FIREMAN

Framework for the
Identification of
Rare
Events via
MAchine learning and IoT
Networks

D2.2 System Architeture Design

v1 Oct. 2019

Introduction

- Leader: LUT
- Partners: SEAT, Oulu, AIT
- Planned: Propose a general framework based on the 3-layer model with a step-by-step procedure.
- Relation to the other WPs: this will provide the framework to all FIREMAN activities

Results

- Propose a 3-layer model of industrial cyber-physical systems
- Propose a step-by-step questionnaire that will define the boundary conditions for designing a tailored rate-event detection
- We applied this framework to three different cases available at LUT

3-layer model for the system architecture design

General framework for specific applications

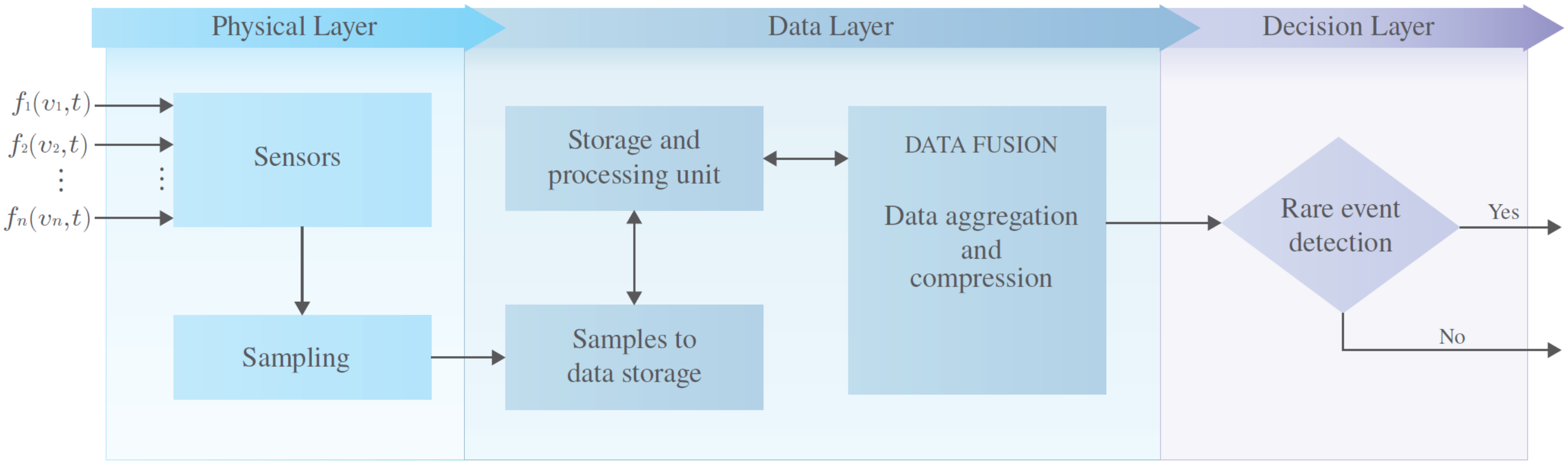


TABLE I: Related questions that can be answered by the proposed framework

S.no	Topic	Related Question
Q0	Anomaly	What is the problem? Is the Rare Event known or unknown?
Q1	Sensors	What kinds of sensors will be used? How many of each can be used and where they can be located?
Q2	Sampling	Which type of sampling will be used? Periodic, event driven or mixed (hybrid)?
Q3	Communication	Which type of communication system (access and network technologies) will be used?
Q4	Data storage	Where the data from sensors are stored and processed?
Q5	Data fusion	How the data should be clustered/aggregated/structured/suppressed?
Q6	Event detection	How to make the ultra-reliable rare event detection based on ML algorithms?

Tennessee Eastman Benchmark

Case 1

Guiding questions (1)

What is the problem?

- Lack of accuracy to detect an event from measurements
- Check the previous table

Is the rare event known?

- Yes, there are 21 types of events (step changes, random variation, valve sticking etc),
- Some are rarer than others and difficult to detect.

A. What are the sensors? How many of each should we use and where they should be located?

- Each of the 52 variables are sensors, no other can be added (benchmark).
- However, we can remove them and analyze the data set with less variables.

Guiding questions (2)

B. Type of sampling? If periodic, Δt? if event driven, how to define the event that will trigger sampling, if hybrid same as above?

- Periodic sampling; all 52 variables are sync.
  - Variables 23 – 36: frequency 0.1 h = 6 min
  - Variables 37 – 41: frequency 0.25 h = 15 min
  - Other variables (1-22; 42:52): frequency 0.05 h = 3 min
  - In all data sets, the sampling frequency is periodic, with Δt = 3 min; eg, the measurements for variables 23 – 36 are repeated twice.
- We cannot get more samples, but we can try to reduce the number of samples and design "events" accordingly.

Guiding questions (3)

C. **What communication technologies can be used?**

- There is no mention about communications in this benchmark.
- Communications is then assumed perfect.
- We can test different communication techniques.

D. **Where the data from sensors are stored and processed?**

- There is no mention about data storage and processing in this benchmark.
- We have freedom to test different topologies/architectures (e.g., cloud, fog, edge, etc.)

Guiding questions (4)

E. How data should be clustered/aggregated/structured (fused)?

- As a benchmark, the data are already structured and synced.
- We could restructure them if we change the number of variables and/or the sampling strategy/rate.
- We have some freedom to play.

F. How to make the ultra-reliable rare event detection for our problem?

- The anomalies are known, and PCA is the “technique” to be compared to.
- Other techniques are: PLS, FDA and CVA
- Our method needs to win against these methods.
- It is good to set the false alarm rate at 1%.

Tennessee Eastman Benchmark

Description of the use case (incl. KPIs)

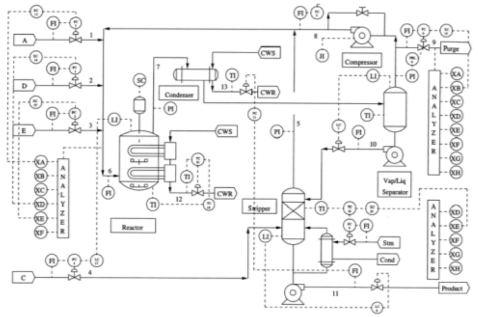
Please briefly describe the use case and the relevant KPIs here.

The process have five main equipments and produces two products from four reactants.

KPI's:

- Product G and H quality
- Operating constraints (low and high normal/shutdown limits)
- Missed detection rate
- Time to detection

Photo that represents the process



Description of the data we need

Please describe the data we need for the FIREMAN Use Case. Think of perfect conditions.

As a benchmark scenario, we have all data that is possible to have.

Description of the data we have (incl. used technology)

Please describe the data we already have access to or we will get access to.

52 measurements including 12 manipulated, which include:

- Feed flow
- Compressor recycle & purge valves
- Separator pot liquid flow
- Stripper liquid product flow

Description of the problems regarding data gathering and preparation

Please describe all problems regarding data gathering and preparation.

As a benchmark scenario there is no need for data gathering and preparation.

Low voltage direct current microgrid

Case 2

Guiding questions (1)

What is the problem?

- Consumer interruption
- Correct battery storage dimensioning to ensure power supply for majority of interruptions

Is the rare event known?

- Yes, it is frequently associated with external nature conditions
- Some are associated with electrical stability

A. What are the sensors? How many of each should we use and where they should be located?

- Overhead branches should count with tree status monitoring (camera, radar, UAV, etc). At least 1 sensor in each branch
- Precise weather information (rain and snow) at overhead lines.
- Underground cable monitoring (humidity, insulation state, etc), sensor along each branch.

Guiding questions (2)

B. Type of sampling? If periodic, Δt? if event driven , how to define the event that will trigger sampling, if hybrid same as above?

- Periodic sampling; 25+ variables are sync.
  - 1 minute each sample
- Sampling could be triggered by “risky” weather and external conditions.
- Periods of most probable event occurrence.

Guiding questions (3)

C. What communication technologies can be used?

- Wired ethernet based
- Wireless 4G, 5G and beyond could be implemented

D. Where the data from sensors are stored and processed?

- Data is stored from LAN to microgrid servers and then to cloud.
- Previous solution is not as fast as required, since real-time monitoring is necessary.
- Wireless+Cloud architecture could be tested

Guiding questions (4)

E. How data should be clustered/aggregated/structured (fused)?

- Data should be aggregated like other variables are stored, with same time stamp.
- Depending of the information of variables monitored, some of them could be suppressed

F. How to make the ultra-reliable rare event detection for our problem?

- Events should be able to detect, when adding variables that monitor physical condition of the grid
- Our method needs to detect reliable beforehand

LVDC Microgrid

One pager
FIREMAN Use Cases

Description of the use case (incl. KPIs)

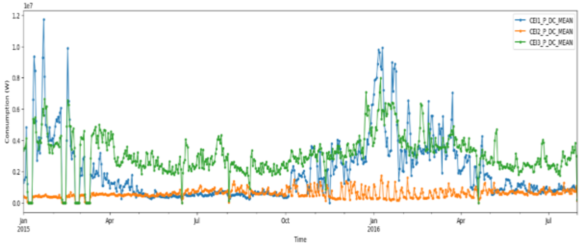
Please briefly describe the use case and the relevant KPIs here.

The microgrid supplies energy to 3 different customers, it counts with battery storage

KPI's:

- Reliability
- Islanding capability
- Microgrid downtime

Photo that represents the anomaly



Description of the data we need

Please describe the data we need for the FIREMAN Use Case. Think of perfect conditions.

Precise climate conditions at the location of the microgrid, wires and other elements.
External conditions sensors such as state of tree branches.

Description of the data we have (incl. used technology)

Please describe the data we already have access to or we will get access to.

More than 25 electrical variables, being the most important, Power consumption, battery storage charge state, voltage profile and currents.

Description of the problems regarding data gathering and preparation

Please describe all problems regarding data gathering and preparation.

Connection to data gathering at metering points is sometimes lost which doesn't give any variable output and might be related with occurrence of events. No effective data gathering solutions.

Adaptive Robotic Gas Metal Arc Welding of Steels

Case 3

Guiding questions (1)

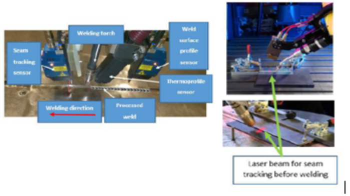
What is the problem?

Is the rare event known?

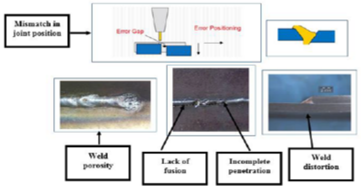
A. What are the sensors? How many of each should we use and where they should be located?

One sensor that acquires all measurements described in the last slide.

Technology used



Weld problems and defect



System considerations when employing AI for welding control

Welding process control	Robot control	Joint geometry control	Sensor control	Monitoring control	Machine setup control
Heat input (voltage, current, wire feed rate, welding speed)	Manipulator dexterity	Joint fit-up Eliminate joint fit-up errors	Weld shape	Weld shape	Grounding of the machine and the work-piece
CTWD (electrode stick out)	End effector maneuvering	*Joint position *Eliminate position errors	Weld profile (throat thickness)	Weld profile (throat thickness)	Work cable stability
Shielding gas flow rate	Motion and trajectory planning		Weld penetration (weld root)	Weld penetration (weld root)	Long welds and tandem arc welding
Mode of metal transfer	Axis control		Weld distortion	Weld distortion	vibrations
	Linkages control		Weld spatter and porosity	Weld spatter and porosity	
	synchronization		Compliance with welding procedure specifications (WPS)	Compliance with welding procedure specifications (WPS)	
	Torch angle				
		Big Data Analysis			
Welding data analysis	Robot motion analysis	Joint geometry analysis	Sensor control analysis	Monitoring control analysis	Machine stability analysis

Guiding questions (2)

B. Type of sampling? If periodic, Δt? if event driven , how to define the event that will trigger sampling, if hybrid same as above?

- Periodic sampling, $\Delta t = 0.04$ s (it might be defined)
- Event-driven might be directly possible
- Possible to generate new datasets with very low time granularity to synthetically test event-driven approaches

Guiding questions (3)

C. What communication technologies can be used?

- Communication technology is not used now.
- USB driver to take the data stored in the laser monitoring device.
- Possibility to use available solutions and real-time monitoring.

D. Where the data from sensors are stored and processed?

- Local storage and processing in PCs
- No clouds, fogs etc.

Guiding questions (4)

E. **How data should be clustered/aggregated/structured (fused)?**
- Data is directly saved in csv, and treated "manually" in excel

F. **How to make the ultra-reliable rare event detection for our problem?**
- Usual statistical approaches to detect peaks, which may mean anomalies.
- Check thesis: https://lutpub.lut.fi/handle/10024/158437

One pager
FIREMAN Use Cases

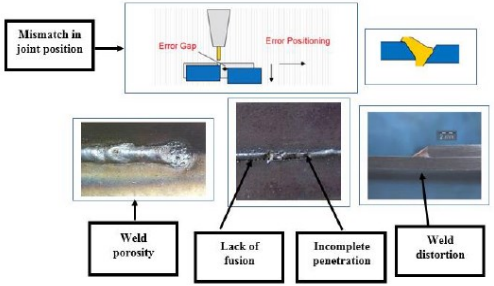
Adaptive GMAW of Steel

Description of the use case (incl. KPIs)

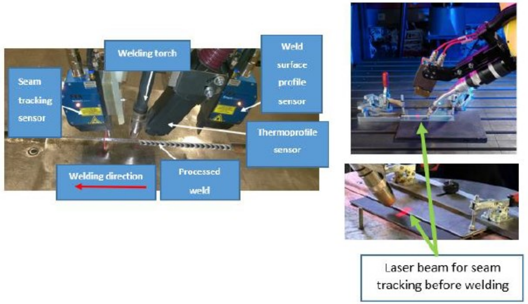
Please briefly describe the use case and the relevant KPIs here.

In Gas Metal Arc Welding (GMAW), the filler wire and shielding gas are key variables of the process. The wire feed rate and the gas flow rate need to be monitored as they constitute to the generate of heat input and mode of metal transfer. The heat input is determined according to the efficiency of the power source in addition to welding voltage, current and welding speed. The contact to work distance (CTWD) which comprise the filler wire stick-out and diameter of the wire need to be measured since they contribute to weld depth penetration. When GMAW process is adapted to robot, essential parameters like torch angle and working angle need to be set accurately. In cases where off-line technology is used, the calibration between the digital twin and the real robot must be set too. The trajectory and movements of the robot and weaving technique of the welding process causes instability to the welding cables and connections which leads to loose-grounding-contact and vibration issues. The relationship between these components and welding parameters create rare events which lead to weld quality problems (see figure on side) as well as weld productivity challenges. The need to monitor, identify, detect and prevent such rare events is imperative.

Photo that represents the anomaly



Description of the data we need



Description of the data we have (incl. used technology)

Please describe the data we already have access to or we will get access to.

Welding parameters, which in this case referred as data such as welding voltage measurements, current, filler wire diameter, gas flow rate, wire feed rate, temperature reading, CTWD measurements, and torch position. These are already available. The technology used is automated robot and laser-based infrared thermography thermoprofile scanner (TPS).

Description of the problems regarding data gathering and preparation

Please describe all problems regarding data gathering and preparation.

Problems of weld defects and imperfections (e.g. weld porosity, lack of fusion, lack of penetration, weld distortion, heat affected zone softening, hydrogen inclusion, cracking) are bound to occur due to incorrect welding parameters settings, material preparation, pre-weld and post-weld heat treatments, etc. Data gathering and preparation is time consuming. Devices used for data gathering must be set appropriately and within defined range. Data captured can have inconsistencies due to wrong device setups and malfunction in sensor reading. Possible anomalies can arise from loose-grounding contact, vibrations during welding due to machine instability.