FIREMAN

WP6 Deliverable 6.3
Demonstration descriptions

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

Title of Deliverable:	Demonstration descriptions
Status-Version:	Final-1.0
Delivery Date:	31/01/2022

Contributors:	Hirley Alves, Jean Sant'Ana, Daniel Gutierrez Rojas, Pedro Nardelli, Pavol Mulinka, Charalampos Kalalas
Reviewers:	Internal Review
Approved by:	All Partners

Document Revision History

Version	Date	Description
v0.1	16 November 2021	Table of Contents
v0.2	24 December 2021	Main contents added
v0.3	10 January 2021	Additional contents added from stable cases
v0.5	20 January 2021	First version delivered for internal review
v0.8	28 January 2021	Second version delivered for internal review
v1.0	31 January 2022	Approved form of Deliverable D6.3

Table of Contents

1 Introduction

1.1 Objective of the document

The objective of "Deliverable 6.3 demonstration descriptions" (D6.3) is to provide the descriptions concerning the experimental deployment of four test cases of FIREMAN approach; thus, D6.3 serves as a reference for the demonstration work to be carried throughout the last stages of FIREMAN. The document describes four use cases initially discussed in "D6.1: Plan the Demonstrations and Targets" [1]. We rely on the guiding questions, introduced in "D2.2: System Architecture Design" [2], to guide the planning and development of the planned demonstrations.

Moreover, the first description is the 5G Test Network (5GTN) at University of Oulu, which is a 5G IoT network with around 300 devices with different sensors. The second describes the microgrid network structure from ERIGRID, that will serve as a first test network for LUT data reduction algorithms, that is expected to apply their results to LUT's real-time simulator at a later stage. The third is a setup at Aalborg University, Denmark, for anomaly detection and classification in a collaboration between LUT and CTTC. Finally, the SEAT use case, which is FIREMAN industrial partner, where we aim at applying FIREMAN solutions on its paint shop drive chain and spindle fault detection.

1.2 Structure of the document

The document is structured as follows:Section 2 presents the description of University of Oulu 5GTN. Section 3 shows the description of LUT's work with ERIGRID's microgrid network. Section 4 describes the test setup at Aalborg University for electric faults in grid converters carried out by LUT and CTTC. Section 5 shows the work done at SEAT, together with some insights of FIREMAN front-end. Finally, Section 6 presents the conclusions.

2 5GTN Use Case

A 5GTN dataset is first introduced on "D6.1: Plan the Demonstrations and Targets" [1]. Furthermore, we investigate techniques to apply some of the FIREMAN techniques presented in previous documents to apply on this dataset as a first insight and demonstration of what could be done at the 5GTN case. Similar to "D2.2: System Architecture Design" [2], we answer the guiding questions for the 5GTN dataset as we have it now.

- **What is the problem?**

  Identify if room's maximum capacity is being attended, malfunction in the ventilation of the rooms, problems with AC.

- **Is the rare event known?**

  In parts. We expect peaks in the CO_2 measurements for bad ventilation, and correlated peaks in the sound, temperature and CO_2 for crowded rooms, however these values are still uncertain.

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

  There are three types of sensor devices: CO2 devices with light and motion; Sound devices, with light and motion/ and underground Moisture devices, measuring soil moisture and pressure. All devices also measure temperature and humidity. There are a total of 452 deployed devices.

- **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, one sample each 15 minutes.

- **What communication technologies can be used?**

  Class A LoRa devices are deployed with the sensors, all using the same spreading factor and carrier frequency.

- **Where the data from sensors are stored and processed?**

  Data is stored on ".csv" files and made available by the responsible team to FIREMAN. Files are originally split into devices, but the same can be easily merged into a single time-series for the whole network. Right now we have access to the data collected from August 2020 to February 2021, with possibility to gather new data in the future.

- **How data should be clustered/aggregated/structured (fused)?**

  We have total freedom to play with the variables, either considering correlation, removing redundant information, reducing and/or increase the number of samples.

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

  The network itself has a 20% error rate. We aim at reducing the number of transmissions to increase the reliability of important, *i.e.*, events, transmissions. We aim at a higher

false alarm rate to diminish true negative. False alarms are not a huge problem compared to production lines in industry, for example.

2.1 Fitting into Analytical Framework

The first step we consider is on fitting the 5GTN network into the analytical framework for LoRa networks from [3], that is also presented on "D3.3a: Design and Evaluation of the Data Acquisition Framework [PART A]" [4]. This is important to allow us to have fast first insights on any methods techniques applied to the network, since gathering results from the analytical equations is much faster from any simulation setup. The downside from it is that the results could be hard to adapt from the real network setup. This happens specially because the model consider a network with uniformly distributed devices. On the other side, our use case consider that all devices present the same transmission period, which is the other strong consideration the framework does. Finally, this approach can give us leads whether a technique would present interesting results as we expect, saving us time for the future practical experiments.

To fit the real network into the framework, we use two key statistics. The first is the outage probability or outage rate, which in practice we will be looking at the collision probability. As we dig a bit into the transmissions attributes as received power or SNR, we see that these values are way above the LoRa characteristics thresholds for the identification of the signal by the gateway. This is also intuitive since these networks support long distances above kilometers, and the 5GTN network presents a couple of hundreds meters distance long. Thus, we can imply that the majority of losses at the gateway are due to collisions. If we check the packets at the server, we can see that there are some loses due to the decoding of the packet at the application layer. However, since the framework does not consider it, we will not take it into account. The second statistic is the network usage, that basically takes into account the number of devices and the activity time of the devices. As we have both of these values, and the LoRa attributes as spreading factor and packet size, we can tune the environment aspects to fit the network.

2.2 Applying the Event-Driven Approach

The Event-Driven approach was introduced in the FIREMAN with scope on "D3.1: Report on Physical Process Data Modelling" [5] applied to sampling reduction and later on "D4.2: Report on Data Reduction Techniques" [6] as a data reduction technique for event detection. Here, we plan to apply these techniques on the sensor devices to reduce the amount of transmitted data. As stated on D4.2, the data reduction provided by the approach increases the rare event detection due to the reduction of redundant information fed to the detector. Here, we are using a ALOHA-like LoRa network. The reduction of transmissions implies on decreasing the amount of traffic generated by the network. This results in a smaller amount of collisions, which as previous stated, is the main contributor to the outage rate.

Altogether with the threshold from the Event-Driven approach, some rules of transmissions are to be deployed. One example is considering that devices transmit at least one message each period of time, even if no event is detected by the approach. Others are in considering past knowledge for each device, since each room has its own normal behavior state. Finally, applying these techniques is not exactly feasible to the analytical framework. However, we

plan to first apply it to the dataset and update the network usage with the percentage of data reduced with the technique.

2.3 Data Imputation after communication layer

Even after data reduction at the sampling and transmission phase, the loss of some packets is still natural to LoRa-like networks. Also, we are not taking into account acknowledge packets, thus the device has no information whether a packet was successful transmitted or not. Thus, a data imputation technique is expected after the communication layer to deal with these losses. Reconstruction techniques similar to the ones implemented in WP3, e.g., [7], are expected to be tested.

3 ERIGrid

ERIGrid dataset was obtained through a Real Time Digital Simulator (RTDS) by placing faults with different parameters on the low voltage CIGRE microgrid benchmark. For this test bed, we designed the topology in the RTDS as a three-phase diagram, including all the control of variables and switches to change the parameters. At the same time, another member of the group was working on translating the automatic data collection algorithm into C++ as required by RTDS. Once the diagram was plotted, we ran different simulation scenarios as an initial test in order to adjust the energy balance in the grid when it was island mode. Furthermore, we checked the current level was in sync with the theory. In Fig. 1 is depicted the topology with all the fault points. Also, in Fig. 2, we can see the hardware elements involved in the whole real time simulation process.

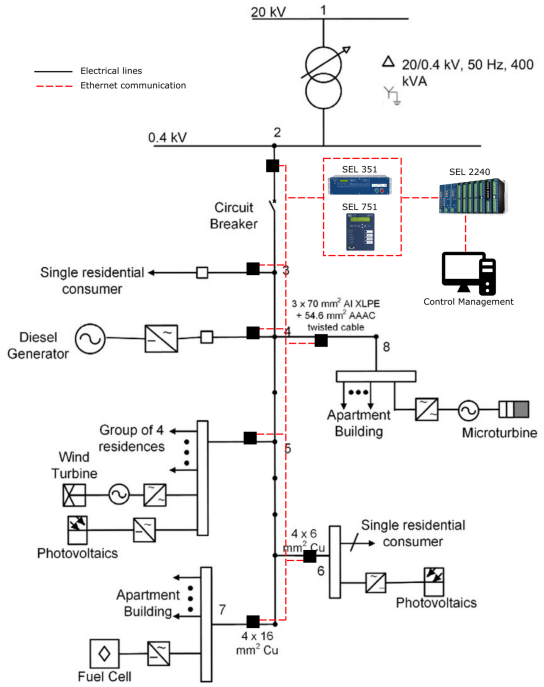

Figure 1: Low voltage CIGRE benchmark microgrid

- **What is the problem?**

  When a high resistance fault occurs in microgrids, it can be challenging to detect as the current levels measured by the sensor change according to the topology of the system. If a fault occurs and is not detected, it might lead to hard of the physical elements of the grid.

- **Is the rare event known?**

  Yes, rare is event is classified as a high resistance fault in the grid. Electrical fault have different parameter and nature, and high resistance are among the rarest.

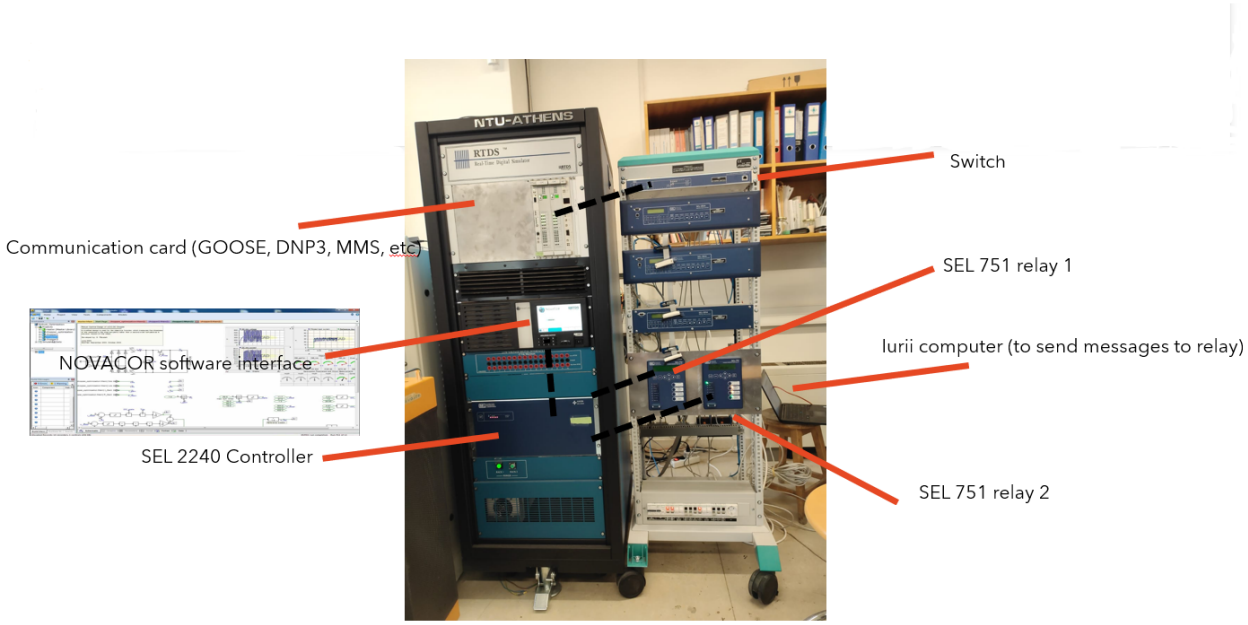

Figure 2: Project team and elements used during research

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

  The sensors are located in each bus of the microgrid where the relays are placed, but they all yield similar results, that's why any measurement device located after the main switch is enough for the application.

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

  The sampling can be done constantly in time series until an event happens. However, in order to save memory from the sensors the best way is once a fault occurs in the system due to change of variables in the sensors, the recording of measurements will be triggered for a specific time window.

- **What communication technologies can be used?**

  Usually for this setup, wired communication technologies are used (Ethernet, optic fiber and coaxial wire) but also newer cellular Long Term Evolution (LTE) and 5G communication could be used to send the data across the network. A communication infrastructure of the setup involves the controller and the sensors (relays or intelligent electronic devices) is seen in Fig. 3.

- **Where the data from sensors are stored and processed?**

  In microgrids, there is a central microgrid management controller that takes care of protective and control actions. In this system is where the measurements coming from the relays are stored, or they could also be sent to the cloud and process remotely.

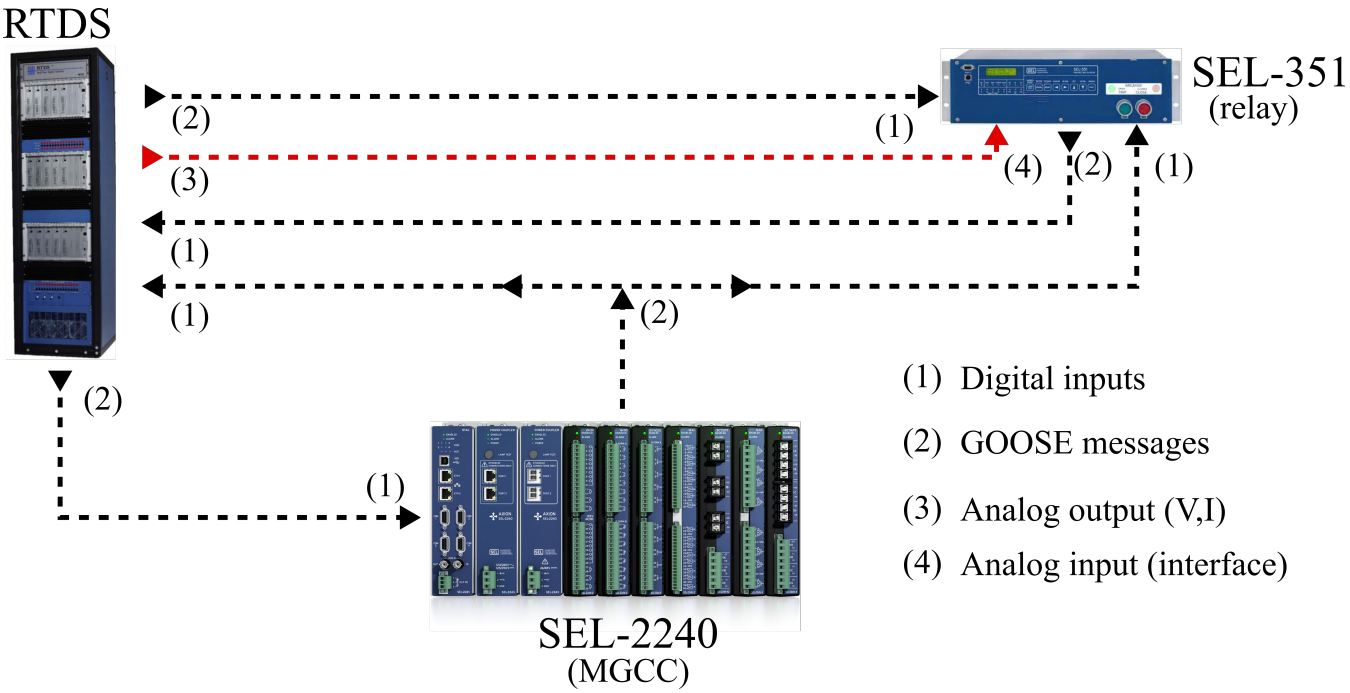

Figure 3: Communication schematic.

- **How data should be clustered/aggregated/structured (fused)?**

  When the data goes to microgrid central management controller from multiple relays, the data is organized to check that the measurements are consistent. The variables are independent and do not need aggregation.

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

  The aim is to be able to detect fault when sudden and small changes in the current happens in the system with high, at least 90% accuracy. These changes are currently very hard to detect for traditional protections, and the need for an automated and self-learning technique is a must.

3.1 Description of the tests

The focus was on collecting fault data from RTDS by using an algorithm that changed simulation parameters for every run and then collect the data. As simulation parameters there were 5 different fault points, 8 different types of fault, 0.1-100 Ohms fault resistances in step of 5 and 3 microgrid scenarios. Data from microgrid was collected from the output files or RTDS, in most of the cases were current for data but also some tripping and logic signal coming from the devices to the switches located in the microgrid were recorded. The total data size was around 9 GB, so we upload to private server cloud for further processing. This data will be used to feed machine learning-based algorithms for hard fault detection in microgrids.

In order to fit the system with the framework, we follow each of the steps for rare event detection. The measurements are collected in real time and based on a pre-trained decision factor, a model will computed whether the signal being sensed and transmitted to the central

management controller is seen as a fault. It is important to also classify the events seen by the model to separate the "common" events from the "rare" events. There might also be instability in the system due to variation in the supply from the distributed energy resources and might change the current levels even if the microgrid is not in fault. Under these scenarios, the model should have capability to identify such changes and frame them as "non-fault."

The goal is to use the framework and incorporate the models into a decision-making application that can be integrated in the devices to act in real-time, bringing more resilience to the microgrid.

3.2 Applying the Event-Driven Approach

Since the measurement are periodic sampling, event-driven will only play a role on when these measurements are recorded. The event-driven approach in the framework should be able to classify the events in categories. This will have big impact in the data traffic and storage. Periodic sampling of 80 point per 60 Hz cycles is expensive when there are around 10 different measurements in each relay location, and they all have to be sent to the central controller for processing. This is why each relay should have a fault recording element that allows each of them to record only when a fault is present in the system and therefore process in the central controller for selection. The fault recording element must have in-build the data model to make correct fault discrimination even in the presence of fault classified as "rare."

4 Grid-Forming Converter in AAU

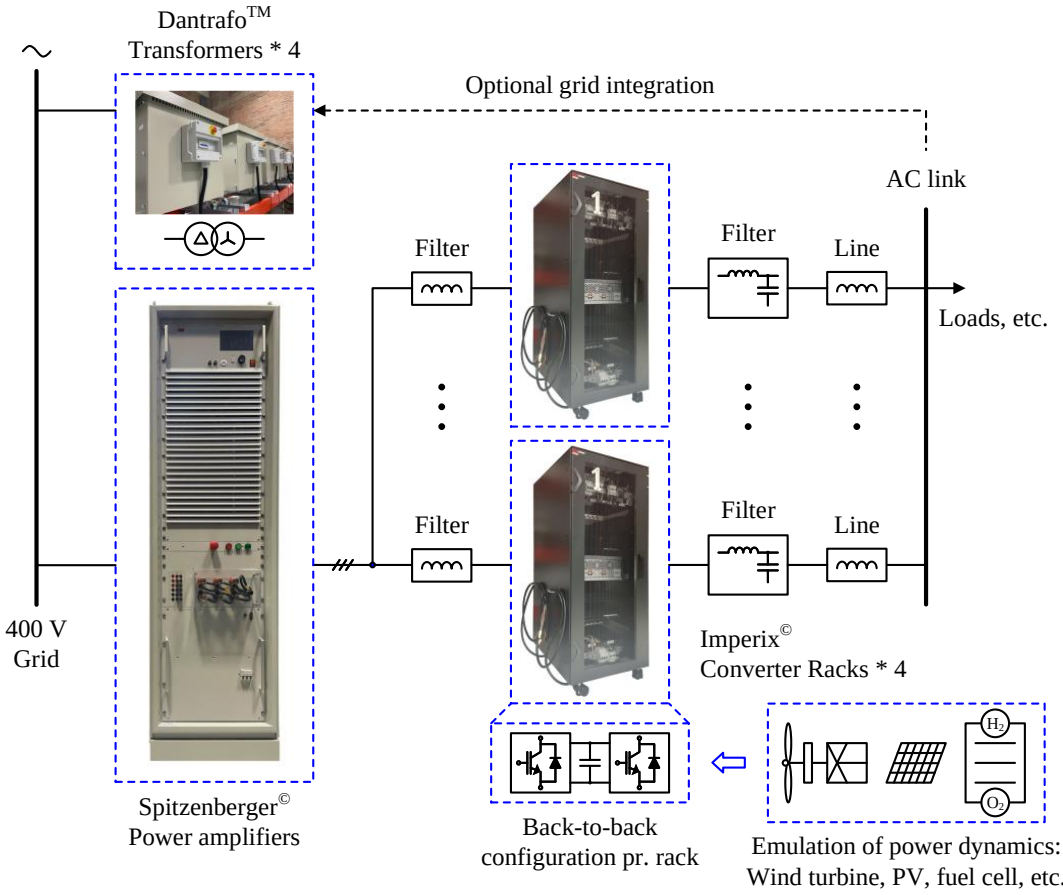

Figure 4: Test setup in Aalborg University (AAU) for data-driven anomaly detection and classification.

In Fig. 4, a state-of-the-art infrastructure is shown which can be used for grid-tied converter level characterization and testing under abnormal grid conditions and frequency scanning-based stability validation. It can be used for test case studies on microgrids and active distribution networks, providing the flexibility for any power topology. It comprises a high-fidelity grid linear amplifier of capacity 45 kVA, 400 V (L-L) having a large signal bandwidth of 30 kHz. It can be used for testing of grid compliance standards, such as IEEE 1547, UL 1741. Furthermore, power hardware-in-the-loop (PHIL) for real-time distribution network studies can also be accommodated. Considering the network expansion of power electronic dominated grids, anomalies can be hard to classify for grid forming converters due to limited current contribution. Hence, relevant data has firstly been extracted from the test setup under various faults and grid parameters and then investigated upon extensive training to formulate a data-driven anomaly classification tool.

- **What is the problem?**

Electrical fault under the test setup seen in Fig. 4 are hard to detect or/and select due to low contribution of current from the grid forming converters.

- **Is the rare event known?**

  Yes, the problem is known. Electrical faults in a custom-made grid with many non-linear power electronic devices.

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

  The sensors are intelligent electronic devices, relays, CT or VT connected to Phasor Measurement Unit (PMU) or central controller that is in charge to store and process the data. They can be located at each node on the grid.

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

  The sampling is extracted as time series.

- **What communication technologies can be used?**

  Communication technologies used for this application normally run a wired-based Ethernet cable, however, LTE 4G/5G technologies can be applied since they can meet the reliability and latency requirements.

- **Where the data from sensors are stored and processed?**

  The data is sent from sensors to a management computer software, where it is processed and converted to the desired format.

- **How data should be clustered/aggregated/structured (fused)?**

  The data obtained from each node of the grid is aggregated only using the relevant variables, such as voltages and currents.

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

  To make reliable detection, the method should be able to detect small changes in the target features during faults.

5 SEAT Paint Shop Drive Chain

The tests are related to the paint shop drive chain, as presented in D6.1 [1]. Fig. 5 presents a schematic picture of the scenario considered.

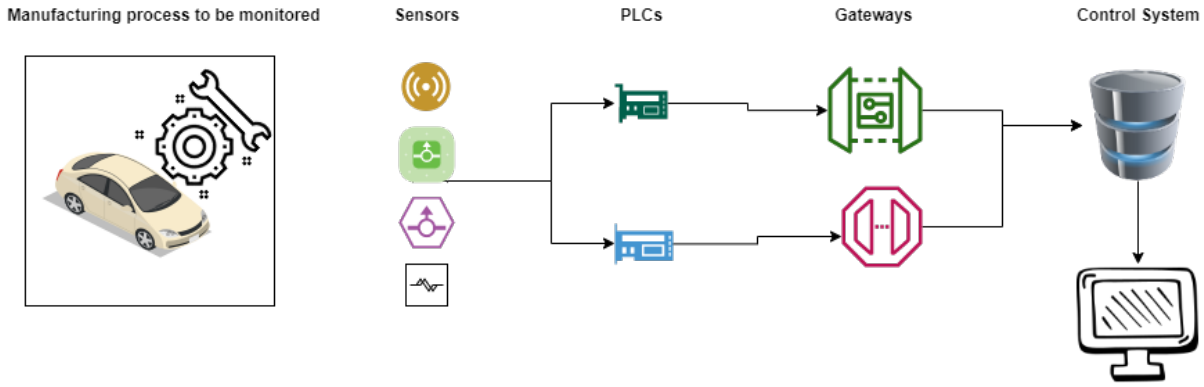

Figure 5: Network setup for condition monitoring [1].

- **What is the problem?**

  Failures in mechanical components (bearing) of the drive chain in paint shop, causing axial displacement which, in turn, results in significant production stops and corresponding loss of bodies.

- **Is the rare event known?**

  Yes, see above.

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

  There are two different types of sensors: a) inclination, sending signals only above fixed threshold values; b) tension, measuring tension of the drive chain. According to the SEAT maintenance team, most of the issues arise from the horizontal movement/misplacement of the drive chain wheels. We have further proposed the installation of new sensor types: a) Inclination sensors that would work in parallel with current sensors and provide measurements with higher granularity; b) Vibration sensors that would be used to monitor failures, in motors pulling the drive chains, associated with axis tilt due to motor bearings. Each gateway can support up to 100 sensors within a range of 350 meters, non-line-of-sight.

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

  Different data acquisition strategies are possible. Aware state heartbeat: 1min, Assessments per heartbeat: 10 (sampling every $\Delta t = 6$sec), Aware state threshold: 15°.

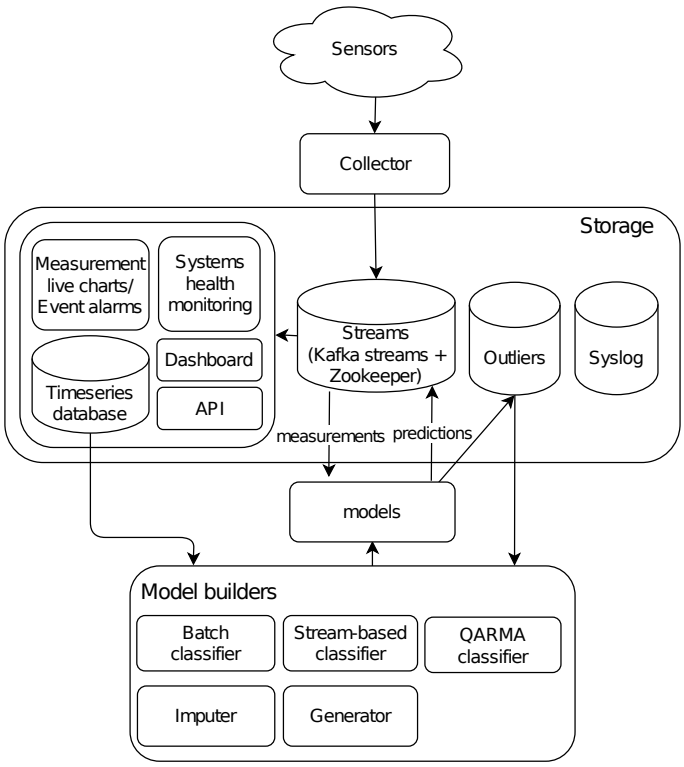

Figure 6: Building blocks of our modular monitoring solution, where the involved interactions (data flows) between individual components are displayed [8].

- **What communication technologies can be used?**

  Sensors use WiFi connectivity, with multiple antennas distributed along the drive chain serving as access points. The gateway can be either Ethernet-based or cellular-based.

- **Where the data from sensors are stored and processed?**

  The gateway collects the sensor data and sends the information online to a cloud-based sensor portal, where a user has the ability to access the information from any web-enabled device and setup notifications through SMS texts, emails and voice call alerts. Historical data can be kept for one year and are available to any number of users. The proposed monitoring solution is explained in details in Sec. 2.2.2 in D6.1 [1], see also Fig. 6.

- **How data should be clustered/aggregated/structured (fused)?**

  The proposed monitoring solution is explained in details in Sec. 2.2.2 in D6.1 [1], see also Fig. 6.

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

  The proposed monitoring solution is explained in details in Sec. 2.2.2 in D6.1 [1], see also Fig. 6.

6 Conclusions

This document presented FIREMAN's four demonstration descriptions. It serves as a base for Task 6.2 Final Project Demos, and for Deliverable 6.2 First Demonstration Trials, which will be released in August 2022, and "D6.4 Final demonstrations and report", which will be released in April 2023. Thus, D6.3 serves as a baseline on the work to be done until April 2023.

References

[1] J. M. de Souza Sant'Ana, E. Eldeeb, H. Alves, P. Mulinka, C. Kalalas, D. Gutierrez-Rojas, and P. J. Nardelli, "D6.1: Plan the demonstrations and targets," *Framework for the Identification of Rare Events via MAchine learning and IoT Networks*, 2021.

[2] LUT, SEAT, OULU, AIT, "D2.2: System architecture design," *Framework for the Identification of Rare Events via MAchine learning and IoT Networks*, 2019.

[3] J. M. d. S. Sant'Ana, A. Hoeller, R. D. Souza, S. Montejo-Sánchez, H. Alves, and M. d. Noronha-Neto, "Hybrid coded replication in ¡sc¿lora¡/sc¿ networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5577–5585, 2020.

[4] P. Nardelli, A. Sena, D. Carrillo, E. N. Tominaga, J. M. de Souza Sant'Ana, and C. Kalalas, "D3.3a: Design and evaluation of the data acquisition framework [part a]," *Framework for the Identification of Rare Events via MAchine learning and IoT Networks*, 2020.

[5] D. G. Rojas, P. Nardelli, C. Kalalas, and C. Papadias, "Report on physical process data modeling," *Framework for the Identification of Rare Events via MAchine learning and IoT Networks*, 2020.

[6] M. Dzaferagic, I. Macaluso, N. Marchetti, C. Kalalas, P. Mulinka, P. J. Nardelli, D. G. Rojas, and I. Christou, "D4.2: Report on data reduction techniques," *Framework for the Identification of Rare Events via MAchine learning and IoT Networks*, 2021.

[7] C. Kalalas and J. Alonso-Zarate, "Sensor data reconstruction in industrial environments with cellular connectivity," in *2020 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC '20)*, August 2020.

[8] P. Mulinka, C. Kalalas, M. Dzaferagic, I. Macaluso, D. G. Rojas, P. J. Nardelli, and N. Marchetti, "Information processing and data visualization in networked industrial systems," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2021, pp. 1–6.