FIREMAN

WP6 Deliverable 6.1
Plan the demonstrations and targets

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

Title of Deliverable:	Plan the demonstrations and targets
Status-Version:	Final-1.0
Delivery Date:	30/07/2021

Contributors:	Jean Michel de Souza Sant'Ana, Eslam Eldeeb, Hirley Alves, Pavol Mulinka, Charalampos Kalalas, Daniel Gutierrez-Rojas, Pedro Juliano Nardelli
Reviewers:	Internal Review
Approved by:	All Partners

Document Revision History

Version	Date	Description
v0.1	07 May 2021	Table of Contents
v0.2	20 May 2021	Table of Contents Updated
v0.3	30 June 2021	First content added
v0.4	24 July 2021	Additional content
v0.5	29 July 2021	Deliverable internal review
v1.0	30 July 2021	Approved and sent to upload on website

Summary

The research is performed within the context of the EU CHIST-ERA project FIREMAN (grant CHIST-ERA-17-BDSI-003).

Table of Contents

1 Introduction

1.1 Objective of the document

The objective of Deliverable 6.1 is to present the first insights and plans of FIREMAN Task 6.1, related to experimental deployment of test cases of FIREMAN approach. The document works around three use cases, where it shows the basics of each use cases, with some description of the network, devices and characteristics we have access. The SEAT use case, which is FIREMAN industrial partner, where we aim on applying FIREMAN solutions on its paint shop drive chain and spindle fault detection. The second is the 5G Test Network (5GTN) at University of Oulu, which is a 5G IoT network with around 300 devices with different sensors. Finally, the microgrid network structure from ERIGRID, that will serve as a first test network for LUT data reduction algorithms, that is expected apply their results to LUT's real-time simulator at a later stage.

1.2 Structure of the document

The document is structured as follows: Section 2 shows the work done at SEAT, together with some insights of FIREMAN front-end. Section 3 presents the description of University of Oulu 5GN. Section 4 shows the description of LUT work with ERIGRID's microgrid network. Finally, Section 5 presents the conclusions.

2 SEAT use cases

2.1 Overview

In the context of WP6, experimental activities related to SEAT predictive maintenance use cases aim to provide new monitoring solutions for:

1. Paint shop drive chain: Monitoring solutions aim to timely detect 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. The key objectives are the reduction of the mean time to repair, the increase of the mean time between failures and the achievement of zero production loss. Predictive maintenance models aim to detect mechanical components' degradation by using large amounts of data to create an early warning system that identifies problems before an axial displacement occurs.

2. Computer Numerical Control (CNC) machines of the machining center: Monitoring solutions aim to predict failures in spindles of CNC machines, which generate defects in components of the gearboxes produced in the SEAT Components plant. The spindle is an element (kind of actuator) of a CNC machine that ensures lineal movement over a surface. The spindle misbehavior is detected after quality losses or failure. The main objectives are the reduction of quality losses, the improvement of the operational equipment efficiency, and the avoidance of unplanned stops and production of not-okay (NOK) parts. Therefore, predictive maintenance models are required to detect the spindle failure and avoid it.

2.2 Activities/Progress

Preliminary data-analysis tasks include features variance, distribution, collinearity, missing values, same values, label imbalance, etc. The available datasets for the spindle-failure use case are composed of over 50 variables, such as axis positions, engine energy consumption, pressure, temperature, fluid levels, preventive control points, failures, quality reports, etc. However, the conducted analysis revealed that the datasets do not include identified failures from previous years. Therefore, and in accordance with SEAT maintenance team, the main focus so far has been on the paint shop drive chain monitoring scenario which is more prone to failures. It is worth noting that the implemented monitoring software solution, comprising the monitoring dashboard, analysis notebooks and fault-prediction machine learning model pipelines described in [1] and [2], is applicable to both use cases.

The network topology for the paint shop drive chain is illustrated in Figure 1. Sensors are sending data only after a certain threshold has been breached. This monitoring solution saves battery, but offers only reactive monitoring, i.e., the maintenance team knows about an issue only when it is already happening or moments before that. According to the provided historical measurements, the sensors send data periodically (every 6s) in a heartbeat manner. The information is sent to a gateway using a proprietary RF protocol. The gateway can be either Ethernet-based or cellular-based. Each gateway can support up to 100 sensors within a range of 350 meters, non-line-of-sight. The gateway collects the sensor data from the wireless

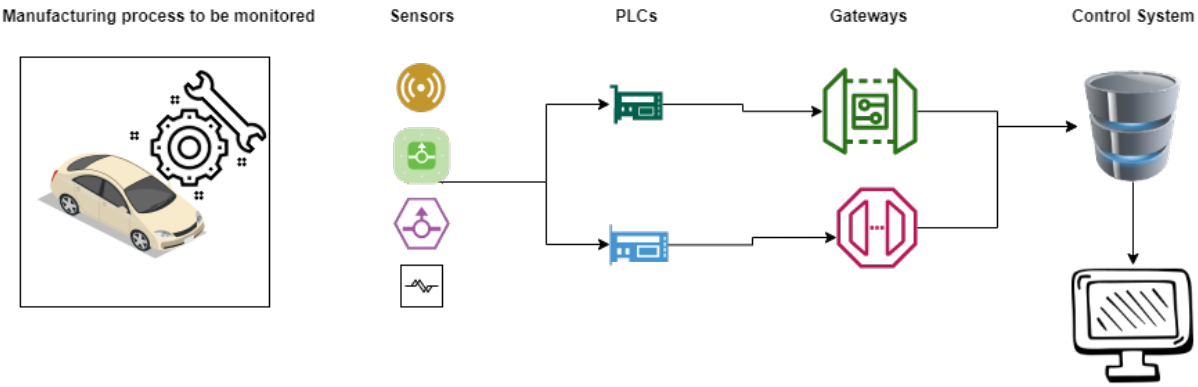

Figure 1: Network setup.

sensor 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.

2.2.1 Sensor installation

Two different solutions have been proposed so far for the installation of new sensor types:

- Inclination sensors that would work in parallel with current sensors and provide measurements with higher granularity;
- Vibration sensors that would be used to monitor failures, in motors pulling the drive chains, associated with axis tilt due to motor bearings.

Inclination, or tilt, measurement can be performed by inclinometers or accelerometers which are used for the same measurement, but are optimized for different purposes and environments[1]. We suggest a low-cost IoT Arduino with accelerometer for tilt measurement, e.g., esp8266 low-cost Wi-Fi microchip with mpu6050 six-axis motion tracking device (accelerometer and gyroscope). Battery consumption testing could be performed on a simple esp8266 chipset testing solution which is currently available, i.e., refer to Figure 2. Vibration sensors would be used to measure possible faults happening inside the motors. The possible implementation of these sensors must be clarified with SEAT maintenance team.

One of the key issues for the installation of new sensors refers to their energy consumption and battery life. Depending on the used (i) microchip, (ii) antenna and (iii) frequency of the measurements, the battery life can vary significantly. Four ways to increase the battery life are (i) put sensor in sleep mode, (ii) decrease the frequency of measurements, (iii) decrease the frequency of communication and (iv) increase the battery capacity (which increases also the size). Depending on the sensor hardware and expected measurement frequency, we can make preliminary battery life assumptions using a simple battery life calculator[2].

[1] https://www.murata.com/en-us/support/faqs/products/sensor/inclinometer/char/0003
[2] https://www.geekstips.com/battery-life-calculator-sleep-mode/

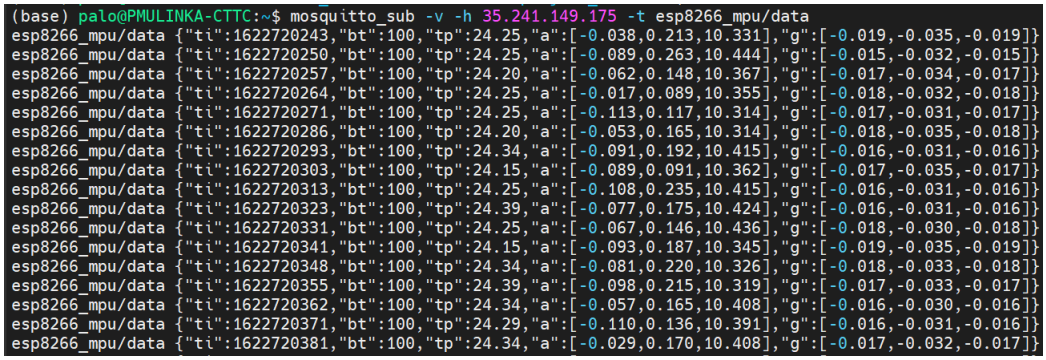

Figure 2: Measurements screenshot example from esp8266 chipset.

2.2.2 Monitoring solution

In networked industrial systems, data visualization tools offer valuable insights for the real-time performance of monitoring processes. A modular and containerized end-to-end solution has been implemented for the purpose of evaluation, implementation and integration of data processing and ML methods into an Industrial Internet-of-Things (IIoT) ecosystem [1].

Our scalable testbed platform is implemented in Docker[3] containers and it is publicly available on GitHub [2]. Due to repository space restrictions, we have included only a portable UCI Spambase dataset [3] in the testbed. However, incorporation of large-scale industrial datasets including normal and anomalous system behavior is also possible. Figure 3 illustrates the monitoring solution design and the interactions between the incorporated modules. In particular, the key building blocks of our end-to-end solution include:

- **Generator**: A Python script sending rows from the UCI Spambase dataset to a collector using the HTTP POST method. This script mimics the communication between sensors and the fusion center.
- **Collector**: A RESTful application program interface (API) implemented in Python Flask web framework used to authenticate connecting clients and validate received data.
- **Storage**: A module including databases and applications for data storage (database and queue) and visualization.
- **Models**: A module representing a trained machine learning (ML) model currently used for classification of events.
- **Model builders**: A framework for tracking of the ML model experiments.

To support communication, we adopt (i) HTTP POST method with enabled authentication for connectivity between the sensors and the monitoring solution; and (ii) Kafka[4] streams, with a role of distributed message queue and storage of intermediate results, for communication between applications. The **Storage** block constitutes the epicenter of the proposed monitoring solution and further consists of the following components:

[3]Docker [Online]: https://www.docker.com.
[4]Kafka [Online]: https://kafka.apache.org.

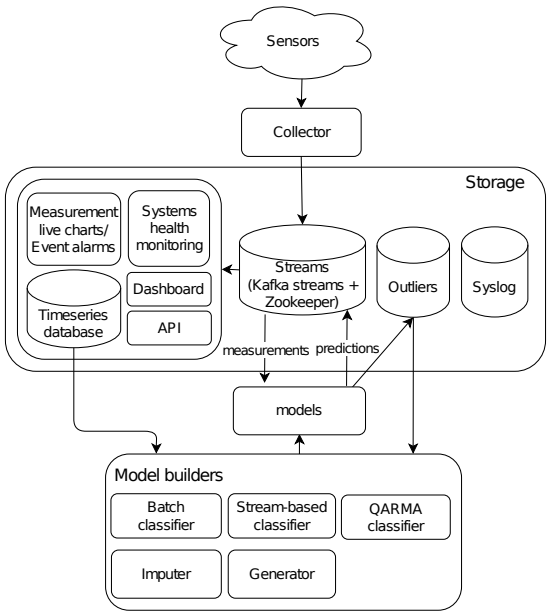

Figure 3: Building blocks of our modular monitoring solution where the involved interactions (data flows) between individual components are displayed.

- **Syslog**: A container or localhost for aggregation and storage of all log files.
- **Outliers**: A database to store detected outliers. As the outliers accumulate over time, we periodically check if they can form a new class or cluster in the measurements and use them in the training of the ML models.
- **Streams**: This block consists of Kafka brokers and Zookeeper to maintain and coordinate the distributed Kafka streams reliably.
- **Timeseries** and corresponding components: In our case, we use the InfluxDB 2.x time series database[5] and corresponding components, i.e., Telegraf, dashboard and alerts, to monitor sensors and Docker containers health. Telegraf is a plugin-driven server agent for collecting and reporting KPIs, and it can be jointly used with Kafka consumer client to forward the data from Kafka streams to time series database and InfluxDB dashboard (shown in Figure 4) with alerts for measurement live charts and notifications (shown in Figures 5 and 6).

For data prediction tasks, we incorporate a set of models built for the ML pipeline, as illustrated in Figure 7. Individual data analysis components are publicly available in the form of Jupyter notebooks on GitHub [4]. In particular, we have so far incorporated diverse approaches for event detection in data streams which can be further used for IIoT anomaly detection scenarios, i.e.,

- **Batch classifier**: Traditional ML models trained on fixed size datasets.
- **Stream-based classifier**: Stream-based ML models continuously adapting to data streams.

[5]InfluxDB [Online]: https://www.influxdata.com.

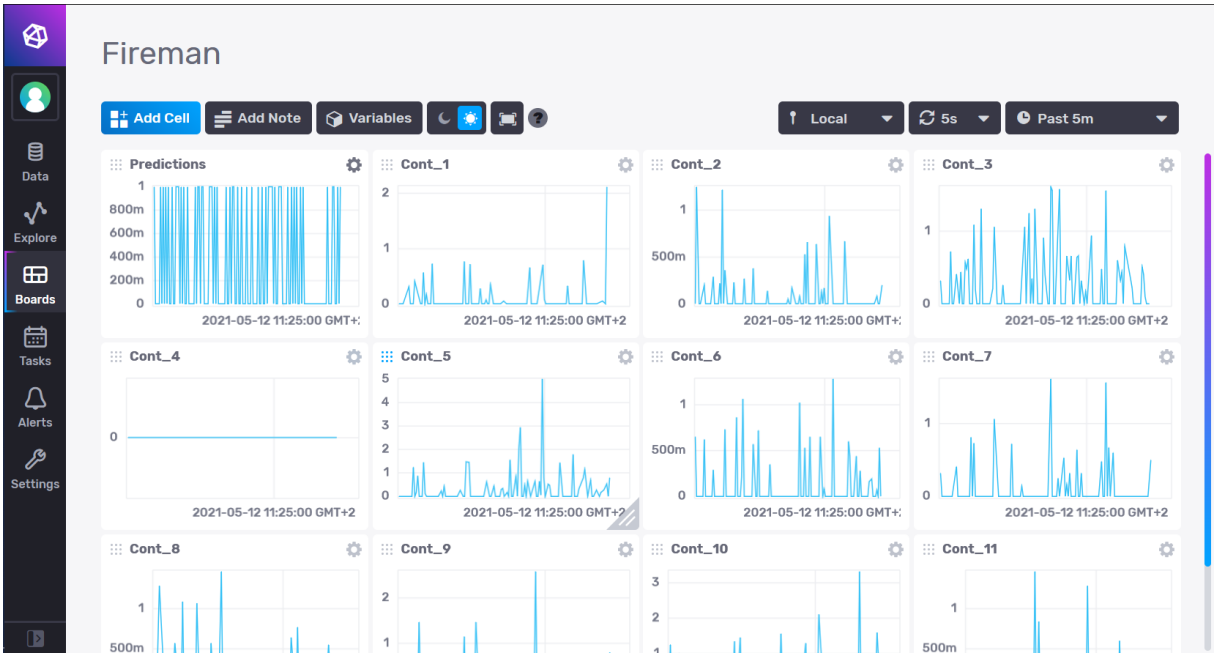

Figure 4: Dashboard with live charts in InfluxDB, where binary classification predictions are illustrated with live charts for dataset features.

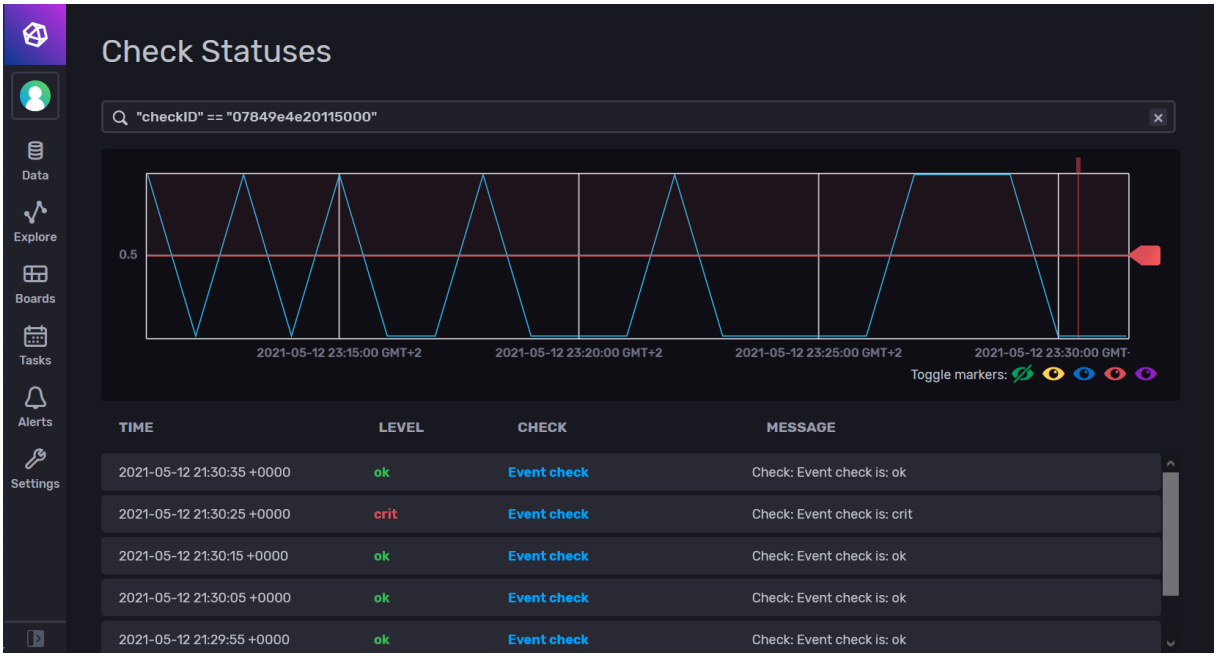

Figure 5: Alert definition. InfluxDB allows the definition of threshold values for alerts and notification endpoints. We can easily define values of interest in predictions, e.g., 0.5 in case of binary classification for no event vs event occurrence.

- **QARMA** [5]: Classifier based on quantitative association rules mining.
- **Imputer**: Methods and models for the imputation of missing measurements.

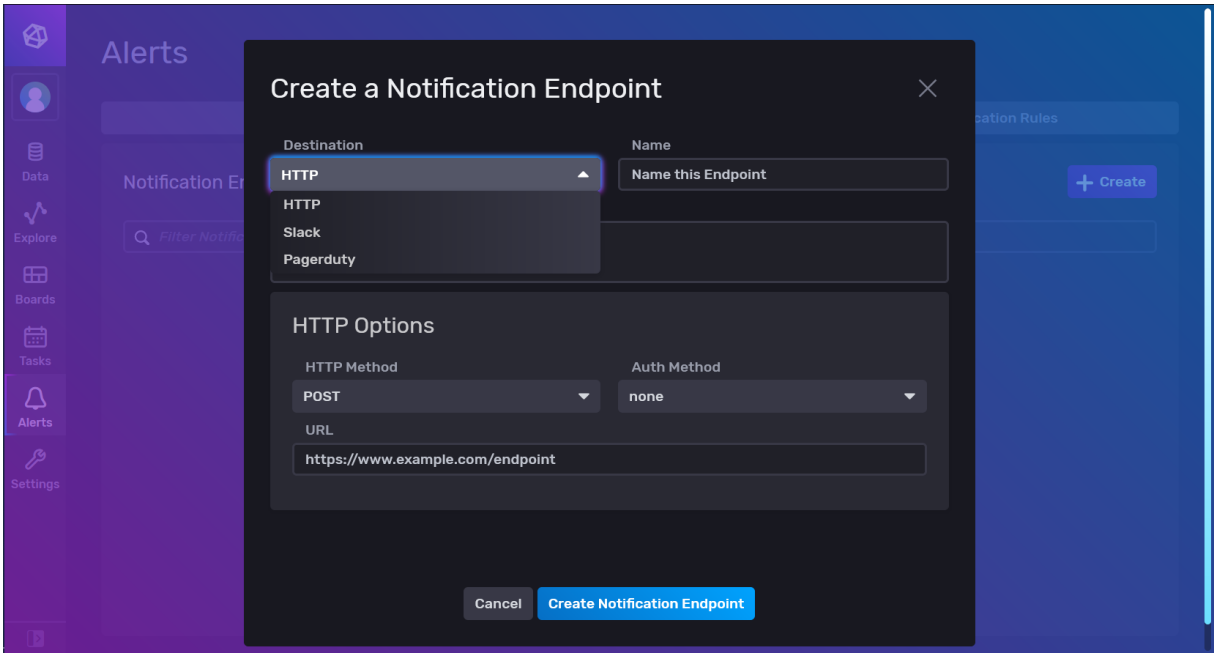

Figure 6: Notification endpoint.

- **Generator**: Methods and models for the generation of new data to deal with imbalanced datasets.

In addition, the core ML pipeline used for the predictions of events consists of the following blocks:

- **Normalization/standardization**: Scaling of the dataset.
- **Imputer**: Imputation of the missing data.
- **Feature engineering**: Analysis, selection of features and reduction of feature space.
- **Outlier detection**: Specific methods and models for outlier detection.
- **Dataset balancing**: Oversampling/undersampling synthetic data generation methods for imbalanced dataset.
- **Classification, evaluation, tuning**: Evaluation and tuning of classification algorithm for best results.

In our end-to-end platform, we have selected the combination of Python libraries, RESTful API, Kafka streams and InfluxDB time series database to keep the solution open-source, modular, scalable and robust. All discussed components constitute pieces of publicly available open-source projects and can be easily interchangeable or removable.

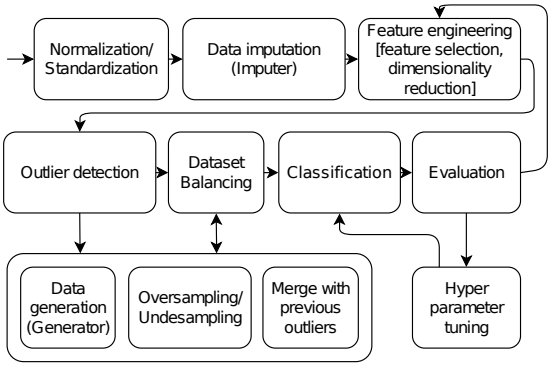

Figure 7: ML pipeline used for data processing tasks, e.g., predictions of events.

3 5G Test Network (5GTN)

The 5GTN is a wide IoT sensor infrastructure, which is coordinated by the smart campus program at the university of Oulu (UOulu) in Finland [6]. The smart campus program aims to provide smart services that support the collaboration between academic and industrial scenarios. In addition, it is progressing towards the 6G test network by 2030. This project mainly targets active research areas, such as wireless connectivity, distributed computing, and machine type communication (MTC).

3.1 Network Description

The 5GTN consists of 452 LoRa nodes distributed over the campus in Linnanmaa area as shown in Fig. 8. The network started to work on June 2020. In addition, a single gateway covers all the nodes. Each device has a device extended unique identifier (DevEUI), which is a 64-bit globally-unique Extended Unique Identifier (EUI-64) assigned by the manufacturer, or the owner, of the end-device According to [7]. The location of each device, their global positioning system (GPS) coordinates, the status of the devices, their time of installation, and their types are documented and saved at the subjected server. The nodes are classified according to the attached sensors into three types as follows: *(i) CO2 devices*, which are mainly responsible of measuring the CO2 levels. In addition, they also senses light and motion. *(ii) Sound devices*, which are mainly measuring average and peak sounds. They are also able to sense light and motion. *(iii) Moisture devices*, which are located under the soil to measure soil moisture and pressure. Furthermore, all devices are measuring temperature, humidity and their battery consumption.

3.2 Available Dataset

In order to receive a sensed message, it passes by two phases as follows: *(i) It passes from the LoRa node to the gateway*, then *(ii) It passes from the gateway to the server*. Normally, each node transmits their data every 15 minutes. There are two available datasets, where the first dataset describe the physical parameter sensed by the sensors, then transmitted to the gateway and the server, respectively, whereas the second dataset provides some information about the physical communication layer of the nodes themselves. The first dataset, which

Figure 8: Map of the distribution of the devices in the campus of the university of Oulu.

describes the second phase, where the successful received message at the gateway is being forwarded from the gateway to the server, consists of 12 columns, where the DevEUI number of each node is presented in the first column, and the time of transmission is presented in the second column. Furthermore, the following physical parameters are presented in the rest of the columns according to the type of the device: battery, temperature, humidity, CO2, light, motion, average sound, peak sound, pressure and moisture. To have raw data at the server, we should refer that the received payload at the server should be decoded to have the actual physical parameters. Fig. 9 presents an example of a CO2 device measures CO2 levels in a room.

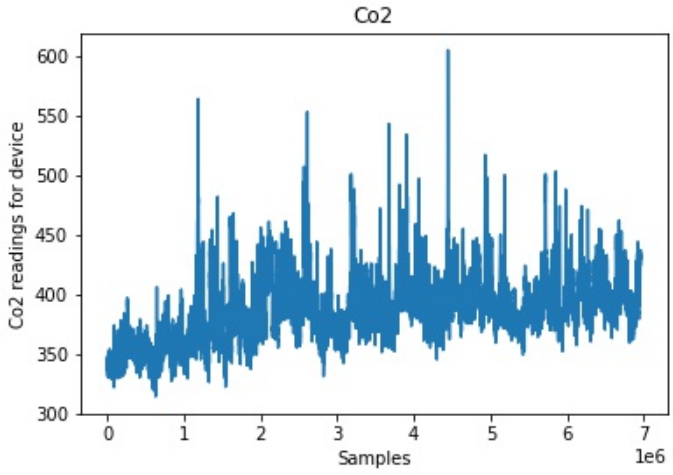

Figure 9: An example of CO2 levels for a CO2 device.

The second dataset describes the communication physical layer parameters of the LoRa nodes. As the first dataset, it has the DevEUI identifier of the devices and the time of transmission. In addition, it describes the used spreading factor (SF) , channel (7 available channels) and the signal-to-noise ratio (SNR) in transmission. It also presents the port number (12 available ports) used in transmission. The received signal strength indication (RSSI), which is the received signal power in milliwatts and is measured in dBm, is also available in the dataset. In addition, the dataset presents the frame counters (FCNT), the sequence number, and the RF chain (RFCH) used by the receiver. This dataset also provides some other useful information, such as the data rate, coding, modulation used in the transmission and the payload message itself. Fig. 10 presents the RSSI parameter used by a certain node while transmitting its packets to the gateway.

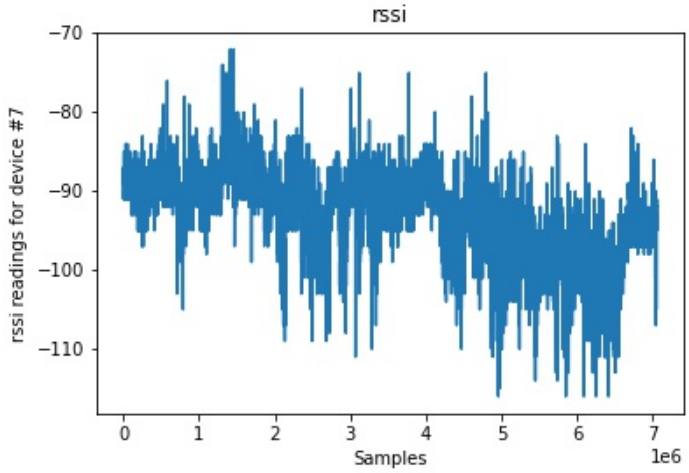

Figure 10: A visualization for RSSI parameter for a device.

The available datasets are collected from August 2020 until February 2021. Both of them provide around 7 millions data rows. The second dataset has around 90000 more data than the first one, due to the lost messages from the LoRa nodes to the gateway as a result to different factors, such as bad channel, interference, and other hardware errors. As a result, during the whole period of the collected dataset, some devices have more data than others.

3.3 Acessibility

According to [6], the 5GTN dataset is available for partners of the university of Oulu. However, it is not available at the moment, but it is planned to be open-source for Oulu partners soon. Although, the collected dataset is within the period August 2020 - February 2021, the network itself is working everyday, and it is possible to access the network for new data to test and validate the developed work on the old data.

3.4 Demonstration Plan

The first experimentation with 5GTN is expected to be taken in the second semester of 2021. The first idea is on using the gathered dataset to identify certain events and apply

FIREMAN algorithms on it. We plan on doing it using sensor data (temperature, CO^2 levels, sound), using the correlation between them, together with wireless medium characteristics, as packets SINR and rssi, to reduce or create redundancy to increase the information reliability. For the final application, we aim on getting the results from this approach and apply it on the 5GTN, doing modifications on devices to serve our expectations.

4 LUT test case

LUT study case will be based on the access of a microgrid infrastructure available through the European-funded program ERIGrid 2.0.[6] The host institution is the *Institute of Communication and Computer Systems* (ICCS) is linked to the *National Technical University of Athens* (NTUA). The aim is to build three representative experiments to generate datasets using the expertise of the host team in order to employ the FIREMAN approach to detect rare events that require action of the adaptive protection scheme. The following next step is to deploy a similar infrastructure at LUT (which is currently acquiring a real-time simulator RTDS).

4.1 Adaptive protection for microgrids based on communication

4.1.1 Objectives

The aim of this study is to contribute on decentralization of energy technologies by investigating technical aspects such as performance of new adaptive protection schemes in microgrids employing communication systems. Specifically, our approach is to apply the framework developed in the previous WPs to predict potential faults (rare operational events) and act accordingly using the three-layer framework and the specific data processes involved (acquisition, transmission, fusion and analytics). The contribution will rely on the performance of communication systems (e.g. ethernet, 5G, IoT etc.) in electrical protection for microgrid and their capabilities to address control and protection tasks of electrical systems using ultra-reliable links with bounded latency.

The research will focus on first simulations and then tests in a real-time environment a microgrid setup to run different topologies in which overcurrent protection will adapt according to the current state of the microgrid (open switches, high DER generation, low DER generation, etc.). To do this, the switches and protective devices must communicate to each other. Therefore, formulation of communication between the devices is seen as one of the possible channels to increase the accuracy and flexibility. The research aims to study three (3) different adaptive protection cases to obtain data that could be also used as benchmark for further research not only related to predictive actions associated with the adaptive protection mechanism to avoid faults but also as a detection or classification of faults that do occur in microgrids with and without the adaptive protection. Besides, as the outcome of the research we will be able to determine a suitable communication system to perform adaptive overcurrent protection in microgrids given the actual constraints of a real deployment. This is achieved by measuring the related communication key performance indicators (KPIs) identified in WP2.

Through the research carried out through this testbed experiments, we will be able to answer questions such as: Is it possible to perform adaptive protection using decentralized communication? What are the variable boundaries for reliable communication in adaptive protection using decentralized communication? How fast is the system able to adapt once the microgrid scenario changes? Is it possible to detect faults when there is low current contribution from DER? All these demonstrate the generality of the proposed FIREMAN approach.

As a specific impact, such questions are significant because of new emerging technologies in both power system and communications engineering domains. The tests performed in standard

[6] http://erigrid2.eu/

microgrid topologies will provide a baseline for comparison to evaluate the project outcomes in an easy way by other researchers in the field of protection and communications. The research work aims to solve these questions through the completion of the following activities.

4.1.2 Planned activities for WP6 tests

- **Activity 1 (A1) – 8.2021** – Model definition using FIREMAN 3-layer model and its implementation on python of LV CIGRE benchmark (see Fig. 11) which will be implemented prior to the visit to NTUA (which shall happen in November 2021). The definition should contain the voltage, number of load and therefore number of relays protecting each segment. Also, DER modelling is based on commonly used models like controlled voltage/current sources.

- **Activity 2 (A2) – 9.-10.2021** – Overcurrent protective settings, these will be studied based on the possible scenarios that might happen in the microgrid and how they affect the setting in each of the relays. The number of scenarios is yet to be defined but at least two scenarios should be implemented to show how the setting adapt accordingly to microgrid state. The different scenarios will be chosen prior to visit.

- **Activity 3 (A3) – 11.-12.2021** – Model implementation, carried out in RTDS software environment, containing all the elements of the microgrid. Also, hardware relays will be connected to the DRTS.

- **Activity 4 (A4) – 11.-12.2021** – Measurement's collection; these results will be gathered from the communication signals as well from fault tests performed to prove the validity of the method under a real case scenario with fault conditions where the response of the protective system should be adequate for the microgrid status.

- **Activity 5 (A5) – Year 2022** -– Replicate the deployment at LUT.

- **Activity 6 (A6) – Year 2022** — Documentation; dissemination of results and uploading obtained dataset in FIREMAN Github.

4.1.3 Use cases

The proposed use cases will be carried out on the topology LVCIGRE benchmark (Fig. 11) using communication setup in Fig. 12 that will be constructed as a cyber-physical energy system following the 3-layer model [8] proposed in WP2 and the techniques developed in WP3-5.

1. Data involving the elements that are communicating with each other are gathered. We will measure variables such as latency and reliability of the communication link. This test is performed with any communication network attached to the relays to perform adaptive protection tasks. Then, electrical variables data is obtained, and we will evaluate the scenario of the microgrid and based on a decentralized system based on cooperation that will send updates in the settings of the overcurrent relays so they will adapt to this new condition. Finally, a test performed under this category consist of given a specific

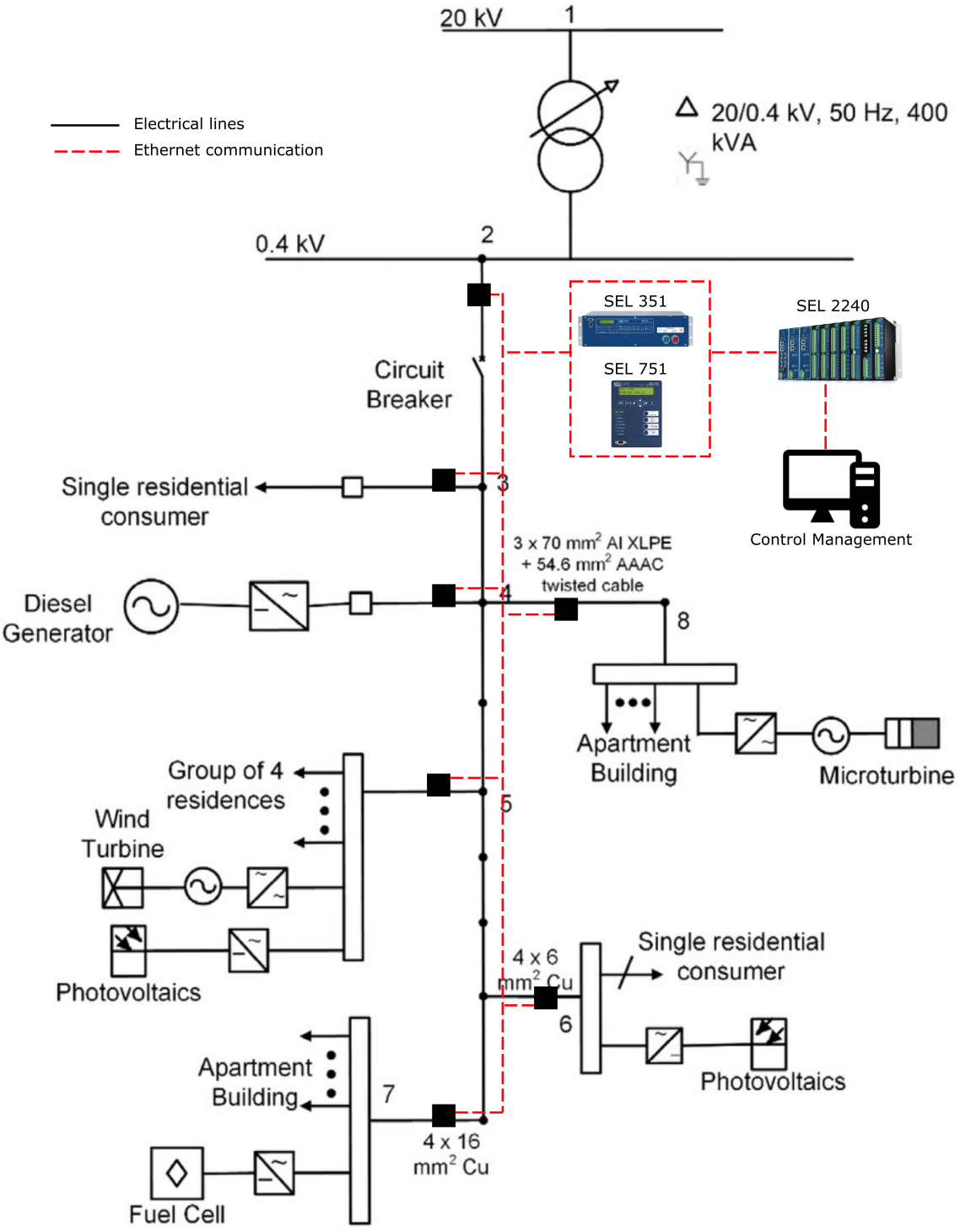

Figure 11: CIGRE topology

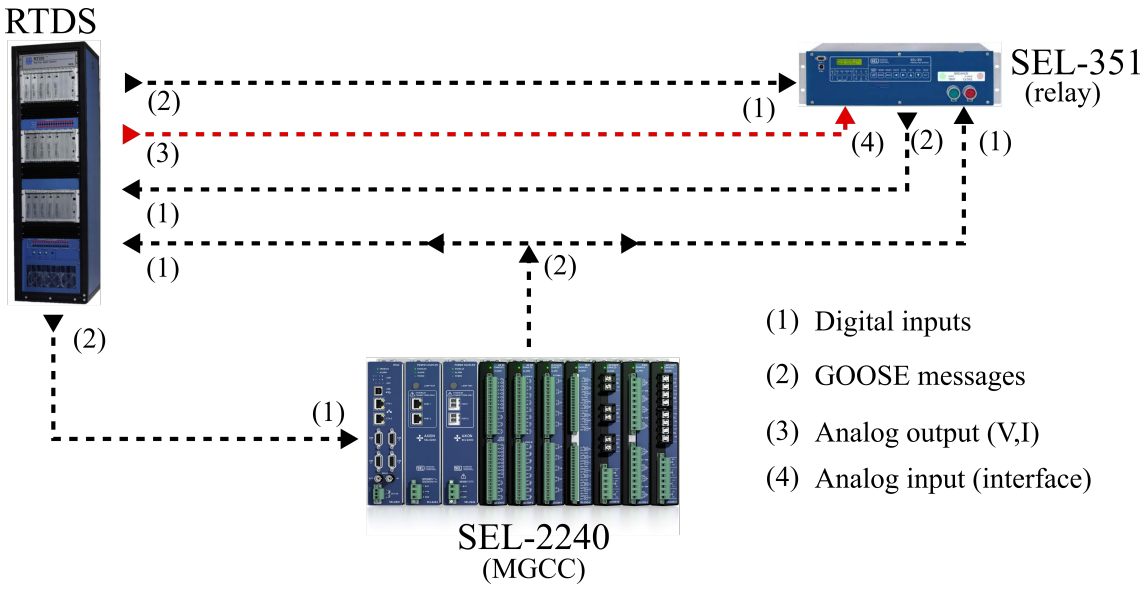

Figure 12: Communication setup for simmulation-hardware environment

microgrid scenario, an electric fault will be introduced in one of the lines to check that the operation times and fault clearing from the relay are according to those used in the standards to effectively protect the microgrid elements. This use case will allow us to test adaptive protection setup (two or more set ups) for communication systems.

2. Consist of using the model from the first use case insert faults in the system changing parameters as location of the fault and fault resistance to create a benchmark that allows different algorithms to detect faults even in the presence of high resistance faults or topologies that due to the low inertia coming from the DER sources, modern relays are not able to detect. In this use case, the data collected can be used to test different algorithms for fault location and classification. These algorithms are not part of this project, but existing ones are to be used and tested.

3. Same model from the two use cases above will be used to rapidly change the scenarios of the grid and therefore trigger communications signals between the devices to adapt to new conditions. Once these changes are introduced in a fast way and periodically, we will be able to evaluate how communication systems perform and what are the consequences to the protection system in the microgrid. The difference with use case 1, is that in this experiment there will be a sequence of microgrid state changes instead of only one. This use case proposes a benchmark scenario to collect data that can be evaluated in data imputation methods to be used together with protection technologies.

5 Conclusions and future directions

This document presented an introductory description of FIREMAN experimental use cases. It servers as a base for Task 6.2 Final Project Demos, and for Deliverable 6.2 First Demonstration Trials, which will be released in January 2022. Thus, we plan on using the discussion here presented to serve as a baseline on the work to be done during the second semester of 2021.

References

[1] P. Mulinka, C. Kalalas, M. Dzaferagic, I. Macaluso, D. Gutierrez-Rojas, P. 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*, 2021, pp. 1–6.

[2] *FIREMAN monitoring solution repository*, [Online]. Available at: https://github.com/5uperpalo/FIREMAN-project_frontend.

[3] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[4] *FIREMAN analysis scripts and documentation*, [Online]. Available at: https://github.com/5uperpalo/FIREMAN-project.

[5] I. T. Christou, E. Amolochitis, and Z.-H. Tan, "A parallel/distributed algorithmic framework for mining all quantitative association rules," *arXiv preprint arXiv:1804.06764*, 2018.

[6] "5GTN website," https://smartcampus.fi/oulu/, 2020.

[7] https://lora-developers.semtech.com.

[8] D. Gutierrez-Rojas, M. Ullah, I. T. Christou, G. Almeida, P. Nardelli, D. Carrillo, J. M. Sant'Ana, H. Alves, M. Dzaferagic, A. Chiumento, and C. Kalalas, "Three-layer Approach to Detect Anomalies in Industrial Environments based on Machine Learning," in *2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS)*, vol. 1, 2020, pp. 250–256.