Orchestration and optimization in industrial IoT (Internet of things)

ARCTOS LABS

3 Apr 2023

1 Abstract

This paper outlines how multi-domain orchestration and optimization capabilities can contribute to increased availability and reliability of industrial IoT systems by making operational capabilities more autonomous. By adding intelligence and optimization into software deployment and failure remediation on a holistic and multi-domain layer, orchestration, and optimization capabilities will enable operational teams to be more responsive and achieve higher availability and reliability of their industrial IoT systems. These capabilities further help enterprises to manage their set of digital information technology (IT) / operational technology (OT) assets, explore more data-driven operations, and achieve a higher level of integration of IT and OT systems towards Industry 4.0 visions.

2 Background

Our world is floating on an ocean of data. In 2022, more than 2.5 quintillion bytes of data were created every day adding up to a staggering 97 zettabytes by the end of the year [1]. According to the same source, 85% of that data is copies of other data. The amount of data processed daily will grow exponentially in the foreseeable future. How data is stored, processed, and utilized is an important issue for most enterprises defining their need for cloud and edge environments.

Industrial enterprises utilize a large span of different IT and OT environments to support their business operations, which are increasingly fuelled by data. These digital systems span all the way from public clouds, wide-area networks, on-premises data centers at various locations, wireless wide-area networks (aka mobile broadband), work-from-home environments, secure access service edge, local area networks, industrial automation systems, sensors, and actuators, etc.

Internet of Things (IoT) is a term that captures the concept of connecting “things” into a digital environment whereby the data produced by these “things” can be easily consumed from central locations or the “things” be centrally controlled. The “things” can be a multitude of various devices that create all kinds of data thereby being an integrated part of enterprises’ digital systems.

The software implementation of industrial automation systems is often divided into industrial IoT service components that may be distributed across the enterprise digital environment thus being dependent on various IT and networking capabilities.

Industrial automation systems have very high requirements on reliability, availability, and performance such as latency. This has led to diverse communication technologies e.g., PROFIBUS, or Ethernet-based solutions such as PROFINET and EtherCAT. Each of these solutions meets a specific set of requirements but as a result, they are only compatible on the physical layer of communication. This lack of interconnectivity hampers the evolution of Industry 4.0 and led to the arrival of the TSN Ethernet standards to support more integrated scenarios, also extended to be integrated into 5G environments [2].

3 Industrial IoT services

Industrial IoT services have different missions depending on the verticals in which they are applied, but they are often built up following a common pattern.

Data is generated from a variety of sources, ranging from simple measuring devices that only measure e.g., a temperature, to cameras or other devices that generate a large amount of streaming data. Sources of data may include business IT systems or even IT systems operated by enterprise suppliers. The data is then processed (potentially in several steps).

Once data is refined it often feeds into a control loop. That control loop is either automatic with some constraints on the speed of control, or manual where human beings are supposed to take an action (in case of some values out of bounds, or thresholds reached). The control loop creates outputs that are then fed into an actuator. Information from the control loop and the recorded data is sometimes required to be stored for later reference.

A typical industrial enterprise has a multitude of such industrial IoT services that are part of the complete industrial IoT system built in a heterogeneous environment with sensors, PLCs, machines, computer hosts, etc.

These environments are migrating towards Hybrid IT and Multi-cloud technologies, which implies that resources are shared between industrial IoT services, but also with other business IT systems.

These resource domains are a mix of on-prem resources that are typically operated by the enterprise itself and resources that depend on an external service provider such as a telecom operator or a public cloud service provider.

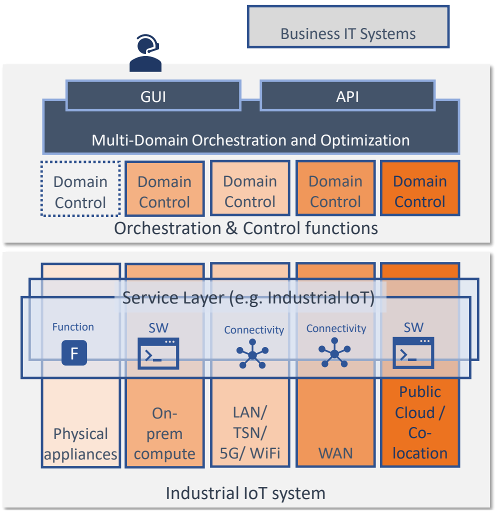

Figure 1: Multi-domain Industrial IoT systems

Physical appliances are devices that contribute functionality but expose limited management and configuration capabilities.

On-prem computing may take many different forms, for example, standard compute servers, PLCs, IoT gateways, single board computers (such as Jetson, and Raspberry Pi).

4 The need for automation to address complexity across domains

As complexity grows, so is the demand for solutions related to automated orchestration and optimization to assure that the entire industrial IoT system with all its industrial IoT services operates with satisfying availability and reliability to support its missions.

Orchestration and optimization, therefore, need to take a holistic approach across all the resource domains and all the industrial IoT services.

A desired characteristic of the industrial IoT system, including its orchestration and control functions, is its resilience to failures. This translates into the capability to launch actions to restore industrial IoT service parameters in case of such failures.

Remediation actions can be launched by control loops inside the industrial IoT services themselves (e.g., re-transmissions, downgrading QoS to keep industrial IoT services available, using alternative communication paths, etc), to minimize industrial IoT service impact from faults. Those control loops are thereby part of the industrial IoT service.

Remediation actions can also be taken by resource domain controllers that contain their own orchestration control loops. These orchestration control loops try to maintain the desired state (intent) of the resource domain and its industrial IoT service components by allocating resources to industrial IoT service components from its resource pool. For example, Kubernetes will strive to re-schedule Pods managed by a ReplicaSet due to a failure of a worker node. SDN controllers may re-route a flow should a certain path become broken.

These different control loops (inside industrial IoT services and in resource domain controllers) may lead to an uncontrolled competition of resources, oscillation, or non-optimal use of resources, and in the worst case further reduced overall industrial IoT performance.

Therefore, a holistic multi-domain orchestration layer is required. This layer will be responsible for coordinating the different domains thereby avoiding competition and achieve a holistic optimization given various operational situations. The capabilities are further outlined in coming sections.

A benefit of having orchestration control loops in externalized controllers (i.e., not in the industrial IoT service itself) is that their behaviour can be tailored to specific enterprise goals and environments.

The multi-domain orchestration functionality also become important to enable cloudification and integration of industrial IoT services towards Industry 4.0 visions with its increased use of data. Capabilities include placing industrial IoT service components at optimal locations, and achieve interconnection between the industrial IoT service components, based on industrial IoT service constraints and current network and compute status and utilization.

To enable appropriate actions these orchestration control loops, require an accurate understanding of the industrial IoT system state to achieve the intended state transitions.

5 Capabilities of a multi-domain orchestrator in industrial IoT

This section contains a list of basic capabilities.

Deploy distributed industrial IoT services and their components

Industrial IoT systems contain a range of different industrial IoT services. The multi-domain orchestration can deploy, update, and retire the industrial IoT services upon request from API/GUI. Such capability includes for the optimizer element to optimally select the location for each industrial IoT service component along the edge-to-cloud continuum whilst fulfilling any industrial IoT service constraints.

The placement needs to consider:

1. The amount of data that is generated and the cost and efficiency of collecting and transporting data and exposing it to processing. I.e., where the data is made available, the efficiency and cost of data transport (network) including potential wireless paths, requirements on saving all raw data, data exposure, and any data brokering.
2. How and where to apply data processing to create information and feed into the decision/action part of the industrial IoT service. Depending on how data is routed, candidate locations for processing, processing requirements for running AI/ML workloads, the requirements on reliability and availability are issues that need to be optimized.
3. How to achieve needed constraints on the industrial IoT service in case that apply? I.e., usage of low-latency and deterministic communication (such as time-sensitive networking – TSN), placement and execution environments for industrial IoT service components that implement data processing functionality, redundancy aspects of those, and the connection to actuators.

Coordinate data communication establishment

As an integral part of a deployment of industrial IoT services, there needs to be a corresponding establishment of data communication. The selection of locations for industrial IoT service components depends on available communication capabilities and characteristics between candidate locations.

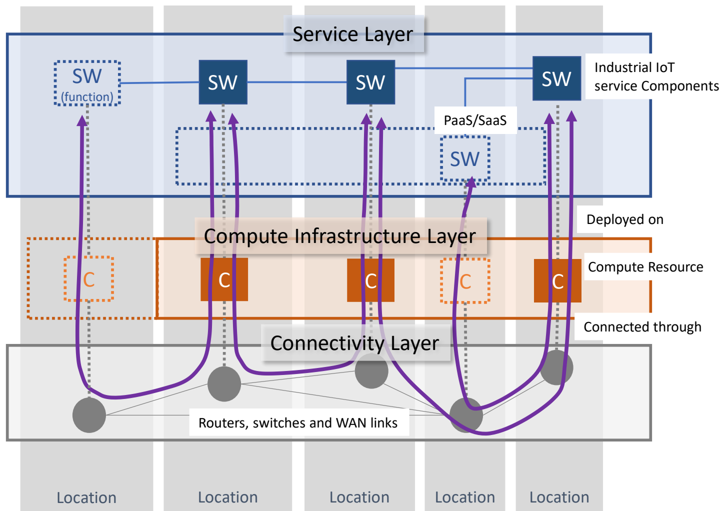

Figure 2: Industrial IoT service traffic traversing the network

Once industrial IoT service component locations have been selected, the multi-domain orchestration needs to assure that data communication between selected locations is available, and if necessary, initiate required actions.

Remediation in case of operational failures

Availability and reliability requirements in industrial IoT systems are normally very high. It may mean that solutions have been designed for redundancy so they can be invoked by individual resource domain controllers (or inside the industrial IoT service itself).

The multi-domain orchestration act on its holistic view and can initiate actions following two directions:

1. Deciding to move a resource domain intent to that of another resource domain, should the initially selected resource domain fail to restore its intents.
2. Alternatively, the multi-domain orchestrator can intelligently back off from some of the industrial IoT services intents and thereby optimize remaining capacity and quality in case there is no solution available.

Act on industrial IoT service parameters that indicate services are impaired or unavailable

Industrial IoT systems typically expose a large range of metrics related to the availability and reliability of the industrial IoT services. To correctly understand those metrics and how they reflect industrial IoT services, the observability capability of the multi-domain orchestration is crucial.

Based on such knowledge, the multi-domain orchestrator can decide to act which could be any of the actions outlined by previous chapters or to instruct a resource domain controller by modifying resource domain intents for industrial IoT service as outlined by the previous chapter.

Mitigating security threats

As industrial IoT services are business-critical, they are sensitive to security threats. The multi-domain orchestrator can apply various measures to deal with threats such as Denial of Service (DoS) attacks. These actions are based on that the observability element can determine the anomaly.

The above actions need to be subject to optimization to determine if the consequences of trying to isolate infected devices or network segments imply a greater reduction in QoS of the industrial IoT system compared to the QoS delivered during the attack.

6 Arctos Labs Edge Cloud Optimization (ECO) as an optimizer in the multi-domain orchestration layer

ECO is a model-driven optimizer that can fulfill the capabilities of the optimization element as outlined by this paper.

It can therefore be an important part of a full multi-domain orchestration layer as outlined below:

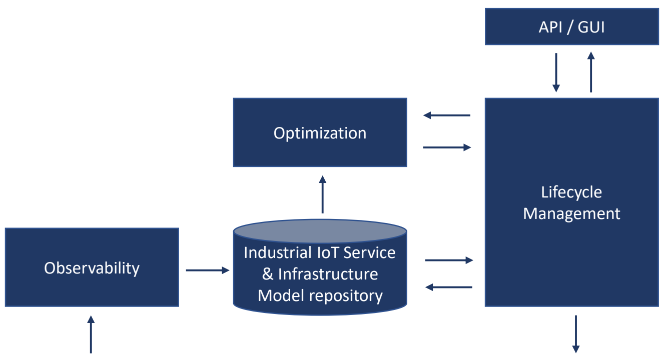

Figure 3 : Architecture of a multi-domain orchestration solution

It uses descriptors for the industrial IoT services as a declarative input describing the composition of industrial IoT services from components and their required compute and data communication characteristics.

ECO uses an optimization model abstraction inside which makes it easy to adapt to various optimization criteria and scenarios.

ECO is also capable of consuming knowledge and related metrics reflecting the underlying communication topology and distribution of compute locations so that it can decide on the placement of industrial IoT service components and optimize for the wanted criteria.

ECO can through these capabilities also sort out resource competition scenarios where multiple industrial IoT services compete on scarce resources within a resource domain.

7 Acknowledgement

This work is performed in the context of the Distributed Artificial Intelligent Systems project supported by the ECSEL Joint Undertaking.

DAIS (https://dais-project.eu/) has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 101007273. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and Sweden, Spain, Portugal, Belgium, Germany, Slovenia, Czech Republic, Netherlands, Denmark, Norway, and Turke

8 References

[1] How Much Data Is Created Every Day in 2022, Webtribunal.
https://webtribunal.net/blog/how-much-data-is-created-every-day/#gref

[2] Integration of 5G with Time-Sensitive Networking for Industrial Communications, 5G-ACIA.
https://5g-acia.org/whitepapers/integration-of-5g-with-time-sensitive-networking-for-industrial-communications/

Arctos Labs Scandinavia AB, Bangårdsgatan 14, 972 35 Luleå, Sweden. E-mail: info@arctoslabs.com. Web: www.arctoslabs.com. Reg. no: 556743-9152