ZEROSWARM

*ZERO-enabling Smart networked control framework for Agile cyber physical production systems of systems*

# D3.2 - CPSoS zero touch network & service management of 5G resources

Project funded by Horizon Europe, Grant Agreement #101057083

| | |
|---|---|
| **Topic** | HORIZON-CL4-2021-TWIN-TRANSITION-01-08 |
| **Project Title** | ZERO-enabling Smart networked control framework for Agile cyber physical production systems of systems |
| **Project Number** | 101057083 |
| **Project Acronym** | Zero-SWARM |
| **Deliverable No/Title** | D3.2 CPSoS zero touch network and service management of 5G resources |
| **Contractual Delivery Date** | M14 |
| **Actual Delivery Date** | M14 |
| **Contributing WP** | WP3 – Shaping 5G in the Industrial Domain |
| **Project Start Date, Duration** | 01/06/2022, 30 Months |
| **Dissemination Level** | Public |
| **Nature of Deliverable** | R |

| **Document History** | | | |
|---|---|---|---|
| *Date* | *Version* | *Author* | *Description* |
| *10-02-2023* | *0.1* | *i2CAT* | *ToC definition* |
| *24-03-2023* | *0.2* | *i2CAT & all* | *Collaboration distribution* |
| *25-05-2023* | *0.3* | *i2CAT & all* | *First draft* |
| *26-06-2023* | *0.4* | *i2CAT, NXW* | *ToC updated* |
| *10-07-2023* | *0.5* | *i2CAT & all* | *Final ToC updates and additional contributions* |
| *20-07-2023* | *0.6* | *i2CAT, ACC, OPT, NXW* | *Final contributions provided. First internal review provided and comments addressed* |
| *28-07-2023* | *0.7* | *Huawei, i2CAT* | *Final internal review comments provided and addressed* |
| *28-07-2023* | *1.0* | *CERTH* | *Final submission* |

## Authors List

| **Leading Author (**Editor) | | | |
|---|---|---|---|
| *Surname* | *Initials* | *Beneficiary Name* | *Contact email* |
| *Carmona* | *EC* | *i2CAT* | *estela.carmona@i2cat.net* |

| **Contributors** (in alphabetic order) | | | | |
|---|---|---|---|---|
| *#* | *Surname* | *Initials* | *Beneficiary Name* | *Contact email* |
| *1* | *Andolfi* | *MA* | *Nextworks* | *m.andolfi@nextworks.it* |
| *2* | *Camps* | *DC* | *i2CAT* | *daniel.camps@i2cat.net* |
| *3* | *Chackravaram* | *KC* | *Accelleran* | *kiran.chackravaram@accelleran.com* |
| *4* | *Darroudi* | *MD* | *Neutroon* | *mahdi.darroudi@neutroon.com* |
| *5* | *Gozalvez* | *JG* | *UMH* | *j.gozalvez@umh.es* |
| *6* | *Guerra-Gomez* | *RG* | *Neutroon* | *rolando.guerra@neutroon.com* |
| *7* | *Kazmi* | *SK* | *Opticoms* | *adil.kazmi@opticoms.de* |
| *8* | *Krzikalla* | *RK* | *SICK* | *roland.krzikalla@sick.de* |
| *9* | *Lucas-Estañ* | *MCLE* | *UMH* | *m.lucas@umh.es* |
| *10* | *Parker* | *SP* | *Accelleran* | *stephen.parker@accelleran.com* |
| *11* | *Zeljkovic* | *EZ* | *Accelleran* | *ensar.zeljkovic@accelleran.com* |
| *12* | *Khodashenas* | *PSK* | *Huawei* | *pouria.khodashenas@huawei.com* |
| *13* | *Krendzel* | *A* | *Huawei* | *andrey.krendzel@huawei.com* |

## Reviewers List

| **List of Reviewers** (in alphabetic order) | | | | |
|---|---|---|---|---|
| *#* | *Surname* | *Initials* | *Beneficiary Name* | *Contact email* |
| *1* | *Agdaci* | *KA* | *Opticoms* | *kerim.agdaci@opticoms.de* |
| *2* | *Khodashenas* | *PSK* | *Huawei* | *pouria.khodashenas@huawei.com* |
| *3* | *Krendzel* | *A* | *Huawei* | *andrey.krendzel@huawei.com* |
| *4* | *Lazaridis* | *GL* | *CERTH* | *glazaridis@iti.gr* |

**DISCLAIMER OF WARRANTIES**

This document has been prepared by Zero-SWARM project partners as an account of work carried out within the framework of the contract no 101057083.

Neither Project Coordinator, nor any signatory party of Zero-SWARM Project Consortium Agreement, nor any person acting on behalf of any of them:

- makes any warranty or representation whatsoever, express or implied,
    - with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or
    - that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or
- that this document is suitable to any particular user's circumstance; or
- assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if Project Coordinator or any representative of a signatory party of the Zero-SWARM Project Consortium Agreement, has been advised of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

Zero-SWARM has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057083. The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed in the deliverable lies entirely with the author(s).

# Table of Contents

# List of Abbreviations

| Abbreviation | Description |
|---|---|
| 3GPP | Third Generation Partnership Project |
| 4G | Fourth Generation |
| 5G | Fifth Generation |
| 5GC | 5G Core |
| AAI | All Active Inventory |
| AF | Application Functions |
| AGV | Automated Guided Vehicle |
| AI | Artificial Intelligence |
| AMF | Access and Mobility Management Function |
| API | Application Programming Interface |
| APN | Access Point Name |
| AUSF | Authentication Server Function |
| BBU | Baseband Unit |
| BSF | Binding Support Function |
| BSS | Business Support Systems |
| CapEx | Capital Expenditure |
| CN | Core Network |
| COTS | Custom Off The Shelf |
| CP | Control Plane |
| CPE | Customer Premise Equipment |
| CPS | Cyber Physical System |
| CPSoS | Cyber Physical Systems of Systems |
| CSMF | Communication Service Management Function |
| CU | Central Unit |
| DDoS | Distributed Denial of Service |
| DHCP | Dynamic Host Configuration Protocol |
| DU | Distributed Unit |
| eMBB | Enhanced Mobile Broadband |
| E2E | End-To-End |
| EPC | Evolved Packet Core |
| HW | Hardware |
| GDPR | General Data Protection Regulation |
| gNB | gNodeB |
| GST | Generic network Slice Template |
| GUI | Graphical User Interface |
| IaC | Infrastructure-as-code |
| ICS | Industrial Control Systems |
| IMSI | International Mobile Subscriber Identity |
| IIot | Industrial Internet of Things |
| IoT | Internet of Things |

| | |
|---|---|
| IT | Information Technology |
| KPI | Key Performance Indicator |
| KPM | Key Performance Measurement |
| L2 | Layer 2 |
| L3 | Layer 3 |
| LAN | Local Area Network |
| LSTM | Long Short-Term Memory |
| LTE | Long-Term Evolution |
| MANO | Management and Orchestration |
| MCC | Mobile Country Code |
| MDAF | Management Data Analytics Function |
| MES | Manufacturing Execution Systems |
| mMTC | Massive Machine Type Communication |
| ML | Machine Learning |
| MLOps | Machine Learning Operations |
| MNC | Mobile Network Code |
| MNO | Mobile Network Operator |
| MSG | Managed Service Providers |
| N3-IWF | Non-3GPP Inter-Working Function |
| Near-RT RIC | Near-Real-Time RAN Intelligent Controller |
| Non-RT RIC | Non-Real-Time RAN Intelligent Controller |
| NEF | Network Exposure Function |
| NEST | Network Slice Template |
| NF | Network Function |
| NFV | Network Function Virtualization |
| NGAP | Next Generation Application Protocol |
| NGRAN | Next Generation RAN |
| NMS | Network Management System |
| NPN | Non-Public Network |
| NRF | NF Repository Function |
| NSA | Non-Stand-Alone |
| NSI | Network Slice Instance |
| NSMF | Network Slice Management Function |
| NSSF | Network Slice Selection Function |
| NSSMF | Network Slice Subnet Management Function |
| NWDAF | Network Data Analytics Function |
| OAM | Operations Administration and Management |
| OAS | OpenAPI Specification |
| OPc | Operator code |
| OPC-UA | Open Platform Communications Unified Architecture |
| OpEx | Operational Expenditure |
| O-RAN | Open Radio Access Network |
| OSS | Operations Support System |

| | |
|---|---|
| OT | Operational Technology |
| PCF | Policy and Charging Function |
| PDU | Packet Data Unit |
| PLC | Programmable Logic Controller |
| PLMN | Public LAN Mobile Network |
| PNF | Physical Network Function |
| QCI | QoS Class Identifier |
| QoS | Quality of Service |
| RAN | Radio Access Network |
| rApps | Radio Apps |
| RB | Resource Block |
| REST | Representational State Transfer |
| RIC | RAN Intelligent Controller |
| RNIB | Radio Network Information Base |
| RT | Real Time |
| RU | Radio Unit |
| SA | Standalone |
| SBA | Service-based architecture |
| SCADA | Supervisory Control and Data Acquisition |
| SCP | Service Communication Proxy |
| SD | Slice Differentiator |
| SDN | Software-Defined Networking |
| SLA | Service-Level Agreement |
| SMF | Session Management Function |
| SMO | Service Management and Orchestration |
| SoTA | State-of-the-art |
| SST | Slice Selection Table |
| SW | Software |
| TAI | Tracking Area Code |
| TCP | Transmission Control Protocol |
| TDD | Time Division Multiplexing |
| TN | Transport Network |
| TSN | Time Sensitive Networking |
| Tx.y | Task x.y |
| UC | User Control |
| UDM | Unified Data Management |
| UDR | Unified Data Repository |
| UE | User Equipment |
| UENIB | UE Network Information Base |
| UP | User Plane |
| UPF | User Plane Function |
| URLLC | Ultra-Reliable Low-Latency Communication |
| VLAN | Virtual Local Area Network |

| | |
|---|---|
| VNF | Virtual Network Function |
| vRAN | Virtual Radio Access Network |
| VSB | Vertical Service Blueprint |
| VSD | Vertical Service Descriptor |
| WP | Work Package |
| xApps | eXtended Apps |

# List of Figures



# List of Tables

# Executive Summary

Effective operation of 5G private networks requires a deep cellular expertise, hindering the fast adoption of this technology in the manufacturing sector. For this purpose, self-configuration, self-monitoring, self-healing, and self-optimization are essential to automate service provisioning and network operations. 5G networks are complex infrastructures featuring 5G radio heads, virtualized RAN and Core Network functions running on COTS servers, alongside management frameworks such as ETSI MANO that allow to manage virtualized network and application functions. Due to the complexity of this technology stack, tasks such as configuration, provisioning, and troubleshooting are performed by high-skilled technicians in state-of-the-art deployments; in contrast, the manufacturing sector would greatly benefit from the availability of configuration mechanisms that are simple enough to be applied by OT operators. In addition, the 5G private network stack generates a tremendous amount of information in terms of: i) base station and core network counters, alarms, traces and logs, ii) hypervisor and server compute related metrics, and iii) telemetry generated by transport network devices; such information can be leveraged by AI/ML algorithms to provide automated management, diagnosis and adaptation of 5G networks, services, and general operation.

The focus of T3.2 is to deliver mechanisms to simplify and enable the automated configuration and operation of 5G private networks and connectivity services, thus eliminating the need for deep technical expertise in the provisioning and configuration phases, while providing predictions (and corrective actions) of network congestion and suboptimal resource management. The outcomes of T3.2 will enable simple and fast provisioning, operation and diagnosis of 5G private networks. Industrial service demands will be met through data (industrial and 5G)-based and AI-driven zero-touch management mechanisms for proactively adapting the 5G NPN configuration and operation (slices, traffic flows and QoS profiles, computing resources). To this aim, the mechanisms under development need to closely coordinate with the open and SDN-based programmable controllers of the local edge industrial network, in order to guarantee end-to-end and near real-time reliable, scalable, and low latency service provisioning.

The Zero-SWARM 5G zero-touch network and service management components will be demonstrated in the project trials.

# 1 Introduction

## 1.1 Document purpose & relationship with other deliverables

This deliverable provides the initial technical contributions of consortium members to the task T3.2: CPSoS zero touch network and service management of 5G resources. This document presents the architectural building blocks for the functionalities under development in T3.2, along with the relevant technical developments achieved so far, whilst trying to highlight the main innovation takeaways and the added value of the proposed solution.

The relationship between D3.2 and other deliverables is as follows. Deliverables D2.1 (Definition & analysis of trials, KPIs & GDPR compliance) and D2.2 (Eco designed architecture, specifications & benchmarking) have been used as inputs to T3.2. Particularly, these have been employed to define the driving requirements for the design and implementation of the system for CPSoS zero touch network and service management of 5G resources. The outputs of T3.2 and D3.2 serve as inputs for (i) Task 6.1 (Integration guidelines & continuous integration), regarding the integration of the developed system into the trials of interest, and (ii) Task 6.3 (Validation and demonstration), regarding the evaluation of such trials.

In addition, T3.2 outcomes will be used as inputs for WP7 to disseminate and communicate the task findings in relevant venues and if proper even push them into the standardization bodies such as ETSI.

## 1.2 Structure of this document

This document is structured as follows. The remainder of Section 1 presents the context of the activity of T3.2, in terms of the importance of the topics addressed within the Zero-SWARM project and, more generally, in Industry 4.0 scenarios. In addition, 5G-related terminology of interest is introduced for non-familiarized readers. Section 2 provides a thorough overview of some of the challenges arising from the integration of 5G networks in the OT domain. Section 2 does not provide insights about T3.2 technical developments; instead, it introduces relevant technicalities for readers who wish to acquire additional 5G-related technical background. Section 3 and Section 4 present T3.2 technical contributions in the scope of zero-touch network and service management of 5G resources. Finally, Section 5 provides a clear summary of the main innovations and foreground technologies developed in T3.2.

## 1.3 Context

The 5G network infrastructure aims to overcome the limitations of previous cellular generations by addressing the challenges associated with managing, controlling, and orchestrating a fully software-based networks and services. Thus, the 5G network infrastructure is designed to accommodate end-to-end 5G-enabled vertical applications. This results in an infrastructure capable of supporting

widespread services, all while meeting the performance and commercial needs of various stakeholders.

5G can be thought of as a multi-service network capable of addressing the connectivity requirements of any application in the manufacturing domain [1]]. 5G networks enable developers and applications to request and configure services, such as network slicing, quality of service (QoS) settings, and resource allocation. In this regard, 3GPP has specified 5G networks to support capabilities such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable and low-latency communication (URLLC). Moreover, 5G networks are expected to provide very high levels of flexibility, enabling the cost-effective delivery of new services, thanks to enabling technologies such as softwarisation, virtualisation, network slicing, and edge computing capabilities. Through these advanced functionalities, 5G networks can be flexibly provisioned to cater for the specific needs of vertical industries, such as smart factories, thus enabling the deployment of vertical applications.

The deployment of vertical applications over 5G networks relies on certain auxiliary services provided by the network, which are consumed by the vertical applications. The term **service** in 5G can refer to specific network slices, QoS settings and network capabilities (such as eMBB, mMTC and mMTC), type of traffic (data, voice, SMS, or video) and also to the aforementioned auxiliary services, such as location management, identity management, resource allocation management, etc [2]].

Network Function Virtualization (NFV) in 5G enables the separation and distribution of essential network functions, such as directory services, file sharing, and IP configuration, across different environments. By decoupling these functions from their original physical machines, they can be bundled together and assigned to specific network environments. This virtualization of networks significantly reduces the reliance on physical components like switches, routers, servers, cables, and hubs, thereby enabling the creation of multiple independent networks. In this regard, 5G architectures leverage the deployment of distributed computing resources across the network infrastructure to host virtualized network functions (VNFs) both from vertical services and from the various network segments. Moreover, by adopting the use of VNFs, network slicing enables the creation of virtual network instances tailored to specific service requirements. Each slice can offer customized connectivity, performance characteristics, and security parameters. As such, dedicated network slices can be provided for different use cases; this is especially relevant in industrial environments, where a wide variety of use cases with very different QoS requirements exists.

In addition, 5G networks can integrate *edge computing* resources, enabling services that leverage computing resources at the network edge. In this manner, the use of edge computing enables additional capabilities, such as faster content delivery with low and ultra-low processing latencies, which are usually required in industrial automation-related use cases.

In the context of connected industries and automation applications, the aforementioned services and capabilities can be provided by both 5G non-public networks (NPNs) in stand-alone mode, which are the focus of this task, as well as by NPNs deployed and operated by mobile network operators (MNOs). Interested readers may refer to [3]] for further insights about the aforementioned deployment scenarios. In addition, NPN and PN hybrid models are also becoming quite feasible and attractive for the industries, and 3GPP also addresses it through the 5G PNI-NPN concept.

5G NPNs provide communication services between wireless devices, and between wireless devices and wired data networks. Industrial devices and NPNs may belong to multiple domains, such as the operational technology (OT) production domain and the information technology (IT) enterprise domain. Network management and configuration services required by industrial applications are made available by the 5G network provider, and these apply to all corresponding stages in the 5G network life cycle, such as network installation, initial configuration and deployment, and network decommissioning. The capabilities and services supported by 5G NPNs must be exposed to industrial applications, with the main goal of supporting the transmission of application data. Moreover, according to [1]], the implementation of the aforementioned 5G NPN capability exposure reference points should embrace a design philosophy based on the following aspects:

- **Usability and simplicity:** the levels of abstraction provided must be suitable for both IT and OT professionals, who might not have an in-depth knowledge of 5G systems and architectures. In this regard, zero-touch management mechanisms offer a high level of automation in the provisioning of 5G networks and applications, thus offering great advantages in terms of usability and simplicity.
- **Modularity and extensibility:** some reference point functions should be optional, while it must be possible to enrich available reference points with new functions, with full backward compatibility. Interactions among different components and layers of the 5G network architecture enable flexible network configuration, thus providing additional abilities, e.g. the deployment of network slices.
- **Service-based interfaces:** these expose network services to consumer applications, and thus should be implemented in a service-oriented manner, e.g. through the use of open representational state transfer (REST) application programming interfaces (APIs).

In this regard, the use of OpenAPI Specification (OAS) [4]] has recently gained much attraction within IT communities, as it provides a standardized framework for the definition, implementation and functionality description of RESTful APIs, thus enabling an easy transfer of knowledge from API providers (such as the 5G NPN) to API consumers (such as industrial applications and developers).

In the scope of T3.2, 5G network services and capabilities are offered for consumption to the vertical applications through RESTful APIs. Moreover, the principles of modularity and extensibility are also

built upon the use of RESTful APIs, which enable the necessary integration between upper orchestration and management layers with other building blocks of the 5G network stack. Therefore, the technical contributions presented in this deliverable embrace all of the above design principles, as well as the application of OAS principles. The application of such design principles to the building blocks of the T3.2 architecture is explained in more detail throughout the deliverable as the different blocks are introduced.

## 1.4 Scope of zero touch network and service management mechanisms

Below, the concept of zero touch management and service management of 5G resources is briefly presented for better comprehensibility. In subsequent sections, the challenges inherent to implementing such mechanisms in the OT domain are explained, and the technical contributions of T3.2 are presented.

Zero-touch management of 5G networks and services entails the automation and optimization of 5G network operations without the need for manual intervention or human interaction. Zero-touch management mechanisms usually leverage the use of technologies such as artificial intelligence (AI), machine learning (ML), and software-defined networking (SDN).

In the context of zero-touch management of 5G networks and services, a service refers to a specific capability or functionality provided over the 5G network infrastructure. As explained in the previous section, 5G services include mMTC, URLLC, eMBB, network slicing, and any other network capabilities such as identity management and resource allocation management. Each service within provided by the 5G network requires management and optimization to ensure its QoS, reliability, and an efficient use of network resources. Zero-touch management automates various aspects of service provisioning, monitoring, and optimization, allowing NPN providers to deliver and maintain high-performing services without the need for manual intervention.

Below, some key terms related to zero-touch management mechanisms of 5G networks and services are presented:

- **Automation:** zero-touch management relies on automation to handle various network operations, including configuration, provisioning, monitoring, troubleshooting, and optimization. Automated processes are designed to minimize human errors, reduce operational costs, and enhance network efficiency.
- **Orchestration:** the orchestration of applications, networks and resources plays a crucial role in zero-touch management. It involves the coordination and management of radio, network and computational resources, services, and functions, across distributed architectures, using software-based controllers. Orchestration mechanisms enable the dynamic provisioning,

scaling, and optimization of services in real-time, based on network and/or application demands and policies.

- **Self-healing and fault management:** zero-touch management sometimes includes self-healing capabilities that allow the network to automatically detect and resolve faults or issues. ML algorithms can be applied to identify or even anticipate network failures and initiate corrective actions without the need for human intervention. This reduces downtime and improves the overall reliability and availability of 5G services.

Zero-touch management mechanisms necessarily incorporate policy-based approaches, where predefined policies guide the automation and decision-making processes. Policies define rules and guidelines for service provisioning, resource allocation, security measures, and other aspects of network operations. Automated systems can enforce policies to maintain compliance and adhere to service-level agreements (SLAs). In this regard, AI and ML are also considered enabling technologies for the zero-touch management of 5G networks and services. AI- and ML-based techniques can be used to analyze large amounts of network data, and to gain insights into network performance, traffic patterns, and application behavior.

The application of AI- and ML-based techniques helps in enforcing established policies, predicting and detecting anomalies, optimizing available resources, and making intelligent decisions for proactive network management. In this manner, service QoS levels are maintained. It is noteworthy that the implementation of AI and ML mechanisms is empowered by the utilization of both cloud and edge computing resources. While edge resources provide lower transmission latencies but lower computational power and storage capacity, cloud resources provide higher transmission latencies but higher computational power and storage. Depending on the requirements of a specific AI or ML mechanism in terms of latency and processing needs as well as offline or online operation, this can be executed over the most suitable compute node across the distributed infrastructure.

All in all, by implementing zero-touch management of 5G networks and services, several benefits are achieved, including improved operational efficiency, reduced costs, faster service deployment, enhanced network reliability, and superior application performance. Moreover, zero-touch management mechanisms reduce complexity from the perspective of the end user. In the context of smart industry settings, this addresses the design requirement related to usability and simplicity, providing suitable levels of abstraction for both IT and OT professionals by eliminating the need to manually perform complex technical tasks such as network provisioning and troubleshooting.

Moreover, zero-touch management mechanisms take advantage of the design principles related to modularity and extensibility. A specific mechanism normally has to interact with one or very few blocks of the 5G network stack. However, by applying modularity and extensibility principles, much complexity is eliminated during the design, the implementation, and the integration of zero-touch

management mechanisms into the network. In addition, such integration is further simplified by the use of service-based interfaces and the use of OAS, through the implementation of standardized RESTful APIs in all components to be integrated.

Last but not least, zero-touch management mechanisms can be applied to the concept of network slicing, by dynamically provisioning, allocating and managing network slices according to specific use case requirements. The application of zero-touch management mechanisms ensures an efficient utilization of resources, and enables the rapid deployment of applications with specific QoS requirements.

## 1.5 Positioning within the Zero-SWARM reference architecture

Below, the aim and scope of T3.2 is introduced within the scope of the overall Zero-SWARM reference architecture. Specifically, Figure 1 depicts the mapping of the T3.2 sub-system to the Zero-SWARM high level architecture.

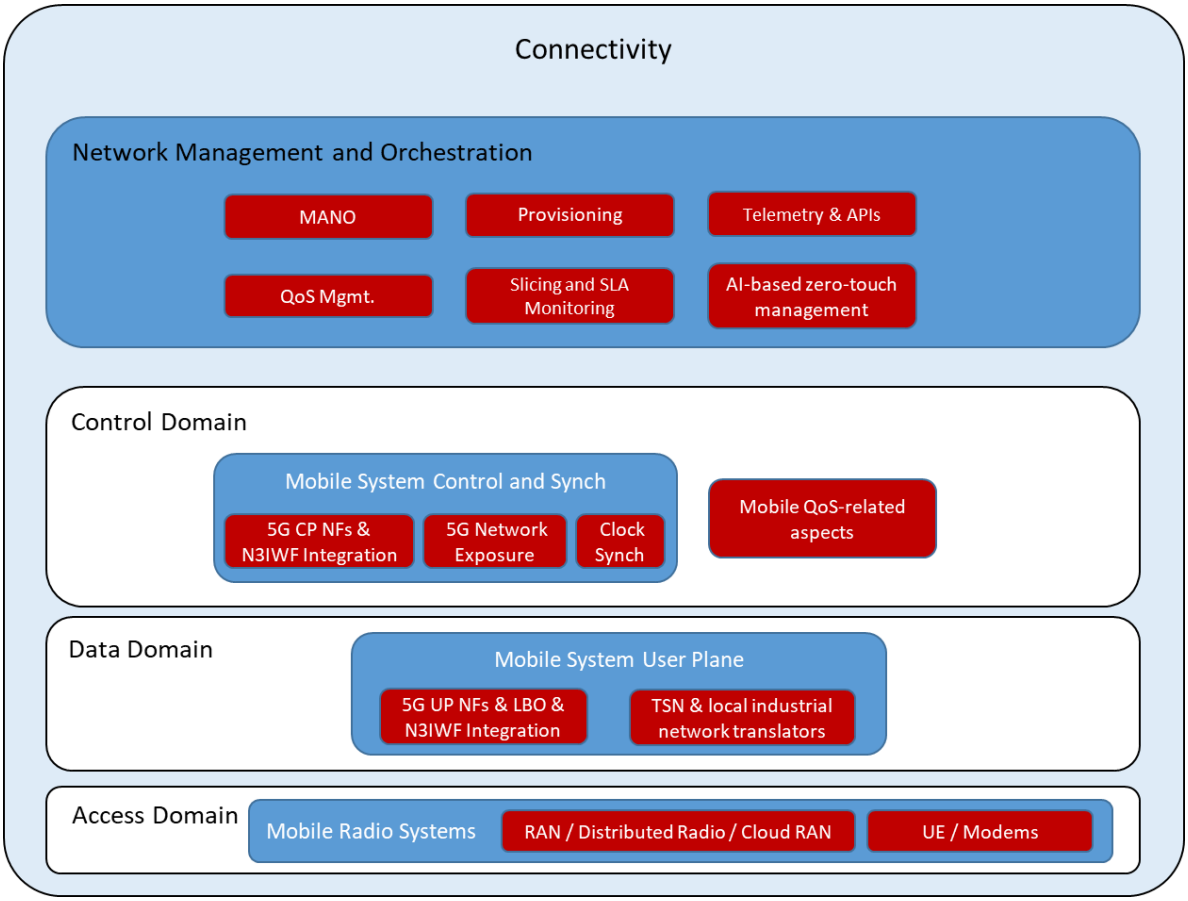


*Figure 1: Mapping of T3.2 sub-system to Zero-SWARM high level architecture*

The building blocks of the T3.2 sub-system fall largely within the *network management and orchestration* layer, as these mainly perform actions related to the network and service management as well as orchestration of 5G resources and services, such as the provisioning of the 5G network, deployment of network slices, application of QoS mechanisms, and so on. Nevertheless, the network

management and orchestration components necessarily interact with other components across the various layers of the 5G network stack in the access, data and control domains.

Figure 2 represents a more holistic view of the T3.2 sub-system within the Zero-SWARM project, where the relationship between the T3.2 sub-system with other architectural blocks, both within the IT and the OT domains, is also represented. While it is out of the scope of this task, the necessary integrations between the IT and OT domains must be considered for the design and implementation of the 5G network. This must also include its interactions with devices in the OT domain, and the subsequent repercussion in the design, implementation and integration of the zero-touch network and service management mechanisms.

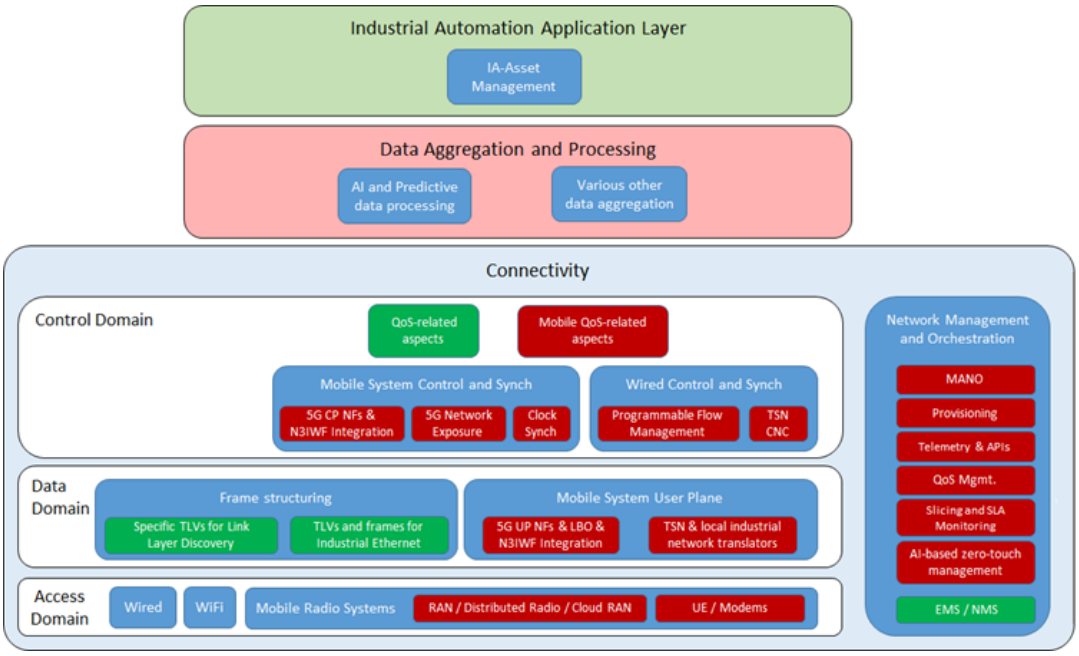


*Figure 2: Relationship of the T3.2 sub-system with other Zero-SWARM architectural blocks*

For example, it is critical to consider QoS-related aspects of the industrial applications, as these must be catered for during the provisioning of 5G network resources. In addition, OT data to be transmitted needs to be converted to a suitable protocol for transmission over the 5G network, and therefore it is critical to consider, for example, the frame structures of the data transmitted through the industrial Ethernet. Some interactions are also expected with the data aggregation and processing layer, particularly in use cases involving the transmission of IIoT data through the 5G network for aggregation and processing purposes. The industrial automation application layer is also of interest, as the requirements of the 5G network and service provisioning are determined by the vertical applications to be deployed on top. It is also envisioned that T3.2 sub-system components might be integrated within the OT domain's network management system (NMS) at a later development stage.

# 2 Challenges and requirements in the introduction of 5G networks in OT processes

The underlying infrastructure and technical architecture that enables wireless communication using the 5G technology standard is referred to as a 5G network. It includes the physical infrastructure, such as base stations, antennas, servers, and backhaul connections, as well as the software network components, such as the 5G Core (5GC) and radio access network (RAN). The 5G network provides the infrastructure required for various 5G services, such as high-speed data transfer, low-latency communication, and massive connectivity. It is in charge of sending and receiving data from and to devices in the 5G ecosystem, allowing for seamless connectivity and communication. A 5G service, in particular, refers to the specific applications, capabilities, and functionalities that are enabled by leveraging the capabilities of the 5G network. Examples of 5G services include enhanced mobile broadband (eMBB) for high-speed internet access, ultra-reliable low-latency communication (URLLC) for mission-critical applications, and massive machine-type communication (mMTC) for IoT deployments.

In the OT domain, a 5G network is the physical infrastructure composed of hardware and software components, and communication architecture that provides wireless connectivity utilizing the 5G technology standards. It includes the installation of 5G base stations, antennas, and other network components in order to offer predictable coverage and reliable connectivity within the OT environment. It addresses the unique needs of OT systems such as industrial control systems, critical infrastructure, and manufacturing processes, and provides dependable and secure wireless communication. A 5G service in the OT domain, in particular, refers to the specific applications, capabilities, and functionalities enabled by leveraging the 5G network. More specifically, the underlying 5G offers seamless connectivity that facilitates the necessary communications of OT services. Some examples of 5G-enabled services in the OT domain include remote monitoring and control of industrial processes, real-time asset tracking, predictive maintenance using machine learning algorithms, or augmented reality-based maintenance and training applications. These OT services leverage the features or services provided by the 5G network, such as high reliability, low latency, high bandwidth, and network slicing, to enable efficient and secure communication within the OT infrastructure.

One of the requirements defined in D2.1 is the need for full control and management skills on non-5G end devices existing on the shop floor in the industrial environment. In particular, the variety of end devices, i.e., 5G integrated devices, non-5G (WiFi, ethernet), etc., on the shop floor requires access, continuous monitoring and maintenance to enable smooth and cost-effective operations. OT

applications such as Automated guided vehicles (AGVs) on the factory shop floor, which we will describe in detail in Section 2.1, require 5G services like low latencies, reliable communication, enhanced mobile broadband, continuous monitoring of the location, and maintenance to avoid collisions within the factory shop floor. For this reason, first, in Section 2.2, we depict an ideal 5G model and its interconnection with OT infrastructure, with particular emphasis on standalone NPN deployments. Section 2.3 introduces general aspects related to the end-to-end service provisioning of 5G NPN, including the network management and orchestration layer, and how these can be used to enable seamless, reliable service delivery with the QoS demanded by industrial applications.

## 2.1 Typical OT processes and underlying infrastructure in the manufacturing sector

OT refers to the hardware and software systems that are used to monitor and control the physical processes and devices in industries, including manufacturing, energy, transportation, utilities, and so on. Industrial processes frequently involve the interplay of physical components and computerized systems, and OT solutions are created to control and optimize these operations. These systems may use a variety of technologies, including programmable logic controllers (PLCs), supervisory control and data acquisition (SCADA) systems, industrial control systems (ICS), and other specialized hardware and software. They play a crucial role in industries by monitoring and controlling processes, collecting and analyzing data, and facilitating automation and optimization.

The OT use cases can be classified based on various factors, one factor could be the real-time (RT) requirement, which is a key feature in the 5G domain. In the OT sector, RT refers to the requirement for rapid or near-instantaneous response and processing of data or events. Specifically, it refers to the permitted latency for a specific OT application scenario. Based on this, we present some of the OT use cases of interest in Zero-SWARM below:

1. Connectivity, automation, and flexibility for the factory floor (Hard RT, latency required typically within milliseconds or microseconds).
2. Integration of wired and wireless components for motion control (Hard RT).
3. Mobile robots and automated guided vehicles (AGVs) (Soft RT, latency typically one second).
4. Process automation monitoring (non-RT, latency required typically several seconds).

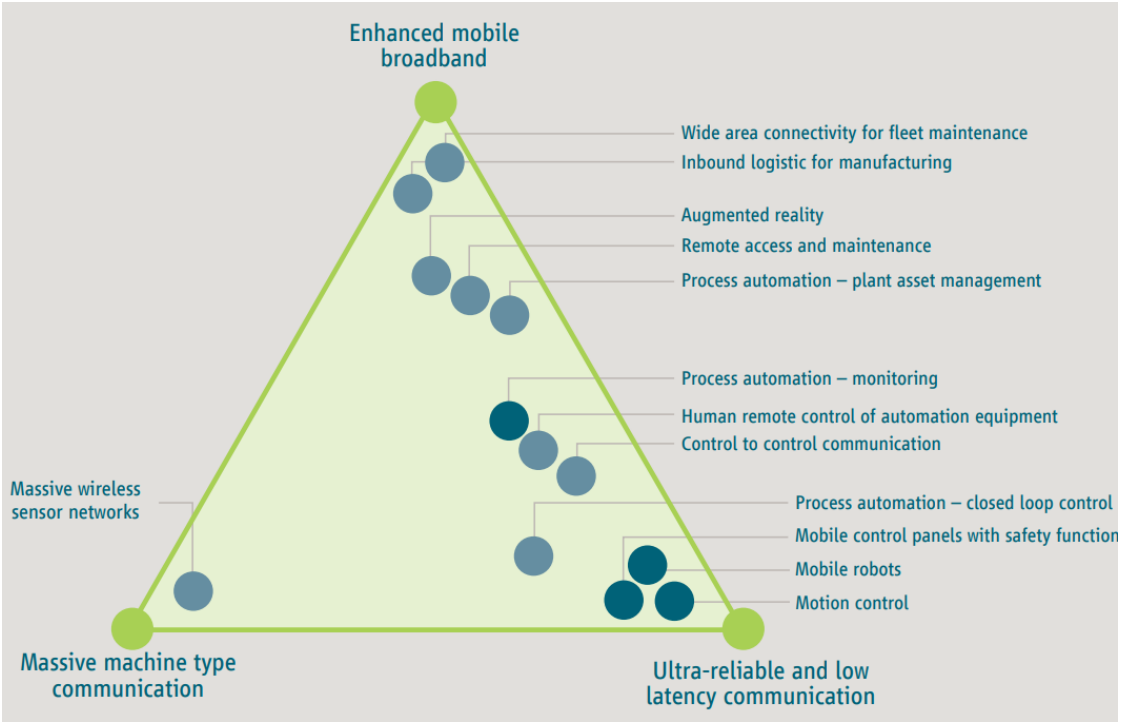


*Figure 3: Industrial Use Cases with their Service Requirements*

**Error! Reference source not found.** depicts additional examples of how the benefits of 5G can be e xploited in the OT domain. Such OT use cases have specific 5G communication requirements that are essential for their successful implementation and operation, such as quality of service (QoS), reliability and availability, and brownfield support. In the OT domain, it is desirable to take advantage of 5G self-configuration and self-management mechanisms, such that the aforementioned requirements are fulfilled seamlessly, without no need to perform complex network and radio configurations according to the specific requirements of each use case.

A specific application of interest in Zero-SWARM, is automated good transport with AGVs in the manufacturing sector. Here, an important task is to create transparency about neuralgic areas of shop floors in order to enable more efficient processes. By combining data from sensors in the infrastructure as well as from the moving AGVs, the situation in a neuralgic area can be interpreted in more detail. Based on this data, actions such as slowing down the AGV, can be deduced. The situation on shop floors with moving objects changes quickly and constantly and, the danger of collisions represents a safety risk. Thus, low latencies and availability for the transmission and processing of data is key. In the neuralgic area represented in Figure 4, data from one infrastructure sensor and one AGV each will be transmitted via 5G, in order to enable such stringent requirements. In this regard, it is necessary to enable (i) simple registration of 5G-enabled AGVs in the network, and (ii) intent-based mechanisms to seamless configure the network to provide a suitable QoS with low latency.

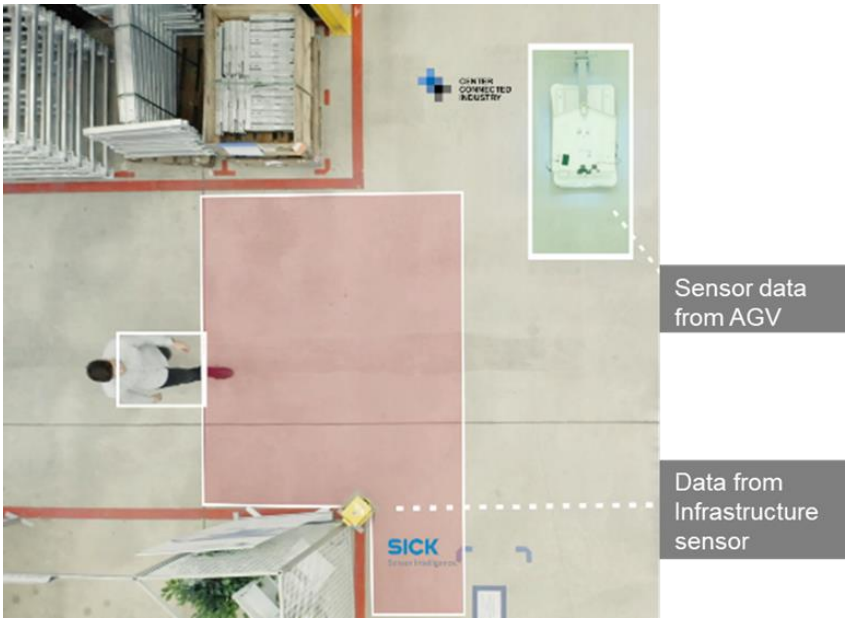


*Figure 4: AGV use case neuralgic area*

In general, a key requirement in OT use cases is the need of a complete concept of network planning, integration (because we still need cables), coverage (indoor and outdoor), guided and supported by automated processes for the setup and administration. Without these, 5G will have difficulties to handle a ramp up in industry.

## 2.2 Ideal 5G NPN and underlying infrastructure to be integrated into OT domain in the manufacturing sector

Industrial OT domains include a large variety of heterogeneous devices and software with stringent communication protocols and complex topology. A 5G NPN solution for manufacturing sector should consider many aspects such as QoS requirements and inter-connectivity.

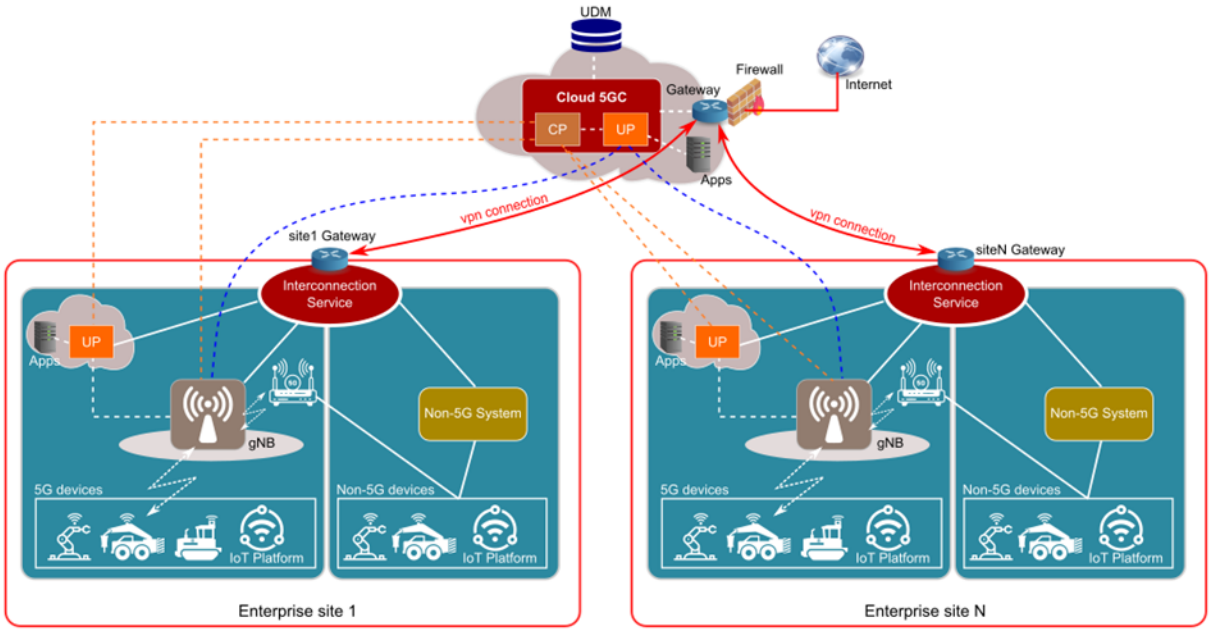


*Figure 5: Ideal 5G standalone NPN Solution for manufacturing sector*

Furthermore, factory floor operators require full control of underlying network and high security levels to protect sensitive data. Considering all these requirements, 5G NPN solution should be a standalone NPN. Figure 5 illustrates a deployment proposal of a standalone NPN. In the following, the main 5G technology components and their configuration requirements are discussed separately.

**Radio Access Network (RAN)**

Some countries provide already spectrum dedicated for industrial use, thus paving the way for the deployment of standalone NPNs. 5G network configuration mechanisms must enable the simple configuration of the radio communication channel. However, RAN system deployment has many other challenges in an industrial environment.

The high number of connected devices leads to the installation of several Radio Units (RUs) in the OT domain causing high interference. This is of particular interest in steel plants, where there might be interferences that can actually be easily overlooked during the radio planning phase. Thus, the implementation phase on the customer's side must ensure a stand-alone system administrated by the customer with an easy-to-use end-to-end network administration tool.

Furthermore, in order to have an open, flexible, scalable and manageable RAN system, the open RAN (O-RAN) concept might be opted in the proposed standalone NPN solution. There are telecom vendor closed solutions available but O-RAN comes with extra openness. This might be considered positive or negative, from one side it brings more control but from the other side it requires extra knowledge and efforts and full responsibility of handling all RAN related situations (scalable after sales customer support). Depends on the customer informed choice O-RAN might be a solution to deploy. In case of having O-RAN, the central unit (CU) component of O-RAN should be deployed on-premise cloud, whilst the distributed unit (DU) must be deployed close to RUs to ensure low latency requirement, as per Figure 5. The selection of the proper user plane (UP) entity depends on the user profile and it is based on the 5G slicing concept. The UP entity on-premise is used by users requiring low latency, high throughput and sending sensitive data. However, the placement of the UP requires expert knowledge about 5G communication protocols and complex configuration aspects. By providing intent-based configuration mechanisms, an OT operator can specify QoS requirements in a more familiar language, which automatically translate into a set of configuration-related parameters and an automated configuration process where the UP can be configured as needed. For more details about UPF and the deployment options please check D3.1.

**Core Network (CN)**

In D3.1, we have discussed mainly telecom vendor's professional 5GC solutions coming on specialized and dedicated hardware. Here we focus on the open-source solutions as an alternative case. Please

note that the raw open-source solutions might not be always 100% ready as off the shelf solutions with full features ready to deploy. It is due to the nature of open-source communities that develop solutions on demand and on best effort basis. It is evident that a tailored solution with a good after sales support might be a good case if there is an absolute need to build up a NPN network out of Open source 5GC (as an analogy we can think of a Red Hat like company built on top of the Linux open source). Nevertheless, since D3.1 covers mainly telecom vendor's perspective, here we put some light on open source 5GC which will be used in the Zero-SWARM trials.

Since the control plane functions of 5GC can potentially be deployed on private or public cloud environment, it is necessary that the software architecture should consider at least one of the following use cases: (i) container-based deployments, where the 5GC functions are deployed as containers, and a container orchestrator such as Kubernetes is required; (ii) virtual machine (VM)-based deployments, where the 5GC functions run on virtual machines; some vendors allow the capability to run the CP and UP functions separately, in different VMs. In the Zero-SWARM 5G standalone NPN stack used in one of the project trials, a solution based in VMs is adopted.

There are many commercial 5GC software available in the market supporting the use cases of interest. There are also open-source software solutions that have shown enormous flexibility and stability, Open5GS being one of those. It is a C language-based 5G core that complies with 3GPP Release 16 and complies with the GNU AGPLv3.0 license, and the commercial license is owned by NextEPC. It is used by a large community and therefore has a rich documentation describing many use cases. The Open5GS 5G SA Core contains the following functions:

- NRF - NF Repository Function
- SCP - Service Communication Proxy
- AMF - Access and Mobility Management Function
- SMF - Session Management Function
- UPF - User Plane Function
- AUSF - Authentication Server Function
- UDM - Unified Data Management
- UDR - Unified Data Repository
- PCF - Policy and Charging Function
- NSSF - Network Slice Selection Function
- BSF - Binding Support Function

Interactions with the 5GC functions are exploited in the AI-driven zero-touch mechanisms presented in Section 4 e.g., for the automated configuration of network slices according to specific requirements.

In addition to the functions listed above, other 3rd party components (such as N3IWF) from other open-source initiatives can also be incorporated into Open5GS architecture. The N3IWF (Non-3GPP Inter-Working Function) is an important component which should be deployed to connect Non-3GPP RANs (e.g WiFi) to the 5G core, to provide compatibility existing communication solutions. Co-existence of 5G and WiFi wireless communication technologies brings some new requirements such as location-based traffic prioritization on certain wireless communication technology. In this regard, Section 3 presents an initial concept for a management solution that will enable provisioning of non-3GPP connectivity services.

### IIoT Platform

Industrial IoT (IIoT) platforms, as one integral component of broader systems such as Manufacturing Execution Systems (MES), are employed in manufacturing industry in order to increase the decision-making abilities in the manufacturing processes. Such platforms gain insights on what is going on in the shop floor (production quality, speed, downtime, stocks and waste) by leveraging the data coming from the shop floor through i.e. 5G NPN to be able to instantly decide on most effective operational actions in MES systems. IIoT platforms require the almost uninterrupted use of compute, storage, and communication systems in the on-site IT. In this regard, Section 3 introduces the Zero-SWARM concept of active reconciliation loops that can detect and revert configuration drifts, thus providing a necessary mechanism for the high availability of the 5G network required in IIoT platforms.

### Edge Computing

Edge computing, that brings the data computation work as close as possible to the data source in order to improve the speed of data operations, is an emerging technology by the efforts of various industry players and several standardization organizations. In this regard, Sick will contribute to Zero-SWARM with the development of 5G capable edge computing device. The main motivation is to increase the speed and accuracy of industrial decision-making systems by bringing data source and various data operations as close to each other as possible. Thus, integration with a direct access down to the sensor over the edge devices through 5G is needed. To achieve this, on-device SIM cards should be managed easily and virtually by OT operators. Such capability should be provided in the management system of the Zero-SWARM 5G NPN stack. After a 5G-enabled CPE is on-boarded to the 5G network, it can be used, for example, to deploy applications with low-latency requirements.

### Baseline model of 5G NPN and OT domain interconnection

The 5G SNPN solution enables communication between the OT and IT networks, in particular, between the wireless devices as well as between wireless devices and wired data networks. The major function

of exposure interfaces is to control the 5G standalone NPN (5G SNPN)'s user plane, which supports the transmission of industrial application data. A typical industrial environment includes both 3GPP and non-3GPP network technologies e.g., industrial Ethernet or WiFi-based networks. Therefore, the 5G SNPN solution should coexist with non-3GPP networks and requires integration with a non-5G OT network.

Furthermore, non-3GPP devices can alternatively access the 5G technology by connecting through 5G CPEs. Zero-touch solutions for the co-existence of 3GPP and non-3GPP technologies are proposed in Section 3.

Network slicing is another key feature of 5G technology that allows the virtual partitioning of a single physical network into multiple logical networks, which are commonly known as slices. Network slicing allows for the allocation of dedicated network resources, such as bandwidth, latency, and quality of service, to each slice, assuring optimal performance and service delivery to prearranged service-level agreement (SLA) of various industrial applications. AI-based zero-touch mechanisms for the slicing of the RAN and the core of the 5G network resources are presented in Section 4.

## 2.3 End-to-end service provisioning of Zero-SWARM 5G NPN

In this section, we discuss the management aspect of the integrated 5G SNPN solution with OT and IT infrastructure, which is shown in Figure 6.

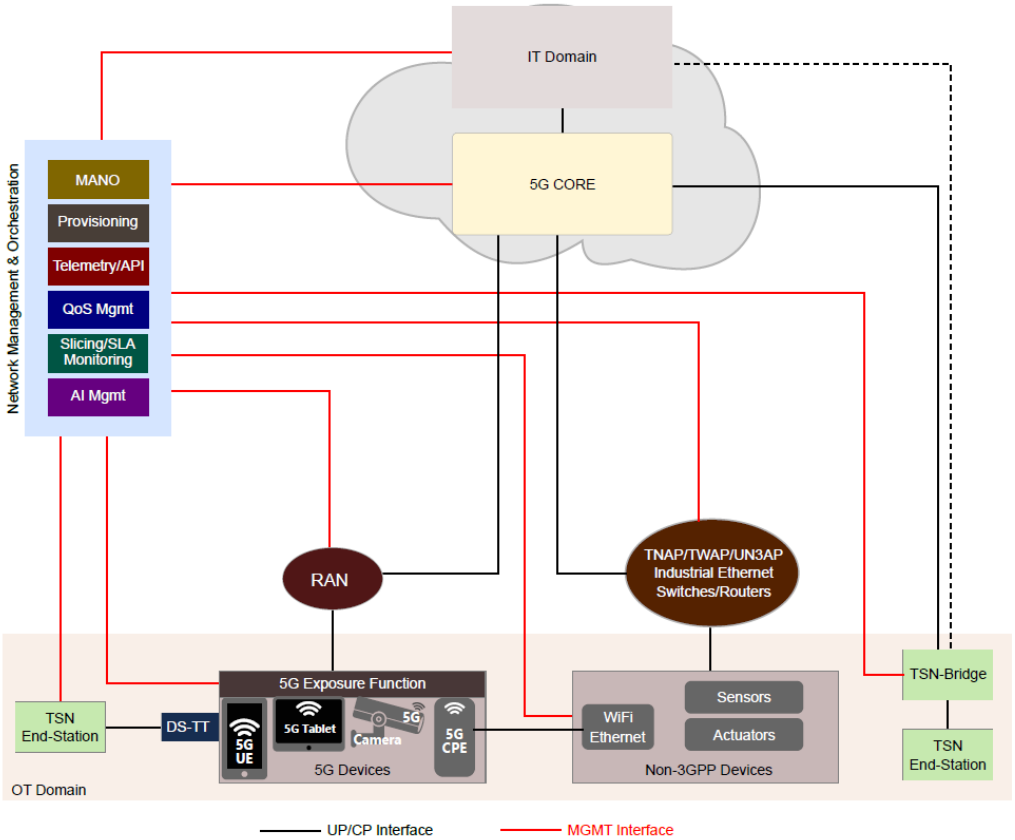


*Figure 6: Management Aspect of 5G NPN and OT Integration*

In order to achieve end-to-end (E2E) service provisioning of 5G resources, the Network Management and Orchestration layer has been placed under Connectivity layer of the Zero-SWARM architecture.

**MANO**

The European Telecommunications Standards Institute (ETSI) Network Function Virtualization (NFV) Management and Orchestration (MANO) framework is a key enabler for the deployment, management, and orchestration of 5G NPN virtual infrastructure (compute, storage, networking), helping to virtualize all the various appliances and services in the network, including security, from the RAN to the 5GC and beyond, and empowering services such as network slicing. This is a must-have to support many applications and use cases, especially as far as low latency and bandwidth efficiency are concerned. NFV MANO is responsible for interacting with operations and business support systems (OSS/BSS) to deliver benefits such as rapid service innovation, flexible network function deployment, improved resource usage, and reduced CapEx and OpEx costs. From the perspective of OT use cases, open source NFV MANO tools such as OSM might provide sufficient lifecycle management of VNF and NS instances to fulfill service requirements in a transparent manner.

**Provisioning**

Provisioning is an important lifecycle phase in providing a 5G NPN network service or a network slice, because a network service provider must map service requirements into commissionable 5G NPN deployment scripts or typical value of generic network slice template (GST) attributes and to deploy resources and network functions correctly in order to satisfy the required SLAs. Thus, it is key to provide mechanisms for easiness of deployment, to reduce the vector for human error upon deployment. When operational, each 5G NPN network service or network slice must be monitored to ensure that its service quality meets the SLA and that there is a mechanism in place for when it fails to do so. In Sections 3 and 4, Zero-SWARM's mechanisms for the provisioning and self-monitoring of network services and slices are presented.

**Telemetry & APIs**

The Network Exposure Function's (NEF's) Northbound APIs (represented in Figure 7) enable the 5G NPN to be more accessible, controllable and programmable. In order to effectively expose the 5G services, flexibility and compliance between NEF and the other 5GC NFs is essential, since NEF interacts with many of them (i.e., through southbound interfaces) that the Service Based Architecture realizes. An example is the Network Data Analytics Function (NWDAF), providing network analytics to assist an external application to make efficient decisions. This is exploited in Section 4 for the development of Zero-SWARM's zero-touch management mechanisms in the context of network slicing.

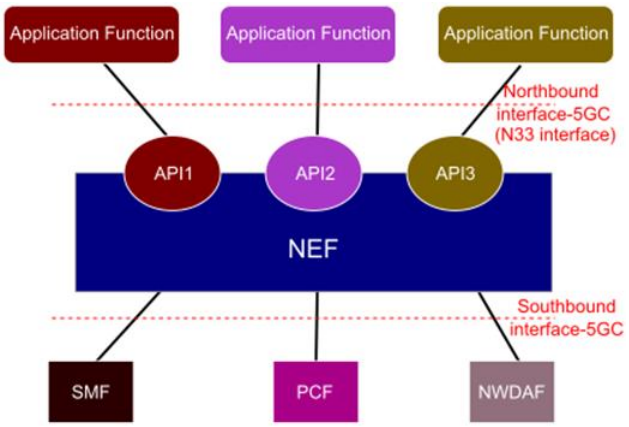


*Figure 7: Network Exposure Function*

### QoS Management

QoS management of 5G NPN is an inevitable need to meet the strict service requirements in the manufacturing industry, such as latency and availability. The traditional failure, service disruption or quality degradation mechanisms are ineffective or inapplicable in NFV environments. Section 3 presents the concept of Zero-SWARM's active reconciliation loop that is established between the Zero-SWARM management system and the 5G infrastructure, capable of detecting configuration drifts in the 5G infrastructure.

### Slicing and SLA Monitoring

The core concept of network slicing enables the provisioning of virtualized but mutually isolated customized networks over a shared virtualized infrastructure platform. This is illustrated by the example given in the following figure where three network slice instances owned by different service verticals, also referred to as tenants, are deployed over a shared virtualized infrastructure. As represented in Figure 8, each slice instance is composed of virtual network functions (VNFs) interconnected by virtual links to deliver a specific network service, including virtual 5GC, virtual radio access network (vRAN), and so on. Infrastructure resources (e.g., compute, network, storage) are automatically abstracted and dynamically allocated to each slice. The network services offered by the slices are consumed by higher-level (application) service instances pertaining to specific verticals.

A credible network slice MANO system is required to manage the infrastructure resources, network slices, and service instance(s) to ensure reliable service delivery within the quality of service (QoS) boundaries and SLAs. This is the focus of Section 4, where Zero-SWARM's mechanisms for the management of network slices are presented.

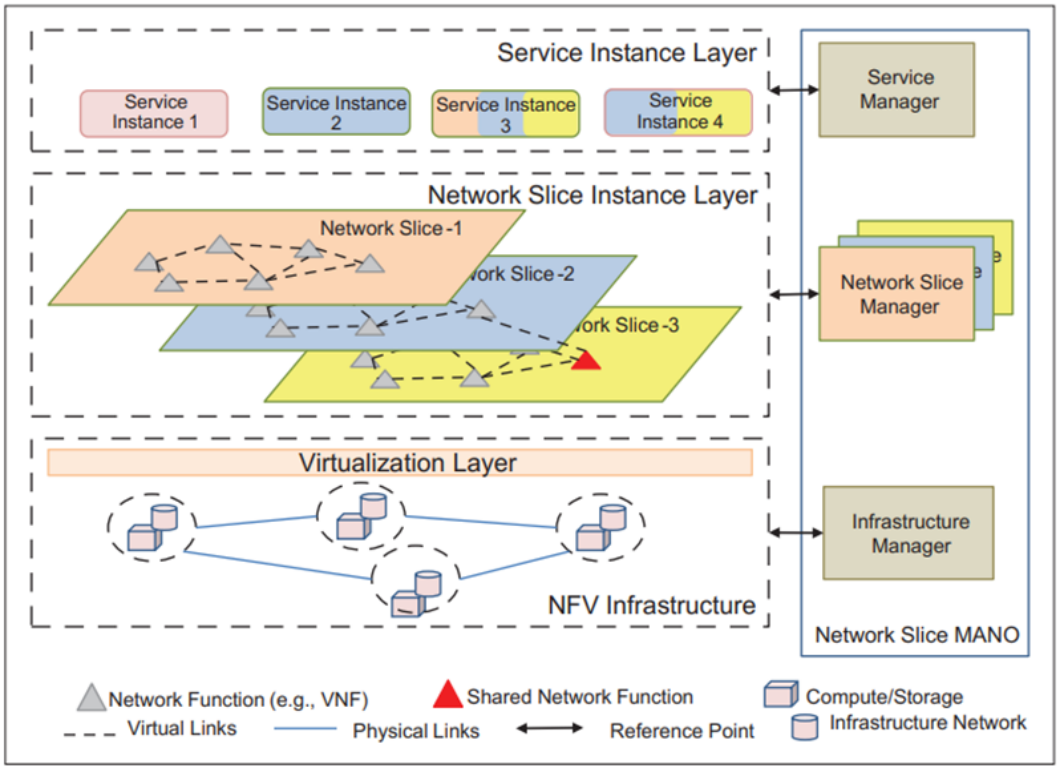


*Figure 8: Network slicing concept*

**AI Based Zero-Touch Network Management**

Besides the fact that the 5G network exposes its capabilities through the NEF, sometimes the applications can assist the network by providing application related information. The 3GPP core network offers several services that are particularly useful to UEs hosting applications. N33 interface exposes APIs that enable applications to update parameters that are already configured in 5GC. Applications may consume these APIs and provoke the 5GC to take specific actions (however, it is worth to note that all UE requests could not be necessarily answered. 5GC needs to check conflictive situations where the request of one or a group of UE doesn't jeopardize the performance of other connected UEs / running services). In addition, necessary features can be configured based on parameters that are statically created in the UEs initial subscription to the 5G network. Although this is not directly addressed in T3.2, Zero-SWARM's zero-touch network management mechanisms, if extended, may serve as enablers for such use cases.

# 3 Zero-SWARM 5G zero-touch network and service management subsystem

This Section presents the developments of T3.2 in terms of 5G NPN configuration, zero-touch connectivity service provisioning, and zero-touch network operation. Aspects related to the zero-touch diagnosis of 5G NPN are introduced as well.

Attention is also paid to the zero-touch configuration, deployment, monitoring and recovery of O-RAN architectures, as a means to provide a flexible, scalable and manageable RAN system in the context of industrial applications.

## 3.1 5G NPN Reference Management Architecture

Figure 9 represents a break-down of Zero-SWARM's reference management architecture of the 5G NPN which, at a very top level, includes a network service provisioning subsystem, a network telemetry subsystem, and an AI engine subsystem for life-cycle management.

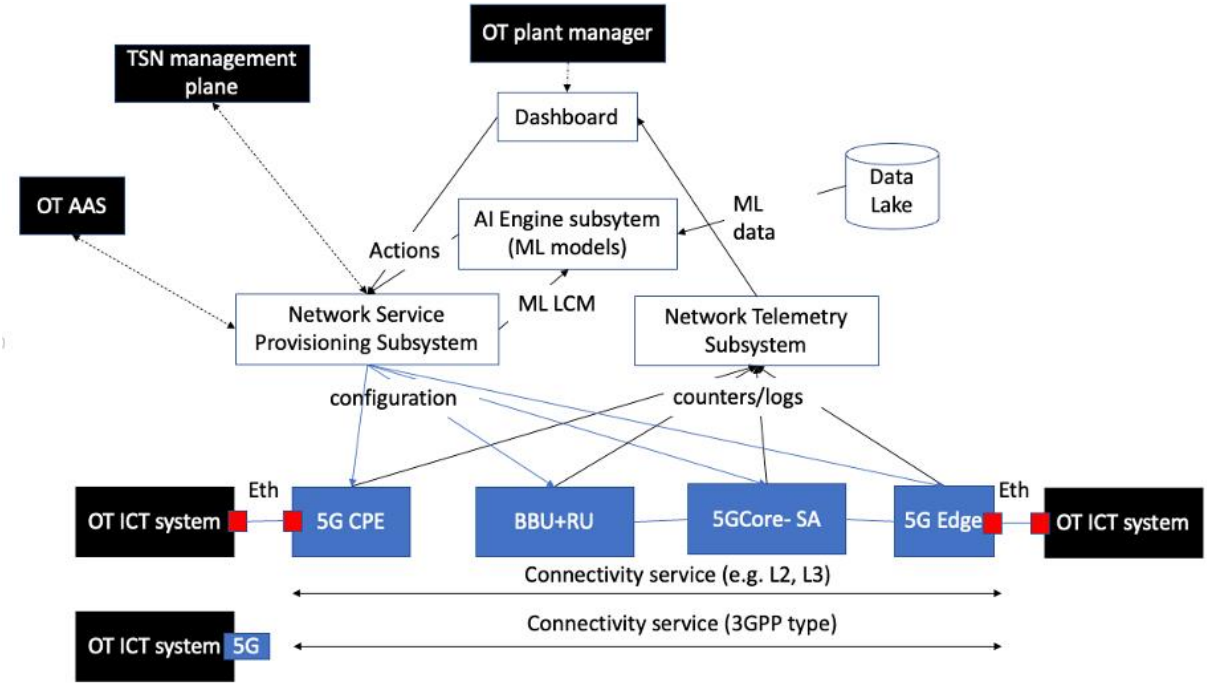


*Figure 9: Reference Management Architecture*

The network service provisioning subsystem should enable capabilities related to the configuration of the various components of the 5G NPN stack, namely the baseband unit (BBU), radio unit (RU), standalone (SA) 5G core, 5G edge computational resources, and CPE. A network telemetry subsystem should be in place to gather metrics from all the segments across the 5G NPN stack. A centralised data lake may be needed to facilitate the processing of data that cannot be analysed where it is generated due to limitations in terms of computational power. The AI engine subsystem should provide AI and ML mechanisms to take decisions based on relevant KPIs, and it should interact with the network service provisioning subsystem in order to execute such decisions.

Potential integrations with other components of the Zero-SWARM network architecture (e.g., OT management systems) are not directly addressed in D3.2. However, all developments in T3.2 follow an Open API approach, to enable and facilitate their latter integration into an overall system architecture, thus the project trials.

## 3.2 Provisioning of 5G NPN

The provisioning of 5G NPN is a fundamental phase of the network deployment process. This involves not only the configuration of the different network components to be able to work together, but also the configuration of parameters which may play a key role in the performance of the network, which need to be adapted to specific requirements.

In state-of-the-art deployments, 5G NPN network configurations are performed manually. This manual process results in lengthy deployment times, due to the large number of configurations to be performed. In addition, the manual nature of this process often yields configuration errors that could have a severe impact over the network performance. Moreover, the provisioning of 5G NPN networks requires a deep technical expertise, thus hindering the uptake of the technology in organisations lacking such expertise. For these reasons, one of the key objectives of this task is to provide strategies to facilitate the network provisioning with the final objective of obtaining a zero-touch approach, thus providing a level of abstraction that is suitable for both IT and OT professionals.

Figure 10 exhibits the general Zero-SWARM 5G NPN deployment. Below, some important configuration aspects of both the 5G core and the radio are introduced.

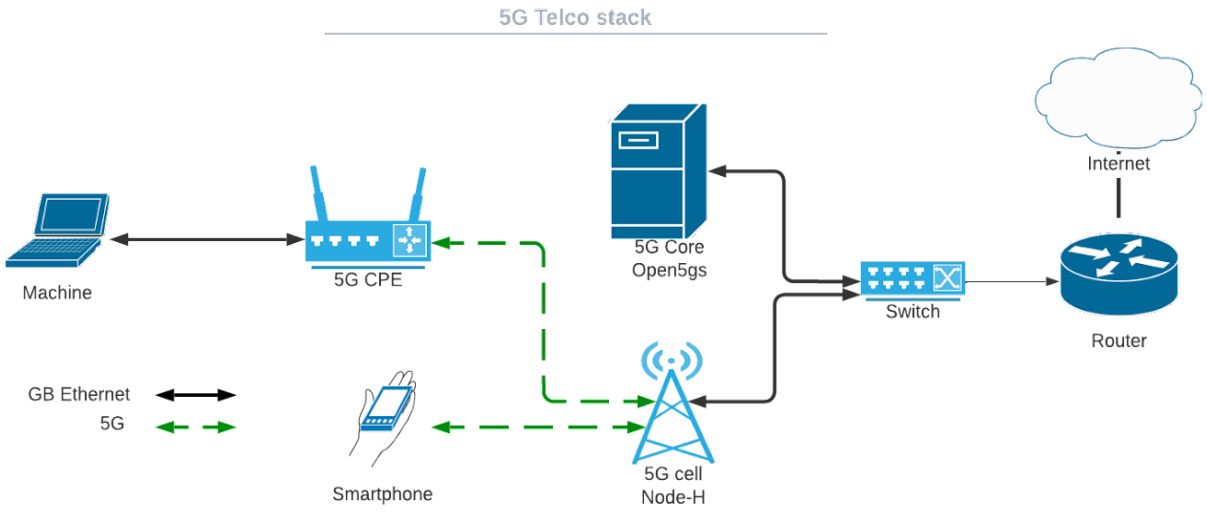


*Figure 10: Basic 5G scheme*

As it depicted in Figure 10, the 5G NPN stack contains a 5G core and a gNB which is connected to the core through a switch. In the represented diagram, the 5G core is implemented using Open5GS [9], and the gNB uses a Node-H cell. Note, however, that other open source solutions exist and/or vendors that provide similar alternatives. The switch has access to internet/data network by using a border router. User equipment in the 5G network can either connect directly to the 5G network (like the smartphone in Figure 10) or through and CPE to relay traffic from other network technologies into 5G

network. Devices with no 5G capabilities, such as legacy robots and factory machines, can use the CPE, industrial routers, or 5G M.2 modules to connect other devices and/or servers in the 5G network. When the provisioning phase is executed manually, network failures due to wrong configurations are likely to occur. The process of troubleshooting these failures to set up the connectivity is a lengthy process that needs to be carried out by highly qualified engineers. Figure 11 shows an example of a typical troubleshooting procedure when the machine on Figure 10 does not have 5G connectivity. As represented in Figure 11, the fault-finding process can be very time consuming, even for a technician with 5G-related expertise. From the OT perspective, it is crucial to minimize this procedure and ideally to provide a transparent IT solution.

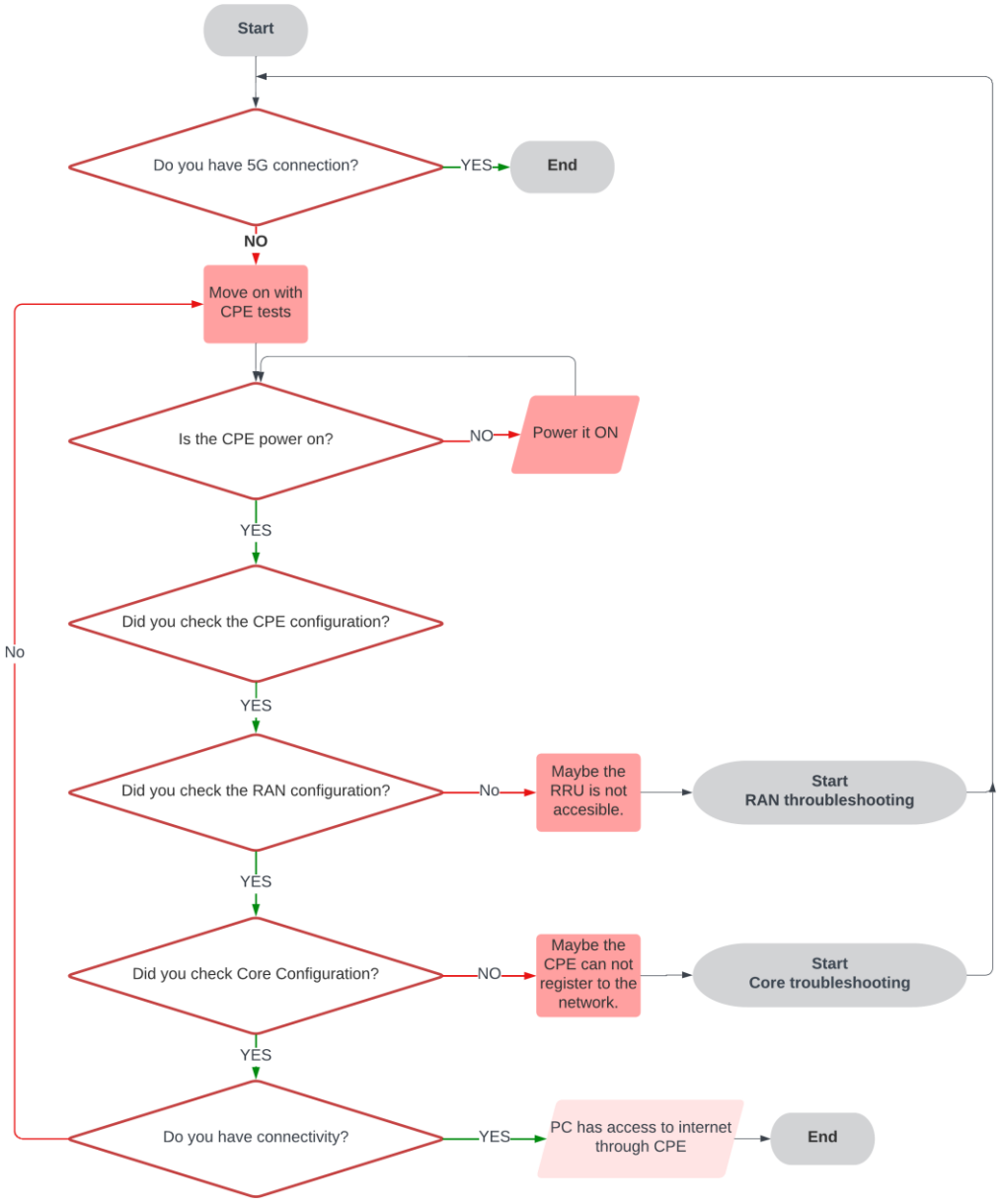


*Figure 11: Example of the complex troubleshooting process*

The remainder of Section 3.2 describes the main aspects of the configuration of the RAN and the core network of the considered 5G stack. The manual configuration of the 5G RAN and core demands a planning phase beforehand to determine the proper configuration; such aspects are explained in Section 3.2.1 and Section 3.2.2. In Zero-SWARM, the configuration of these elements is automated through APIs to facilitate interconnection and, potentially, automation. The final aim to improve the deployment procedure, for example by through the development of a user interface.

### 3.2.1 5G Core Configuration Aspects

The 5G core is responsible for providing advanced networking capabilities and efficient service delivery. The 5G core architecture is designed to be highly flexible, scalable, and capable of handling diverse requirements. At the heart of the 5G core architecture lies the Service-Based Architecture (SBA), which is based on a modular and distributed approach. The SBA allows network functions to be deployed as software services that communicate with each other using well-defined APIs. This service-based approach enables agility, scalability, and the ability to introduce new services and features with ease.

One notable implementation of the 5G core is Open5GS [9]], already described in Section 2. Open5GS offers a range of network functions implemented as software modules, allowing users to build their own 5G core network using off-the-shelf hardware and cloud-native environments. Overall, Open5GS provides a flexible and customizable framework for building 5G networks, fostering the development of new services and applications that leverage the capabilities of 5G, such as low-latency applications, IIoT services, and network slicing. As an example, in the Zero-SWARM's Spanish trials, the Open5GS core is installed in a bare metal server.

Figure 12 shows the general architecture of the Open5Gs core. As represented, this core can be configured in both 5G non-stand-alone (NSA) and stand-alone (SA) modes. In an NSA deployment, the 5G network infrastructure is built upon an existing 4G Long-Term Evolution (LTE) network. In a SA deployment, the 5G network operates independently without relying on an existing 4G infrastructure. Both the radio access network (RAN) and the core network are specifically designed for 5G, therefore allowing for full 5G functionality and capabilities. SA deployment provides enhanced performance and enables the utilization of advanced 5G features, such as network slicing and low-latency communications. In the case of Open5GS, the SA deployment means that the Open5GS core functions constitute the entirety of the 5G core network. It handles all the control and user plane functionalities, including AMF, SMF, UPF, and NSSF, among others.

The configurations applied in Zero-SWARM and described in this Section is based on the 5G SA deployment, which provides the necessary functionalities needed during the provisioning and initial configuration phase.

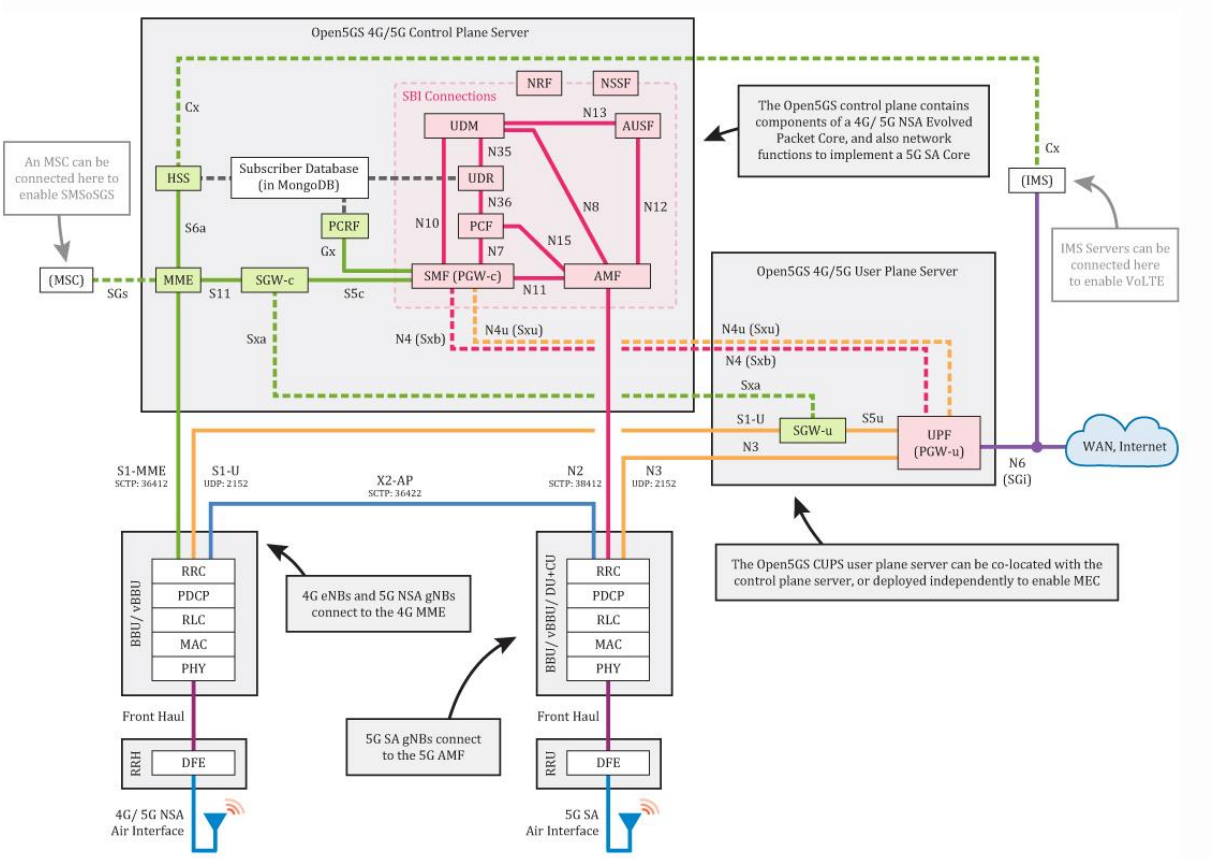


*Figure 12: General Diagram of Open5GS Core [10]]*

As a representative example, Figure 12 also shows the network function level view of an Open5GS implementation of the 5G core. This implementation is applied in Zero-SWARM's Spanish trials, where Open5GS is installed in a bare metal server. In such an installation, it is possible to access the configuration files of each network function, as portrayed in Figure 12, where an SA configuration is represented. Figure 12 also shows the interfaces inside the core, which are internally configured when the core is installed. Interested readers can find the default configuration of each interface on the official Open5GS code repository [11]].

Open5GS implements a database to register the UEs using mongoDB [12]], and it exposes a web-based user interface, allowing the addition and editing of subscriber data at http://localhost:3000, where localhost represents the IP address in which the core is installed. In order to register and add a UE to the core, a number of parameters are required. These include the International Mobile Subscriber Identity (IMSI), security context parameters such as the K-subscriber key, operator code (OPc), and authentication management field. Additionally, slice configuration details like the Slice Selection Table (SST) and Slice Differentiator (SD), as well as session configuration settings like Access Point Name (APN) and QoS Class Identifier (QCI), are required.

Table 1 shows the configuration diagram of the AMF network function. Parameters given in italic style represent the endpoints that need to be configured. For instance, in the current status of the core both the Mobile Country Code (MCC) and the Mobile Network Code (MNC) need to be input three times each in order to configure the Public LAN Mobile Network (PLMN).

*Table 1: AMF Configuration Scheme*

| | | |
|---|---|---|
| NGAP | *IP address* | |
| Metrics | *IP address* | |
| GUAMI | PLMN id | *MCC* |
| | | *MNC* |
| | AMF id | *Region* |
| | | *Set* |
| TAI | PLMN id | *MCC* |
| | | *MNC* |
| PLMN support | PLMN id | *MCC* |
| | | *MNC* |

The correct configuration of the aforementioned parameters is crucial. In particular, the IP address of the Next-Generation Application Protocol (NGAP) is important, because it is used by the gNB to connect to the 5G core [13]].

On the other hand, the User Plane Function (UPF) should be configured to add the protocol GTP-U IP address and enable packet transmission and reception services to user plane entities. Additionally, the sub-networks that should be used for each packet data unit (PDU) session also need to be provided in the UPF configuration file, where it is possible to specify different sub-networks per each access point name (APN). Note that a PDU session provides end-to-end user plane connectivity between the UE and the data plane.

## 3.3 RAN Configuration Aspects

To provide the reader with some context regarding the configuration aspects of the 5G RAN in NPNs, the specific requirements of Node-H RAN are used as a representative example. Node-H RAN cells will

be deployed in Zero-SWARM's Spanish trials. Node-H cells can host radio units of different vendors such as T&W or Askey. The device should be initially configured using a configuration file. When the device is configured, it exposes an API where the current configuration could be checked and modified in a more efficient and automatic manner.

Generally, across different vendors, the RAN configuration is divided up into the Central Unit (CU) configuration and the Distributed Unit (DU) configuration. The CU configuration contains parameters such as: PLMN id, AMF IP address and mobility parameters, among others. While in the DU configurations is possible to find: channel bandwidth in resource blocks, frequency band in terms of the absolute radio frequency channel number, time-division duplexing (TDD) patterns that indicate the dedicated slots for downlink and uplink, transmit power and the tracking area code (TAI), among others. Additional details about the configuration parameters of Node-H are presented in Appendix B. Such configuration parameters can be configured to provide a given QoS according to the target OT application. Moreover, these parameters could be read and configured using an API, opening the door to introducing automation in the configuration, decision making and alarm system in order to reduce deployment times and increase the performance.

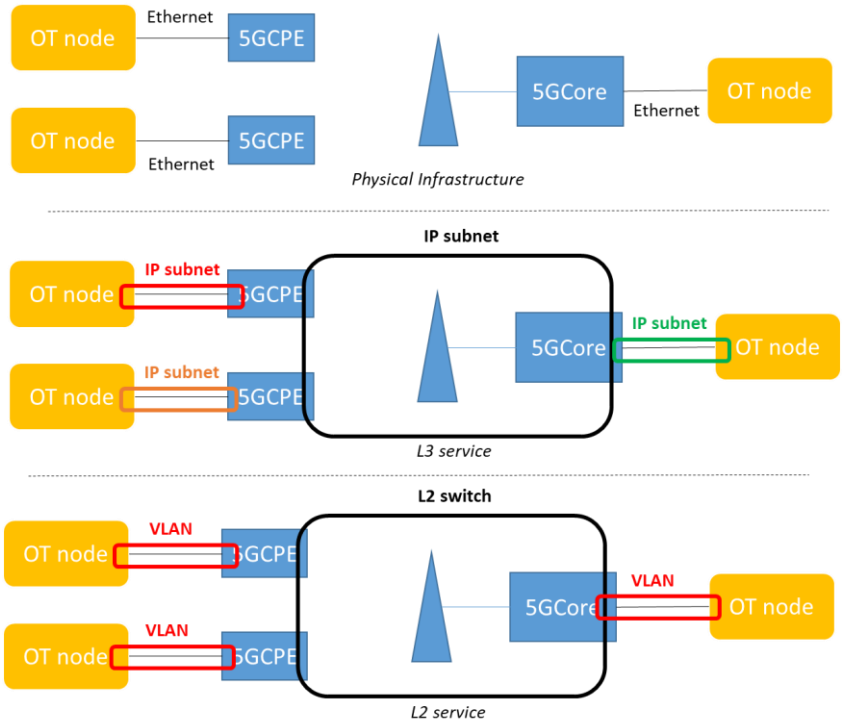


*Figure 13. Physical infrastructure (top), L3 service model (middle) and L2 service model (bottom)*

## 3.4 Zero-Touch end-to-end connectivity service provisioning

A solution for the provisioning of both L2 and L3 connectivity services is presented in Figure 13. Such solution enables OT operators to deploy the required service without the need to reconfigure any OT devices.

The top part of Figure 13 depicts the physical infrastructure involved in a 5G NPN deployment, where 5G components are highlighted in blue, and in OT components are highlighted in yellow. It can be seen that OT devices (or OT nodes, as shown in Figure 13) connect to the 5G system through an Ethernet interface. Moreover, OT devices can connect to the 5G system through the 5G Core network, or through a 5G customer premise equipment (CPE) with an embedded 5G modem. Note that, in practice, an OT device can also have natively embedded 5G connectivity; this case can be considered as a particular implementation of the architecture shown in **Error! Reference source not found.**, in which c ase the 5G CPE functionality would be deployed within the OT device. Thus, without loss of generality we consider hereafter the architecture depicted in Figure 13.

Based on the proposed 5G architecture, the 5G NPN can be seen as a system to interconnect OT devices, where two basic connectivity services can be provided:

- A Layer 3 (L3) connectivity service: Depicted in the middle part of Figure 13. In this service each OT device belongs to a different IP subnet, and the 5G NPN system acts as an IP router that is able to route IP packets from any IP subnet.
- A Layer 2 (L2) connectivity service: Depicted in the bottom part of Figure 13. In this service the OT devices all belong to the same L2 virtual local area network (VLAN), and therefore to the same IP subnet. In this set-up, the 5G NPN system acts as a layer 2 switch, meaning that broadcast and multicast Ethernet frames generated by the OT devices can reach all the other OT devices in the same VLAN. Under this service model, the 5G NPN system should be able to transport multiple L2 VLANs simultaneously, properly segregating L2 broadcast domains.

In the context of the L2 and L3 connectivity services described above, the notion of "*Zero-Touch Service provisioning*" can be understood from the perspective of the OT operator, who does not need to reconfigure any of the OT devices, i.e. the yellow boxes in **Error! Reference source not found.**, to p rovision the desired L2 or L3 service. Thus, the following functional requirement for the Zero-SWARM 5G-NPN management system is addressed:

*The Zero-SWARM 5GNPN management shall allow an OT operator to provision an end-to-end L2 or L3 connectivity service across the OT devices that connect to the 5G NPN, without requiring any intervention on the connecting OT devices.*

Future work in Task 3.2 will be devoted to designing management tools that can address this requirement.

## 3.5 Zero-Touch Operation of 5G NPN

After a connectivity service has been provisioned through the Zero-SWARM management service, a number of events can occur that disrupt a running service. Relevant examples include:

- Unintended infrastructure reboots e.g., due to a power cut, which can potentially revert the RAN or core network back to an undesired configuration.
- Manual configurations executed outside of the Zero-SWARM management system, e.g. through the dashboard of one of the network element managers, which the potential to disrupt and break down the intended configuration state.
- Additional configurations provided through the Zero-SWARM management system that conflict with pre-established configurations.

Zero-touch operation mechanisms are therefore required that are able to maintain the intended configuration state, avoiding any of the errors listed above. Recently, solutions have appeared for zero-touch operations in cloud systems, based for example on the Kubernetes Operator framework[1]. In Task 3.2 we will study how to apply these technologies to the management of 5G NPNs.

Figure 14 depicts two fundamental properties of Zero Touch Operations mechanisms for the Zero-SWARM 5G NPN management system, namely:

- A *declarative service interface* is supported in the north-bound, allowing the NPN operator to express the desired service configuration, without providing low-level configuration details. The service definition expressed through this interface becomes the ground truth desired configuration.
- An *active reconciliation loop* is established between the Zero-SWARM management system and the 5G infrastructure. Without any intervention from the NPN operator, the active reconciliation loop detects configuration drifts in the 5G infrastructure and reverts them back to the desired ground truth configuration.

Hence, the following requirement for the Zero-SWARM 5G NPN management system in terms of zero-touch operation is addressed:

*The Zero-SWARM 5G NPN management shall offer a declarative interface to the NPN operator, which will represent the ground truth configuration state. Through an active reconciliation loop, any configuration drift will be automatically reverted to the ground truth, without requiring intervention from the NPN operator.*

[1] https://operatorframework.io/

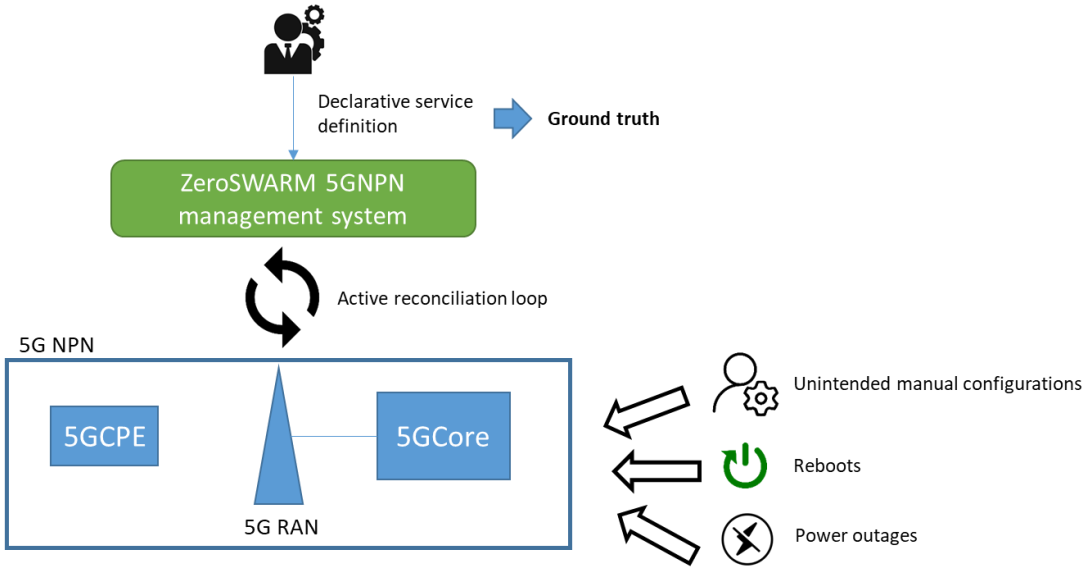


*Figure 14: Declarative service definition and active reconciliation loop*

## 3.6 Zero-Touch Diagnosis of 5G NPN

As previously mentioned, the integration of 5G NPN in the OT domain represents a significant challenge. In particular, it is important to mention the lack of telecommunication engineers in factories or manufacturing floors. For this reason, it is fundamental to design a 5G NPN system that includes automatic diagnostic features to minimize the human intervention and the network support periods. Self-healing, self-configuring and self-monitoring are desired features to facilitate the 5G integration in the OT domain.

As the initial stage to pave the road, it is crucial to identify common errors that could appear not only during the configuration process but also in the operational phase of the 5G network. It is also worthy to identify possible solutions or troubleshooting procedures to solve the mentioned issues. Some examples of common issues are listed as follows:

1. The radio access network is not registered to the core network.
2. The UEs are not able to see the 5G network
3. The UEs cannot register to the 5G network.
4. The UE is registered in the network, but it does not have a data session.
5. There are crashes or serious alarm and/or performance issues in the RAN or core.

These failures can be caused by multiple errors not only on the provisioning or configuration phase but also during the network operation. The solutions are often high time demanding, because they require an expert troubleshooting process. Especially, it may reduce the productivity and performance of the device on the OT domain. For this reason, the zero-touch diagnostic has a significant role, which means

to introduce automatic capacity to analyze the possible causes of the network failures and providing an advanced monitoring system. In order to accomplish this task, a self-monitoring system will have a huge impact because it has the goal of analyzing the performance of the network in order to detect the possible failures. Such telemetry capabilities are normally available in current RAN and core implementations; interested readers can find typical RAN and gNB parameters in Appendix A, where Node-H parameters are provided as an example.

After failures are detected, multiple solutions can be proposed to process the network status in order to provide alarms and suggest the possible cause of the failures, reducing the recovery time and the human interaction. Ideally, an AI-based system can provide possible actions to do in the network configuration to solve the problem.

Figure 9 describes a general architecture that can manage the zero-touch diagnostic procedure, where the network telemetry subsystem must extract the significant metrics, traces and logs to have the whole view of the network. This subsystem contains a huge complexity, as it has to monitor each network element with different strategies and formats to expose the metrics.

After extracting the network status, the AI-based engine subsystem will detect the anomalies (alarms, performance degradations, and so on), and take specific actions to solve the problem. Finally, the network service provisioning subsystem should execute the actions on the corresponding network elements.

## 3.7 O-RAN Logical Architecture

O-RAN introduces RAN disaggregation and adds flexibility to the overall RAN network. The O-RAN architecture offers significant advantages in the smart industry context; by providing flexibility and scalability, it allows organizations to adapt and scale their 5G networks efficiently. In addition, the interoperability and vendor neutrality of O-RAN foster innovation by enabling integration with solutions from different vendors. Moreover, the deployment of edge computing capabilities reduces latency and supports near-real-time processing for time-sensitive applications.

However, with this flexibility comes the complexity of managing such a network, especially in the context of zero-touch. The O-RAN logical architecture is introduced below to provide the reader with some necessary background information. Then, in Section **Error! Reference source not found.**, Zero-S WARM's approach to O-RAN zero-touch configuration and installation is presented. In addition, the approach to O-RAN zero-touch DU and RU monitoring is presented in Section 3.9.

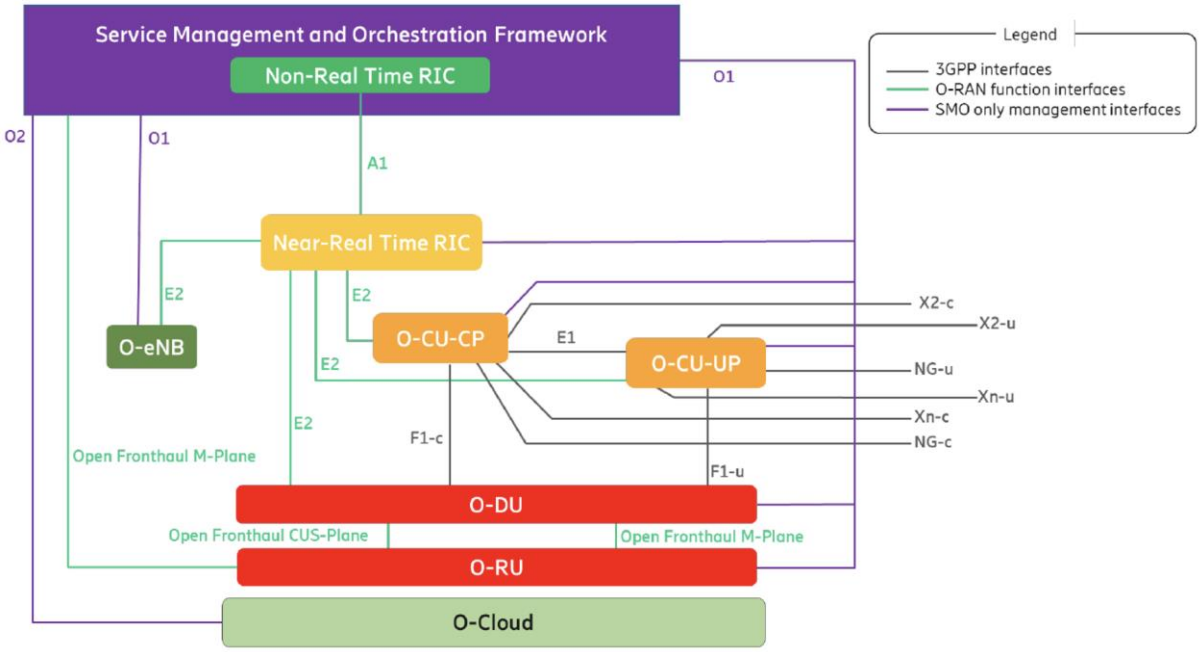


*Figure 15: O-RAN Logical Architecture*

The O-RAN logical architecture consists of disaggregated RAN components (O-CU-CP, O-CU-UP, O-DU and the O-RU), the near-real time (RT) RAN intelligent controller (RIC) and the non-RT RIC interconnected using 3GPP interfaces and O-RAN defined interfaces, as shown in Figure 15 above.

These Network components are deployed and managed by the Service Management and Orchestration framework layer (SMO) integrated with Non-RT RIC using O2 and O1 interface respectively. The Near RT RIC supports telemetry and control interfaces towards RAN Nodes using E2 interface. The E2 interface terminating near-RT RIC is used by the eXtended Apps (xApps) that can implement machine learning models to gather telemetry information and perform control loops with a latency in the order of 10ms to 1 second. The open fronthaul interface between the O-DU and O-RU runs a control loop in the order of less than or equal to 10ms. The Non-RT RIC gathers telemetry data from RAN elements and performs control loops of the order of greater than 1 second latency. All control loops in the O-RAN architecture aim to optimise the performance and user experience of the UEs connected to the network.

The Non-RT RIC can train the ML models based on collected data and deploy them as radio Apps (rApps) to Non-RT RIC or xApps in Near-RT RIC. The RIC will support radio network information base (RNIB) and UE network information base (UENIB) that capture the cell specific information and the RIC UE id mappings.

The telemetry data received from all RAN nodes via the E2 interface can help in pervasive monitoring of performance, scaling and optimizing the resource usage in a mobile network. The A1 interface

configures the policies and enrichment information necessary for xApps to optimize the network performance and user experience of the UEs.

## 3.8 O-RAN Zero-Touch Configuration and Installation

O-RAN zero touch configuration and installation involves physical network function like O-RU performing vendor specific actions[2] on startup and contacting operator specific virtualized core software server on service management and orchestration (SMO) to send a 'physical network function (PNF) Registration Event' [29]]. This Physical Network Function Registration Event contains all parameters necessary to identify the serial number, equipment type, model number and vendor specific details. When this event is received in the SMO, the SMO can start zero touch provisioning of all dependent Network functions that are required to service this PNF. The SMO maintains an internal datastore of AAI (All Active Inventory) to verify the identity of the PNF and to retrieve the necessary configuration templates with necessary meta data to configure the service. If the Network functions are already deployed only delta changes to configuration to include the new PNF are performed through Netconf. In case of Network functions deployed directly by SMO, the SMO can query the deployment management service[3] for the IP address and ready state of the network service, before attempting to configure it via O1 interface or Netconf.

Figure 16 below, shows a sample plug and play sequence diagram for O-RAN Split 2 DU or Split 7.2x architecture when the DU/RU unit is powered on to obtain IP address and bootstrap configuration (optional) via dynamic host configuration protocol (DHCP). It assumes that the dependent Virtualized Network Functions (VNF) packages (e.g., Helm charts) and configuration templates are onboarded to SMO by the operator beforehand (including meta data to identify the authorized PNFs) that can connect to the network. The PNF registration event from the DU drives the plug and play deployment of the service(s) including the configuration from SMO.

[2] This could be something like DHCP provisioning or contacting vendor specific server address provisioned in the factory

[3] This could be Kubernetes for example.

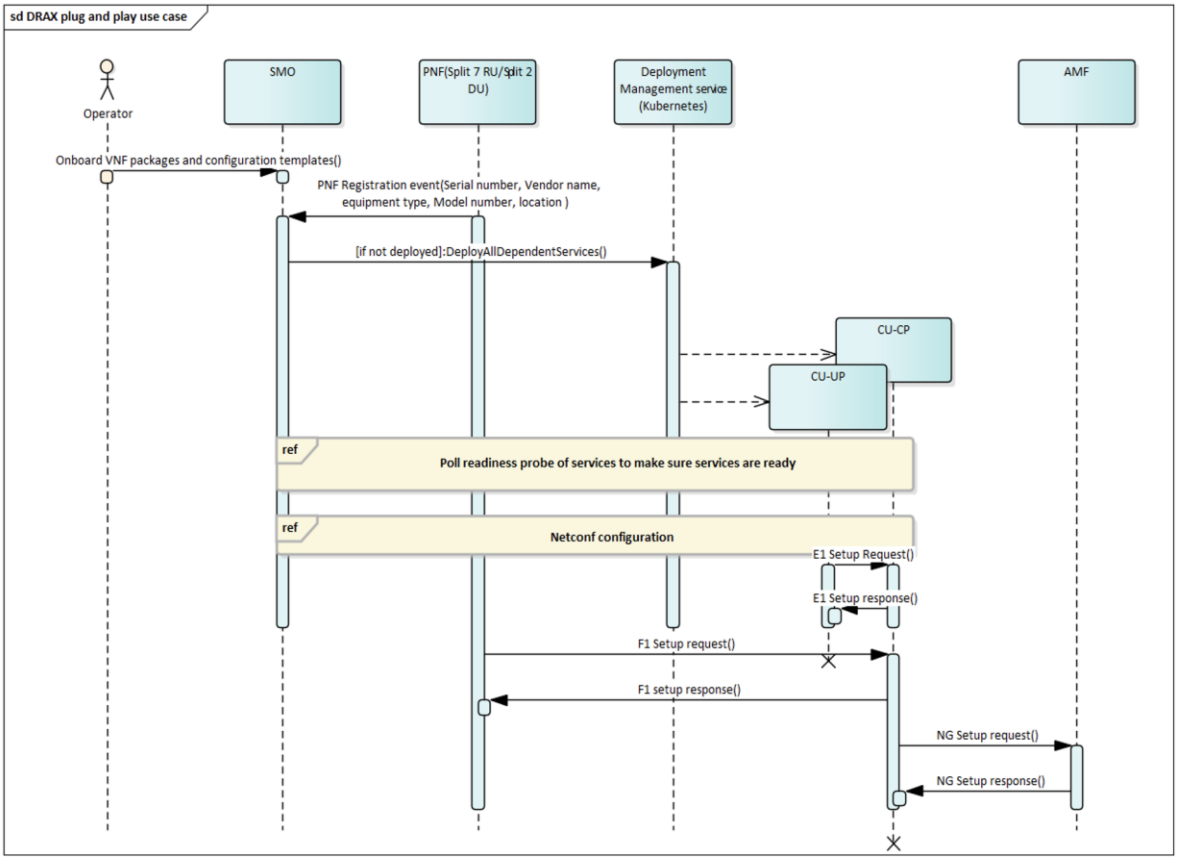


*Figure 16: O-RAN Plug and Play Use Case*

In summary, O-RAN envisions the ability to discover newly plugged-in RAN components, provision them into the network, and finally configure them. In the O-RAN disaggregated world, this means configuring the RU, DU and CU to interconnect with each other in order to bring a 5G cell on air. In order to complete the previously mentioned tasks, the RAN components need to have all the necessary O-RAN interfaces implemented. However, current RAN vendors do not always implement all the O-RAN interfaces which creates a challenge.

In the Accelleran dRAX product, we envision the ability of provisioning the RAN components and configuring them. To be able to do so, we also have to take into account RAN vendors with limited O-RAN interface support. Accelleran builds its own 5G CU component. The Accelleran CU is fully 'softwarised' into Docker images. These Docker images are deployed using Helm Charts on top of Kubernetes. This allows full flexibility in terms of their configuration on deployment and allows us to create the necessary interfaces for provisioning the CU as well as configuring it via NetConf. In order to integrate this into the Accelleran dRAX product, including the RIC and SMO, we plan to create the necessary abstractions. The dRAX API is a gateway to all the API endpoints present in the system. Extending this dRAX API, we will create APIs which allow provisioning of the Accelleran CU, as well as configuring it. To be able to do so, we will create YANG Data Models for provisioning and configuration.

To further simplify the provisioning of the configuration of the CU, we will also adapt our dRAX Dashboard, which is effectively a graphical user interface (GUI) for the dRAX SMO. The dRAX Dashboard implements a front-end GUI input formwhich will call the dRAX API provisioning and configuration API endpoints in the background. With the dRAX API endpoint we aim to be able to programmatically provision and configure the CU and use the dRAX Dashboard to provide a user-friendly visual way of doing the same.

The Accelleran dRAX product interacts with 3rd party DU and RU RAN components. Therefore, there is a need to provision and configure them in a similar, simple way as with the Accelleran CU. However, the added complexity is that the RAN vendors often don't implement the official interfaces. Therefore, in the context of the Zero-SWARM project, we envision creating an abstraction layer called the Cell Wrapper. The Cell Wrapper will implement on its southbound the custom interface towards the DU and RU components. On the northbound we will create a Generic YANG Data Model inspired by the O1 O-RAN interface. This will allow us to tap into the Cell Wrapper from the northbound in a unified way, which will then get translated into RAN vendor specific interfaces on the southbound, effectively abstracting the complexity.

To integrate the Cell Wrapper into the dRAX product, we envision further extending the dRAX API, to create the necessary provisioning and configuration interfaces that interact with the Cell Wrapper. This will allow a programmatic approach to provisioning and configuring DUs and RUs. On the other hand, extending the dRAX Dashboard to also create a visual input form that uses those APIs will allow us to have a user-friendly visual interface as well.

In this way, we look to create the building blocks of fully provisioning and configuring the CU, DU and RU components of the 5G RAN.

## 3.9 O-RAN Zero-Touch DU and RU Monitoring

O-RAN gathers telemetry to monitor the DU performance via E2 interface Key Performance Measurement (KPM) service model. The E2 nodes advertise supported RAN functions, associated E2 Service model styles and the E2 node configuration via E2 setup request. The E2 KPM service model allows the xAPPs to subscribe to specific metrics supported by the E2 node on node, cell, slice and UE basis. These metrics include physical resource block (PRB) utilisation in terms of percentage (both DL and UL), number of connected UEs, average UE throughput (both UL and DL), etc., that are standardised in [30]]. The DU alarms are collected via the O1 interface. The RU telemetry and alarms are collected at the SMO using the O1 interface.

However, not all RAN vendors implement the full E2 and O1 O-RAN interfaces. Parts of the functionality can be exposed over custom interfaces, while other parts can be missing altogether. To tackle this challenge, as mentioned earlier, we plan to introduce the Cell Wrapper concept - an abstraction layer

that will extend the functionality of the underlying RAN components. The Cell Wrapper is envisioned to operate in a microservice oriented architecture, allowing different modules to interact with each other in order to perform the global abstraction. As mentioned previously, O-RAN envisions collecting alarms and telemetry via the O1 interface from the DU and RU. The Cell Wrapper is envisioned to include the Auto Repairer module, which will monitor the DU and RU. By collecting information, the Cell Wrapper Auto-Repair module will be able to also react to alarms. Hence, we plan to create functionalities that allow self-healing, such as restarting the connection between the DU and RU in case it's lost.

By collecting telemetry information from the DU and RU, the Cell Wrapper is also able to know the state of the network. This information is useful for the higher up layers, such as the RIC and the SMO. Therefore, we plan to investigate the ability of exposing such information via heartbeat or beacons. This can be a tool to allow discovery of RAN components.

The Accelleran CU implements a similar mechanism of exposing telemetry data to the RIC and SMO, as well as heartbeats to allow for its discovery.

# 4 Data-based & AI-driven zero-touch management mechanisms

Zero-touch management mechanisms refer to a set of processes and technologies that aim to automate and simplify the management and maintenance of 5G networks without any human intervention. These mechanisms enable the intelligent and autonomous operation of 5G networks by automating various tasks and reducing the need for manual configuration and monitoring.

## 4.1 Overview of Zero-Touch Management Mechanisms

One key aspect of the zero-touch management mechanism is self-configuration. With this mechanism, limited manual inputs are needed for tasks such as device registration, authentication, and network parameter optimization. Self-configuration enables quick deployment and minimizes human errors in the network setup process.

When we talk about zero-touch management mechanism, we can mention the self-optimization. In 5G, this aspect allows for analysis of both real-time and historical data to make intelligent decisions that optimize performance. Depending on specific requirements and priorities, these decisions can be time-critical or more flexible. The time aspect of a zero-touch management mechanism in 5G primarily involves the speed and agility with which actions can be performed. The ability to execute actions in real time or in a more flexible way provides several benefits, such as: troubleshooting a network issue or optimizing network performance that can be executed in real time, allowing for quick resolution

and minimal disruption to services. On the other hand, actions can be executed in a more dynamic and adaptable manner. For instance, network elements can be reconfigured or optimized based on changing network conditions or traffic patterns that can be identified thanks to the historical data. By efficiently utilizing network resources, self-optimization ultimately enhances the overall user experience.

Zero-touch management also involves self-healing capabilities. 5G networks can detect and diagnose faults or anomalies in their operation and take corrective actions automatically. For example, if a base station or network element malfunctions, the network can self-heal by reconfiguring its resources, re-routing traffic, or triggering appropriate actions to restore normal operations. By reducing downtime and improving network reliability, self-healing mechanisms enhance the overall network performance.

In addition, the network can anticipate changes in the environment, traffic generation, or performance, or even anticipate failures or anomalies to proactively adjust the network configuration to the anticipated situation. This proactive management based on an anticipated knowledge allows to seamlessly guarantee the QoS requirements of industrial services and applications demanding very-stringent latency and reliability requirements.

Zero-touch management includes also self-security mechanisms. 5G self-contains the security in itself as a black box; however, the main security threats here are related to the integration of 5G NPN with existing IT and OT domains with other technologies, and thus the variety of standards, lack of best practice / blueprints, and vulnerability of IT / OT systems, which indirectly makes the 5G insecure. Therefore, 5G networks need to be secure from various threats, such as unauthorized access, data breaches, or distributed denial of service (DDoS) attacks. The self-security mechanisms continuously monitor and detect potential security breaches and automatically take preventive or corrective actions to secure the network. This can include blocking suspicious traffic, applying security patches, or activating additional security measures to mitigate security risks.

As described in Section 2, in 5G networks, the introduction of network slicing enables operators to efficiently share their infrastructure by creating multiple logical partitions called network slices. These slices can be customized to meet the specific business or technical requirements of their customers or services. By offering different levels of isolation, network slices can be differentiated and updated independently. Additionally, network slices can be adjusted in terms of mobile connectivity, virtual functions, and computing and storage resources.

### 4.1.1 The data driven closed loop concept

The zero-touch management mechanism can be implemented using the closed loop approach depicted in Figure 17. This approach refers to a feedback mechanism that enables dynamic adaptation,

optimization and control of various network parameters. It involves continuously monitoring and analysing the network performance metrics in real-time and making adjustments in order to maintain and enhance the overall network quality.

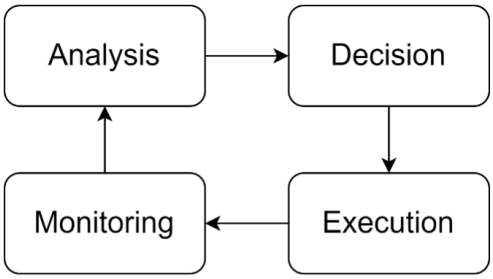


*Figure 17: Closed loop concept*

The procedure initiates with the monitoring phase that make a continuous monitor of the 5G network and permits to retrieve all the data that will be used in the next phases. For this particular scenario, the monitoring metrics originate from several distinct data sources that needs to be gathered and unified:

- End-device monitoring data
- RAN monitoring data
- NWDAF monitoring data
- IoT platform monitoring data
- Edge/Cloud infrastructure monitoring data
- Application monitoring data

Following the Monitoring step, we can find the Analysis step which analyses the collected data to identify potential issues, bottlenecks, or areas requiring improvement. The analysis step involves evaluating the performance of the system and making decisions based on the collected data. This step typically involves analysing various metrics and parameters, such as signal quality, interference levels, traffic patterns, and user experience. The goal of the analysis step is to identify or anticipate any issues or areas of improvement in the system's operation and make necessary adjustments or optimizations. This could include adjusting transmission parameters, applying a network slice or a network slice subnet reconfiguration, or implementing better resource allocation strategies to enhance the overall performance of the 5G network.

Based on the Analysis step, the Decision step makes decisions on necessary reconfiguration/optimization to enhance the performance of the system. Finally, the Execution step executes the decision that comes from the previous step, requesting the system to reconfigure itself.

### 4.1.2 Zero-touch Management Mechanism high level architecture

In the context of 5G networks, the application of network automation enables the control and management systems to dynamically adjust and adapt the infrastructure, resource allocation, network slices, and running services based on real-time data. This data-driven approach involves making decisions using various metrics and KPIs such as network conditions, infrastructure load, resource utilization, and service demands. As described in Section 4.1.1 these parameters serve as inputs for the steps that follows the Monitoring step, where algorithms, potentially utilizing AI/ML techniques, are used to analyse, make decisions, and optimize the network triggering automated actions within the control and management system.

The high-level architecture for NPN zero-touch management mechanism is depicted in Figure 18. The lower part of the picture is the 5G NPN already explained in Section 3.1.

The upper part, the 5G management system, is composed by several functional elements defined by the 3GPP TR 28.801 specification [27]. This specification identifies three functional elements: the Communication Service Management Function (CSMF), the Network Slice Management Function (NSMF), and the Network Slice Subnet Management Function (NSSMF).

The CSMF is responsible for processing requests for new communication services (CS) and managing CS instances provided by a network operator. It translates CS requirements into network slice characteristics, such as the SST, required capacity of mobile connectivity, and QoS requirements. The CSMF interacts with the NSMF to request the creation of the related Network Slice Instance (NSI).

The NSMF is responsible for the management and end-to-end orchestration of NSIs based on requests received from the CSMF. It splits the NSI into internal Network Slice Subnet Instances (NSSIs) according to the Network Slice Template (NEST) and manages their lifecycle. The NSMF, in coordination with the closed loop functions, makes decisions regarding the composition of a NSI, including the re-usage of pre-existing NSSIs shared among multiple NSIs and the coordination of provisioning, scaling, and configuration. The actual implementation of these decisions is delegated to the NSSMFs, which are responsible for the management and orchestration of each NSSI.

As outlined in the 3GPP TS 28.533 specification [28], which follows the SBA pattern, a typical deployment of the 3GPP management system consists of domain-specific NSSMFs. These NSSMFs are specific to the RAN, Core Network (CN), or Transport Network (TN) domains. They are customized to meet the requirements and technologies of their respective domains.

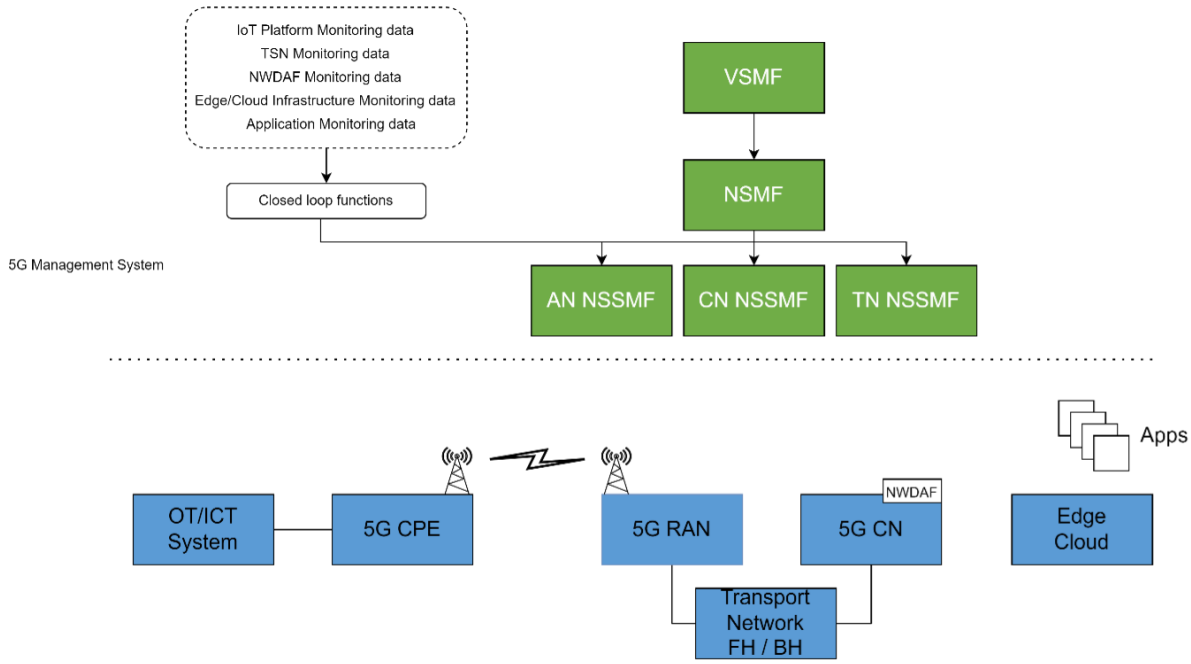


*Figure 18: High level architecture for NPN zero-touch management mechanism*

Data analytics is crucial for automating 5G networks, as it allows for the collection and processing of data from various sources to develop the most efficient re-optimization strategies.

Regarding the network monitoring data, the 3rd Generation Partnership Project (3GPP) has introduced a framework for data analytics and network automation in 5G infrastructures, specifically through the Network Data Analytics Function (NWDAF) [16]] and Management Data Analytics Function (MDAF). The NWDAF collects data from other network functions and external sources, providing data analytics services. Similarly, the MDAF supports management decisions through data analytics. In Release 17, there is a greater emphasis on these functions, with the introduction of models that support AI/ML techniques and frameworks for closed-loop control and management. This evolution is explored in the standards surrounding NWDAF, MDAF, and the closed-loop control framework, as well as in the standardization outcomes from other organizations like ETSI ENI.

In particular, the NWDAF output analytics data may be used to take data-driven network automation decisions and trigger the related actions. In other words, the data generated at the 5G Core Network and related to the network performance constitutes one of the inputs for the AI/ML algorithms that drives the automated re-optimization of the 5G infrastructure.

### 4.1.3 NWDAF

The NWDAF is the 5G Core Network (CN) function responsible for collecting and analysing data. It was first introduced in the 3GPP TR 23.791 report [16]] and can operate in either a centralized or distributed model with either global or per-slice scope. It can collect metrics and data analytics information from various sources such as other 5G CN NFs, Application Functions (AF), the

management system (particularly the Operations Administration and Management – OAM), and external data repositories. By processing this data and utilizing aggregation mechanisms and prediction algorithms, the NWDAF creates analytics data that can be used by other NFs or AFs. The resulting analytics data can inform network automation decisions and trigger related actions. This approach can be utilized in Zero-SWARM to create a zero-touch management mechanism where data generated by the 5G CN regarding network performance is used as input for AI/ML algorithms that automate the re-optimization of the 5G NPN infrastructure.

Beginning in Release 16, the 3GPP TS 23.288 [20]] specification is standardizing the interfaces and procedures of NWDAF, which allow other NFs/AFs to utilize data analytics services. This support includes both query/reply and subscribe/notify models. The corresponding Open APIs can be found in the 3GPP TS 29.520 specification [21]]. Recently, the 3GPP TS 23.288 has been improving the data analytics framework in relation to Release 17. In this update, the NWDAF deployment model is transitioning towards hierarchical and distributed architectures. This involves the introduction of additional interfaces and procedures for communication between NWDAFs. Some of these functionalities include the discovery of NWDAFs, transferring analytics subscriptions to different NWDAFs, and designating a specific NWDAF (known as the Aggregator NWDAF) responsible for combining and aggregating output analytics from multiple instances of NWDAF.

## 4.2 Data-based and AI-driven resource management mechanisms

Network slicing was introduced in 5G and exploits the virtualization and softwarization of networks to create different logical network instances over a common network infrastructure. Each instance can be configured and tailored to support specific Quality of Service (QoS) profiles so that network slicing can simultaneously support several services with diverse requirements.

Network slices are referred to as Network Slice Instance (NSI) in the 3GPP technical specification TS 28.530 [15]. Following [15], each NSI can be integrated by one or more Network Slice Subnet Instances (NSSI). An NSSI represents a group of network functions and their corresponding resources that allow managing each group of network functions independently of the network slice instance. An NSSI may contain instances of the CN managed functions or instances of RAN managed functions (see the examples of NSI and NSSI supporting different services in Figure 19). As defined in [15], NSSIs at the CN or at the RAN can be managed independently. First efforts in the research community were focused on Network Slicing at the CN level (see for example, [22]]-[24]]). This is because the concept of network slicing was initially proposed for the 5G CN. By exploring software defined networking (SDN) and network function virtualization (NFV) principles, a fully virtualized CN instance can be optimized for each function.

The 3GPP also introduced the slicing concept to the RAN in [25]]: the RAN may need specific functionality to support multiple slices or even partitioning of resources for different network slices. To date, much less efforts have been devoted to design solutions for RAN slicing. This can be problematic (in particular for latency-sensitive services) because if the RAN is not optimized all the benefits gained with network slicing at the CN level can be lost at the radio access network level. In this context, Zero-Swarm will work on the design of innovative solutions to advance the current state of the art. The proposals will focus on the allocation and management of radio resource to the different RAN slices. These proposals will take into account the current state of standardization activities [14]][31]]. In addition, it is important emphasizing that RAN slicing solutions do not need to be tightly coupled with network slicing solutions at the CN level since as indicated by the 3GPP in [15] NSSIs at the CN or at the RAN can be managed independently.

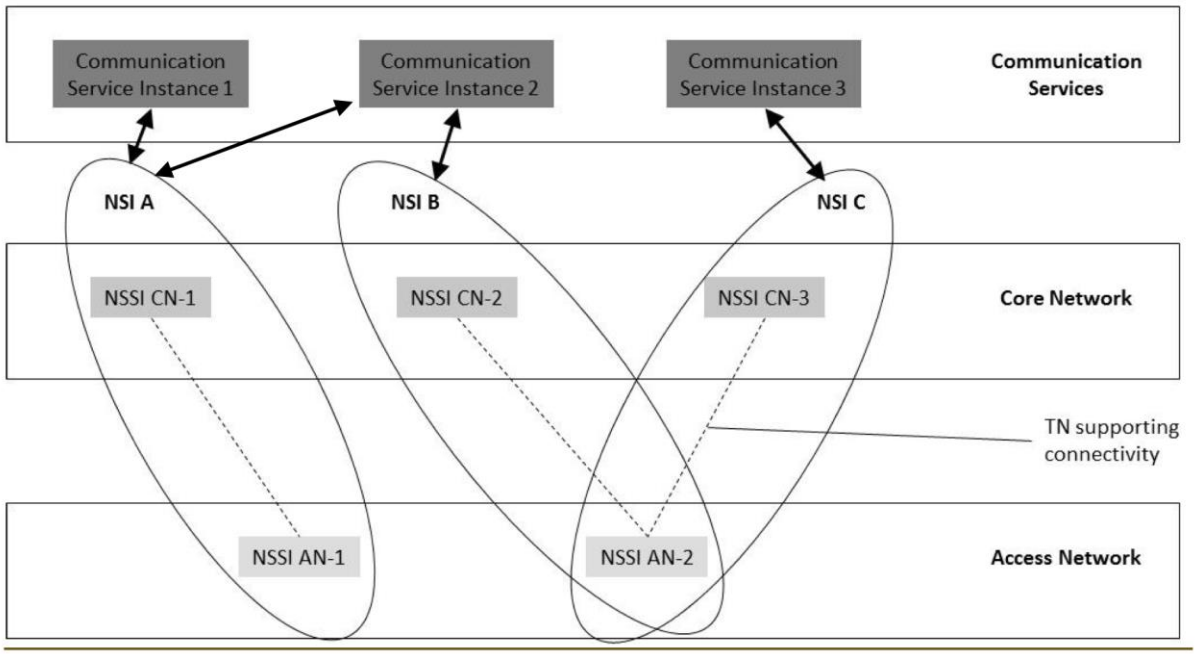


*Figure 19: A variety of communication services instances provided by multiple NSIs [15].*

The lifecycle of a Network Slice is defined by the 3GPP in [15]. The lifecycle of a Network/RAN Slice includes the following four main phases that are illustrated in Figure 20. The phases are:

- Preparation. In this phase, the slice does not exist. This phase evaluates the service requirements that will have to be supported by the slices. It also includes the design of the network slices. The network environment is prepared and other preparations that are necessary before creating a slice are executed in this phase.
- Commissioning. This phase creates the slices and partitions (or allocates) the resource blocks (RBs) among the slices. During the creation process, RBs are allocated to slices using a

partitioning scheme. The allocation is maintained (at least) for the duration of the allocation window. It can be maintained for longer if conditions do not change.

- Operation. The operation phase includes several management tasks such as supervision and reporting, and resource planning and modification of slices. During this phase, we monitor the performance achieved by the slices and report their main KPIs (Key Performance Indicators). Resource planning computes the usage of the radio resources and requests modifications of the slices if the KPIs are not satisfactory.
- Decommissioning. This phase terminates the slices and releases the RBs. RBs can be allocated to new slices with potentially different configurations.

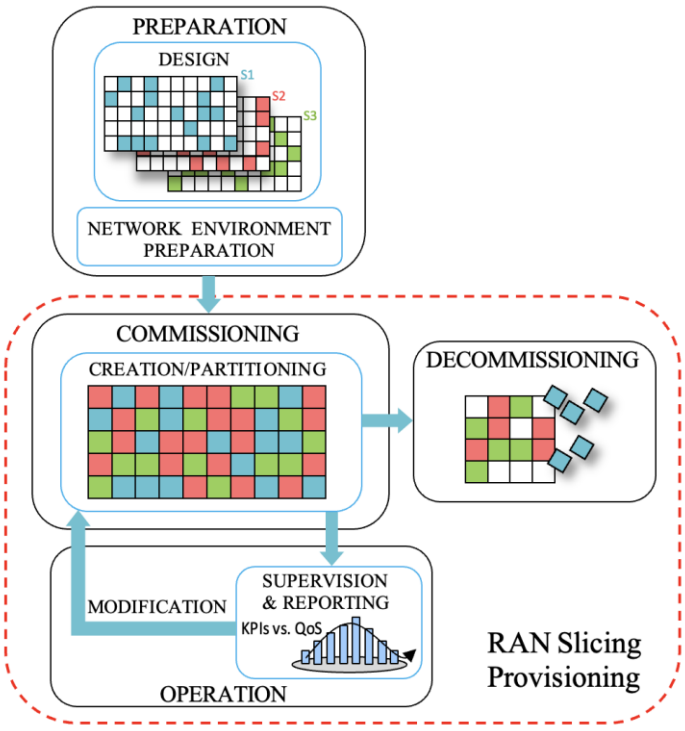


*Figure 20: RAN slicing and lifecycle of slices [26].*

In Zero-Swarm, we will design novel zero-touch RAN slices management solutions to be applied for the initial resource provisioning of the RAN slice during the commissioning phase, and for runtime modifications during the operation phase of the slice. Particularly, we will focus on the partitioning and management of radio resources in the different RAN slices to guarantee the communication requirements of the services supported by the different RAN slices. The initial resource provisioning of the RAN slice will be based on reasonable long-term estimations of service demands and the slice capacity, while the runtime management will be based on short-term predictions and fluctuations of the traffic demands. The designed solutions will be based on the use of Artificial Intelligence and/or Machine Learning techniques and will exploit the data collected by the network to extract and anticipate knowledge about the network performance and/or traffic demands. Based on this

knowledge, the radio resources allocated to the RAN slices will be adapted to seamlessly satisfy the stringent latency, reliability and bandwidth requirements of industrial applications.

## 4.3 Solution implementation

### 4.3.1 Monitoring platform

Data plays a critical role in the management of a 5G network without human intervention in the context of zero-touch network management. According to [19]], a federated data infrastructure, such as the ones in the Zero-SWARM project, ensures that data remains local to its original source and allows for decentralized computations and analyses, eliminating the need for data transfer or centralization.

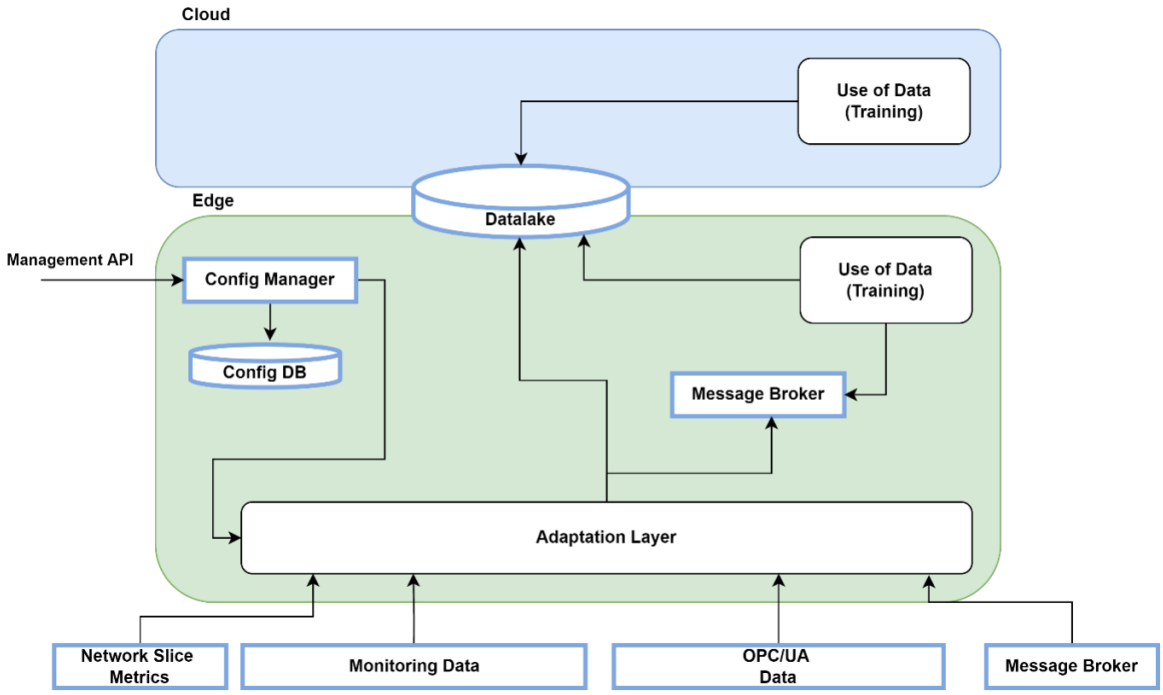


*Figure 21: Collecting Platform Architecture*

Figure 21 illustrates the Collecting Platform, which serves as a key enabler for zero-touch network slice management. It is capable of storing and utilizing various types of data, including metrics and data analytics from the NWDAF, to train AI/ML algorithms. A realistic use case for using the Collecting Platform in the closed loop described before is the possibility to query the NWDAF for a specific network slice ID in order to obtain the network data and share and store them in order to be analysed and used to make decisions for the zero-touch management mechanism. More details about the presented collecting platform can be found in D4.4.

### 4.3.2 Analysis & Decision platform

The MLOps framework described in [19]] and depicted in Figure 22 can then be employed as an Analysis & Decision platform. The ML Operations (MLOps) can perform several operations, such as:

- analysing the data that comes from the Monitoring step,
- deploying the decision algorithms on the edge server, near the shop floor,
- enabling the generation of new vertical service blueprint (VSB) and vertical service descriptor (VSD) or providing instructions directly to the NSFM (**Figure 18: High level architecture for NPN zero-touch management mechanism** Figure 18) for the self-configuration of the network slice management.

A possible use case for using the MLOps Framework in the closed loop described before is the possibility to query the Collection Platform specific networks data and using them to make decisions for the zero-touch management mechanism.

The solution's architecture, depicted in Figure 22 , considers the security and privacy concerns related to sensitive data transmission (see also D4.4). It achieves this by keeping the monitoring data within the local edge, ensuring that no runtime data is transmitted outside of it. The data required for training/retraining and testing of the time forecasting ML models is transmitted to a cloud server only when necessary. Furthermore, additional measures like data encryption or pseudonymization can easily be included in the solution for added precaution.

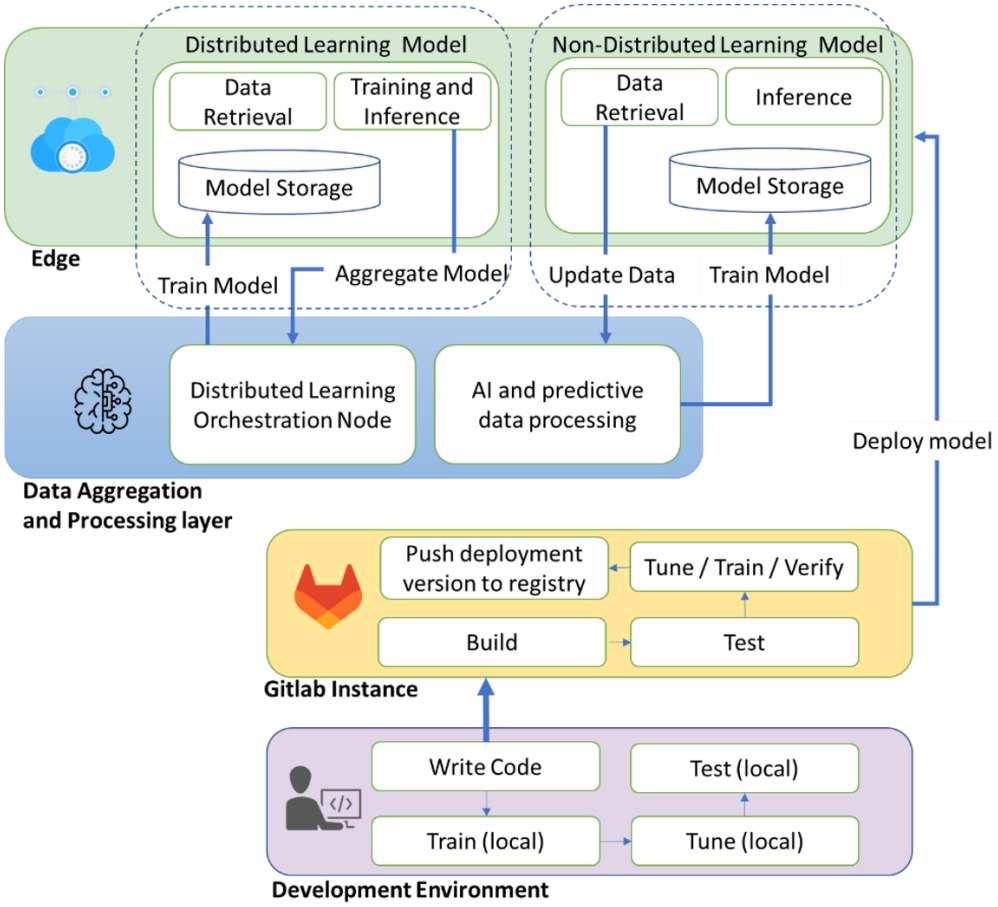

*Figure 22: MLOps Framework*

Industrial applications have unique requirements in terms of network performance, reliability, latency, security, and other factors. Network slicing provides a way to allocate dedicated virtual network resources to meet these specific requirements.

However, in order to create an appropriate network slice, it is necessary to understand and translate the industrial requirements into the corresponding network slice parameters. This translation process involves analysing the industrial application's needs and determining the necessary network characteristics.

For example, if an industrial application requires low latency, high reliability, and high bandwidth, the network slice would need to be configured with parameters that allow for these characteristics. This could involve allocating dedicated resources, applying specific Quality of Service (QoS) policies, or configuring network functions and protocols to meet the desired requirements. The translation process may also involve considering factors such as scalability, mobility, security, and energy efficiency, depending on the specific industrial use case. It requires a close collaboration between industrial application providers and network service providers to ensure that the network slice configuration aligns with the industrial requirements.

The collection platform described in [19]] and depicted in Figure 21 can be utilized to convert the industrial requirement into specific network slice parameters. This platform can bring together the industrial requirement and the data obtained from network monitoring, enabling them to be used for self-configuring the network.

***AI-Based automatic end-to-end slice scaling on a new UPF***

In the realm of network exploration, cutting-edge technology is paving the way for predicting the future state of network metrics. With the help of a long short-term memory (LSTM)-based predictive model, network experts are now able to determine the future state of these metrics more effectively than ever before. This predictive model allows for the analysis and prediction of multiple metrics simultaneously, resulting in a comprehensive understanding of the network's future state. This breakthrough development has significantly aided in minimizing uncertainties and enhancing decision-making processes.

One particular application where the multi-metric future state comes into play is in determining virtual UPF stress levels (please note that here we are not referring into specialized UPF system coming with particular hardware deployment – assigning enough physical UPF is a network planning and dimensioning problem which is out of the scope of this subsection). UPF, or User Plane Function, serves as a crucial element in a network, responsible for processing and forwarding user data. However, the

stress levels experienced by virtual UPFs can fluctuate, leading to potential network congestion or inefficiencies. To tackle this challenge, a fuzzy logic controller is employed, utilizing the multi-metric future state as input. The fuzzy logic controller takes into account various parameters, such as network traffic, available resources, and historical data, to assess the stress levels that virtual UPFs may encounter in the future.

Once the future virtual UPF stress levels are accurately forecasted, an integral part of the system comes into play: scaling the associated slice. A request is automatically generated and sent to the Vertical slicer (introduced in section **Error! Reference source not found.**), triggering the necessary adjustments t o accommodate the forthcoming stress levels. This ensures optimum performance, smooth functionality, and prevents any potential disruptions within the network.

With this advanced predictive model and its subsequent integration into the network infrastructure, challenges related to network congestion and suboptimal resource management can be effectively addressed. By proactively scaling the relevant network slices, the network can intelligently adapt to future stress levels, providing an optimized experience for users and maximizing the efficiency of network operations.

The described solution for the AI-based automatic end-to-end slice scaling is depicted in Figure 23.

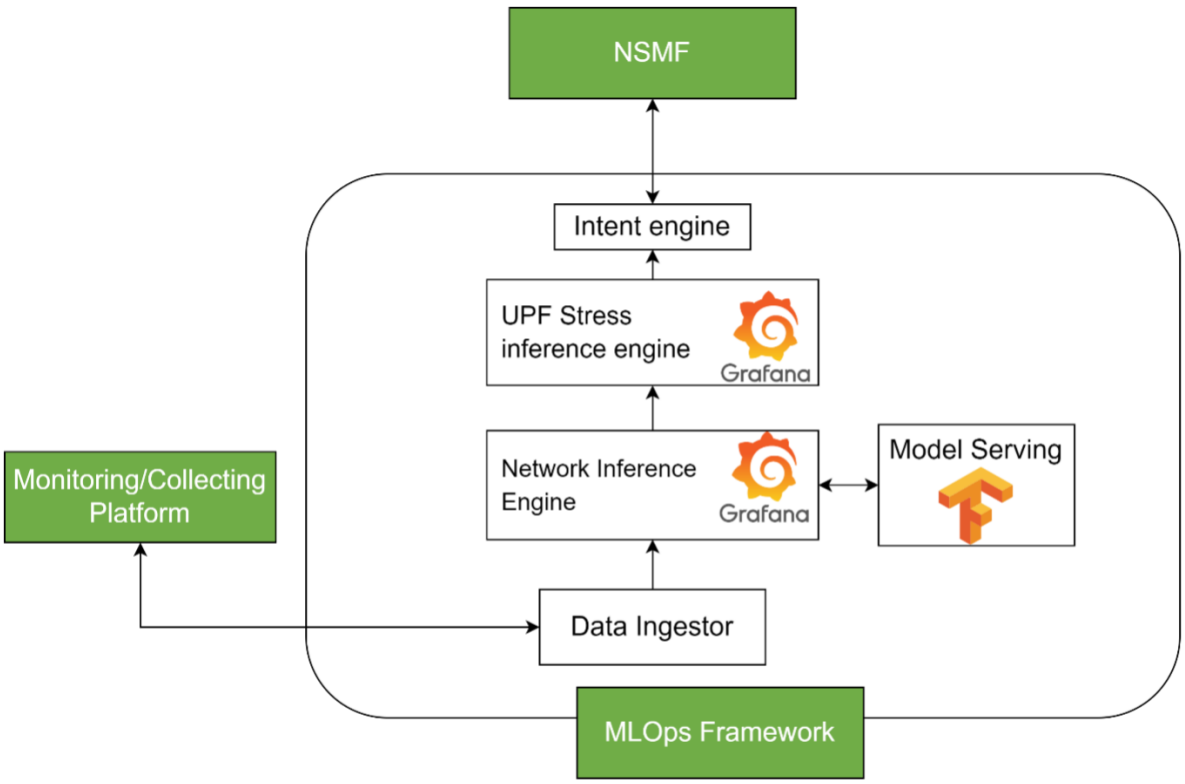


*Figure 23: AI-based automatic end-to-end slice scaling on a new UPF*

After applying a smoothing method to minimize the noise interference from the monitoring data collected in the Monitoring Platform, it is possible to infer the future state of the network, leveraging a multivariate Machine Learning Neural Network. Primarily, we can utilize a LSTM model integrated with a convolution graph. This combination allows us to capture the intricate relationships within the network data and predict its future behavior more accurately. Using the predicted future state as an input, the end-to-end slice scaling of Figure 23, employs an AI controller designed to determine the stress level of the network. This AI controller operates using a fuzzy logic approach, which enables it to make nuanced decisions based on the available information. By considering various factors and their degrees of importance, the fuzzy logic controller accurately assesses the network's stress level.

When the stress level surpasses a predefined threshold, our system triggers the scaling process to ensure optimal network performance and stability. This activation is achieved through a post request to the Vertical Slicer (explained in Section **Error! Reference source not found.**). By initiating scaling p rocedures in response to high stress levels detected by the AI controller, we ensure the network's resilience and its ability to handle greater demand efficiently.

The intent engine is in charge of translating the Industrial requirements into appropriate slice intent in order to send command to the Network Slice Management Function (NSMF) described in Section **Error! Reference source not found.**).

### 4.3.3 Execution platform

In a closed loop approach, the execution step refers to the implementation of actions or tasks based on the feedback received during the previous step.

The 5G system infrastructure introduced the concept of network slicing, as specified in the 3GPP TS 23.501 [14]]. A network slice is like a virtual network that offers specific network capabilities and characteristics. A Network Slice Instance (NSI) consists of a set of Network Functions (NF) that have their own computing, storage, and networking resources. NFs can be implemented as Physical Network Functions (PNF) running on dedicated hardware or as Virtual Network Functions (VNF) on a shared computing infrastructure, such as the cloud. In a 5G network, an NSI includes NFs for the control and user planes of the 5G Core Network, as well as Next Generation RAN (NGRAN) functions for the 3GPP mobile access network.

### 4.3.4 Service & Network orchestrator in cloud/edge continuum

We present an execution platform that is a network slicing solution and that can be used as a reference for the current project, called Service and Network orchestrator in clod/edge continuum.

The Service and Network orchestrator in cloud/edge continuum solution, proposed in this section is developed by Nextworks and is a slice management platform that originated from the 5G Transformer project [17]].

Figure 24 illustrates the key components of the Service and Network orchestrator in cloud/edge continuum. In this context, the VSMF, which stands for Vertical Service Management Functions serves as the main entry point for a Vertical that wishes to create its own vertical service on top of a specific Network Slice. We can use an AI/ML algorithm as a source of commands for the VSMF to configure/manage the network slices.

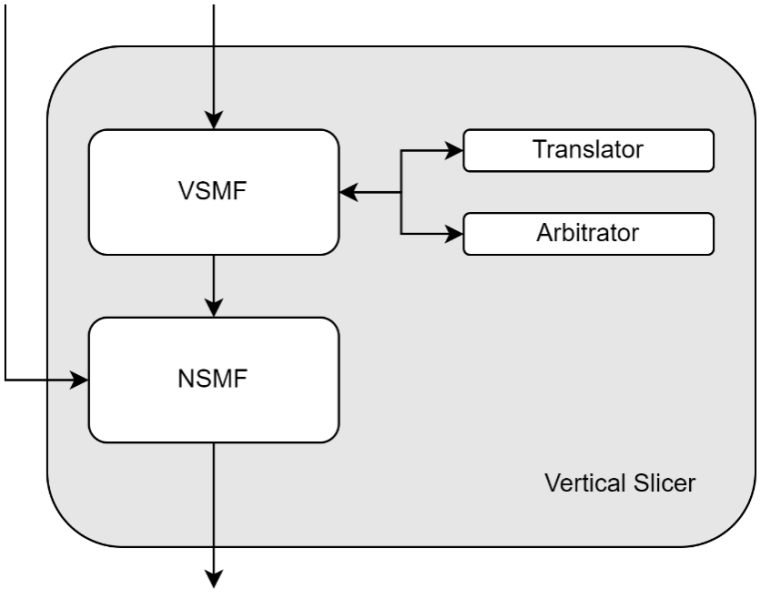


*Figure 24: Vertical Slicer: High level architecture*

The VSMF defines a vertical service using two non-standard information models known as the Vertical Service Blueprint (VSB) and Vertical Service Descriptor (VSD). The VSB acts as a template and represents a category of vertical services, while the VSD is created when the VSB is filled with specific values that characterize a particular service in the category. The VSB includes a range of parameters provided by the VSMF, such as the list of virtual functions and vertical applications to be deployed, dimensioning parameters for a slice (e.g., number of users), and connectivity settings (e.g., L2/L3), QoS, intra and inter slice connection points, etc.

Once the VSD is created, it is translated into a Network Slice Template (NST) by a dedicated Translator module. The NST models the network slice where the vertical service will operate. At the same time, an Arbitrator module determines if the slice can be deployed based on available resources. The NST is then used by the NSMF for the deployment of the network slice.

The NSMF is a monolithic element. Its role is to manage the lifecycle of end-to-end network slices exclusively, without considering the underlying network segmentation and related sub-slices. The NSMF breaks down network slices into Network Services and requests the NFV orchestrator (OSM is one of the supported NFV orchestrators) to create, update, and terminate them.

A VSMF has the capability to interact with multiple instances of NSMF simultaneously and also with other VSMFs. This allows for the deployment of a federation of Service and Network orchestrator in cloud/edge continuum across multiple administrative domains.

# 5 Novelty, innovation, and take-aways

The main objectives of T3.2 are related to the development of zero-touch mechanisms for the management of 5G network and services. This section presents a summary of the main novelty, innovation and take-aways of the foreground technologies currently being developed in the scope of T3.2.

In first instance, this task aims to deliver mechanisms to simplify the provisioning of the 5G NPN, using as a baseline a 5G NPN stack based on existing commercial and open-source hardware and software network components. While in modern 5G NPNs the provisioning of the network requires a large number of configurations and integrations across the various 5G NPN stack elements, this development intends to leverage the use of pre-configured scripts (developed during a pre-planning phase), that can be executed and perform all required configurations in a manner that is agnostic to the user, potentially through a user interface. This eliminates the need for a deep technical expertise in the provisioning of the network and, critically, minimizes the probability of human configuration errors during provisioning, which are hard to troubleshoot, requiring the skills of an expert engineer.

Second, a solution for the provisioning of both L2 and L3 connectivity services has been presented and is currently being developed; in addition, it will be tested in the scope of WP6 trials over the coming months. Such solution aims to enable OT operators to provision an end-to-end L2 or L3 connectivity service across a number of OT devices connected to the 5G NPN, without the need to reconfigure any OT devices. In this manner, novel use cases can be enabled, such as those related to the reconfigurability of production lines or production cells, with the potential to enhance the flexibility and efficiency of production processes.

Third, a solution for the zero-touch operation of 5G NPN has been introduced, with the objective of providing a declarative service interface and an active reconciliation loop for the maintenance of desired configuration states. In the state-of-the-art, service disruptions and misconfigurations usually require a manual troubleshooting effort by an expert NPN operator. Under the proposed zero-touch

operation solution, active reconciliation loops can automatically revert any configuration drift back to the desired state, without requiring intervention from a network operator.

Fourth, an investigation is currently being carried out in terms of the most common issues and errors that can arise during the configuration process and the operational phase of the 5G NPN, such as the RAN not being registered to the core network, or the UEs not being able to register to the 5G NPN. The outcomes of this investigation will pave the way for the future design of future zero-touch diagnosis mechanisms for 5G NPN.

Fifth, developments related to O-RAN zero-touch configuration and installation are proposed. Specifically, APIs for the provisioning and configuration of Accelleran's CU will be developed; a GUI will also be provided from which an operator can initiate the provisioning and configuration of the CU in a user-friendly and visual way. Zero-SWARM also envisions the creation of an abstraction layer (the Cell Wrapper) to enable interactions with third party DU and RU RAN components. The envisioned developments focus on the creation of custom interfaces towards the DU and RU components, along with a generic YANG data model in the northbound that will abstract underlying complexities and provide a unified entry point to apply configurations. The necessary APIs will be created for provisioning and configuration and, in addition, an intuitive user-friendly visual interface will be developed for the deployment and configuration of the DU and RU.

Sixth, an additional development related to O-RAN aim to provide zero-touch mechanisms for the monitoring of the DU and RU, with the objective of providing interoperability with third party RAN vendors. This development also leverages the use of the Cell Wrapper concept with an additional Auto Repairer module for the monitoring of the DU and RU, providing the ability to react to alarms. This development aims to provide functionalities to allow self-healing. Additional developments will focus on leveraging the collection of telemetry to know the state of the network, and to develop a tool to allow the discovery of RAN components.

Seventh, an intent engine is being developed for the translation of industrial requirements into an appropriate slice intent, in order to enable the creation of an industrial vertical slice with the required connectivity capabilities for a given industrial application. This key feature abstracts technical 5G-related jargon away so that, when a new slice is to be provisioned, OT operators can express their needs in a language that is more natural to them. As a baseline, an execution platform with a built-in service and network orchestrator in cloud/edge continuum is used as background component; this component offers the capability to configure the network slice over an existing 5G network, in an automated manner. Network slicing is also supported by a Collecting Platform background component, which can collect network data of a specific network slice that can be used to make decisions by zero-touch management mechanisms.

Another ongoing effort in T3.2 is related to the development zero-touch resource management mechanisms, based on AI/ML techniques, to be applied for the initial resource provisioning of the RAN slice during the commissioning phase, and for the runtime modifications during the operation phase of the slice. The initial resource provisioning of the RAN slice will be based on reasonable long-term estimations of service demands and the slice capacity, while the runtime management will be based on short-term predictions and fluctuations of the traffic demands. Based on this knowledge, the radio resources allocated to the RAN slices will be adapted to seamlessly satisfy the stringent latency, reliability and bandwidth requirements of industrial applications.

Last, an advanced LSTM-based predictive model is proposed for addressing challenges related to network congestion and suboptimal resource management, which have an impact over the network's virtual UPF stress levels. By applying this model, network slices are proactively scaled; therefore, the network can adapt intelligently to future stress levels, and provide an optimized experience for users by maximizing the efficiency of network operations. This functionality is supported by an NPN zero-touch management architecture that tackles security and privacy concerns related to sensitive data transmission, by keeping monitoring data within the local edge and ensuring that no runtime data is transmitted outside of it; data required for training and/or testing of the forecasting ML model is transmitted to a cloud server with additional security mechanisms (such as encryption or pseudonymization) only when strictly necessary.

# References


[1] 5G-ACIA, Exposure of 5G Capabilities for Connected Industries and Automation Applications. 5G-ACIA White Paper, February 2021.

[2] S. P. Shah, B. J. Pattan, N. Gupta, N. D. Tangudu and S. Chitturi, "Service Enabler Layer for 5G Verticals," 2020 IEEE 3rd 5G World Forum (5GWF), Bangalore, India, 2020, pp. 269-274.

[3] 5G-ACIA, white paper, 5G Non-Public Networks for Industrial Scenarios, July 2019.

[4] OPENAPI Initiative, OPENAPI. Available online: https://www.openapis.org. Retrieved: May 29th, 2023.

[5] DIN SPEC 27020:2020-03 requirements and Reference architecture of a security Gateway for the exchange of industry data and service

[6] Karsten Schweichhart, 2018, Reference Architectural Model Industry 4.0 (RAMI 4.0), https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf

[7] Industrial Internet Reference Architecture v1.9 Jun 2019; https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf

[8] Bauer, M.; Boussard, M.; Bui, N.; Loof, J.D.; Magerkurth, C.; Meissner, S.; Nettsträter, A.; Stefa, J.; Thoma, M.; Walewski, J.W. IoT Reference Architecture. In Enabling Things to Talk; Functional-decomposition viewpoint of the IoT-A reference architecture. Source: IoT-A, Springer: Berlin/Heidelberg, Germany, 2013; pp. 163–211, https://link.springer.com/chapter/10.1007/978-3-642-40403-0_8

[9] Open5GS. Available online: https://open5gs.org. Retrieved: June 8th, 2023.

[10] Open5GS Quickstart. Available online: https://open5gs.org/open5gs/docs/guide/01-quickstart/. Retrieved: June 8th, 2023.

[11] Open5GS Github. Available online: https://github.com/open5gs. Retrieved: June 8th, 2023.

[12]What is MongoDB? Available online: https://www.mongodb.com/docs/manual/. Retrieved: June 8th, 2023.

[13] 5G; NG-RAN; NG Application Protocol (NGAP) (3GPP TS 38.413 version 15.3.0 Release 15), Technical Specification, May 2019.

[14] 3GPP TS 23.501, “System architecture for the 5G System (5GS); Stage 2 (Release 17)”, v17.1.1, June 2021

[15] 3GPP TS 28.530, “Management and Orchestration; Concepts, use cases and requirements, (Release 17)”, v17.1.0, March 2021

[16] 3GPP TR 23.791, “Study of Enablers for Network Automation for 5G (Release 16)”, v16.2.0, June 2019

[17] 5G-TRANSFORMER - 5G Mobile Transport Platform for Verticals, http://5g-transformer.eu/

[18]Vertical Slicer, open-source code available online: https://github.com/nextworks-it/slicer

[19] Zero-Swarm “D4.4 Federated data infra & toolkit for data-driven model.v1”

[20] 3GPP TS 23.288, “Architecture enhancements for 5G System (5GS) to support network dataanalytics services (Release 17)”, v17.1.0, June 2021

[21] 3GPP TS 29.520, “5G System; Network Data Analytics Services; Stage 3 (Release 17)”, v17.3.0, June 2021

[22] W. Lee, T. Na and J. Kim, "How to Create a Network Slice? - A 5G Core Network Perspective," in *Proceedings of the 21st International Conference on Advanced Communication Technology (ICACT)*, PyeongChang Kwangwoon_Do, Korea (South), 2019, pp. 616-619.

[23] D. Sattar and A. Matrawy, "Optimal Slice Allocation in 5G Core Networks," *IEEE Networking Letters*, vol. 1, no. 2, pp. 48-51, June 2019.

[24] D. A. Chekired, M. A. Togou, L. Khoukhi and A. Ksentini, "5G-Slicing-Enabled Scalable SDN Core Network: Toward an Ultra-Low Latency of Autonomous Driving Service," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 8, pp. 1769-1782, Aug. 2019.

[25] 3GPP; Technical Specification Group Services and System Aspects; Study on Communication for Automation in Vertical Domains (Release 16), 3GPP TR 22.804 v16.2.0, December 2018.

[26] J. García-Morales, M. C. Lucas-Estañ and J. Gozalvez, "Latency-Sensitive 5G RAN Slicing for Industry 4.0," IEEE Access, vol. 7, pp. 143139-143159, 2019.

[27] 3GPP TR 28.801, “Study on management and orchestration of network slicing for next generation network (Release 15)”, v15.1.0, January 2018

[28] 3GPP TS 28.533, “Management and Orchestration; Architecture framework (Release 16)”, v16.7.0, March 2021

[29] "Event Registration." ONAP VNF Requirements and Qualification Test Suite (VNFRQTS) 7.1 Specification. The Linux Foundation, n.d. Web. 11 Jul. 2023. <https://docs.onap.org/projects/onap-vnfrqts-requirements/en/latest/Chapter8/ves7_1spec.html#event-registration>.

[30] “TS 23.501 – System Architecture for the 5G System:” 3GPP, n.d. Web. 11 Jul. 2023. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3413>.

[31] ONAP 5G RAN Network Slicing. Available at https://wiki.onap.org/download/attachments/10784151/ONAPNetworkSlicingv3_25Jan2018.pptx?api=v2.

[32] J. García-Morales, M. C. Lucas-Estañ and J. Gozalvez, "Latency-Sensitive 5G RAN Slicing for Industry 4.0," IEEE Access, vol. 7, pp. 143139-143159, 2019.

# Appendix A: RAN, NodeH, gNB parameters

| Object Name | Parameter Name | Read/Write | Description |
|---|---|---|---|
| FaultMgmt | **FaultMgmt** | R | This object contains parameters relating to Fault/Alarm Management. |
| FaultMgmt | SupportedAlarmNumberOfEntri | R | The number of entries in the SupportedAlarm table. |
| FaultMgmt | MaxCurrentAlarmEntries | R | The maximum number of entries allowed in the FaultMgmt.CurrentAlarm.{i}. table. |
| FaultMgmt | CurrentAlarmNumberOfEntries | R | The number of entries in the CurrentAlarm table. |
| FaultMgmt | HistoryEventNumberOfEntries | R | The number of entries in the HistoryEvent table. |
| FaultMgmt | ExpeditedEventNumberOfEntri | R | The number of entries in the ExpeditedEvent table. |
| FaultMgmt | QueuedEventNumberOfEntries | R | The number of entries in the QueuedEvent table. |
| SupportedAlarm | **SupportedAlarm** | R | Supported Alarm Entries which can be raised by the device. |
| SupportedAlarm | EventType | R | Indicates the type of event. |
| SupportedAlarm | ProbableCause | R | Qualifies the alarm and provides further information than EventType . |
| SupportedAlarm | SpecificProblem | R | Provides further qualification on the alarm beyond EventType and ProbableCause. |
| SupportedAlarm | PerceivedSeverity | R | Indicates the relative level of urgency for operator attention, see [ITU-X.733]. |
| SupportedAlarm | ReportingMechanism | RW | Indicates the reporting mechanism setting of the alarm. |
| CurrentAlarm | **CurrentAlarm** | R | Contains all currently active alarms. |
| CurrentAlarm | AlarmIdentifier | R | Identifies one Alarm Entry in the Alarm List. |
| CurrentAlarm | AlarmRaisedTime | R | Indicates the date and time when the alarm was first raised by the device. |
| CurrentAlarm | AlarmChangedTime | R | Indicates the date and time when the alarm was last changed by the device. |
| CurrentAlarm | EventType | R | Indicates the type of event. |
| CurrentAlarm | ProbableCause | R | Qualifies the alarm and provides further information than EventType . |
| CurrentAlarm | SpecificProblem | R | Provides further qualification on the alarm beyond EventType and ProbableCause. |
| CurrentAlarm | PerceivedSeverity | R | Indicates the relative level of urgency for operator attention, see [ITU-X.733]. |
| CurrentAlarm | AdditionalText | R | This provides a textual string which is vendor defined. |
| CurrentAlarm | AdditionalInformation | R | This contains additional information about the alarm and is vendor defined. |
| HistoryEvent | **HistoryEvent** | R | Alarm events added or updated in FaultMgmt.CurrentAlarm.{i}. are simultaneously entered into this table. |
| HistoryEvent | EventTime | R | Indicates the date and time when the alarm event occurs. |
| HistoryEvent | AlarmIdentifier | R | Identifies one Alarm Entry in the Alarm List. |
| HistoryEvent | NotificationType | R | Indicates the reason for the specific alarm notification event. |
| HistoryEvent | ManagedObjectInstance | R | Specifies the instance of the Informational Object Class in which the alarm occurred. |
| HistoryEvent | EventType | R | Indicates the type of event. |
| HistoryEvent | ProbableCause | R | Qualifies the alarm and provides further information than EventType . |
| HistoryEvent | SpecificProblem | R | Provides further qualification on the alarm beyond EventType and ProbableCause. |
| HistoryEvent | PerceivedSeverity | R | Indicates the relative level of urgency for operator attention, see [ITU-X.733]. |
| HistoryEvent | AdditionalText | R | This provides a textual string which is vendor defined. |
| HistoryEvent | AdditionalInformation | R | This contains additional information about the alarm and is vendor defined. |

| Object Name | Parameter Name | Read/Write | Description |
|---|---|---|---|
| RAN | **RAN** | R | This object contains parameters relating to configuring the 5G FAP. |
| RAN | CUNumberOfEntries | R | The number of entries in the CU table. |
| RAN | CUUPNumberOfEntries | R | The number of entries in the CUUP table. |
| RAN | DUNumberOfEntries | R | The number of entries in the DU table. |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1. | **CU** | RW | This object contains a table of all tunnel instances configured for the LTE FAP. |
| CU | Valid | RW | Denote if CU is configured |
| CU | CuOwnIPAddress | RW | IP address of the CU |
| CU | AmfIPAddresses | RW | IP address strings for AMF. Comma-separated list of strings , maximum length 6 |
| CU | CuupIPAddresses | RW | IP address strings for CUUP. Comma-separated list of strings , maximum length 6 |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.GlobalGnbId. | **GlobalGnbId** | RW | This object contains the GNB Identity params for CU |
| GlobalGnbId | PlmnIdentity | RW | PLMN ID consists of Mobile Country Code (MCC) and Mobile Network Code (MNC). |
| GlobalGnbId | GnbIdBitsUsed | RW | GNB identity number of bits used, this can be configured to use between 22 bits to 32 bits |
| GlobalGnbId | GnbId | RW | GNB Identity, this can be configured to use between 22 bits to 32 bits |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.Security. | **Security** | RW | This object contains the security params for CU |
| Security | SecurityAlgo | RW | Security algotihm |
| Security | CipherEnable | RW | Cipering enable |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.Mobility. | **Mobility** | RW | This object contains the security params for CU |
| Mobility | ANRThresholdRSRP | RW | Threshold to enter and leave ANR Mode |
| Mobility | IntraNRThresholdRSRP | RW | Threshold to enter IntraNR Mode |
| Mobility | InterNRThresholdRSRP | RW | Threshold to enter InterNR Mode |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.Mobility.Resele | **Reselection** | RW | This object contains the security params for CU |
| Reselection | SNonIntraSearchP | RW | Signalled Value in SIB2, actual value = signalled value * 2, value 32 = don't broadcast |
| Reselection | SNonIntraSearchQ | RW | Signalled Value in SIB2, value 32 = don't broadcast |
| Reselection | ThreshServingLowP | RW | Signalled Value in SIB2, actual value = signalled value * 2 |
| Reselection | ThreshServingLowQ | RW | Signalled Value in SIB2, value 32 = don't broadcast |
| Reselection | CellReselectionPriority | RW | Signalled Value in SIB2 |
| Reselection | QRxLevMin | RW | Signalled Value in SIB2, actual value = signalled value * 2, value |
| Reselection | QRxLevMinSUL | RW | Signalled Value in SIB2, actual value = signalled value * 2, value value -71 = don't broadcast |
| Reselection | QQualMin | RW | Signalled Value in SIB2, value -44 = don't broadcast |
| Reselection | SIntraSearchP | RW | Signalled Value in SIB2, actual value = signalled value * 2 |
| Reselection | SIntraSearchQ | RW | Signalled Value in SIB2, value 32 = don't broadcast |
| Reselection | Pmax | RW | Signalled Value in SIB2, value 34 = don't broadcast |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.CarrierList. | **ListCarrier** | R | This object contains parameters relating to the neighbor list. |
| ListCarrier | NRCellNumberOfEntries | RW | The number of entries in the NRCarrier table. |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.CarrierList.NR | **NRCarrier** | RW | Table containing the LTE EUTRA (i.e. intra-RAT) cell list. |
| NRCarrier | SsbArfcn | RW | SsbArfcn of NRCarrier |
| NRCarrier | Scs | RW | Sub-Carrier Spacing in khz (15,30,60,120,240) . |
| NRCarrier | QRxLevMin | RW | Signalled Value in SIB4, actual value = signalled value * 2, value |
| NRCarrier | QRxLevMinSUL | RW | Signalled Value in SIB4, actual value = signalled value * 2, value value -71 = don't broadcast |
| NRCarrier | QQualMin | RW | Signalled Value in SIB4, value -44 = don't broadcast |

| Object Name | Parameter Name | Read/Write | Description |
|---|---|---|---|
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.CarrierList. | **ListCarrier** | R | This object contains parameters relating to the neighbor list. |
| ListCarrier | NRCellNumberOfEntries | RW | The number of entries in the NRCarrier table. |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.CarrierList.NR | **NRCarrier** | RW | Table containing the LTE EUTRA (i.e. intra-RAT) cell list. |
| NRCarrier | SsbArfcn | RW | SsbArfcn of NRCarrier |
| NRCarrier | Scs | RW | Sub-Carrier Spacing in khz (15,30,60,120,240) . |
| NRCarrier | QRxLevMin | RW | Signalled Value in SIB4, actual value = signalled value * 2, value |
| NRCarrier | QRxLevMinSUL | RW | Signalled Value in SIB4, actual value = signalled value * 2, value value -71 = don't broadcast |
| NRCarrier | QQualMin | RW | Signalled Value in SIB4, value -44 = don't broadcast |
| NRCarrier | Pmax | RW | Signalled Value in SIB4, value 34 = don't broadcast |
| NRCarrier | ThreshXHighP | RW | Signalled Value in SIB4, actual value = signalled value * 2 |
| NRCarrier | ThreshXLowP | RW | Signalled Value in SIB4, actual value = signalled value * 2 |
| NRCarrier | ThreshXHighQ | RW | Signalled Value in SIB4, value 32 = don't broadcast |
| NRCarrier | ThreshXLowQ | RW | Signalled Value in SIB4, value 32 = don't broadcast |
| NRCarrier | CellReselectionPriority | RW | Signalled Value in SIB4 value 8 = don't broadcast |
| NRCarrier | QOffsetFreq | RW | To offset the ranking of this neigbour frequency by that amount. |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.NeighborList. | **NeighborList** | R | This object contains parameters relating to the neighbor list. |
| NeighborList | NRCellNumberOfEntries | RW | The number of entries in the NRCell table. |
| InternetGatewayDevice.Services.FAPService.1.CellConfig.NR.RAN.CU.1.NeighborList.N | **NRCell** | RW | Table containing the LTE EUTRA (i.e. intra-RAT) cell list. |
| NRCell | SsbArfcn | RW | SsbArfcn or NRCell |
| NRCell | Scs | RW | Sub-Carrier Spacing in khz (15,30,60,120,240). |
| NRCell | OperatingBand | RW | Band Definition |
| NRCell | PlmnIdentity | RW | PLMN ID consists of Mobile Country Code (MCC) and Mobile Network Code (MNC). |
| NRCell | TrackingAreaCode | RW | Tracking Area Code. Corresponds to TrackingAreaCode 3 Bytes as specified in [3GPP-TS.38.331]. |
| NRCell | CellIdentity | RW | Contains the 36-bit NR Cell Global Identity (NCI). |
| NRCell | GnbIdBitsUsed | RW | GNB identity number of bits used |
| NRCell | PhyCellId | RW | Physical cell ID, as specified in [Section 7.4.2.1/3GPP-TS.38.211]. |
| NRCell | QOffsetCell | RW | To offset the ranking of this neigbour cell by that amount. |
| NRCell | QRxLevMinOffsetCell | RW | Signaled Value in SIB3 actual value = signalled value * 2, value 0 = don't broadcast |
| NRCell | QRxLevMinOffsetCellSUL | RW | Signaled Value in SIB3 actual value = signalled value * 2, value 0 = don't broadcast |
| NRCell | QQualMinOffsetCell | RW | Signaled Value in SIB3, value 0 = don't broadcast |

| Object Name | Parameter Name | Read/Write | Description |
|---|---|---|---|
| DU | **DU** | RW | This object contains the DU config params |
| DU.General | **General** | RW | This object contains the general config params for DU |
| General | Enable | RW | Enables or disables this entry. |
| General | CellIdx | RW | Cell Index |
| General | NrMode | RW | Currently only TDD supported |
| General | PhyCellId | RW | Physical cell ID, as specified in [Section 7.4.2.1/3GPP-TS.38.211]. |
| General | ChannelNrArfcn | RW | ARFCN points at center RE of channel |
| General | ChannelBwRb | RW | Width of channel in terms of RB |
| General | OperatingBandsDl | RW | Comma-separated list of strings. Band Definition Downlink |
| General | OperatingBandsUl | RW | Band Definition Uplink |
| General | MaxMimoLayers | RW | Max number of downlink MIMO Layers |
| General | MaxUlMimoLayers | RW | Max number of uplink MIMO Layers |
| General | TxPowerPerPathDBmx10 | RW | Tx power per path in 0.1dB |
| General | DuOwnIPAddress | RW | IP address of the DU |
| General | SyncSource | RW | Sync Source Selected |
| General | FreqSyncTimeout | RW | Timer value for supervisory of the Frequence Sync duration. 0 meas disabled.On expiry Alarm is raised. |
| DU.MAC | **MAC** | RW | This object contains the MAC related config params of DU |
| MAC | SearchSpace0 | RW | Width of channel in terms of RB |
| MAC | BwpDlNumberOfEntries | R | The number of entries in the BwpDl table. |
| MAC | BwpUlNumberOfEntries | R | The number of entries in the BwpUl table. |
| MAC | NmacTestMode | RW | Test modes for NMAC |
| DU.MAC.TDD | **TDD** | RW | This object contains the TDD relevant parameters of DU / MAC |
| TDD | TDDSlotsPeriod | RW | The number of slots in a single TDD Period |
| TDD | TDDSlotsDL | RW | The number of downlink slots in a single TDD Period |
| TDD | TDDSlotsUL | RW | The number of uplink slots in a single TDD period |
| BwpDl | **BwpDl** | RW | This object contains the MAC related BWP-DL config params of DU |
| BwpDl | Nrb | RW | Width of channel in terms of RB |
| BwpDl | Scs | RW | Sub-Carrier Spacing in khz (15,30,60,120,240) |
| BwpDl | OffsetToCrb0 | RW | Offset to Control Resource Block 0 |
| BwpUl | **BwpUl** | RW | This object contains the MAC related BWP-UL config params of DU |
| BwpUl | Nrb | RW | Width of channel in terms of RB |
| BwpUl | Scs | RW | Sub-Carrier Spacing in khz (15,30,60,120,240). |
| BwpUl | OffsetToCrb0 | RW | Offset to Control Resource Block 0 |

| Measurement Name | Description |
|---|---|
| **Memory Usage** | **Number of memory bytes that are used or free.** |
| NR.Memory.TotalUsed | The total number of kilobytes that are currently used in the on-board RAM memory |
| NR.Memory.TotalFree | The total number of kilobytes that are currently free (unused) in the on-board RAM memory |
| **Flash Usage** | **Number of FLASH bytes that are used or free in the measurement period.** |
| NR.Flash.TotalUsed | The total number of kilobytes that are currently used in the non-volatile FLASH memory |
| NR.Flash.TotalFree | The total number of kilobytes that are currently free (unused) in the non-volatile FLASH memory |
| **CPU Usage** | **Percentage of CPU usage** |
| NR.CPU.AverageLoad | The CPU load in percent |
| **Data Interface** | **Data transmitted and received on the LAN/Ethernet interfaces.** |
| NR.Interface.TotalBytesTx | The total number of kilobytes that have been transmitted (in the uplink direction). |
| NR.Interface.TotalBytesRx | The total number of kilobytes that have been received (in the downlink direction). |
| NR.Interface.DataRateTx | The average data rate for bytes transmitted (in the uplink direction). |
| NR.Interface.DataRateRx | The average data rate for bytes received (in the downlink direction). |
| **Connectivity and Availability Time** | **Various system timers accumulated during the measurement period.** |
| NR.SystemTime.TotalUpTime | The number of seconds that the device has been running for since the last reboot. |
| NR.SystemTime.Total5GAvailabilityTime | The total number of seconds that the 5G has been transmitted |
| NR.SystemTime.Total5GOutageTime | The total number of seconds that the 5G has not been transmitted |
| **Radio resource utilization** | |
| **DL Total PRB Usage** | **Total usage (in percentage) of physical resource blocks (PRBs) on the downlink for any purpose.** |
| RRU.PrbTotDl | Total usage (in percentage) of physical resource blocks (PRBs) on the downlink for any purpose. |
| **UL Total PRB Usage** | **Total usage (in percentage) of physical resource blocks (PRBs) on the uplink for any purpose.** |
| RRU.PrbTotUl | Total usage (in percentage) of physical resource blocks (PRBs) on the uplink for any purpose. |
| **PDU Session Management** | |
| **Number of PDU Sessions requested to setup** | **Number of PDU Sessions by the gNB.** |
| SM.PDUSessionSetupReq | Split into subcounters per S-NSSAI. |
| **Number of PDU Sessions successfully setup** | **Number of PDU Sessions successfully setup by the gNB from AMF.** |
| SM.PDUSessionSetupSucc | Split into subcounters per S-NSSAI. |
| **Number of PDU Sessions failed to setup** | **Number of PDU Sessions failed to setup by the gNB.** |
| SM.PDUSessionSetupFail.Sum | Sum of all failure causes |
| SM.PDUSessionSetupFail.RadioNetworkUnspecified | Failue cause RadioNetworkUnspecified |
| SM.PDUSessionSetupFail.RadioNetworkSliceNotSupported | Failue cause RadioNetworkSliceNotSupported |
| SM.PDUSessionSetupFail.MultiplePduSessionIdInstances | Failue cause MultiplePduSessionIdInstances |
| **Mobility Management** | |
| **Inter-gNB handovers** | |
| **Number of successful legacy handover executions** | **Provides the number of successful legacy handover executions received by the source gNB.** |
| MM.HoExeInterSucc | Number of successful legacy handover executions received by the source gNB. |
| **Number of failed legacy handover executions** | **Provides the number of failed legacy handover executions for asource gNB.** |
| MM.HoExeInterFail.UeCtxtRelCmd.Sum | Sum of number of failed legacy handover executions for all causes |
| MM.HoExeInterFail.UeCtxtRelCmd.UeCtxtRelCmdAllCauses | Number of failed legacy handover executions for asource gNB for all causes due to UeCtxtRelCmd |
| MM.HoExeInterFail.UeCtxtRelCmd.RrcReestabReq | Number of failed legacy handover executions for asource gNB with cause RrcReestabReq |
| MM.HoExeInterFail.UeCtxtRelCmd.HoExeSupTimer | Number of failed legacy handover executions for asource gNB with cause HoExeSupTimer |
| MM.HoExeInterFail.UeCtxtRelCmd.RetrUeTtxtReq | Number of failed legacy handover executions for asource gNB with cause RetrUeTtxtReq |

| Measurement Name | Description |
|---|---|
| **RRC connection establishment related measurements** | |
| **Attempted RRC connection establishments** | **Number of RRC connection establishment attempts for each establishment cause.** |
| RRC.ConnEstabAtt.Sum | Total number of RRC connection establishment attempts |
| RRC.ConnEstabAtt.Emergency | Number of RRC connection establishment attempts for establishment cause Emergency |
| RRC.ConnEstabAtt.HighPriorityAccess | Number of RRC connection establishment attempts for establishment cause HighPriorityAccess |
| RRC.ConnEstabAtt.MtAccess | Number of RRC connection establishment attempts for establishment cause MtAccess |
| RRC.ConnEstabAtt.MoSignalling | Number of RRC connection establishment attempts for establishment cause MoSignalling |
| RRC.ConnEstabAtt.MoData | Number of RRC connection establishment attempts for establishment cause MoData |
| RRC.ConnEstabAtt.MoVoiceCall | Number of RRC connection establishment attempts for establishment cause mo-VoiceCall |
| RRC.ConnEstabAtt.MoVideoCall | Number of RRC connection establishment attempts for establishment cause mo-VideoCall |
| RRC.ConnEstabAtt.MoSms | Number of RRC connection establishment attempts for establishment cause mo-SMS |
| RRC.ConnEstabAtt.MpsPriorityAccess | Number of RRC connection establishment attempts for establishment cause mps-PriorityAccess |
| RRC.ConnEstabAtt.McsPriorityAccess | Number of RRC connection establishment attempts for establishment cause mcs-PriorityAccess |
| RRC.ConnEstabAtt.Spare | Number of RRC connection establishment attempts for establishment cause spare |
| **Successful RRC connection establishments** | **Number of successful RRC establishments for each establishment cause.** |
| RRC.ConnEstabSucc.Sum | Total number of successful RRC connection establishments |
| RRC.ConnEstabSucc.Emergency | Number of successful RRC connection establishments for establishment cause Emergency |
| RRC.ConnEstabSucc.HighPriorityAccess | Number of successful RRC connection establishments for establishment cause HighPriorityAccess |
| RRC.ConnEstabSucc.MtAccess | Number of successful RRC connection establishments for establishment cause MtAccess |
| RRC.ConnEstabSucc.MoSignalling | Number of successful RRC connection establishments for establishment cause MoSignalling |
| RRC.ConnEstabSucc.MoData | Number of successful RRC connection establishments for establishment cause MoData |
| RRC.ConnEstabSucc.MoVoiceCall | Number of successful RRC connection establishments for establishment cause mo-VoiceCall |
| RRC.ConnEstabSucc.MoVideoCall | Number of successful RRC connection establishments for establishment cause mo-VideoCall |
| RRC.ConnEstabSucc.MoSms | Number of successful RRC connection establishments for establishment cause mo-SMS |
| RRC.ConnEstabSucc.MpsPriorityAccess | Number of successful RRC connection establishments for establishment cause mps-PriorityAccess |
| RRC.ConnEstabSucc.McsPriorityAccess | Number of successful RRC connection establishments for establishment cause mcs-PriorityAccess |
| RRC.ConnEstabSucc.Spare | Number of successful RRC connection establishments for establishment cause spare. |
| **Failed RRC connection establishments** | **Number of failed RRC establishments, this measurmenet is split into subcounters per failure cause.** |
| RRC.ConnEstabFailCause.Sum | Total number of RRC establishment failures |
| RRC.ConnEstabFailCause.NetworkReject | Number of RRC establishment failures for establishment cause NetworkReject |
| RRC.ConnEstabFailCause.NoReply | Number of RRC establishment failures for establishment cause NoReply |
| RRC.ConnEstabFailCause.Other | Number of RRC establishment failures for establishment cause Other |
| **RRC Connection Re-establishment** | |
| **Number of RRC connection re-establishment attempts** | **Number of RRC connection re-establishment attempts.** |
| RRC.ReEstabAtt | Number of RRC connection re-establishment attempts. |
| **Successful RRC connection re-establishment with UE context** | **Successful number of RRC connection re-establishment when UE context can be retrieved.** |
| RRC.ReEstabSuccWithUeContext | Successful number of RRC connection re-establishment when UE context can be retrieved. |
| **Successful RRC connection re-establishment without UE context** | **Successful number of RRC connection re-establishment when UE context can not be retrieved.** |
| RRC.ReEstabSuccWithoutUeContext | Successful number of RRC connection re-establishment when UE context can not be retrieved. |

| Measurement Name | Description |
|---|---|
| **Number of Active Ues** | |
| **Mean number of Active UEs in the DL per cell** | **Mean number of active DRBs for UEs in an NRCellDU.** |
| DRB.MeanActiveUeDl | Calculated per PLMN ID, per QoS level and per supported S-NSSAI. |
| **Maximum number of Active UEs in the DL per cell** | **Maximum number of active DRBs for UEs in an NRCellDU.** |
| DRB.MaxActiveUeDl | Calculated per PLMN ID, per QoS level, and per supported S-NSSAI. |
| **Mean number of Active UEs in the UL per cell** | **Mean number of active DRBs for UEs in an NRCellDU.** |
| DRB.MeanActiveUeUl | Calculated per PLMN ID, per QoS level, and per supported S-NSSAI. |
| **Max number of Active UEs in the UL per cell** | **Maximum number of active DRBs for UEs in an NRCellDU.** |
| DRB.MaxActiveUeUl | Split into subcounters per QoS level, and subcounters per S-NSSAI. |
| **DL PDCP SDU Data Volume Measurements** | |
| **DL Cell PDCP SDU Data Volume** | **Data Volume (amount of PDCP SDU bits) in the downlink delivered to PDCP layer.** |
| DRB.PdcpSduVolumeDL | Calculated per PLMN ID and per QoS level (mapped 5QI) and per S-NSSAI. The unit is Mbit. |
| **UL PDCP SDU Data Volume Measurements** | |
| **UL Cell PDCP SDU Data Volume** | **Data Volume (amount of PDCP SDU bits) in the uplink delivered from PDCP layer to higher layers.** |
| DRB.PdcpSduVolumeUL | Calculated per PLMN ID and per QoS level (mapped 5QI) and per S-NSSAI. The unit is Mbit. |
| **IP Latency measurements** | |
| **General information** | |
| **Average IP Latency DL in gNB-DU** | **Average IP Latency in DL (arithmetic mean) within the gNB-DU.** |
| DRB.RlcSduLatencyDl | Split into subcounters per QoS level and subcounters per S-NSSAI. |