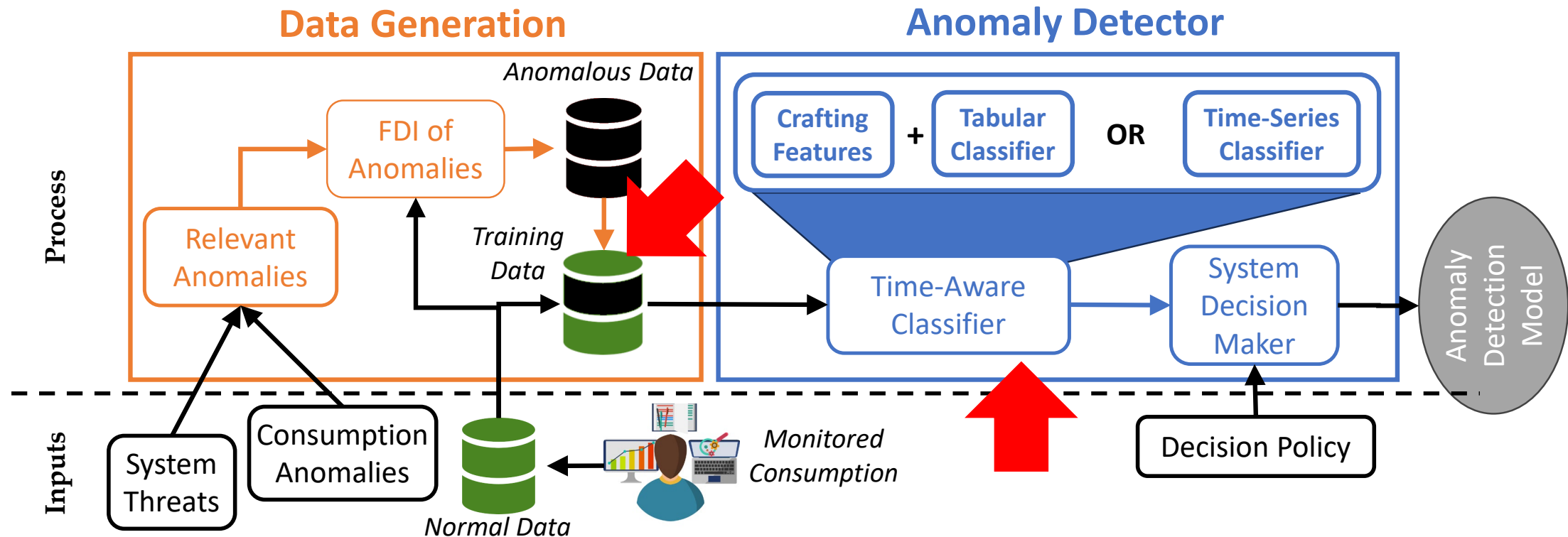
- For any case study, we apply the following:
  - Each anomaly is simulated multiple times using False Data Injection on clean data traces



- For any case study, we apply the following:
  - This creates a labelled dataset of power consumption,



- For any case study, we apply the following:
  - That can be used to craft anomaly detectors



- Means that with just two inputs:
  - Normal power consumption data
  - Threat listing for a specific case study
- It is possible to automatically derive an automatic procedure that outputs the model of an anomaly-based intrusion detector
- Nice, but maybe sounds too easy? Does it work?



- We conducted an extensive experimental campaign using many public datasets, applying our methodology on each of them

Dataset Name	Size (rows)	Monitored Devices	Monitor Frequency	Total Features	Consumption Features	Other Numerical Features	Textual features
PMJ Hourly	>1M	13 devices	Hour	1	1 – MW	0	0
Individual Household	>2M	1 device	Minute	7	3 - global Active, Reactive, Intensity of energy	4 – Voltage, Sub1, Sub2, Sub3	0
Solar Power	136k	2 city areas, 21 and 22 devices each	15 Mins	4	1 - Daily energy yield	3 – AC/DC power, Total yield	0
Steel factory	95k	3 load types	15 Mins	9	3 - Usage kWh, Lagging Reactive kVarh, Leading Reactive kVarh	4 - CO2, Secs from 00.00 am, % Lagging Power, % Leading Power	Weekday, Holiday
Tetuan City	150k	3 city zones	10 Mins	6	1 - City zone power consumption	5 - Temperature, Humidity, Wind Speed, General diffuse flows, Diffuse flows	0
Mississippi MSU-ORNL	70k	4 breakers + controller	unknown	116	29 for each controller	0	0
Goi Energy	6M	600 devices	Hour	1	1 - kWh	1 – Imput	0

Within the old PTR22-24 and for the PTR25-27!

- We conducted an evaluation campaign using different ML algorithms
  - Tree ensembles, NNs, Recurrent NNs (time series-oriented)
  - Using some of the datasets surveyed before
  - Injecting anomalies and seeing how easy it is to detect them
  - Metrics: Accuracy, MCC, False Positives / hr, TPR (Coverage), Detection Delay

Dataset	PMJ Hourly			Household			Solar Power			Steel			Tetouan			Average		
Classifier	GNB	XGB	GRU	GNB	XGB	GRU	GNB	XGB	GRU	GNB	XGB	GRU	GNB	XGB	GRU	GNB	XGB	GRU
ACC	0.836	0.996	0.974	0.522	0.987	0.966	0.519	0.942	0.916	0.523	0.990	0.961	0.779	0.987	0.981	0.636	0.981	0.960
MCC	0.587	0.975	0.808	0.046	0.829	0.592	0.054	0.845	0.455	0.053	0.941	0.796	0.512	0.970	0.933	0.250	0.912	0.717
FPh	0.13	0.00	0.00	56.60	0.02	0.04	0.65	0.03	0.01	0.51	0.00	0.01	0.23	0.02	0.02	11.62	0.01	0.02
TPR	0.691	1.000	0.691	1.000	0.982	0.505	0.607	1.000	0.578	0.481	0.997	0.918	0.793	1.000	0.997	0.714	0.996	0.738
DD (mins)	0.0	19.6	0.0	0.1	4.1	5.7	125.4	21.4	162.3	78.2	13.0	51.2	0.0	1.6	5.7	40.7	9.0	45.0

- But are all anomalies equally difficult to detect? **Hard to tell!**
- We repeat experiments with a multi-class classifier
- Normal vs A1/A2/A3 instead of normal vs anomaly

Classifier	TPR (Coverage, Recall)							DL (avg)						
	LDA	GNB	DT	XGB	ET	RF	LB	LDA	GNB	DT	XGB	ET	RF	LB
A1 (low peak)	0.000	0.667	0.750	0.833	0.750	0.833	0.667	0.00	0.00	26.67	15.00	26.67	33.00	16.88
A2 (noise)	0.000	0.333	1.000	0.967	1.000	1.000	0.833	0.00	22.50	19.50	28.45	27.00	34.50	48.60
A3 (zero)	1.000	1.000	1.000	1.000	0.846	1.000	1.000	0.00	6.92	0.00	0.00	0.00	0.00	0.00
A4 (decrease)	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0.00	15.00	15.00	2.14	3.21	8.57	7.50
A5 (increase)	0.000	0.182	1.000	1.000	1.000	1.000	0.818	0.00	0.00	17.73	15.00	10.91	21.82	46.67
A6 (repeat)	0.000	1.000	0.846	1.000	0.846	1.000	1.000	0.00	9.23	39.55	27.69	15.00	34.62	16.15



- This is as good as it can get, but needs additional work to complete it.
- First (easier): ensure that the overall methodology is implemented in a framework that requires minimal intervention from the user (who may not be domain expert) → **Working on it for the PTR25-27**

<https://github.com/tommyipposz/RAINWATER>



- This is as good as it can get, but needs additional work to complete it.
- Second (more difficult): our power consumption detectors aim at detecting all anomalies as best as they can, but some anomalies may be more frequent than others, because the attacks that manifest as such specific anomaly are more likely than others.
- In other words, we must complement this approach with techniques to devise attack paths, which give a more precise view on how an attack is being conducted and how likely/easy it is

- Widely adopted standards such as ISO/IEC 27005, NIST SP 800-30, and IEC 62443-3-2 define a common, high-level workflow
  1. Identify assets, threats, and threat scenarios
  2. Estimate likelihood and impact
  3. Determine cyber risk
  4. Select and apply countermeasures
  5. Evaluate residual risk against an acceptance threshold
- These standards clearly prescribe what steps must be performed, but intentionally leave open how the analysis should be conducted.

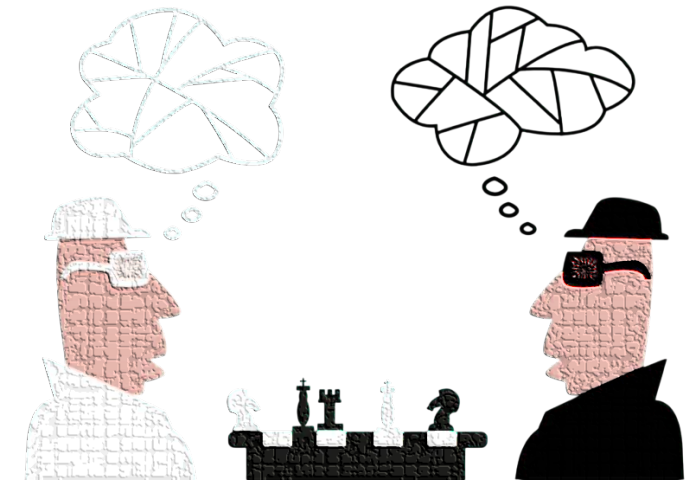


- Widely adopted standards such as ISO/IEC 27005, NIST SP 800-30, and IEC 62443-3-2 define a common, high-level workflow
  1. Identify assets, threats, and threat scenarios
  2. Estimate likelihood and impact
  3. Determine cyber risk
  4. Select and apply countermeasures
  5. Evaluate residual risk against an acceptance threshold
- These standards clearly prescribe what steps must be performed, but intentionally leave open how the analysis should be conducted.

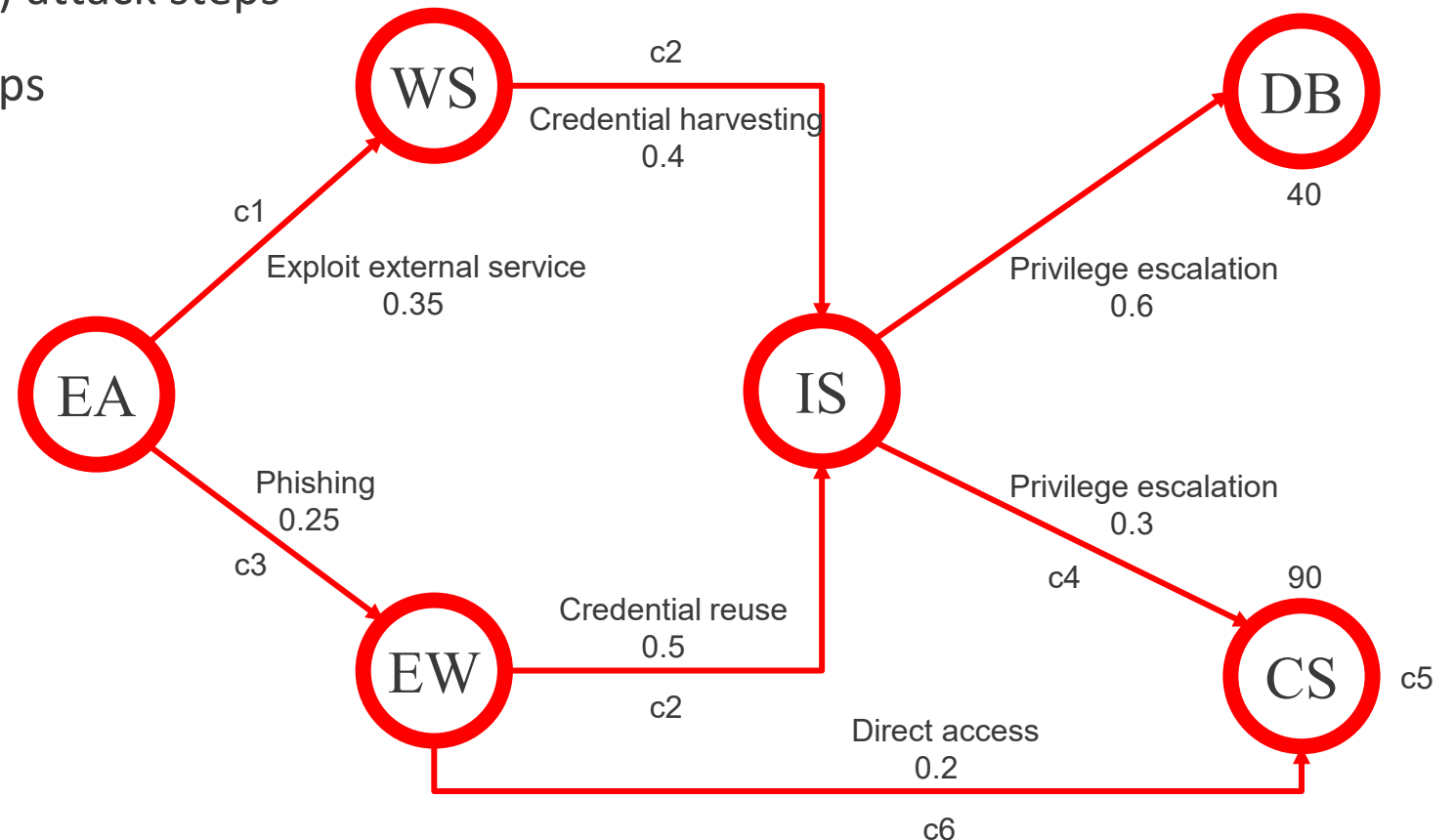


- Limitation 1: Risk is assessed locally, scenario per scenario
  - threat scenarios evaluated independently, no representation of multi-step attacks, dependencies between threat scenarios not modeled, propagation effects ignored
- Limitation 2: Static attacker model
  - risk assessment does not model the attacker as a decision-maker, how does an attacker compare alternative attack paths? how does an attacker react to deployed countermeasures? how does an attacker select a best course of action under a given defensive posture?
- Limitation 3: Risk assessment does not optimize defensive choices
  - traditional risk assessment verifies whether a defensive configuration is acceptable, countermeasures added until residual risk falls below a threshold, costs and resource constraints implicit, no formal trade-off between cost and effectiveness

- A mathematical framework to study strategic interactions among multiple decision-makers
  - agents with possibly conflicting objectives
  - decisions taken under strategic interdependence
  - outcomes evaluated through utility or loss functions
- Using Game Theoretical Approach, but why?
  - Leader–follower strategic interaction, where defender acts before the attack takes place
  - Asymmetry of roles: defender commits to a strategy first, attacker observes and reacts
  - The defender anticipates the attacker's reaction when choosing defenses



- Attack Graph: A system-level representation of multi-step attacks
  - nodes represent system states or assets
  - edges represent feasible (multi-stage) attack steps
  - likelihoods associated with attack steps
  - impacts associated with target states





- Attack Graph: A system-level representation of multi-step attacks
  - nodes represent system states or assets
  - edges represent feasible (multi-stage) attack steps
  - likelihoods associated with attack steps
  - impacts associated with target states
- Players: defender (leader), attacker (follower)
- Strategy spaces
  - defender selects countermeasures (cost/budget)
  - attacker selects an attack path
- Rules
  - defender commits to a strategy first knowing it will be observed by a rational attacker
  - attacker observes the defender's strategy and selects a best-response
- Payoff
  - defender wants to minimize its risk
  - attacker represent a worst-case adversary who wants to maximize the defender's risk



Andrea Bondavalli, Andrea Ceccarelli, Manuel Drago,  
Paolo Lollini, **Tommaso Zoppi**  
RCL Group - University of Florence - Italy  
e-mail: [tommaso.zoppi@unifi.it](mailto:tommaso.zoppi@unifi.it)