FLUIDOS

CREATING A FLUID, DYNAMIC, SCALABLE,
AND TRUSTABLE COMPUTING CONTINUUM.

Use Cases

www.fluidos.eu

VISIT OUR WEBSITE!

The FLUIDOS Ambition

At a first sight, the proliferation of computing devices and resources across the world can suggest that the computing continuum is already a reality. In fact, we have smart edge devices, smartphones and computers, servers at the edge, geographically distributed computing infrastructures for 5G services, as well as data centers and cloud resources, namely a continuum set of computing resources from IoT devices to the cloud.

However, the FLUIDOS project has an even more ambitious target: **enable end-users, developers, resource owners, applications and services to leverage the entire computing continuum, from IoT devices to the edge and cloud, as a unique pool of resources.** Which means, overall, not having to deal anymore with problems such as "Where should I install and run this application?" with multiple answers such as on a specific device, on the edge or on the cloud, or "What happens if the location where my application is running is no longer reachable?", but delegate the solution of these problems to a new type of infrastructure that handles them **transparently**, without requiring external intervention. Simplifying the life of end users, developers, and resource owners in a way that is similar to what happens nowadays when we have to communicate on the Internet: our applications just send data, and the smart network infrastructure just does the magic of delivering the data to the destination.

The FLUIDOS ambition is, in fact, to **design and build a platform**, called "meta Operating System", that **creates a "fluid" computing infrastructure (hence the name FLUID-OS) where applications and services can float, irrespective of their physical location, and are continuously orchestrated to adapt to the ever changing environment, operating conditions, and requirements.** This will require overcoming the barriers posed by the **different technologies** present on the continuum (e.g., an application could be executed in any location, such as a smartphone or a remote server); the challenges of the **ever changing operating conditions** (network latency and congestion; availability of green energy; flash crowds and more); and overcoming the boundaries of **administrative domains**, given that the computing continuum will result from the cooperation of many entities who will share the infrastructure under their control, such as the Internet of today.

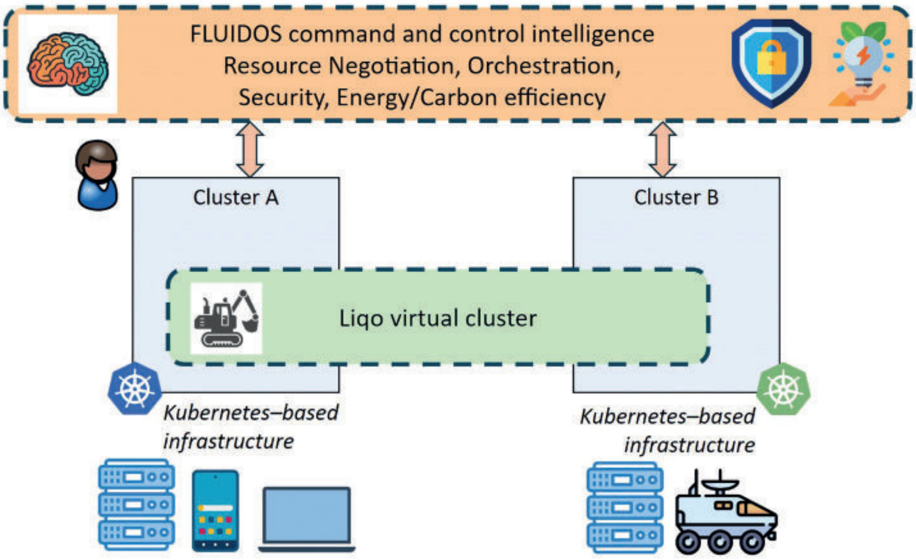
The FLUIDOS Architecture

The FLUIDOS software stack, shown in the figure below, is heavily based on existing components, hence reusing what is already widely accepted by the industry.

Starting from the bottom, FLUIDOS is OS-agnostic, although most of the physical infrastructure we consider runs one of the many flavors of Linux. The next layer introduces the FLUIDOS "big bet", i.e., it assumes that all the devices in the physical infrastructure adopt Kubernetes as a unifying layer. Actually, this is supported by the fact that multiple Kubernetes flavors exist, with different requirements, performance, and scalability properties, although they all expose the same API and provide the same overall functional behavior. In the end, this means that the FLUIDOS stack runs on resource-constrained machines such as Raspberry PI3-class boxes and other embedded devices, robots and drones, individual laptops, desktops and servers, as well as cloud-based clusters.

However, Kubernetes creates individual and separated clusters: Liqo (https://liqo.io), a promising open-source software, bridges this gap through the concept of the virtual cluster, spanning across multiple physical nodes. In fact, Liqo provides the elementary substrate for the FLUIDOS continuum, namely the basic network, storage and service fabric.

Finally, FLUIDOS builds an intelligent layer of command and control software on top of the above building blocks, either by creating new components to address the challenges introduced in the next section, or by extending Kubernetes and/or Liqo.



The FLUIDOS Challenges

To achieve its objectives, FLUIDOS needs to address the following main challenges:

- **Borderless network, storage and service fabric.** A borderless infrastructure must be established among clusters, which enables applications to operate across the continuum as they were set up in a single (Kubernetes) clusters, with no differences.
- **Seamless security within virtual clusters.** Virtual clusters, created on top of the endless computing continuum, must experience the same security properties guaranteed within a single Kubernetes cluster, irrespective of the technology present in the cluster itself (e.g., orchestration system such as K8s vs. K3s; network provider such as Cilium vs. Calico; etc.), and must provide the proper borders and security barriers around the virtual cluster itself.
- **Authentication, authorisation, permissions and trust.** Nodes composing the FLUIDOS continuum can be owned by different actors, which may originate strong security concerns given that nobody is completely in control of the entire infrastructure. Hence, authentication, authorisation, permissions and trust are key components here, including the strong guarantees that workload owners do not harm the hosting infrastructure, and that the resource provider does not alter the third-party software running on its infrastructure.
- **Intent-driven, policy-driven borderless orchestration.** A novel borderless orchestration system, driven by high-level intents and policies, addresses two orthogonal dimensions with highly scalable algorithms: (a) it orchestrates workloads running in each FLUIDOS virtual cluster (i.e., optimised scheduling); (b) it determines *when*, *how* and with *whom* each virtual cluster has to be established (i.e., optimised virtual cluster lifecycle management) on top of the (almost) endless resources of the continuum.
- **Energy and carbon awareness.** Given the importance of sustainability for the future of our planet, FLUIDOS is deeply committed to develop algorithms and components that can leverage the continuum to achieve higher energy efficiency, and/or to privilege green energy. Furthermore, FLUIDOS can reuse more efficiently existing hardware resources thanks to the well-known principle of resource sharing.

The FLUIDOS Use Cases

The FLUIDOS ambition is highly challenging and cannot be achieved without involving external stakeholders, namely an external community of developers and early adopters. Hence, FLUIDOS heavily invested in new use cases that highlight the advantages of the FLUIDOS computing continuum, transparently driven and orchestrated by the underlying infrastructure substrate (the FLUIDOS meta-OS). As possible examples, use cases will demonstrate that the FLUIDOS continuum:

- **achieves better sustainability**
- **enables more robust infrastructures**
- **decreases development and/or maintenance costs**
- **is fast enough to react to the ever changing conditions, and more.**

Those can be found in the following sections of this publication.

USE CASES

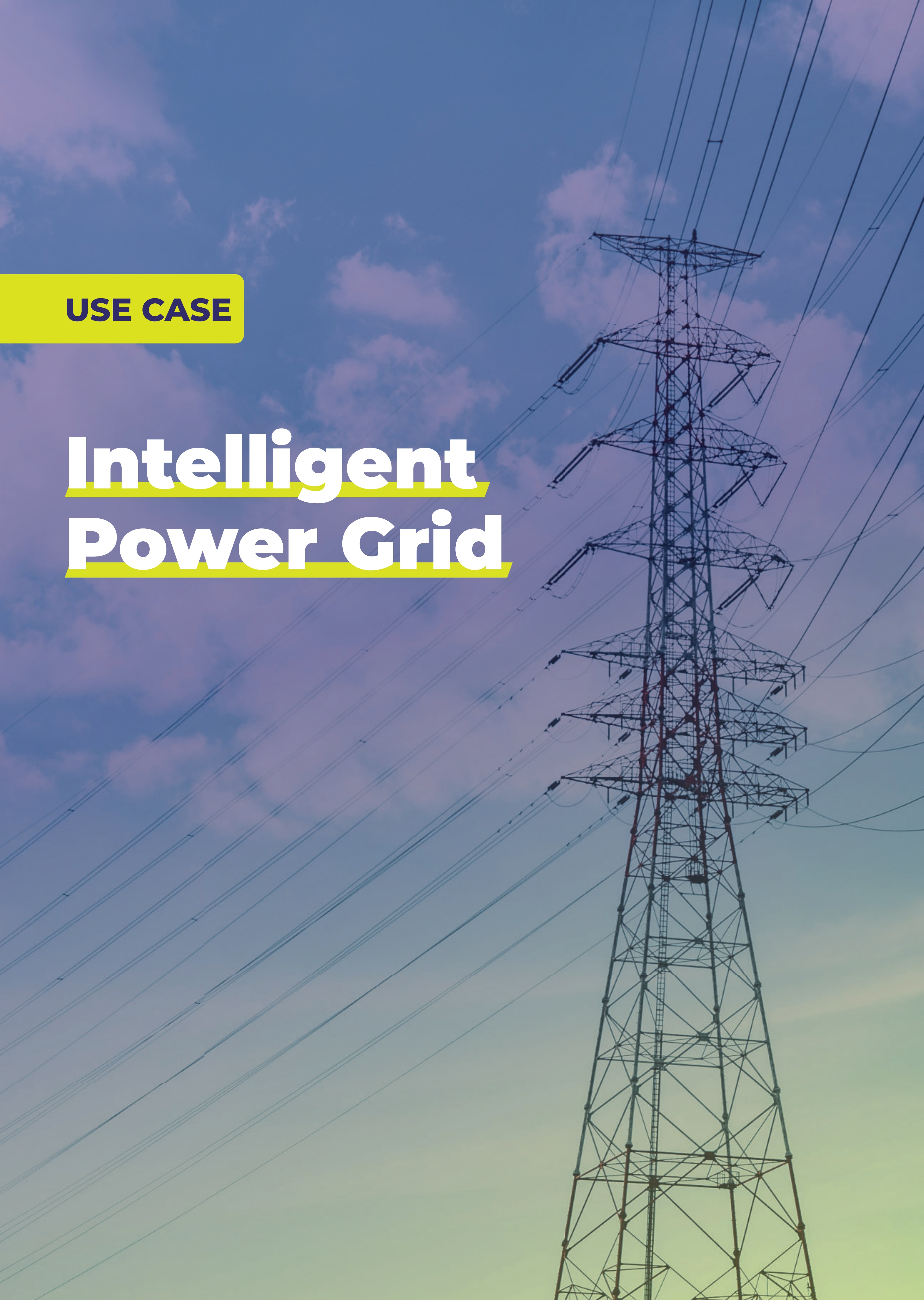
Intelligent Power Grid

Robotic Logistics

Smart Viticulture

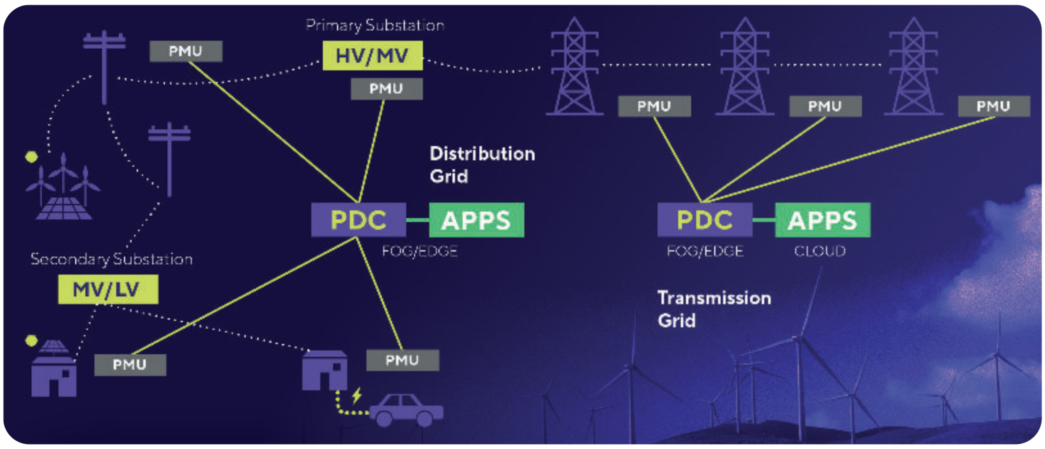
USE CASE

Intelligent Power Grid



Intelligent Power Grid

The introduction of a massive number of renewable energy sources requires a highly monitored distribution grid that can coordinate energy producers and consumers in real-time. The emerging power grid architecture will rely on a large number of measuring devices such as PMUs (Phasor Measurement Units) strategically positioned on the power grid to measure electrical quantities, with real-time data collection and synchronization carried out by PDCs (Phasor Data Concentrators).
The synchronized data collected by PMUs and managed by PDCs serves as crucial input for the Grid State Estimation algorithm, which is the basis for real-time applications such as monitoring, control and protection of the grid, and offline applications such as archiving and offline analysis. Traditionally, PMUs have been deployed within the transmission grid. However, the evolving power grid demands a more comprehensive monitoring approach, including the distribution grid.



PROBLEM

The introduction of PMUs into the distribution network brings about various challenges, including:

- Scalability: The number of PMUs required for the power grid to be observable is significantly higher for the distribution case. While hundreds of PMUs are needed for the transmission network to be observable, the distribution network will require a number of the order of thousands. This implies the need to manage a massive number of distributed devices and their corresponding data flows.

- Resilience: With thousands of devices in play, enhancing the resilience of the ICT infrastructure for data collection is crucial. For instance, managing faults or planned maintenance should be automated, as well as the possibility for the above devices when disconnected from the network and/or unable to access companion services running in the cloud.

- **Latency:** The system must support large-scale data processing in real-time, meeting the specific latency constraints of the targeted applications (of the order of tens of milliseconds for control applications)
- **Cybersecurity:** The power grid is a critical infrastructure for the country and, as such, must be protected from any malicious attack, including the ones targeting monitoring and control services.

CURRENT APPROACH

While the transmission grid already hosts PMUs typically managed centrally by a single or few PDCs, the introduction of PMUs in the distribution grid is still at an experimental level. Nevertheless, the massive number of PMUs and PDCs in the future grid makes this centralized approach (a) hardly scalable, (b) not resilient to network failures or forced disconnections (e.g., to preserve a portion of the network from an ongoing attack).

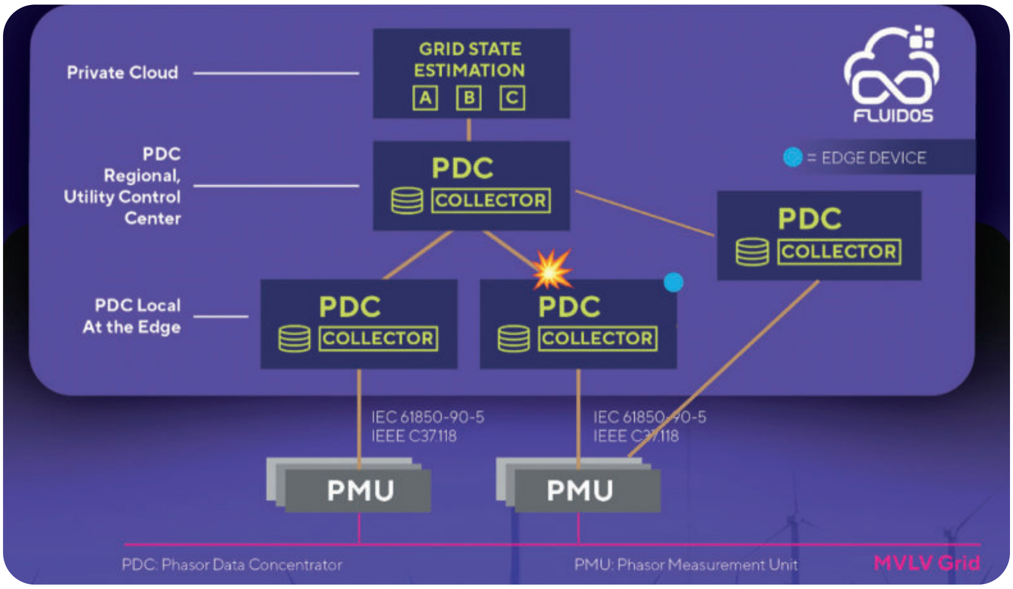
FLUIDOS APPROACH

Traditionally, PDCs were monolithic applications running on dedicated hardware; however, with the increasing computational power available at lower costs, this is changing in recent years. Experimental efforts are underway to virtualize applications and utilize Kubernetes for orchestrating the deployment of PDCs and real-time analysis applications at the edge. This is aimed at reducing latency issues and improving resiliency, avoiding the need of operator physical assistance in case of outages, and paves the way for their usage within a FLUIDOS-based environment.
In fact, FLUIDOS creates a continuum of resources from the edge to the cloud and enables the displacement of workloads, such as data collection and analysis processes, based on specific scenarios (faults, reconfiguration, maintenance). The main features of the approach enabled by FLUIDOS are:

COMPUTING CONTINUUM FLUIDOS would enable PDCs and analysis applications to continue functioning even if communication with control centers is interrupted by migrating PDC services to an adjacent node in case of fault.

CYBERSECURITY FLUIDOS ensures service isolation from other applications on the hosting node with different usage permissions. It also leverages logging and anomaly detection capabilities, and provides survival capabilities in case a portion of the grid is disconnected from the main network, hence preserving its operations in case of a cyber attack.

INTENT-BASED ORCHESTRATION FLUIDOS can automatically orchestrate PDCs based on the latency between the node and PMUs, thereby improving the power grid state estimate or responding to faults.



FLUIDOS ADVANTAGES

- **Scalability:** Effectively managing a high number of PMUs.
- **Resilience:** The capacity to tolerate and respond to faults, forced disconnections, and (planned) maintenance.
- **Application Performance:** Improved performance by satisfying the application requirements.
- **Cybersecurity:** Enhanced security measures.

KEY PERFORMANCE INDICATORS (KPI'S)

Reduction of software update time on N geographically distributed devices: from O(N) duration to O(1)

Support for time-sensitive services also on nodes without GPS hardware: synchronization error < 32us

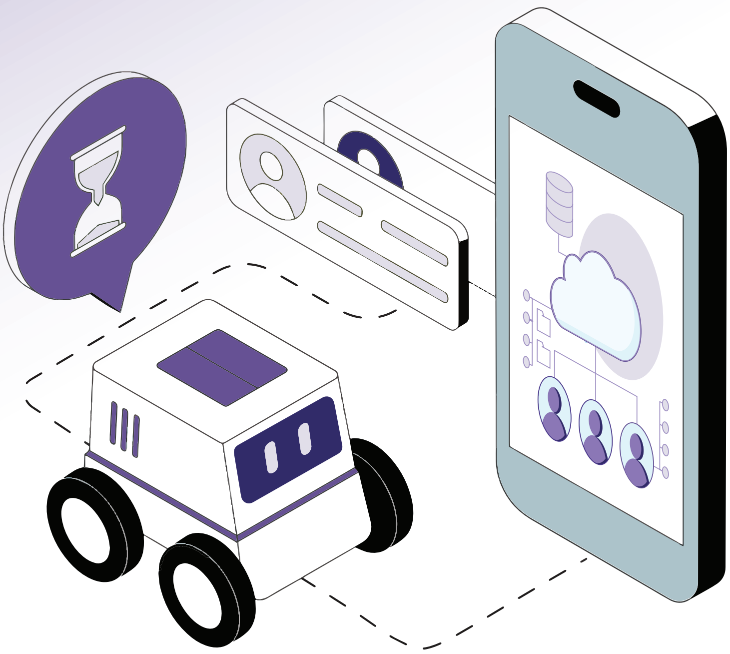
Scalability of phasor data processing at local grid distribution level: >= 1K secondary stations

No electrical downtime in presence of an ICT fault

Reduction of phasor data loss after a processing chain reconfiguration: >= 80%.

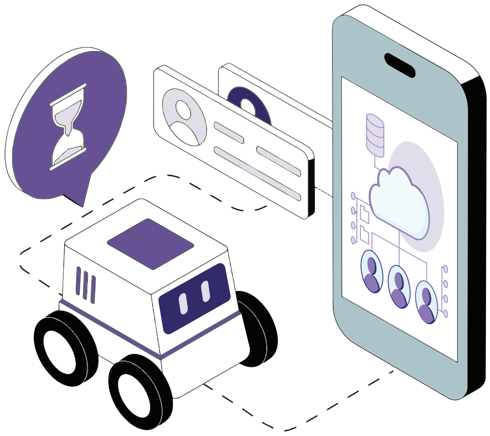
USE CASE

Robotic Logistics



Robotic Logistics

Mobile robots for Industry 4.0, smart logistics, and retail are resource-constrained and battery-powered mobile robots that operate in shared spaces with humans and other IoT devices, such as elevators, automatic doors, and other mobile robots. For autonomous operation, mobile robots are equipped with sensors and high-performance computational algorithms that (1) need considerable computing power and (2) need to be executed at the highest speed.



PROBLEM

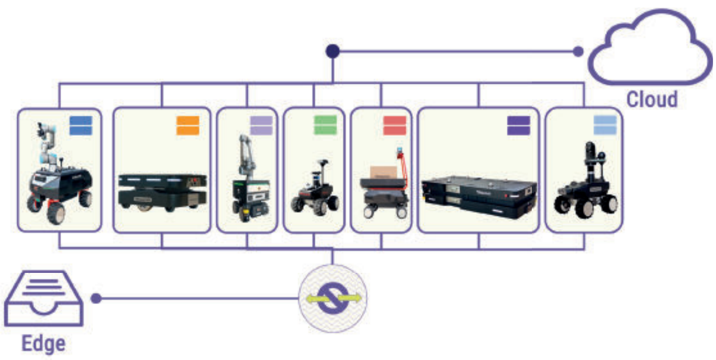
This kind of robotics, by definition, uses battery-powered devices, unlike stationary robot arms. This adds constraints to an already complicated application. In robotics logistics, idle robots are not productive. And intrinsically, robots need to charge their batteries to continue working. During this time, the robot is not available to move things around.

The robot's software is complex and requires heavy computation, so it is mandatory to equip the robot with powerful enough computing engines that, however, should limit the consumed power. Onboarding an ultra-low-resource device will lead to a robot that is not capable of processing the required tasks, and onboarding an ultra-high-resource device will deplete the battery very quickly.

There is an additional relationship between the computer load and the battery drain. Heavy computational tasks drain the battery faster.

CURRENT APPROACH

To solve the above computing problem with the current technology, we can follow either one of the following two approaches: (a) CENTRALIZED PROCESSING, (b) INDIVIDUAL COMPUTING.



a Centralized Processing

The robot fleet operates under a centralized communication architecture, whereby the intelligence (i.e., processing resources and software) is placed in a central component, such as the fleet manager (or orchestration system). All communication between the individual robots and the fleet manager is routed through an intermediary edge or cloud device. In this setup, the central device acts as a hub that (a) implements processing algorithms, and (b) facilitates and controls the communication flow within the network.

This centralized approach may introduce a single point of failure and potential latency issues as all communications must pass through the central hub, and it is not appropriate when robots may experience connectivity issues (e.g., poorly connected area in a factory).

b Individual Computation

In this approach, robots typically compute everything on the device itself. This practice, however, may lead at least to three main problems.

First, computational inefficiency, as it may result in scenarios where some robots remain idle and underused, while others are overwhelmed with computational tasks, nearing the point of overloading their onboard computers.

Secondly, over-dimension of the available computing resources (or, alternatively, the necessity to deploy simplest computing algorithms), given that the necessary hardware resources to comply with peak computing demands must be available on the robot itself; alternatively, to save on the cost of the above hardware, simplest (and less resource-hungry algorithms, but also less efficient) should be used.

Thirdly, excessive battery usage, as all the processing happens on the robot itself, substantially draining the battery power, which is particularly important on small robots (with a small-size battery, while computing power for robot navigation and other algorithms is an invariant with respect to the size of the robot.

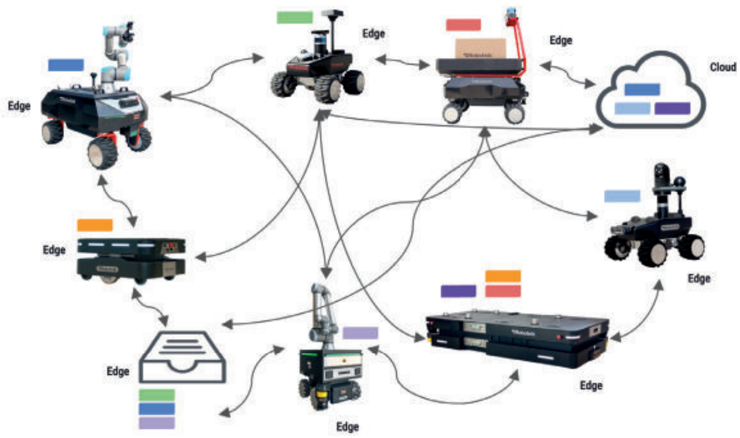
FLUIDOS APPROACH

The FLUIDOS continuum has the capability to transparently use computing resources nearby, increasing the overall system‘s productivity by intelligently and dynamically externalizing the robotics workload to other devices (e.g., a server at the factory premises) and/or using the robot‘s idle time (i.e., when the robot is docked in the battery charging station) to increase the entire system‘s computational capabilities. Each robot can leverage this approach when it is well-connected to the network, while it can rely on its sole onboard computing capabilities (which are turned on only upon necessity) when moving in a poorly connected area.

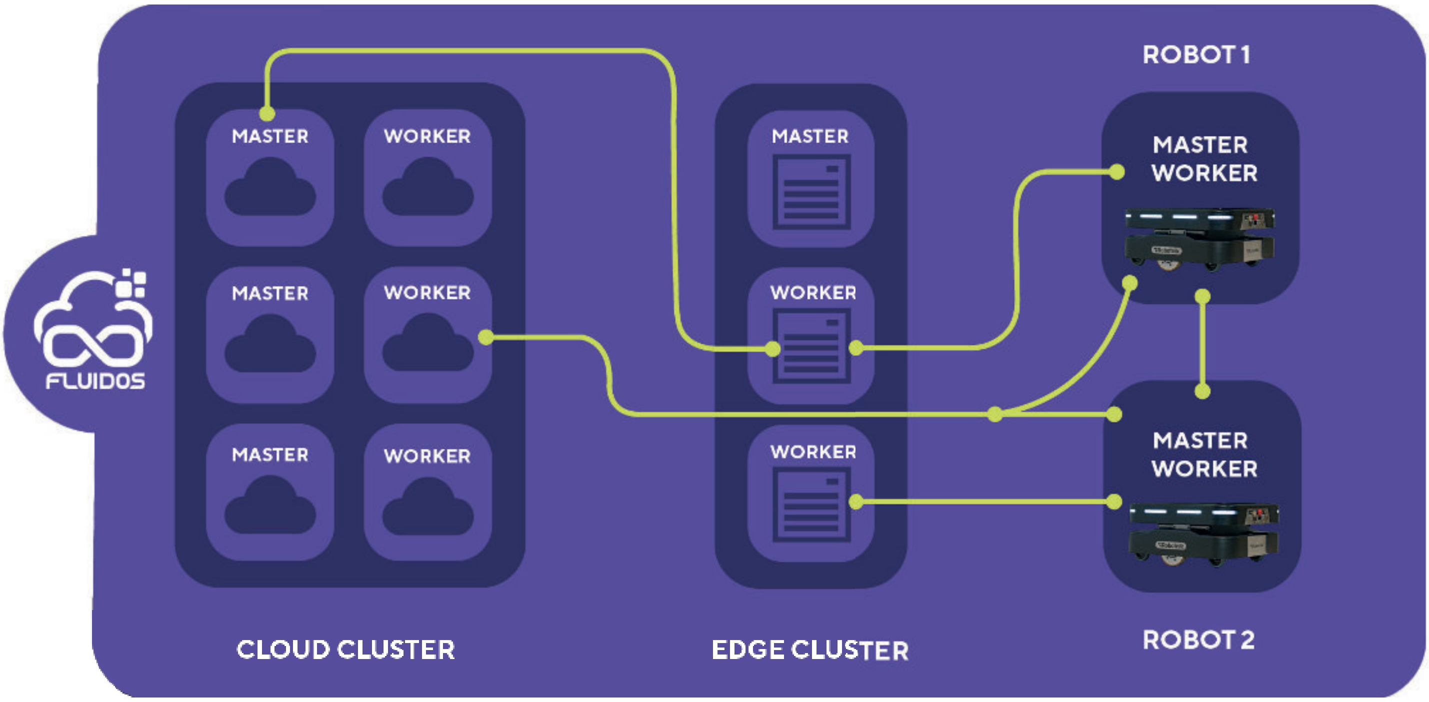
FLUIDOS APPROACH

This approach will lead to a significant decrease in battery usage and an increase in the robot's computational capabilities beyond its onboard limits, with the capability to dynamically adapt its computing behavior (i.e., onboard or offloaded) based on the actual operating conditions.

With FLUIDOS, we can apply the cloud continuum computing approach by considering each robot as an edge device and intelligently and dynamically outsourcing robotics workloads to other robots or devices depending on the environment. This can be achieved without interfering with the robotics task, so robot developers will be able to run their applications without changing their way of working.



Instead of using monolithic bare-metal workloads, the robots will use cloud-native technologies like containerization and Kubernetes to split and dynamically place the workloads in different devices. All robots will be treated as edge devices that can accept or externalize workloads, instead of being isolated devices.



FLUIDOS APPROACH

Given the potential capability of the FLUIDOS intent-based orchestrator to pursue different objectives, workload distribution among the different available systems can be optimized to achieve diverse goals such as:

- Maximize the battery life of the robots.
- Minimize the time it takes for the robots to complete their tasks.
- Ensure that all of the robots are evenly utilized.
- Avoid overloading any individual robot.

Highly dynamic decisions can be envisioned as well. For example, if a robot is low on battery, FLUIDOS might move its workloads to other robots with more battery power. Or, if a robot is overloaded, FLUIDOS might move some of its workloads to other robots that are less busy.
In a nutshell, the FLUIDOS approach to robot workload orchestration enables to improve the performance, efficiency, and reliability of your entire system.

FLUIDOS ADVANTAGES

- Increased robot capabilities by using more powerful stationary devices without compromising battery life.
- Ability to use the robot fleet for non-robotic tasks, increasing the value of the fleet.
- Decreased battery usage and increased robotic fleet productivity.
- Reduced deployment time using cloud-native technologies.
- Reduced costs by onboarding lower-resource computational devices.

KEY PERFORMANCE INDICATORS (KPI'S)

Improved lifetime of running vehicles, due to better usage of the battery: at least 15%.

Increased speed of the moving robots due to the capability to leverage smarter services running at the edge: at least 15%.

Potential reduction of the number of robots required to serve a factory: at least 10%.

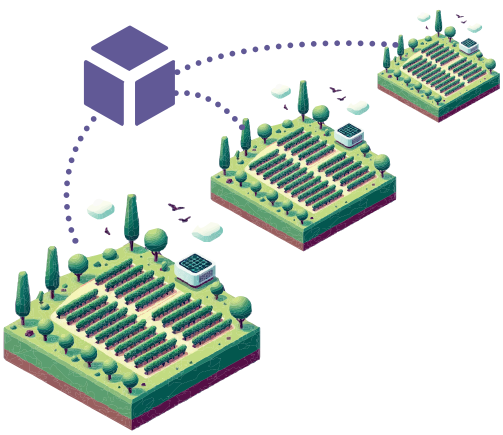
Reduction of the cost of the computing hardware installed on each robot: at least 10%.

USE CASE

Smart Viticulture

Smart Viticulture

Unlocking Efficiency in Viticulture Management



TerraviewOS from Terraview GmbH, is a cloud-native service for agriculture and specifically in viticulture and helps producers manage their property inputs and outputs.
In this use case, software has to be installed and operated at both the edge property premises and in the cloud. The locations carry out different operations (e.g., data gathering at the edge, data processing in the cloud).
Customers operate through „crate“ edge devices, aiming to integrate a central service by FLUIDOS technology to address lspecific problems. These edge devices handle specific TerraviewOS functions, while the central service provides additional capabilities. Each of these devices are FLUIDOS nodes and correspond to a customer managing one or more vineyards, using shared resources from Terraview.

PROBLEM

- No simple and integrated network fabric over cloud continuum: The current landscape lacks a unified network fabric across the cloud continuum, leading to connectivity and interoperability challenges among different microservices installed in different locations.

- No easy way to define and manage a cloud continuum application: Managing applications in cloud continuum environments is currently complex, affecting resource and lifecycle management. In particular, the location of each microservice within the Terraview application can vary based to the current operating conditions, e.g., a cloud-based micro-service should be installed at the edge when the connectivity is weak or absent, which is a per-customer, manual, costly, and error-prone configuration.

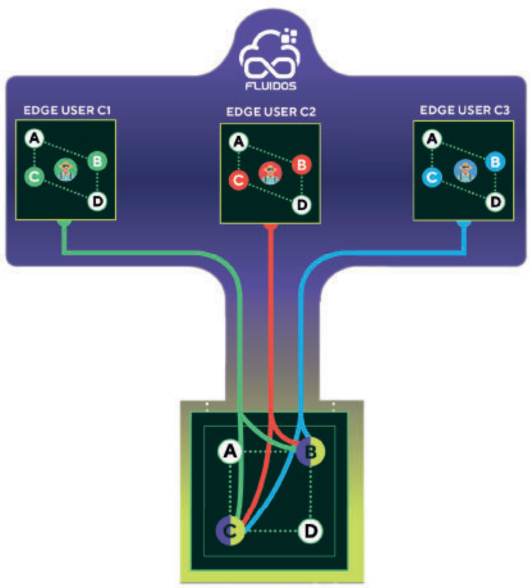
- Deficit of security, privacy or trust: Security, privacy, and trust issues hinder edge/fluid computing‘s efficiency. The Terraview application should benefit from advanced security measures (TEE/TPM, anomaly detection, etc.) for encryption, data protection, and access control.

CURRENT APPROACH

- Everything is centralised: The current centralised architecture, with services in one location, risks failures and latency.
- All data resides with the service provider: Concerns about privacy and security emerge from centralised data storage. The use case aims to create a distributed storage solution to boost security and user control.
- No reliability on the customer side - network dead, no service: Reliance on constant network access causes service interruptions during outages.
- Manual and per-customer configuration. The orchestration of the different components of the Terraview application has to be done per-customer based on its actual operating conditions, manually, which is expensive, error-prone, and not flexible (e.g., cannot adapt in case of a change in the operating conditions).
- Huge data ingress over limited network connection: Large data ingestion over restricted networks leads to congestion and inefficiency.

FLUIDOS APPROACH

The novel FLUIDOS-based TerraviewOS is designed with a focus on efficient and secure operation across multiple customer on-premise sites leveraging the support services operated upon centralised cloud infrastructure. At the core of this architecture lies the concept of the FLUIDOS domain, which encompasses both customer sites and centralised services. Each customer site will host a FLUIDOS node, responsible for running TerraviewOS services. These nodes are designed to ensure security, with data encrypted both at rest and in transit, and will provide multi-tenant access to necessary centralised services. This approach aims to facilitate the secure sharing of workloads and data across FLUIDOS nodes that reside in the cloud. The backend services of TerraviewOS will automatically be executed in the best available location, preferably in the cloud but also at the edge in case of network issues, with a crucial design choice being the avoidance of one backend instance per FLUIDOS node, without any manual intervention from the Terraview engineers.



FLUIDOS ADVANTAGES

- Security: raw data stays local, isolation and trusted execution on the network.
- Decentralisation: local access, no central cloud, reduced network traffic, workloads at edge and core.
- Operating costs: no manual intervention from Terraview engineers to deploy the application; everything is automatically deployed (and re-deployed) by FLUIDOS.
- Business continuity: adapt to network changes, operate without uplink, and replicate for failover.

KEY PERFORMANCE INDICATORS (KPI'S)

30%

Reduction in cost per tenant operations by TER: at least 30% reduction.

Data transferred: reduction of one order of magnitude from GBs to hundred MBs.

Processing time of imagery: unchanged (no increase).

Complete disconnection: zero end-user downtime via state synchronisation.

VISIT OUR WEBSITE!

Join FLUIDOS in the pursuit of creating an open and collaborative ecosystem focused on the development of a multistakeholder market of edge services and applications, promoting European digital autonomy.

www.fluidos.eu

@fluidosproject

company/fluidos

The FLUIDOS project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101070473.