chist-era

FIREMAN
(**F**ramework for the **I**dentification of **R**are **E**vents via **MA**chine learning and IoT **N**etworks)

Mid-term report (5.2019 – 10.2020)

1. Identification

Project Id	
CHIST-ERA Call topic	Big data and process modelling for smart industry (BDSI)
Project acronym	FIREMAN
Project title	**F**ramework for the **I**dentification of **R**are **E**vents via **MA**chine learning and IoT **N**etworks
Start date	May 2019
End date	April 2022
Website	https://fireman-project.eu/

Coordinator of the project	
Name	Pedro Henrique Juliano Nardelli
Organisation	LUT University
Country	Finland
E-mail	Pedro.Nardelli@lut.fi

2. Summary

The Internet of Things (IoT) is creating a new structure of awareness – a cybernetic one – upon physical processes. Industries of different kinds are expected to join this revolution, leading to the so-called Factories of the Future or Industry 4.0. Our considered IoT-based industrial cyber-physical system (CPS) works in three generic steps:

1. **Large data acquisition / dissemination:** A physical process is monitored by sensors that pre-process the (assumed large) collected data and send the pre-processed information to an intelligent node (e.g. aggregator, central controller).
2. **Big data fusion:** The intelligent node uses machine learning techniques (e.g. data clustering, pattern recognition, neural networks) to convert the received ("big") data to useful information to guide short-term operational decisions related to the physical process.
3. **Big data analytics:** The physical process together with the acquisition and fusion steps can be virtualized, building then a cyber-physical process, whose dynamic performance can be analysed and optimized through visualization (if human intervention is available) or artificial intelligence (if the decisions are automatic) or a combination thereof.

The **FIREMAN** focus is on how to optimize the prediction, detection and respective interventions of rare events in industrial processes based on these three steps. The framework is application-independent; however, our demonstrated solution will be designed case-by-case. We will consider the CPS working as a complex system so that these three steps, which operate with relative autonomy, are strongly interrelated. For example, the way the sensors measure the signal related to the physical process will affect what is the best data fusion algorithm, which in turn will generate a certain awareness of the physical process that will form the basis of the proposed data analytics procedure. This analysis of the three-layer model was motivated and initially tested in two smart grid cases.

In specific terms, we have proposed a general framework based on six general guiding questions that define the boundary conditions to provide ultra-reliable detection / prevention of rare events related to pre-determined industrial physical processes. We have identified several potential applications from SEAT at different industrial plants in Spain, and from different LUT University laboratories, as well as University of Oulu 5G Test Network (5GTN) that provides wireless connectivity to its smart campus. Additionally, we have extensively worked with well-known (simulation) benchmark scenario for anomaly detection in industrial processes, called Tennessee Eastman Process (TEP), for testing purposes. For selected applications, we have provided a detailed description of anomalies and related Key Performance Indicators (KPIs); additionally, we have listed the communication system KPIs. We have also applied the proposed framework based on one plus six questions in three applications selected by LUT (TEP, a low-voltage DC microgrid and a welding machine).

For testing purposes, we started our research focusing on TEP, which consists of five major parts: a condenser, a vapor-liquid separator, a reactor, a product stripper, and a recycle compressor. It simulates a normal production process with twenty-one different faulty processes caused by some arbitrary/known disturbances. There are 52 monitoring variables, 41 measured and 11 manipulated. Once a faulty condition is activated, the dynamics of the process are affected, and some values may change. For FIREMAN, we consider that these 52 variables are signals during the acquisition phase. In this phase, those signals can be sampled in different manners: the most usual periodic sampling (i.e., sample at specified periods of time) or event-driven, non-periodic, sampling (sample always that a predetermined event happens, e.g., sample the signal whenever it is above or below a given threshold, or if the variation between two samples is above a given threshold). Our results showed that this approach can significantly reduce the number of samples in the acquisition phase, which also impact the number of transmissions required. Besides, we have shown that the variation approach improves the detection of events that are usually hard to detect using established statistical methods. Different protocols for the event-driven approach were extensively

studied in more details in an electricity metering dataset with much lower granularity, showing excellent results in terms of data compression and reduction of traffic offered to the communication network. This has resulted in one published master thesis (LUT), and one doctoral thesis that is approved for public defence (University of Oulu).

The data transmission and dissemination from sensors to the storage/processing unit(s) relies on an IoT network deployment. Initial results showed the potential of promising existing technologies based on Low-Power Wide-Area Networks (LPWAN) and emerging techniques for beyond-5G systems such as Non-Orthogonal Multiple Access (NOMA), massive multiple-input multiple-output (MIMO), Intelligent Reflective Surfaces (IRS), and grant-free random access. Multi-class ALOHA and Generalized Frequency Division Multiplexing (GFDM) were also evaluated. We have also analysed in a testbed the performance of a private industrial LTE network for one specific control application, leading to poor results due to high values of latency (from 20ms in the perfect scenario to more than 200ms with some network congestion). A new deployment is being implemented at LUT to assess the performance of LTE/5G connectivity in the operation of a low-voltage DC micro-grid that will be the basis for one of our demonstration activities, incorporating new ideas reported in recently published/submitted papers. In addition, we have also published a study of twenty-one factors to support the selection of the most suitable IoT platform based on the applications' needs. In addition, new results on data imputation were obtained to mitigate negative data erasures during the sensors' transmission considering an established big data set from phasor measurement units in power grids.

For data fusion, we have started experimenting with TEP. We have employed some classical methods for dimensionality reduction, such as the Principal Component Analysis (PCA) and mutual information. We obtained preliminary results in a wireless sensor network scenario considering a Graph Signal Processing technique to reduce the dataset dimensions. For the data analytics, we have proved based on another dataset the high performance of the Quantitative Association Rule Mining Algorithm (QARMA) for detecting rare events in industrial environments. For FIREMAN, QARMA is being tested to create association rules for TEP, trying to assess its performance against the existing solutions. The qualitative association rules could also be used in the data fusion since QARMA "tells" which variables are important in detecting anomalies. Additional work on QARMA was carried out to classify faults in power grids. An important feature of QARMA when compared with other (deep) learning methods is its explainability due to explicit associative relations between features, which is useful not only for detecting failures but also for clarifying their causes, helping then to mitigate possible future occurrences. .

The ongoing research based on TEP aims at incorporating all aspects together to assess the performance of the anomaly detection in that benchmark and compare the results with the existing literature in the field. Our work will employ a communication network layer following a model proposed in the literature (including aspects from physical layer and network layer, including advantages and drawbacks of cloud/edge computing in performing the computational tasks needed). We will also start tests in real environments at SEAT, as well as at LUT and University of Oulu, based on the insights obtained during the extensive study of TEP.

Our activities have already proved their impact within the academic community: we have organized two special sessions in top conferences, published several papers in top journals and conferences, and many other papers are currently under review. More than this, we are expecting that our results generate impact beyond academia. For example, predictive maintenance may lead to economic benefits to industries due to improved efficiency in the plant with less downtimes. Besides, FIREMAN may help to reduce accidents in industrial environments. All in all, FIREMAN activities follow the concept of Industry 4.0, which is expected to be a source of a new wave of qualified jobs in Europe. Besides, FIREMAN is also collaborating with another CHIST-ERA consortium called SOON from the same BDSI call to build joint activities to increase the impact of both complementary projects.

More details of FIREMAN (including deliverables and research papers) can be found in our website: https://fireman-project.eu/. Contact information: Pedro.Nardelli@lut.fi (project coordinator).

3. Progress Report

3.1. Project objectives and activities implemented

During the first year and a half of activities, i.e., from May 2019 (M1) to October 2020 (M18), all five technical work packages (WPs) of FIREMAN have been active.

An overview of the progress towards the achievement of the project objectives is provided in the following table.

Project objectives	Progress towards achievement
Model the industrial physical process to retrieve the signal characteristics, e.g., time- vs. event-triggered sampling, from the physical data captured/recorded using sensors.	A framework was defined (D2.1, D2.2) to define possibilities for different data acquisition methods. We showed the event-based acquisition provides advantages in the cases studied by FIREMAN (D2.3, D3.1 and D3.2). Current work in T2.3.
Design a flexible network topology for the scalable deployment of many sensors to securely collect data at a low cost with a proper utilization of the scarce shared resources.	Different cutting-edge transmission techniques for were studied and many papers published focusing on a dynamic network topology for massive machine-type communications (current work in T3.3 and T4.1).
Optimize the communication protocols to efficiently handle mission-critical transmission with stringent QoS requirements (ultra-high reliability, very low-latency) in industrial setups.	Together with the massive connectivity, we have also proposed different new protocols to be employed in LoRa networks to reach the QoS required to industrial setups (current work in T3.3 and T4.1).
Perform preliminary data analysis and processing to transform multi-stream raw data generated by various sources (heterogeneous and high-dimensional) into usable formats.	We have initially focused on the benchmark TEP project (its normal and extended versions), as well as data from power systems. We are currently defining the practical setups at SEAT, LUT and Oulu (current work in T4.2, T5.1, T6.1)
Extract useful information from the monitoring data and guide short-term operational decisions related to the physical process monitored by the sensors.	Extensive study of TEP benchmark to improve the detection “real-time faults” in WP2 and 3. The proposed delta-based technique outperforms the benchmark scenario with much less data transmission. This approach is being extended to the test cases in T6.1.
Achieve proactive (predictive) maintenance via data-driven learning and mining techniques that identify and unlock the potential value of data.	The proposed delta-based technique “filters” data in the acquisition phase, enabling fast fault detection. Currently we are studying the impact of this filtered data to classification in T4.2, T5.1
Offer system situational awareness on the occurrence of rare events for timely detection, effective decision-making and possible issuing of actuator commands.	This is the focus of WP5, including how to visualize rare events to enable effective decision-making. Activity just started.
Develop experimental test-beds to enable the validation and evaluation of the proposed solutions and verification of the overall system performance.	Testbeds are defined and we are organizing with the relevant persons how to deploy the proposed solution. We are following the proposed procedure presented in D2.1 and D2.2. WP6 just started, where we will focus on the testbeds.

In what follows, we provide a summary of the research activities implemented at WP level during this period.

WP2 aims to define the key requirements of the use cases to be studied, build a general framework for providing an ultra-reliable rare-event detection and understanding the fundamental limits of the proposed solutions. The general idea is presented in [C-3].

During Task 2.1 (led by **SEAT**), the focused study cases were defined, and key performance indicators (KPIs) were proposed. **SEAT** proposed a table to be filled with the support of experts in order to precisely identify the application KPIs. **SEAT** has applied the approach in different study cases: drive chain failure in paintshop processes and spindle (kind of actuator) failures in Computer Numerical Control (CNC) grinding machines in the machining centre. **LUT** applied in three cases: micro-grid, Tennessee Eastman Process (TEP) and a welding machine. **CTTC** and **UOULU** proposed a generic set of KPIs related to the wireless communication in industrial setups. As a side note, as previously discussed, the study cases from Nokia and the wind turbine needed to be dropped. **SEAT** provided the technical coordination of this task. This task has been now finalised and the outcome is available in D2.1. In Task 2.2, **LUT** (task leader) further developed the 3-layer model of cyber-physical systems to focus on rare-event detection. The proposed framework is general but can be specified for all use cases studied in the context of FIREMAN (and beyond them). The framework is based on six questions, going from sensors to analytics, covering aspects related to the design of communication and choices in computing architecture. This approach was tested in the three **LUT** cases, supported by **UOULU** and **AIT**. **SEAT** studied their use cases following this approach. This task has been now finalised and the outcome is available in D2.2 and [S-2]. Task 2.3 has so far focused on analysing the performance limits of the Quantitative Association Rule Mining Algorithm (QARMA) in two controlled scenarios: (i) fault classification in power grids; and (ii) TEP anomaly detection and diagnosis. This task involved **AIT** (task leader) and **LUT**. **UOULU's** work involves the effect of communication networks on QARMA, which is currently being finalized. The approach based on theoretical bounds turned out not relevant and we decided to focus on QARMA. Ongoing results are documented in [C-4], [C-14] and [B-1].

Based on the definition of requirements and system architecture in **WP2**, the overall aim of **WP3** is to design a data acquisition framework that encompasses all stages from local data collection and storage to data transmission in large-scale industrial deployments. Since August 2019 (M4), the research activities in **WP3** are focusing on the fulfilment of specific objectives, i.e.,

- Develop accurate analytical/simulation models for the signal characterization of the different industrial processes.
- Develop an agile, cost-efficient, network deployment, enabling seamless and secure connectivity for a sheer scale of sensors and local data collection.
- Design transmit-diversity schemes and optimize protocol operation to guarantee ultra-high reliability and low latency for mission-critical sensor data and control commands.

In the context of **WP3**, the FIREMAN partners are contributing towards several objectives, e.g., data characterization (Task 3.1), local data collection (Task 3.2) and data transmission (Task 3.3). In Task 3.1, **LUT** has investigated different sampling strategies for data acquisition ([C-2]) while **CTTC** and **UOULU** have focused on the derivation of analytical models for traffic characterization based on point processes. In Task 3.2, **LUT** is considering the level of data compression required for periodic and event-driven sampling by considering the reconstruction error. **CTTC** has proposed a data compression strategy to effectively exploit the temporal and spatial correlation in the sensors' observations while **AIT** is investigating the trade-offs that arise when using standard relational database technology to manage the datasets required in FIREMAN versus using NoSQL databases. **AIT** is exploring candidate database platforms which are suitable for the storage system of the FIREMAN architecture. Finally, in Task 3.3, the involved partners have been investigating potential technical enablers to accommodate large-scale data transmission. In this context, **LUT,** together with **AIT,** have studied the applicability of massive multiple-input multiple-output (MIMO) and non-orthogonal multiple access (NOMA) networks in dense industrial setups ([J-1], [J-3], [C-7], [C-11], [C-12]). **CTTC, LUT and AIT** have jointly investigated the potential of the Intelligent Reflective Surfaces (IRS) paradigm for enhancing reliability in ultra-reliable industrial scenarios ([C-9],[J-16]). **UOULU** has also investigated the performance of LoRa technology enhanced with a novel message replication strategy in highly dense industrial communication scenarios ([J-4], [C-10]). **LUT** also provided results of the performance of Generalized Frequency Division Multiplexing (GFDM) and multi-class ALOHA systems ([C-6], [C-13]). Moreover, they started investigating the impact of Successive Interference Cancellation (SIC) in LoRa networks and uplink superposed replications in Low-Power Wide-Area Networks (LPWAN) ([[J-7], [J-11]]). **SEAT** is preparing an internal report about their own existing data acquisition, communication and processing infrastructure to be used by the partners. Finally, **UOULU** investigated a multi-antenna scheme for wireless energy transfer for massive machine-type communication (mMTC) scenarios.

A detailed discussion on the **scientific achievements** per task level for the first year and a half of FIREMAN is provided in Section 3.3 of the report. In addition, the **WP3** research outputs have been consolidated in numerous **scientific publications** which are listed in Section 4.1 of this report. Two FIREMAN **deliverables**, i.e., D3.1: "*Report on Physical Process Data Modelling*" and D3.2: "*Design and Functional Architecture for Data Acquisition*", have been published in July 2020 (M15) where the main outputs of Tasks 3.1 and 3.2 (which ended in M15) are documented. Without significant deviations, **WP3 has** so far been progressing well according to the initial project plan and towards its objectives defined in the proposal. The level of partner collaboration in the context of **WP3** is high as demonstrated already by the multiple joint publications. The bi-weekly telcos organized by **CTTC** (**WP3** leader) provide useful status updates while the monthly reports uploaded in Sharepoint offer a regular overview of the partners' activities.

The goal of **WP4** is to develop advanced machine learning techniques to perform data fusion, mining, and interpretation, which will enable efficient control and awareness of the physical processes. These efforts will allow us to convert the pre-processed raw sensor data to useful information and enable the analysis and short-term operational decisions based on preliminary analysis. Since November 2019 (M7) the **WP4** activities on the evaluation of different ways to aggregate and cluster the heterogeneous data have started. These activities are related to Task 4.1. The involved partners (**UOULU, LUT, CTTC, AIT, SEAT**) have made significant progress. **UOULU** has started a literature review on clustering and data aggregation and has progressed with the data aggregation and scheduling with QoS guarantees research., **UOULU** started working on interference meta-distributions and distributed radio resource management for mMTC scenarios with data aggregation. **CTTC** is exploring various strategies for intra-cluster discovery and cluster-head selection policies aiming to minimize the involved signalling overhead. A preliminary analysis of a cluster formation mechanism based on a flexible structure of the involved peer-discovery signatures is documented in D4.1. At the same time, **LUT** has been focusing on the communication aspects of system, with their work on Massive MIMO NOMA and grant-free random access, as well as GFDM and multi-class ALOHA. **UOULU** finished the initial modification on the Python LoRA Simulator and started comparing results from simulations and theoretical models. **AIT** provided further investigation of NoSQL databases, while LUT has described how to select an IoT platform based on 21 key factors. **SEAT** is currently collecting detailed information about their study cases to be further used by FIREMAN consortium. Task 4.2 focuses on data compression and interpretation. **TCD** is the leader of this task focusing on Generative Adversarial Networks (GAN) for dimensionality reduction and data imputation in a realistic environment. **LUT** is supporting this task by evaluating TEP scenario together with **AIT**, to compere the performance of GAN against QARMA. **UOULU** is evaluating the impact on communication impairments in this scenario. The above-mentioned activities of all involved partners (**TCD, LUT, AIT, UOULU**) fit nicely together allowing us to study the implications that missing measurements can have on the detection of rare events, and the techniques that could allow us to circumvent potential problems. The data imputation techniques would allow us to either impute missing measurements originating from sensing failures, failures in the communication or values missing due to data compression decisions (values not being shared with the aggregator to save energy and communication resources because those values can be imputed at the receiver). In an effort to link WP3-WP4 activities, **CTTC** has leveraged the potential of interpretable dynamical models to extract knowledge from the ambient sensor measurement space under unfavourable network conditions. In particular, **CTTC** is currently investigating the lingering influence of communication impairments to the emergence of missing data among measurement streams at the level of a fusion centre and the preliminary outcomes have resulted in two conference publications ([C-16] and [C-18]).

WP5: This WP has two ongoing tasks. In Task 5.1, **AIT** is the leader and is providing a literature review of different DL techniques with focus on rare event detection, also developing a parallel/distributed Deep Learning library within the context of the Open-Source library popt4jlib (https://github.com/ioannischristou/popt4jlib). **LUT** supported **AIT** to compare QARMA and Deep ANNs in fault classification. **UOULU** initially indicates how communication can affect QARMA performance in detect rare events. Task 5.2 is related to visualisation. **LUT** is leading this activity together with **AIT** is producing a GitHub page (https://github.com/5uperpalo/FIREMAN-project) to support the visualisation of rare events, initially focusing on TEP.

WP6: This WP just started. **UOULU** and **LUT** are defining their test cases and making the first assessment of the existing deployments to run the first benchmark experiments. **SEAT** is defining their own demonstration case.

3.2. Transnational collaboration

LUT and **AIT** have collaborated in a fault classification in power grids to compare the performance of QARMA and deep learning in a controlled task, related to Task 2.3. In relation to WP2 and WP3, **LUT**, **AIT** and **UOULU** are actively collaborating to study and assess the impact of event-driven approaches in different industrial settings based on the proposed three-layer model. In the context of WP3 and WP4 research activities, Dick Carrillo Melgarejo (**LUT**) performed a research visit at **CTTC** premises during November 2019 where he collaborated with Charalampos Kalalas (**CTTC**). The principal research topic (grant-free uplink access) was aligned with the objectives of Task 3.3. The joint research efforts and outcomes resulted in the preparation of a conference paper which was submitted in January 2020 and was accepted for presentation ([C-9]). Besides, **LUT** is also working with **AIT** in the topic of massive MIMO and NOMA related to WP3 and WP4 ([J-3], [J-8], [J-16]). Besides, **LUT** and **UOULU** (although in the same country) have also collaborated in tasks related to event-driven techniques, cyber-physical systems modelling and advanced communications ([J-5], [C-2], [J-7]). **AIT**, **LUT** and **TCD** are currently working on TEP for dimensionality reduction. The whole team collaborate in two publications: [C-3] describing the general idea of FIREMAN and [C-14] presenting the first results related to TEP. Besides, **SEAT** was involved in the selection process of the researcher hired by **CTTC** who will be fully allocated in FIREMAN.

Note: Since March 2020, all face-to-face activities and travel were put on hold due to COVID-19. This created some impairments to some activities, although the overall progress of the research activities is satisfactory.

All in all, the different expertise of the consortium has been successfully channelled to solve the proposed interdisciplinary problem, as well as specific domain problems related to cyber-physical system modelling, data acquisition, fusion and analytics. This convergence has been very positive, making the team as a whole to solve problems that the groups would be hardly capable to solve individually.

As we keep frequent telcos (every second week for technical discussions, and monthly steering group) and target at concrete goals and action points, the transnational collaboration is going smoothly. The only obstacle is the lack of synchrony in the funding agencies, which make face-to-face meeting more difficult (see Section 3.7).

In addition, FIREMAN is collaborating with joint activities with another CHIST-ERA consortium SOON (https://www.chistera.eu/projects/soon). In late June we had our first joint telco and in October we have participated in a joint workshop as part of INTER-ENG conference organized by SOON members called **International Workshop on Smart Technologies in Industry 4.0 (RATIONALITY)**. We have agreed to produce a book in Springer combining the key complementary research topics covered by the two consortia and prepare another joint workshop in one world-class conference of IEEE.

3.3. Significant events and results

The research of **WP2** is divided into three tasks: Two already concluded and one is ongoing. The ended tasks are related to the definition of potential study cases, their KPIs and the general framework to be tested in the particular study cases. The ongoing task focuses on fundamental limits. We hereby highlight the main technical achievements so far **per task level for each individual beneficiary involved**.

Task 2.1: Definition of Key Performance Indicators and Requirements (M1-M4; responsible: SEAT; involved: LUT, CTTC, AIT, UOULU). This task provided the KPIs and requirements to test and validate the proposed approach. **SEAT** defined a special table to be filled to define the KPIs, and filled it considering the following cases: drive chain failure in paintshop processes and spindle (kind of actuator) failures in Computer Numerical Control (CNC) grinding machines in the machining centre. **LUT** has redefined its study cases: TEP benchmark scenario, micro-grid and welding machine; it filled the proposed table. **CTTC** and **UOULU** worked together to define the communication KPIs for industrial environments under consideration. **AIT** provided the technical coordination of this task. The main outcomes of this task are documented in D2.1.

Task 2.2: Scenarios and System Architecture (M3-M6; responsible: LUT; involved: AIT, UOULU, SEAT). This task focused on the development of a general framework based on a three-layer model consisting of physical, information and regulatory layers. **LUT** proposed a three-layer model of industrial cyber-physical systems used to build a general framework based on 1+6 questions in a step-by-step procedure that give the boundary conditions of the particular solution to be taken. **LUT** tested this approach in its three proposed study-cases, also partly looking at potential use in electricity grids [J-5], [J14] and selection of IoT platforms for ML [S-5]. **SEAT** focused on its two test cases. **UOULU** and **AIT** provided support to **LUT** via several interactions in the proposed framework. The main results of this task are documented in D2.2 and in [C-15].

Task 2.3: Fundamental Studies (M5-M21; responsible: AIT, involved: LUT, UOULU). This task focuses on the analysis and research on theoretical bounds for our detection and prediction process according on the "important" events identified in Task 2.1. **AIT** focused on the use of QARMA algorithm in different scenarios [C-4] in collaboration with **LUT** and **UOULU** and with focus on TEP; and in [J-9] in collaboration with **LUT** for fault classification in power grid scenarios. In particular terms, these results indicate the potential to use QARMA for rate-event detection; one remarkable issue here is that QARMA has a high level of explainability of outcomes, also following the approach proposed in [J-15]. In addition, motivated by the COVID-19 pandemia, **LUT** team approached the event detection of epidemics based on machine-learning [J-13]. The main initial outcomes of this task are documented in D2.3.

The **WP3** research activities span over three different tasks which have all been initiated during the first year of the project. Two of the tasks, i.e., Task 3.1 and Task 3.2, have been already finalized (M15) and Task 3.3 is currently the only active one. We hereby highlight the main technical achievements so far **per task level for each individual beneficiary involved**.

Task 3.1: Physical Process Data Modelling (M4-M15; responsible: LUT; involved: CTTC, UOULU). The objective of this task lies on the characterization and classification of the raw and pre-processed data generated by the sensors. In this context, **LUT** (task leader) has investigated the differences between periodic and event-driven data sampling and has developed a flexible event-driven measurement technique based on spike-filtering. In addition, **LUT** has performed a classification of a power grid (electricity metering) dataset to identify which signals perform better with periodic or event-driven sampling and investigate whether the event-driven measurement technique with spike-filtering produces meaningful output for this dataset. **LUT** also applied the proposed event-driven techniques in the TEP dataset. **CTTC** has investigated analytical models for the accurate characterization of sensor traffic streams that exhibit high spatiotemporal correlation. Instrumental in this direction is the use of point processes, e.g., Markovian Arrival Process (MAP), which allow for a tractable expression of the traffic activation probability. The exact derivation of this metric will act as an input for the performance assessment of different communication protocols developed in Task 3.3. The application of this framework has already been studied by **CTTC** for cascade failure modelling in an N-element power system using as datasets the operational (inter-failure) times of the electrical components. Several parameter-estimation strategies have been explored, i.e., method of moments, least squares, maximum likelihood, expectation-maximization algorithm, and the trade-offs between complexity and accuracy have been identified. **UOULU** is also exploiting the MAP framework to model source arrivals at a given node. The main initial outcomes of this task are documented in D3.1.

Task 3.2: Local Data Collection and Storage (M6-M15; responsible: AIT; involved: LUT, CTTC, SEAT). The objective of this task is the development of solutions for data collection and sampling and techniques for efficient data storage. In this context, **LUT** studied the level of data compression required for periodic and event-driven sampling by considering the reconstruction error. A data compression strategy has been also developed by **CTTC** to effectively exploit the temporal and spatial correlation in the sensors' observations. The proposed strategy is composed of two building blocks: (i) an estimation block and (ii) a compression block. In the former, the temporal and spatial correlation of the data is exploited by means of a multidimensional linear filter in order to remove the impact of the sampling noise. In the compression block, the output of the estimation block is compressed by using truncated versions of the Karhunen-Loève expansions of the signals. **AIT** (task leader) has investigated the trade-offs that arise when using standard relational database technology (e.g., using Oracle 18c or MySQL 5.7) to manage the datasets required in FIREMAN versus using NoSQL databases, as it is well-known that the two general approaches can have significant implications on the ability of the project to handle big data. It is currently estimated (given datasets gathered and processed in the EU-funded projects PROPHESY and QU4LITY) that the size of the datasets required by FIREMAN will be in the order of 100MB at minimum to 10GB maximum, spanning 2 orders of magnitude in size difference. Already, working with the "educational" TEP dataset, the size of the data is around 100 MB (when expanded to include extra variables indicating state at previous times.) Even at maximum (10 GB), such dataset sizes are easily handled by traditional RDBMS technology, so Oracle 18c RDBMS should be an excellent candidate. However, depending on the access patterns required by the algorithms to be researched in FIREMAN, NoSQL databases supporting much higher CRUD operations throughputs, such as MongoDB are also valid candidates. In particular, since highly complex ad-hoc queries are not anticipated by the machine-learning algorithms to be used in the project, MongoDB, with its replication, sharding, and map-reduce server capabilities is also an excellent candidate as the storage system of the FIREMAN architecture. **SEAT** is collecting detailed information on how data are currently collected and stored in internal systems including data sampling and compression, and other relevant techniques and parameters. This information will be useful for **AIT** and **LUT** to determine the most suitable approaches for local data collection and storage considering a real industrial environment such as that of **SEAT**. The main initial outcomes of this task are documented in D3.2.

Task 3.3: Large-scale Data Transmission (active since November 2019 (M7); responsible: CTTC; involved: LUT, AIT, UOULU). The objective of this task is the design of agile communication schemes able to achieve scalable and ultra-reliable, low latency transmission of the sensing information. The ever-increasing number of interconnected devices expected in future industrial IoT applications will impose even more demanding requirements on the communication networks that will quickly surpass the capabilities of current technologies. Massive access, ultra-low latency, and high reliability are just a few of the features required to be provided by future communication systems. In this context, **LUT** has already extensively investigated the performance of massive MIMO-NOMA networks in dense industrial deployments [J-1], [J-3], [C-7], [C-11], [C-12],[J-7], [J-8], [J-16]. **LUT** also studied a multi-class communication system based on ALOHA [J-2], the performance of GFDM modulation, which can be a promising solution for large-scale transmissions [C-6], [C-13], and a preliminary demonstration of a 4G network in an industrial environment [C-1]. **LUT** and **CTTC** (task leader) have jointly studied ([C-9]) the applicability of the emerging IRS paradigm for reliable uplink access in an industrial scenario involving ultra-reliable low-latency communication (URLLC). The performance assessment revealed significant gains in terms of reliability, resource efficiency, and capacity and for different configurations of the IRS properties, unveiling the potential of IRS as a promising and low-cost solution to provide link diversity and achieve reliable uplink connectivity in industrial setups. **CTTC** has also proposed a general data-acquisition model composed of a fusion centre and N sensors where the impact of imperfect communication is considered. Using an ON-OFF power allocation strategy for a subset of active sensors, the transmission threshold is determined in such a way that the expected distortion of the best linear unbiased estimator at the fusion centre is minimized with respect to the channel realizations and the number of active sensors. **AIT** focuses on an mMTC scenario, where many power-limited sensor nodes, equipped with a single antenna, need to transmit their data to a hub (base station) with a high computational power and with an antenna array. The framework under study aims to maximize the sum rate of all transmitting sensors, subject to individual power and rate constraints. **UOULU** has already investigated the applicability of Long Range (LoRa) networks in highly dense industrial communication scenarios. The capacity of LoRa Wide-Area Network (LoRaWAN) in terms of the number of sensors under coverage has been evaluated using message replication techniques to exploit time diversity and increase reliability. In particular, a novel hybrid coded message replication scheme has been proposed by **UOULU** ([J-4]) which interleaves simple repetition and a recently proposed coded replication method and achieves improved performance - subject to minimum reliability requirements - without requiring additional transmit power. Moreover, they investigated the reliability gain when using superposed signal decoding at the gateways [J-11]. In addition, **UOULU** has setup a system-level simulator of LoRa technology for testing various protocol solutions in the context of FIREMAN [C-10]. In a collaboration between LUT and UOULU, different scenarios of LoRa were studied for industrial environment utilizing ns3 [J-7].

The work on **WP4** has started on time - November 2019 (M7). All partners involved in **WP4** have been taking part of the regular biweekly meetings related to the activities that are part of **WP3** and **WP2**. The purpose of these meetings is to keep the work harmonized between the tasks and to strengthen the collaboration between the partners. The participation in these meetings helped accelerate the activities related to **WP4**.

<u>Task 4.1</u>: Aggregation of heterogeneous big data (active since November 2019 (M7); responsible: UOULU; involved: LUT, CTTC, AIT, SEAT). The objective of this task is to enhance traditional clustering algorithms so that they account for the communication capabilities (e.g., limited communication infrastructure or provisionally deployed access points). The work on this task resulted in some early work delivered by **UOULU** and **LUT. UOULU** finished the work related to enhanced reliability in Low-Power Wide-Area Networks (LPWAN), in particular in LoRaWAN, with coded replications in November and started working on the data aggregation and scheduling with QoS guarantees, as well as the work on NOMA for enhanced ultra-reliable low-latency communications (URLLC) in coexistence with Enhanced Mobile Broadband (eMBB) traffic. This work resulted in a couple publications ([J-4]). Their work on LPWAN [C10] was presented at 6G Summit 2020. Also, **UOULU** is working on interference modelling in MTC scenarios with aggregation. **LUT** started the work on grant-free random access in collaboration with **CTTC**, which was presented at the 6G Summit 2020 [C-9]. **CTTC** is also exploring various strategies for intra-cluster discovery and cluster-head selection policies aiming to minimize the involved signalling overhead in dense industrial IoT scenarios ([C-8]). **LUT** is also extensively studying massive MIMO – NOMA scenarios deployments [J-1], [J-3], [C-7], [C-11], [C-12],[J-7], [J-8] **LUT** has also started work on spectral and energy efficiency in IRS-NOMA also together with **AIT** [J-16].

Task 4.2: Data compression and interpretation (active since April 2020 (M12); responsible: TCD; involved: LUT, AIT, OULU). The objective of this task is to investigate methods for big data reduction at various levels of the system architecture. The work will focus on machine learning approaches for dimensionality reduction of the aggregated data. Early work on this task resulted in a journal paper that was published in the IEEE Systems Journal [J-6] by **TCD**. Additionally, **TCD** has been focusing on the development of ML algorithms for data imputation in an industrial environment. Similarly, **CTTC** is currently investigating the lingering influence of imperfect wireless sensor transmissions to the emergence of missing data among measurement streams at the level of a fusion centre. The preliminary outcomes have resulted in two conference publications [C-16] and [C-18]. Some early results of the data compression algorithms used by **TCD** (i.e. PCA) have been shared with **AIT** for further analysis on how the compression affects the detection of rare events and its classification. In terms of data compression and interpretation, **LUT** has been considering the level of data compression required for periodic and event-driven sampling while considering the reconstruction error. **AIT**, **LUT, UOULU** and **TCD** are currently collaborating on dimensionality reduction of the TEP dataset. The work of WP5 has started in January 2020 with Task 5.1 focusing on the rare event detection using QARMA and other learning techniques. Task 5.2 recently started in close cooperation trying to build new visualisation focusing on rare events. **Task 5.1:** ML/DM Techniques for Detection and Prediction of Rare Events: The QARMA and R4RE family of algorithms is being expanded with many more features for different Use-Cases.

1. When it is known that a certain sub-set of sensors (attributes) in a dataset plays an important role in detecting the value for a target variable, it is possible to ask the algorithm to create rules based only on this sub-set, instead of working with the entire set of features available in the dataset.
2. The above works in conjunction with the MinExplainSet algorithm that was first implemented in the FIREMAN project a few months ago that searches for a minimal cardinality set of variables (features) that are contained in a rule-set that “explains” most (or all) of the target values of the target variable in the dataset. The MinExplainSet algorithm has been modified in two major directions: (a) a variant allows to extract from a given rule-set a minimal cardinality set of rules that “explain” most (or all) of the dataset instances that have the desired target attribute value, which can be of more benefit in certain production Use Cases and (b) a parallel version of the hybridized Breadth-First-Search and Beam Search has been developed and fully tested for use in cases where the cardinality of the original rule-set is too big to handle in a single CPU core.

Further, a parallell/distributed version of arbitrary-depth Feed Forward Neural Networks has been developed and tested (and is in the process of adding more features) that allows for layers of neural nodes that are any arbitrary mix of nodes (e.g. one node with a linear activation function, another 3 nodes with ReLU activation functions, and another 5 nodes with the GELU activation function, all in the same layer). The software (Open-Sourced and available in https://github.com/ioannischristou/popt4jlib) is being further developed and will eventually be combined with QARMA to provide for explainable predictions as follows: the deep network will make predictions, and whenever an explanation is asked for the prediction, the QARMA-produced ruleset will be searched to extract the rule with the highest support and confidence that supports the decision; if no such rule exists (meaning there is no support in the training dataset for the NN’s decision) the NN decision will be “vetoed” and the QARMA-based decision making will be used instead.

AIT is collaborating with LUT on the above developments. A paper containing a Use-Case of the DNN library within popt4jlib has been submitted to Elsevier Electric Power Systems Research journal, in collaboration with LUT, UOULU and other FIREMAN partners.

Task 5.2: The activities were based on how rare events could be better visualized. We initially focusing on TEP (constantly updated in https://github.com/5uperpalo/FIREMAN-project). Currently, we are working on how to crate a better visualisation of QARMA by indicating its association rules and mapping then in network plot. This is associated with its explainability.

WP6 just started its activities and we are preparing the testbeds.

3.4. Technology readiness level (TRL)

FIREMAN is mainly located in TRLs 2 and 3. However, for some activities, like study of beyond 5G systems, FIREMAN is TRL 1. Applications to be studied at LUT and SEAT can be associated with TRLs 4 and 5.

3.5. Consortium meetings

Meetings*				
N°	**Date**	**Location**	**Attending partners**	**Purpose**
1	May 17, 2019	LUT University	ALL	Online kickoff
2	Aug. 29, 2019	Oulu University	ALL	Kick-off and special session at ISWCS
3	May 15	Online	ALL (expected)	Yearly workshop

* All face-to-face meetings are cancelled until COVID-19 situation becomes solved. Monthly plenary meetings are held instead.

3.6. Deliverables

Deliverables*					
N°	**Title**	**Nature**	**Delivery date (month)**		**Partner in charge**
			Contractual	**Actual**	
7.1	Website	Dissemination	M2	M1	CTTC
7.2	Plans for Dissemination, Communication and Standardisation	Dissemination	M3	M2	TCD
1.1	Quality Assurance Plan	Management	M4	M3	LUT
2.1	User requirements and key performance indicators	Research	M4	M4	SEAT
1.2	Data Management Plan	Management	M6	M6	AIT
2.2	System architecture design	Research	M6	M4	LUT
1.3a	Yearly report	Management	M12	M12	LUT
7.4	Draft exploitation plan	Dissemination	M12	M12	SEAT
2.3	Important Rare Events Identification	Research	M12	M12	AIT
3.1	Report on Physical Process Data Modelling	Research	M15	M15	LUT
3.2	Design and Functional Architecture for Data Acquisition	Research	M15	M15	AIT
4.1	Initial Results on Heterogeneous Big Data Aggregation	Research	M18	M18	UOULU
1.4	Mid-term report	Management	M18	M18	LUT

*Available in the website. D2.2 will be updated when SEAT allows to publicly upload more detailed internal information related to test cases.

4. Documentation of results, valorisation, impact

4.1. Scientific publications (conferences/workshops, book chapters, etc.)

Scientific publications			
Publication Id		**Multi-partners (Yes/No)**	**International* (Yes/No)**
Journals with impact factor (IF)			
[J-16]	A. S. de Sena, D. Carrillo, F. Fang, P. Nardelli, D. B. da Costa, U. S. Dias, Z. Ding, C. B. Papadias, W. Saad, "*What Role Do Intelligent Reflecting Surfaces Play in Non-Orthogonal Multiple Access?*", accepted to IEEE Wireless Communication Magazine, July 2020. (IF: 11.391)	Yes	Yes
[J-15]	A. Jung and P. H. J. Nardelli, "An Information-Theoretic Approach to Personalized Explainable Machine Learning," in IEEE Signal Processing Letters, vol. 27, pp. 825-829, 2020, doi: 10.1109/LSP.2020.2993176. (IF: 3.105)	No	Yes
[J-14]	H. M. Hussain, A. Narayanan, P. H. J. Nardelli and Y. Yang, "*What is Energy Internet? Concepts, Technologies, and Future Directions*," in IEEE Access, doi: 10.1109/ACCESS.2020.3029251. (IF: 3.745)	No	Yes
[J-13]	R. L. Rosa et al., "*Event Detection System based on User Behavior Changes in Online Social Networks: Case of the COVID-19 Pandemic*," in IEEE Access, doi: 10.1109/ACCESS.2020.3020391. (IF: 3.745)	No	Yes
[J-12]	D. Gutierrez-Rojas, P. H. J. Nardelli, G. Mendes and P. Popovski, "*Review of the State-of-the-Art on Adaptive Protection for Microgrids based on Communications*," in *IEEE Transactions on Industrial Informatics*, July 2020. (IF: 9.112)	No	Yes
[J-11]	J. M. S. Sant'Ana, A. Hoeller, R. D. Souza, H. Alves, S. Montejo-Sanchez, "LoRa Performance Analysis with Superposed Signal Decoding," in IEEE Wireless Communications Letters, Early Access, July 2020, doi: 10.1109/LWC.2020.3006588. (IF: 4.66)	No	Yes
[J-10]	W. Chen, H. Ding, S. Wang, D. B. da Costa, F. Gong and P. H. J. Nardelli, "*Ambient backscatter communications over NOMA downlink channels*," in *China Communications*, vol. 17, no. 6, pp. 80-100, June 2020, doi: 10.23919/JCC.2020.06.007. (IF: 2.024)	No	Yes
[J-9]	M. Ullah, P. J. Nardelli, A. Wolff and K. Smolander, "*Twenty-one key factors to choose an IoT platform: Theoretical framework and its applications*," in IEEE Internet of Things Journal, June 2020. (IF: 9.936)	No	No
[J-8]	A. S. de Sena, F. R. M. Lima, D. B. da Costa, Z. Ding, P. H. J. Nardelli, U. S. Dias, C. B. Papadias, "*Massive MIMO-NOMA Networks with Imperfect SIC: Design and Fairness Enhancement*," in IEEE Transactions on Wireless Communications, June 2020. (IF: 4.951)	Yes	Yes
[J-7]	F. H. C. dos Santos Filho, P. S. Dester, E. M. G. Stancanelli, P. Cardieri, P. H. J. Nardelli, D. Carrillo and H. Alves, "*Performance of LoRaWAN for Handling Telemetry and Alarm Messages in Industrial Applications*", MDPI Sensors 2020, 20, 3061. (IF:3.275)	Yes	Yes
[J-6]	M. Dzaferagic, N. McBride, R. Thomas, I. Macaluso, N. Marchetti, "*Improving In-Network Computing in IoT Through Degeneracy*", IEEE Systems Journal, February 2020. (IF: 3.987)	No	Yes
[J-5]	M. Tomé, P. Nardelli, H. Majid Hussain, S. Wahid, A. Narayanan, "*A Cyber-Physical Residential Energy Management System via Virtualized Packets*", MDPI Energies, 13(3), 699, February 2020. (IF: 2.702)	Yes	Yes

[J-4]	J. M. S. Sant'Ana, A. S. Hoeller, R. D. Souza, S. Montejo-Sanchez, H. Alves and M. De Noronha Neto, "*Hybrid Coded Replication in LoRa Networks*," IEEE Transactions on Industrial Informatics, January 2020. (IF: 9.112)	No	Yes
[J-3]	A. S. de Sena, D. Benevides da Costa, Z. Ding, P. Nardelli, U. S. Dias, C. B. Papadias,"*Massive MIMO-NOMA Networks with Successive Sub-Array Activation*", IEEE Transactions on Wireless Communications, December 2019. (IF: 4.951)	Yes	Yes
[J-2]	P. Dester, P. Cardieri, P. Nardelli, J. M. C. Brito, "*Performance Analysis and Optimization of a N-class Bipolar Network*", IEEE Access, September 2019. (IF: 3.745)	No	Yes
[J-1]	A. S. Sena, D. Benevides da Costa, Z.Ding, and P. Nardelli, "*Massive MIMO–NOMA Networks With Multi-Polarized Antennas*", IEEE Transactions on Wireless Communications, September 2019. (IF: 4.951)	No	Yes
Conference proceedings			
[C-19]	N. Mayedo, O. L. Alcaraz, H. Alves, and M. Latva-aho, "On the SIR meta distribution in massive MTC networks with scheduling and data aggregation," submitted to VTC2021, 2020.	No	No
[C-18]	T. Alexopoulos, C. Kalalas and G. Korres, "On the Imputation of Power System Measurement Streams with Imperfect Wireless Communication," in Proc. of IEEE Global Power, Energy and Communication Conference 2020 (IEEE GPECOM '20), virtual event, October 2020.	No	Yes
[C-17]	C. Kalalas and J. Alonso-Zarate, "Lightweight and Space-efficient Vehicle Authentication based on Cuckoo Filter," in Proc. of IEEE 5G World Forum 2020 (IEEE 5G-WF '20), virtual event, September 2020.	No	Yes
[C-16]	C. Kalalas and J. Alonso-Zarate, "Sensor Data Reconstruction in Industrial Environments with Cellular Connectivity," in Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2020 (IEEE PIMRC '20), virtual event, August 2020.	No	Yes
[C-15]	H. Majid Hussain and P. Nardelli, "*IoT-based Heuristics for Demand Response based on Home Energy Management System*", in Industrial CPS conference, February 2020.	No	Yes
[C-14]	D. Gutierrez-Rojas, M. Ullah, I. T. Christou, G. Almeida, P. Nardelli, D. Carrillo, J. M. S. Sant'Ana, H. Alves, M. Dzaferagic, A. Chiumento, C. Kalalas, "*Three-layer Approach to Detect Anomalies in Industrial Environments based on Machine Learning*", in Industrial CPS conference, February 2020.	Yes	Yes
[C-13]	D. Carrillo-Melgarejo, S. Kumar, G. Fraidenraich, P. Nardelli, D. Benevides da Costa, "*Achievable Sum Rate and Outage Capacity of GFDM Systems with MMSE Receivers*", in International Conference on Communications (ICC), Dublin, Ireland, June 2020.	No	Yes
[C-12]	A. S. de Sena, F. R. M. Lima, D. Benevides da Costa, and P. Nardelli, "*Enhancing User Fairness in Massive MIMO-NOMA Networks*", in International Conference on Communications (ICC), Dublin, Ireland, June 2020.	No	Yes
[C-11]	A. S. de Sena, D. Benevides da Costa, Z. Ding, P. Nardelli, U. S. Dias, C. B. Papadias, "*Successive Sub-Array Activation for Massive MIMO-NOMA Networks*", in International Conference on Communications (ICC), Dublin, Ireland, June 2020.	Yes	Yes
[C-10]	A. S. Hoeller, J. M. S. Sant'Ana, J. Markkula, K. Mikhaylov, R. D. Souza, H. Alves, "*Beyond 5G Low-Power Wide-Area Networks: A LoRaWAN Suitability Study*", 6G Wireless Summit 2020 (6G Summit), Levi, Finland, March 2020.	No	Yes

[C-9]	D. Carrillo-Melgarejo, C. Kalalas, A. S. Sena, P. Nardelli and G. Fraidenraich, "*Reconfigurable Intelligent Surface-Aided Grant-Free Access for Uplink URLLC*," 6G Wireless Summit 2020 (6G Summit), Levi, Finland, March 2020.	Yes	Yes
[C-8]	C. Kalalas and J. Alonso-Zarate, "*Massive Connectivity in 5G and Beyond: Technical Enablers for the Energy and Automotive Verticals,*" 6G Wireless Summit 2020 (6G Summit), Levi, Finland, March 2020.	No	Yes
[C-7]	A. S. Sena, D. Benevides da Costa, Z. Ding, and P. Nardelli, "*Massive MIMO-NOMA Networks with Multi-Polarized Antennas*", in Proc. of IEEE Global Communications Conference (GLOBECOM '19), Hawaii, USA, December 2019.	No	Yes
[C-6]	D. Carrillo-Melgarejo, J. M. Moualeu, P. Nardelli, G. Fraidenraich, and D. B. da Costa, "*GFDM-Based Cooperative Relaying Networks with Wireless Energy Harvesting*", in Proc. of International Symposium on Wireless Communication Systems 2019 (ISWCS '19), Oulu, Finland, August 2019.	No	Yes
[C-5]	A. Chiumento, N. Marchetti, I. Macaluso, "*Energy Efficient WSN: a Cross-layer Graph Signal Processing Solution to Information Redundancy*", in Proc. of International Symposium on Wireless Communication Systems 2019 (ISWCS '19), Special Session on IoT in Energy Systems and Industrial Environments, Oulu, Finland, August 2019.	No	Yes
[C-4]	I. Christou, "*Avoid mistaking the hay for the needle in the stack: Improving Rare Event Detection with Online Pruning of Useless Rules in Quantitative Association Rule Mining*", in Proc. of International Symposium on Wireless Communication Systems 2019 (ISWCS '19), Special Session on IoT in Energy Systems and Industrial Environments, Oulu, Finland, August 2019.	No	Yes
[C-3]	P. Nardelli, C. Papadias, C. Kalalas, H. Alves, I. Christou, I. Macaluso, N. Marchetti, R. Palacios, J. Alonso-Zarate, "*Framework for the Identification of Rare Events via Machine Learning and IoT Networks*", in Proc. of International Symposium on Wireless Communication Systems 2019 (ISWCS '19), Special Session on IoT in Energy Systems and Industrial Environments, Oulu, Finland, August 2019.	Yes	Yes
[C-2]	M. Tomé, P. Nardelli, L. C. Pereira da Silva, "*Flexible event-driven measurement technique for electricity metering with filtering*", in Proc. of IEEE International Conference on Industrial Informatics (INDIN' 19), Espoo, Finland, July 2019.	Yes	Yes
[C-1]	F. Polunin, D. Carrillo Melgarejo, T. Lindh, A. Pinomaa, P. Nardelli and O. Pyrhonen, "*Demonstrating the Impact of LTE Communication Latency for Industrial Applications*", in Proc. of IEEE International Conference on Industrial Informatics (INDIN' 19), Espoo, Finland, July 2019.	No	Yes
Book chapters			
[B-1]	R. Sedar, C. Kalalas, F. Vazquez-Gallego, J. Alonso-Zarate, "*Intelligent Transport System as an Example of a Wireless IoT System,*" Chapter in "Wireless Networks and Industrial IoT- Applications, Challenges and Enablers", published by Springer. ISBN: 978-3-030-51473-0. Q3 2020.	No	No

*With the term "International", we refer to journals/conference proceedings with a content of cross-national significance and global soundness, beyond the context of a single country.

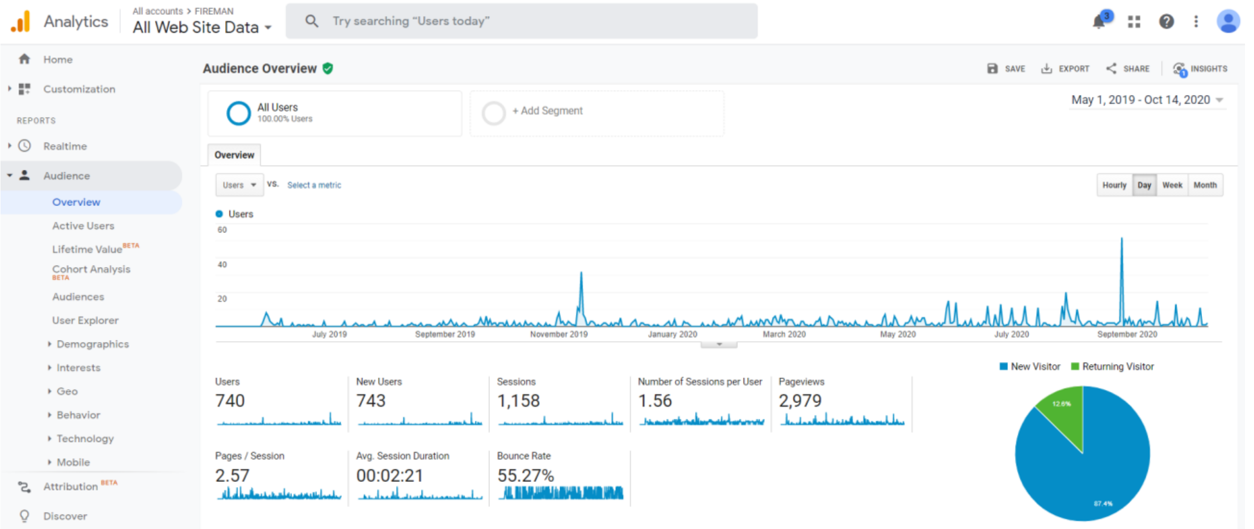
4.2. Valorisation (software, products, spin-offs, etc.)

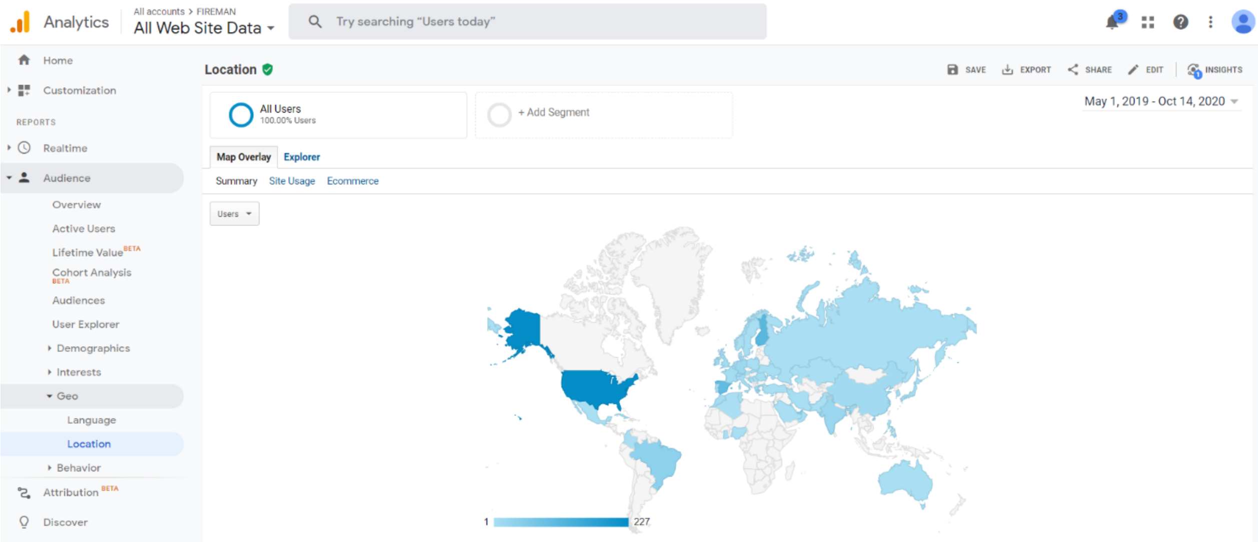
TCD has been working on creating a spin-off company based on the machine learning research they are involved in. They successfully finished two commercialisation training camps organized by the NDRC (National Digital Research Centre). These two programs resulted in the **TCD** team winning two prizes related to the potential of the start-up idea and the commercial roadmap. Through the work related to the commercialisation of their research they have done a thorough customer discovery and validation, which resulted in a complete market and competitor analysis, and allowed them to clearly state the value proposition. Currently, they are preparing an application for commercialisation funding from Enterprise Ireland to continue the business and technology development which should result in a Minimum Viable Product (MVP) and first paying customers (early adopters).

Open repository of FIREMAN: https://github.com/5uperpalo/FIREMAN-project

4.3. Other dissemination of results

- **Project website:** including deliverables, dissemination and events. **CTTC** is managing the [website](https://fireman-project.eu/), including updates related to publications and events. According to Google Analytics (data retrieved on October 14th 2020), the FIREMAN website (https://fireman-project.eu/) has attracted worldwide interest since its creation.





- The project coordinator and Dr. Kalalas (CTTC) are members of the IEEE-IES Technical Committee on Industrial Agents
- Participation in a joint workshop as part of INTER-ENG conference organized by SOON members called **International Workshop on Smart Technologies in Industry 4.0 (RATIONALITY)**.
- The Project Coordinator is currently writing a textbook for Wiley called: **Cyber-Physical Systems: Theory, Methodology and Applications,** which is partly based on the methodology developed within FIREMAN
- Organization of two special sessions in the prestigious ISWCS'19 and PIMRC'20 conferences with 3 papers from FIREMAN in the first, and 1 in the second.
- SEAT presented the project internally in different committees both at the management level and at the operational level.
- The project coordinator presented FIREMAN in the 2019 EU-US Frontiers of Engineering Symposium in Sweden (https://www.naefrontiers.org/188386/2019-EUUS-Frontiers-of-Engineering-Symposium)

4.4. Updated Risk Management

The situation regarding potential project risks, as listed in the FIREMAN proposal, has been regularly reviewed by both the project coordinator and the technical coordinator as well as by the Work Package leaders in project plenary meetings. The project will keep closely monitoring for any potential risk emerging, e.g., due to Covid-19. In the following, an updated risks table is presented where some comments on the state of play for each risk can be found.

Description of risk	WPs involved	Proposed Risk Mitigation Measures	Risk materialised?	State of play
Technical work differs from project objectives.	WP2-WP5	This situation will be avoided via frequent reviews of technical work by FIREMAN WPs' leaders.	No	Work is progressing as planned.
Technical approaches of WP3-WP5 cannot be integrated in a consolidated CPS architecture.	WP2-WP5	This risk will be avoided through early cross-WP communication within FIREMAN consortium and on project management level through tracking of cross-WP activities.	No	Technical solutions/enablers are fully consistent with the system architecture proposed in WP2 and disseminated with the research community. WP3-WP4 interaction is active in several topics, such as missing data imputation (WP4) due to communication impairments (WP3).
Solutions for data acquisition cannot fulfill QoS (reliability / latency / scalability).	WP3	If possible, relaxing the requirements may be considered, e.g., in the sense that not all requirements have to be fulfilled at the same time.	No	There are no indications for this risk to be materialised at the moment.
Technical approaches of WP3-WP5 turn out to be too complex for experimental implementation.	WP3-WP5	The project management will consider a reduction of the complexity of the algorithms to be implemented.	No	Early to assess. Mapping of the functional requirements to technical approaches is ongoing.
No significant events in the dataset to detect.	WP3-WP4	1) Convince the dataset provider to provide better dataset. 2) Produce a simulated (other) dataset	No	The dataset utilized up to now are benchmarks for fault detection or have known events. This will be reassessed when the testbeds are deployed.

		containing events of interest to verify algorithms work as expected.		
QARM algorithms too slow to detect reliably conditions leading to important rare events.	WP5	Since QARMA is inherently parallel and scalable, the consortium will add processing units on-demand.	No	Until now, QARMA has performed as expected.
Template-grid methods taking too much time to run.	WP5	Sub-sampling methods will be used to reduce the search space.	No	Not applicable at this point.
Algorithms for the prediction of rare events cannot cope with large amount of data.	WP5	The use of sparse methods will be explored in order to reduce drastically the amount of data processed by the analytics (e.g. artificial intelligence) algorithms.	No	The proposed data acquisition method based on event-driven sampling is based on a “filtering” protocol for decreasing the datasets. Besides, all algorithms analysed were running properly, without any issues related to the large datasets.
Partner with responsibility for experimental setup drops out.	WP6	FIREMAN will reach other relevant industrial partners that have been already contacted to cooperate.	No	There are no such indications currently.
Insufficient or low-quality disseminatio n.	WP7	FIREMAN targets only high-impact conferences and journals. The WP7 leader will regularly and timely circulate calls for papers to the interested partners.	No	FIREMAN research outputs have been already disseminated in top-notch journals and international conferences.
Departure of partner.	ALL	If possible, other FIREMAN partners take over the activities. Otherwise, a new partner will be recruited.	Partly	This is an ongoing situation: FIREMAN is requesting a transfer from AIT to ALBA as the Greek partner. The researchers involved in FIREMAN moved from AIT to ALBA without affecting the collaboration within the team.

Deliverables plan and quality issues.	ALL	All deliverables are reviewed by the management. If necessary, external reviewers will also be used.	No	Deliverables have been produced on time and with no delay. The appointment of external reviewers in WP3 and WP4 deliverables ensures their quality and completeness.