Decentralized AI and Architectures for Massive Wireless Network Slicing Scalability and Sustainability in 6G

RESILIENT

Grant No. TSI-063000-2021-55

E4: Final release of the use case requirements, KPIs and 6G DAWN architecture (security)

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

UNICO I+D

GOBIERNO DE ESPAÑA

cttc Centre Tecnològic de Telecomunicacions de Catalunya

Abstract

This report develops the 6G DAWN architecture as defined in E3 by focusing on its key concepts by analyzing state of the art and focusing on the architectural contributions of the project beyond state of the art. Though there are global concepts that are transversal to both subprojects, 6G DAWN-RESILIENT puts more emphasis on security/trust management.

Document properties

Document number	E4
Document title	Final release of the use case requirements, KPIs and 6G DAWN architecture (security)
Document responsible	Selva Vía (CTTC)
Document editor	Selva Via (CTTC), Engin Zeydan (CTTC)
Authors	Engin Zeydan (CTTC), Albert Bel (CTTC), Josep Mangues-Bafalluy (CTTC), Farhad Rezazadeh (CTTC), Luis Blanco (CTTC), Sarang Kahvazadeh (CTTC), Farhana Javed (CTTC), Oriol Font (SRS), David Gregoratti (SRS), Ismael Gomez (SRS), Manuel Lorenzo (Ericsson), Saravanan Kalimuthu (Ericsson), Gokhan Turhan (Ericsson), Rodrigo Díaz Rodríguez (ATOS), Hristo Koshutanski (ATOS), Ignacio Labrador Pavón (ATOS), Sonia Castro (ATOS), Jaime Azcorra (Telcaria), Aitor Zabala (Telcaria)
Target dissemination level	Public
Status of the document	Final
Version	1.0
Delivery date	31 October 2024
Actual delivery date	31 October 2024

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

Disclaimer

This document has been produced in the context of the 6G DAWN Project. The research leading to these results has received funding from the Ministerio de Asuntos Económicos y Transformación Digital (MINECO), under grant TSI-063000-2021-55.

All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

For the avoidance of all doubts, the MINECO has no liability in respect of this document, which is merely representing the author's view.

Contents

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

List of Figures

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

UNICO I+D

GOBIERNO DE ESPAÑA

cttc Centre Tecnològic de Telecomunicacions de Catalunya

List of Tables

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

List of Acronyms

3GPP – 3rd Generation Partnership Project

5G-ACIA - 5G Alliance for Connected Industries and Automation

ACT- Actuator

AD – Anomaly Detection

AE – Analytics Engine

AF – Application Function

AI – Artificial Intelligence

B5G – Beyond 5G

CaaS – Container as a Service

CQI – Channel Quality Indicator

CNF – Containerized Network Function

COTS – Commercial Off-The-Shelf

CPU - Central Processing Unit

CSP – Communications Service Provider

CSI – Channel State Information

CU- Central Unit

DE – Decision Engine

DL – Downlink

DMO - Domain Manager and Orchestrator

DU- Distributed Unit

E2E – End To End

EC – Energy Consumption

eMBB – Enhanced Mobile Broadband

ETSI – European Telecommunications Standards Institute

gNB - next Generation Node B (5G node B)

GNSS – Global Navigation Satellite system

GPS – Global Positioning System

ICMP – Internet Control Message Protocol

IDMO- Inter-Domain Manager and Orchestrator

ILE- Infrastructure Layer Emulator

IRU – Indoor Radio Unit

ISPM – Infrastructure Status Prediction Module

KPI – Key Performance Indicator

KPM- Key Performance Measurement

MCData – Mission Critical Data

MCPTT – Mission Critical Push To Talk

MCVideo – Mission Critical Video

MCX – Mission Critical services

ML – Machine Learning

MPLS – Multi-Protocol Label Switching

M&O- Management and Orchestration

MS – Monitoring System

NDT – Network Digital Twin

NEF – Network Exposure Function (3gpp)

NETCONF – Network Configuration Protocol

NF – Network Function

NPN- Non-Public Networks

NR – New Radio

NWDAF – Network Data Analytics Function

OAM – Operations, Administration and Maintenance

O-RAN – Open Radio Access Network

PoC- Proof of Concept

OTA – Over-the-Air

PCF – Policy Control Function

PN – Public Network (3gpp)

PNI-NPN – Public Network Integrated Non-Public Network (3gpp)

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

PPDR – Public Protection & Disaster Relief

PT-Paquete de Trabajo (Work package)

QoE- Quality of Experience

QoS – Quality of Service

RAN – Radio Access Network

RAN WG3 – RAN (Radio Access Network) Workgroup 3

R- RESILIENT

RB – Resource Block

RC – Radio Controller

RDI – Radio Dot Interface

RDS – Radio Dot System

RIC - RAN Intelligent Controller

RF – Radio Frequency

RLC – Radio Link Control

RT- Real Time

RTT - Round-Trip Time

SA1 – System Aspects Workgroup 1

SDN – Software Defined Networking

SDR-Software Define Radio

TDD – Time Division Duplex

TETRA – Terrestrial Trunked Radio

TS – Technical Specification

TR – Technical Report

UC- Use Case

UE – User Equipment

UL – Uplink

UPF- User Plane Function

URLLC – Ultra-reliable Low-Latency Communication

USA – United States of America

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

USRP - Universal Software Radio Peripheral

VPN – Virtual Private Network

WG- Working Group

xApp-Cross Application

YANG – Yet Another Next Generation

ZDM – Zero-Defect Manufacturing

ZSM – Zero-touch Service Management

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

Executive Summary

Document E3 presented the "First release of the use case requirements, KPIs and 6G DAWN architecture." This represented the framework over which the project evolved, and has proven to have been correctly defined, since no modifications were needed after that. Therefore, after E3, the work focused on confirming this initial framework and on further developing its key concepts by carrying out a detailed analysis of state of the art and progress beyond. This outcome is then fed into the architecture and the various PoCs of the project. In this sense, this document presents the state of the art of the key technical concepts that the project deals with, presenting their evolution because of the 6G DAWN contributions. Concepts that are considered global are presented both for 6G DAWN ELASTIC and 6G DAWN RESILIENT subprojects, but there is also a specific section for the technical concepts for the RESILIENT subproject, with security one of the main themes.

The following list presents the key concepts the project tackles, both globally, or specifically for each sub-project.

Global key concepts:

- NPN Digital Twin System
- Extreme Edge
- AI/ML agent for control loops
- xApps in O-RAN
- Relation of vertical KPIs with the network configuration
- Inter(a)-slice reconfiguration and massive slicing
- NEF instance for KPI data and configuration capabilities exposures of NPNs
- AI/ML methods for reducing energy consumption

RESILIENT specific key concepts:

- Develop standard-compliant network interfaces to support AI driven network fault management systems at NPN
- Decentralized Intrusion Detection to integrate novel trust-based evaluation mechanisms
- Blockchain in 5G
- Identification and handling of abnormally heavy energy components

The main contribution in each of the architectural key concepts is discussed in detail in the corresponding section and all contributions are summarized in the conclusions section.

The key concepts are also mapped to the PoCs in which they will be applied in section 4.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

1 Introduction

This document is entitled "Final release of the use case requirements, KPIs and 6G DAWN architecture". The previous deliverable E3 "First release of the use case requirements, KPIs and 6G DAWN architecture" [6GDE3] presented the initial architecture that the project implements in the various Proof of Concepts (PoCs). The requirements and the KPIs per use case were also detailed. No significant refinement has been identified in the use cases, architecture, and PoC definition, including requirements and KPIs, hence considering what it was presented on [6GDE3] as final. As a consequence, this document focuses on developing in detail the key concepts that stem from previous work in the project so as to finalize all the details of the architectural work and to identify the 6G DAWN contributions beyond state of the art.

Technical key concepts that are considered global to both subprojects 6G DAWN ELASTIC and 6G DAWN RESILIENT are presented in Section 2. Section 3 deals with the specific topics in the RESILIENT sub project having security, trust and resilience as core themes. Section 4 presents the mapping among PoCs and concepts, identifying which proof of concept deals with a given technical concept. Finally, section 5 presents the conclusions.

2 6G DAWN Global Key Concepts

In Section 2 we are going to present the global key concepts that have been studied in 6G DAWN project, both in ELASTIC and RESILIENT, analyzing the state of the art when the project has started, and 6G DAWN contributions beyond this state of the art.

2.1 NPN Digital Twin System

The Network Digital Twin (NDT) concept we refer to in this project embodies the realization of a Digital Twin for 5G Network systems, and, more specifically, a Digital Twin for a Non-Public Network (NPN) [5GACIA] gNB system. NDT sits, thus, in the convergence of Digital Twin and Mobile Networks concepts [NDTERI], where the Digital Twin approach is applied to different use cases of network: from R&D to network operation (management, deployment, and site engineering). The aim is to achieve an accurate real-time representation of a physical network entity creating a virtual image of a real network element and then use it to simulate different scenarios and identify the most suitable one to the requirements and patterns of traffic of the users/devices enjoying mobile network connectivity services in an NPN.

Within this scope and context, the Network Digital Twin overall abstract model represented in Figure 1 is characterized by a set of complementary aspects of the physical system and environment of a specific NPN system, which the corresponding NPN digital instance is aware of, namely, Spectrum Assignment, Spectrum usage technique, Radio Unit type, gNB configurations, air conditions, KPIs and Traffic Models.

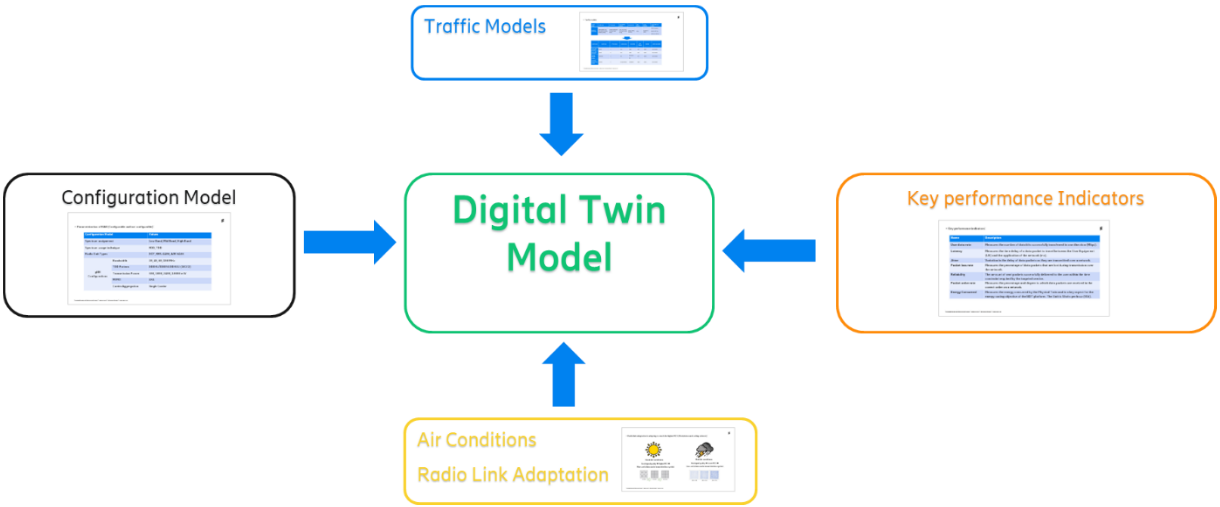


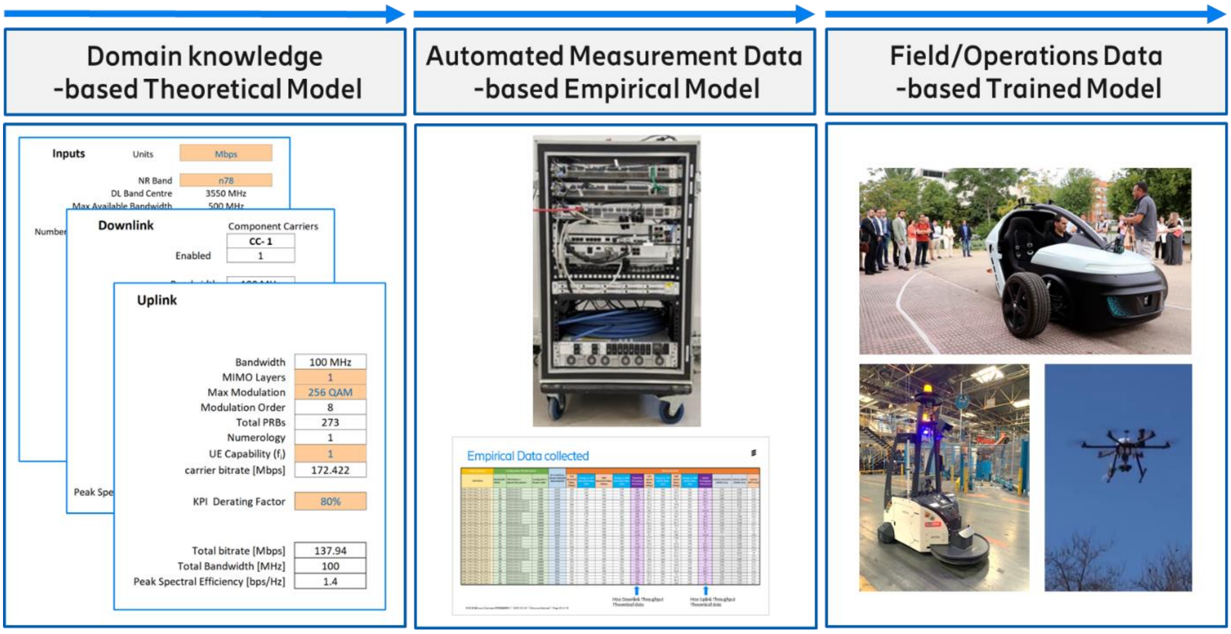
FIGURE 1: NETWORK DIGITAL TWIN

When it comes to the key use cases supported by the NDT platform they involve i) Configuration proposal for Physical Twin based in service requirements, that delivers rank ordered list of feasible configurations that match the requirements of user service, ii) Configuration implementation in the Physical Twin based in service requirements, securing the actual implementation and necessary

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

reconfiguration of the system accordingly, and iii) Performance comparison between Physical and Digital Twin performance, providing rich information about how the service is running vs expectations.

Then, a foundational design principle for the NDT platform is the integration of three sources of knowledge for making up robust and accurate models of the twinned NPN instances.


Domain knowledge -based Theoretical Model
Automated Measurement Data -based Empirical Model
Field/Operations Data -based Trained Model
Empirical Data collected

Inputs Units Mbps
NR Band n78
DL Band Centre 3550 MHz
Max Available Bandwidth 500 MHz
Number
Downlink Component Carriers CC- 1
Enabled 1
Peak Spe
Uplink
Bandwidth 100 MHz
MIMO Layers 1
Max Modulation 256 QAM
Modulation Order 8
Total PRBs 273
Numerology 1
UE Capability (f_j) 1
carrier bitrate [Mbps] 172.422
KPI Derating Factor 80%
Total bitrate [Mbps] 137.94
Total Bandwidth [MHz] 100
Peak Spectral Efficiency [bps/Hz] 1.4

FIGURE 2: FROM DOMAIN KNOWLEDGE MODELLING TO MODEL TRAINING FOR NDT

As illustrated in Figure 2, the NDT platform is proposed to be fed with i) a domain knowledge model, ii) an empirical model, and iii) a data-driven trained model. That combined strategy allows for respectively i) dealing with expectable ranges for performance, ii) predicting performance levels for multiple potential configurations and finally iii) accurately predicting performance for the system in its current configuration, condition, and usage, based on training.

Finally, the proposed NDT platform embodies a set of services that may be either directly used by the CSP or invoked and composed with other services, by several stakeholders of the NPN it serves. This provides the NPN ecosystem with a powerful tool to gain observability of the NPN and devise new services for the customization, monetization, optimization, and sustainability of the NPN. We propose to i) base the exposure of NDT services on interoperable standard interfaces whenever possible (in practice, whenever a similar service is already defined by 3GPP), and ii) provide open access APIs to access NDT platform services that lay beyond the scope and plans of mobile networks interoperable standards, for dealing with NPN NDT specific services exposure.

2.1.1 State of the Art

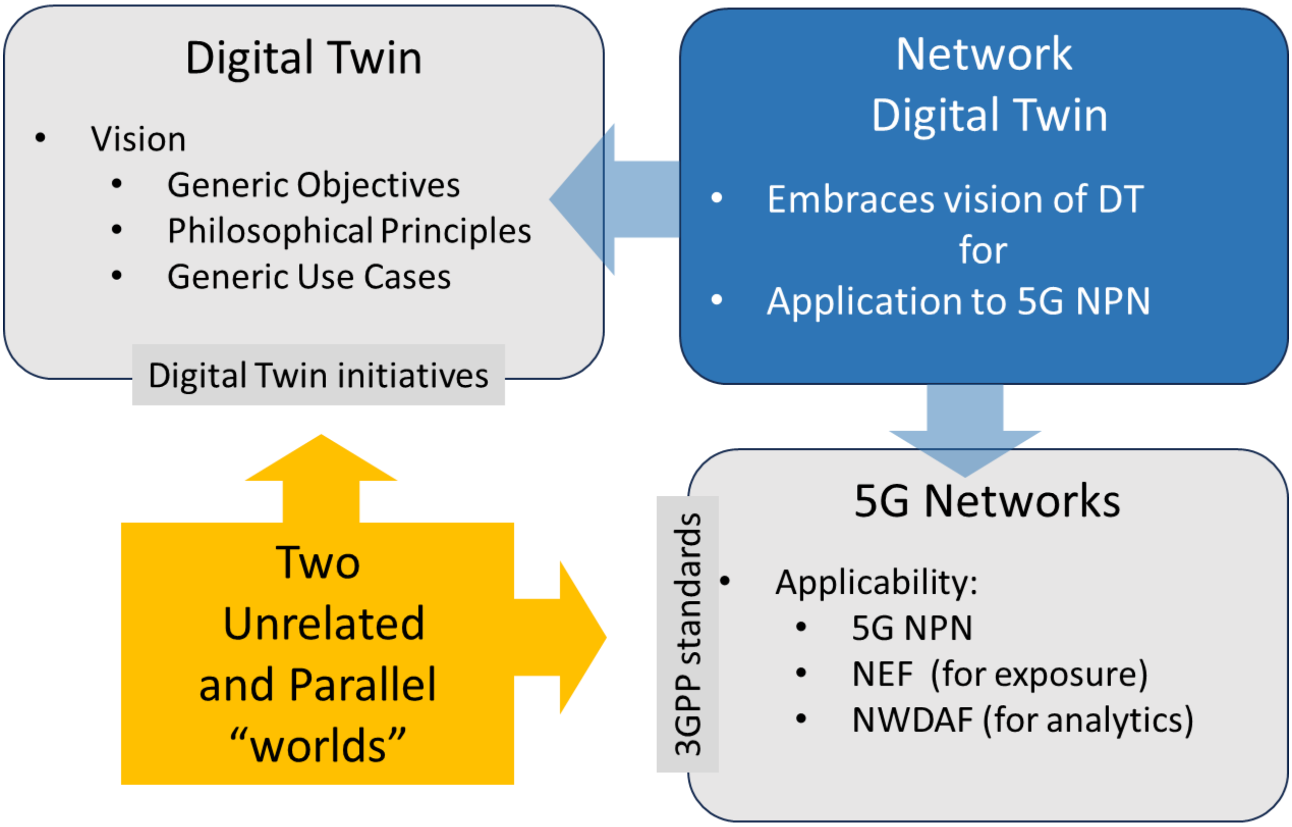


FIGURE 3: SOTA CONTEXT FOR NETWORK DIGITAL TWIN

As represented in Figure 3, the "worlds" of Digital Twin (principles) and Network technologies (standards) are indeed unrelated, evolving in parallel tracks, and governed by different dynamics. So, the Network Digital Twin concept is the first convergence of those two worlds and their trends into an innovative concept that stemming from and embracing the vision of Digital Twin, it is meaningful, feasible, and applicable, in particular, for 5G NPN systems.

For a structured analysis, trends in the two complementary worlds (Digital Twin, and Networks) that are instrumental to the overarching concept of Network Digital Twin, shall be analyzed now.

Analysis from the Digital Twin space

The digital twin concept could be defined as a digital representation of a real-world object synchronized at a specified periodicity and fidelity. This original concept where systems were mirrored to monitor unreachable physical spaces was used by NASA in 1970 for the Apollo missions, among others. This potential of virtual and simulated models has been considered as precursors of Digital Twin (DT), however, are not considered proper DTs, as lack of seamless connections and real-time data exchange allowing the periodic "twinning" of the digital to physical [GRIEVE].
The DT concept evolved and, in 2002, Michael Grieves described an accurate real use case applied to manufacturing [OLCO]. In this white paper, the DT concept is defined as containing the following

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

three primary elements: the physical object in real space, the simulated object in virtual space (with virtual sub-spaces), and the link for data flow between the real and virtual spaces.
Digital Twin is used, today, in industries such as aerospace, automotive, energy, and manufacturing helping to improve processes and products. Despite its wide -and fragmented- applicability and momentum, and because of the many diverse ecosystems the DT approach is embraced in, no common industry standards are addressing a specification of this phenomenon. Instead, only general ambition principles are discussed at several associations and institutes such as Digital Twin Consortium [OLCO], and reflected in articles in journals [BARR].

Analysis from the Mobile Network technologies domain

No 3GPP standards (specs) are defined or planned to be defined in the near future for a full-fledged all-purpose Network DT platform as such. A possible reason behind this could be that an eventual 3GPP specification for a Network Digital Twin platform should cater to creating, maintaining, and exposing Digital Twin instances for any type of 3GPP network, so the complexity involved in such type of functionality for complete (physical) 5G networks, with potentially millions of users, multiple traffic models and heterogeneous usage patterns, thousands of apps and many thousands of square kilometers of coverage is simply overwhelming and clearly not addressable for the near future.
That said, within the family of specs of 3GPP the closest one to be supporting some of the basic requirements of Network Digital Twin is that of NWDAF, since the role of this NF in the 5GS is to analyze QoE and provide insights for enabling automation in the Core for reducing CAPEX and OPEX whilst safeguarding QoE target levels. So, in many ways, core processes assumed to be executed by a NWDAF function, including continuous performance measurement collection and carrying out real-time network data analytics, are also required for a Network Digital Twin implementation. However there's a critical space of requirements that NWDAF specs do not cover vs the intent of NPN Network Digital Twin encompassing i) creating, maintaining, and exposing Digital Twin Instances for simulation and planning purposes (Digital Twin principle), ii) the specific focus of application to NPNs (physical system to be twinned), and iii) generating insights, recommendations and enforcing actuations beyond the core and into the RAN, as major influential network segment for NPN performance optimization (domain of analysis and actuation).
Finally, 3GPP NEF specs are also adjacent to the NDT space, considering the possibility of exposing NDT capabilities to AFs. Here there's a two-way influence. On the one hand, NEF specs for Analytic Exposure are south-bound supported by the NDT platform, for that purpose; and, on the other hand, the capabilities for RAN configuration supported by NDT could be exposed via NEF-type (not standardized) interfaces. In summary, NEF is the proposed facade for interoperable exposure of NDT capabilities to AFs, through the standardized interfaces whenever available, and adequately complemented by new interfaces to be developed for non-3gpp-supported capability exposure in the scope of NDT.

2.1.2 Beyond the State of the Art in 6G DAWN Context

The key technological aspects enclosed in the NDT concept being researched that go beyond the SotA are:

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

- It implements the full set of digital twin functionalities from performance monitoring and analysis to recommendations and actuation -i.e. this NDT platform provides a comprehensive one-stop shop innovative Network Digital Twin service developed ahead of eventual standardization in the mid-term.
- It specifically addresses PNI-NPN scenarios, modeling NPN technology and performance vs the typical usage patterns and environmental conditions applying to this type of environment, i.e. this NDT Platform is designed for dealing with and learning from, specific scenarios and needs not addressed by generic analytics tools. Also, given the NPN context, the NDT platform incorporates and exposes special network configuration capabilities for performance optimization purposes -a service that makes full sense for NPNs as opposed as for PNs, where it is neither applicable nor feasible in practice.
- It leverages three complementary models for NPN performance (theoretical model, empirical model and ML model) corresponding respectively to the model of ranges of performance supported by 5G for a broad set of NPN configurations, the model of actual performance of the system measured in the lab for multiple configurations and usage patterns, and, finally, the ML model derived, and improved continuously, from the actual usage, behavior and performance of the system in the field. (See resources).
- It integrates with 5G network architecture at the 5G Core, as any other 5GC NF, i.e. it is designed to be deployed in the network itself, thus gaining privileged access to other NFs and resources. This approach is similar to -and inspired by- that of other standardized 3GPP NFs (such as NEF and NWDAF) which must interact with both 5G network and AFs in order to effectively fulfil its mission.
- It supports deploying both autonomic close-loop strategies (embedded) and open-ended (co-operative) approaches with upper layer functions and applications (open exposure of APIs) for network observability and smart ML-based optimization, i.e. this NDT platform is versatile, modular and open.

2.2 Extreme-Edge

As already described in the previous Deliverable E3 [6GDE3], the Extreme-Edge domain refers to those resources in the Network Continuum beyond the technical and the administrative domains of a specific stakeholder, also part of that network continuum. As a whole, this "extreme-edge" concept refers to the deployment of advanced computational capabilities at the very edge of a network, in a highly distributed way. This concept goes beyond traditional edge computing, which typically involves processing data close to where it is generated but still within a well-structured and controlled infrastructure that commonly belongs to a specific stakeholder. Extreme-edge computing pushes this boundary further, enabling processing, decision-making, and intelligence in remote, mobile, or harsh environments. Besides, this Extreme-Edge domain may contain physical or virtualized network infrastructure resources (compute, storage, and networking resources) belonging to multiple stakeholders, e.g., multiple network operators, vertical industries, hyper-scalers, neutral-hosts providers, or even end-users. The integration of this extreme-edge domain as part of the network management and orchestration (M&O) processes is considered a relevant innovative topic in the European 6G research flagship projects [HEX][HEX2].

One of the main challenges associated with the integration of this extreme-edge is the high heterogeneity of devices in such domain, which besides the regular networking infrastructure already in the 5G core and edge networks (e.g., data center and distributed servers, routers, load balancers, storage systems, switches, etc.) may also include a very high number and technologically diverse set of end-user devices, such as industrial devices, devices in vehicles, AR/VR devices, IoT devices, home appliances, smart city systems, drones, etc. Many of those devices can be also battery-powered, or with reduced computing or storage capabilities. But beyond accessing those devices to send them simple commands or gather data from them (something already feasible in the 5G technology), the main challenge towards 6G is the possibility to "orchestrate" network service components on those infrastructure resources as it is already done in the 5G edge and core networks. The rationale behind that is that, in many cases, these devices may have significant computing and storage capabilities, which can allow to extend the pool of the available infrastructure resources from the edge (already reachable in 5G) to the extreme edge. This will also enhance the regular edge capabilities in terms of latency, since the extreme-edge devices are even closer to the end users. Also, in the possibility to offload certain computing tasks on those devices (e.g., AI/ML-related intensive computing tasks), which may also help to reduce the data transmission needs (if certain computing task can be performed directly on an extreme-edge device, there is no need to transmit a big amount of data to the regular edge or core networks).

However, this integration of the extreme-edge domain requires to re-think the M&O system towards 6G. We are not talking anymore about deploying (orchestrating) services on a static set of infrastructure resources belonging to a specific MNO and deployed on strictly controlled and supervised facilities. On the contrary, resources at the Extreme-Edge can be highly volatile: they could unexpectedly connect/disconnect, move, or unexpectedly change their available computing/storage/networking resources. Also, this would be a cloud-native-like massive in scale ecosystem and, as mentioned, with devices belonging to a multiplicity of stakeholders (and not just the MNOs). All this obviously poses relevant challenges when it comes to managing and orchestrating network services on this infrastructure with the appropriate QoS and QoE levels. These challenges can only be addressed by evolving the current 5G M&O components and mechanisms, incorporating new features tailored to deal with this huge, diverse, and highly dynamic environment.

2.2.1 State of the Art

Given a generic network with its constituent nodes, the concept of edge is intuitively associated with the subset of nodes that are located close to the external limit of such network. These edge nodes, in some cases, are considered to have special features, e.g., by playing the role of providing access to the network for users outside the network (access nodes) or by being able to enable low latency interactions with external users, because users and edge nodes could be physically located close each other. In this context, the Edge Computing concept is understood as a network architecture concept that enables cloud computing capabilities and service environments, which are deployed close to the UE, promising several benefits such as lower latency, higher bandwidth, reduced backhaul traffic, and prospects for new services compared to the cloud environments [23.558].

In parallel to this concept of the edge, the more specific "extreme-edge" concept is also used to refer to those outermost edge nodes, i.e., those located at the very edge of the network. This is the case, e.g., in the White Paper for Research Beyond 5G [NW-15], where “extreme edge” network functions are early mentioned as playing a role “in the area of access network and user equipment (UE) or user devices (UD)”. Also, in [PWB16], that mentions the extreme edge concept associated with a specific edge computing platform (called ParaDrop) providing computing and storage resources at such extreme edge of the network, to allow developers to flexibly create new types of services. In this case, that extreme edge consisted of the Wi-Fi Access Points (AP) or the wireless gateways through which all end-device traffic (from personal devices, sensors, etc.) was passing through. For this, the justification for using these extreme-edge nodes was because the Wi-Fi AP had unique contextual information about end-devices (e.g., regarding proximity or channel characteristics) that were lost in devices located deeper into the network. This concept of "extreme-edge" referring to components at the extreme edge of the network, or even to components connected to the network already at the user domain is also sometimes referred to as "far-edge" [FUT21], or also as the “fog” [HV19] or the “mist” [ACC-19], trying to convey the idea of a “cloud” that would be in a very close proximity to the end-user.

The same extreme-edge concept has also been used in the IoT technology context, where the IoT devices are understood as the outermost edge nodes in the network. E.g., [PMJ+19] defines the extreme-edge as those IoT devices where the most essential and limited tasks (identified as the sensing and sending data related tasks) are carried out, while the processing of data is performed at the so-called “regular” edge. Other examples of references to the "extreme-edge" concept in the context of the IoT technology are [MMS+20] [RKN23] and [MKZ+17], but many others could be found.

Regarding the telecommunications sector, one of the most relevant works is the ETSI-defined Multi-access edge computing (MEC) concept [MEC003], formerly mobile edge computing, which is intended to enable cloud computing capabilities at the edge of the cellular network. This MEC technology is designed to be implemented at the cellular base stations, or other edge nodes. This MEC concept targets the edge resources as a whole, without specifically mentioning the extreme-edge resources.

Also, in the context of the telecommunications sector, the 5GCity project [5GC16] was early mentioning the extreme-edge concept in the EC research community, proposing to build and deploy a common, multi-tenant, open platform to extend the centralized cloud model to the extreme edge of the network. Later, in the research towards the future 6G networks, the concept of the extreme-edge integration in resource and service orchestration processes has started to be considered as a relevant innovation topic in the EC flagship projects Hexa-X [HEXD61-21] and Hexa-X-II [HEX2D63-24].

All in all, we can say that today, this concept of edge-to-edge is already widely used, being increasingly referred in multiple works in the field of networking technologies, and specifically, in relation to the development of 5G and 6G technologies.

2.2.2 Beyond the State of the Art in 6G DAWN Context

Below are the envisaged 6G DAWN contributions beyond the state of the art in what regards considering the extreme-edge key concept:

a) Consideration of the extreme-edge domain as part of the multistakeholder/multi-domain network continuum.
Rationale:
As explained in the previous sub-section, the extreme-edge concept is commonly defined in the state-of-the-art associated to a "specific network", which is presented as a self-contained entity with a defined number of nodes connected to each other, and where the extreme-edge are those outermost edge nodes in such network.

However, one of the main innovations envisaged towards 6G is the unified orchestration across the so-called *network continuum* [HEXD62], which refers the multistakeholder set of network infrastructure resources (i.e., compute, interconnect, and storage resources), physical and/or virtualised, spanning across the different technological and administrative domains of the network, but exposed to each stakeholder as if they were a single integrated resource. Considering this, the extreme-edge domain for each specific stakeholder participating in the network continuum is considered to be not only the subset of resources at the very edge of its own network, but also all those network resources in the continuum beyond such edge, which may be owned by multiple stakeholders (not only MNOs, but also infrastructure providers, vertical industries, neutral hosts, hyper-scalers, or even end-users), and so, distributed in different technological and administrative domains. This would make the extreme-edge domain as a relative concept: a specific "regular" edge resource belonging to a specific stakeholder X could be considered as an extreme-edge resource for another stakeholder Y. Also, this makes the extreme-edge devices highly heterogeneous. They include not only small-scale devices like Customer Premise Equipment (CPE) or IoT devices, but also medium or large-scale computing and networking resources. These resources may be hosted by relevant stakeholders such as certain industries or hyper-scalers. But most significantly, this is what makes the extreme edge so huge at a cloud-native scale, diverse, and volatile, as it may contain a wide variety of resources beyond a single MNO's own domain. As a whole, this approach would be in line with the mist computing

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

paradigm [ACC-19] (everything computing everywhere in a decentralized and distributed computing mesh), or the long-standing "pervasive networks" concept [ECP+02], but in the context of the mobile communications networks.
This obviously poses new challenges beyond the state-of-the-art regarding the management and orchestration of the network resources and the services deployed on them. As a whole, the approach here is drastically different from the one commonly taken in the previous 5G system, where the services orchestration problem used to focus in a centralized manner on a specific stakeholder scope: the MNO. In this case, however, the orchestration must be natively multistakeholder, to address the deployment and the operation of the network services over the different domains of the network continuum.

b) Implementation of a software component to emulate in a controlled laboratory environment the concepts associated with the extreme-edge domain.

Rationale:
Demonstrations and proofs of concept targeting the extreme-edge integration in the state-of-the-art consider this network domain at a very reduced scale, typically with only a small set of devices with some of the features expected to be found for the extreme-edge devices, e.g., using a very small set of reduced computing devices [HEXPOC51], small-scale robots and drones [HEX2DEM1] [HEXPOC52], or single AR headsets [HEX2DEM2].
However, although this "device oriented" approach provides value in terms of the M&O of the network service components on reduced-scale devices, it is considered that is just part of the story in what regards the extreme-edge domain, since other important features (e.g., the high heterogeneity and potentially high number of devices in a cloud-native scale) remain out of scope, which omits much of the complexity associated with managing and orchestrating resources and services on this extreme edge domain.

So, beyond the state-of-the-art, in 6G DAWN, we consider of paramount importance to have available a realistic infrastructure environment for experimentation able to emulate the most relevant features of the extreme-edge as a whole, i.e., considering its heterogeneity, volatility, and the large number of resources, instead of just relying on a small set of specific devices.

However, having available a realistic environment for experimentation with these features could result in extremely complex and expensive. A realistic setup would consist of having a physical datacenter with a large number of computing nodes of different natures in which the networks of multiple stakeholders could be emulated with their respective volatile and heterogeneous resources (e.g., a considerable number of computing and storage nodes of different scales, personal devices, IoT devices, or other kind of devices like those used in the state-of-the-art demos mentioned above). An infrastructure like this one would obviously be challenging to deploy and maintain.

To overcome this problem, 6G DAWN will develop a software component, called Infrastructure Layer Emulator (ILE) to emulate, in a more affordable and flexible way, the main features of the infrastructure envisaged for the future 6G networks, namely:

1. The deployment of heterogeneous and large number of computing nodes.
2. The emulation of different stakeholders.
3. The emulation of the different network domains under consideration (not only the extreme-edge, but also the regular core and edge domains).
4. For the extreme-edge resources, the emulation of their inherent high-volatility, with devices unexpectedly connecting/disconnecting and/or changing certain properties (e.g., memory or CPU occupancy, battery level in the case of battery power devices, etc.).
5. The capability to deploy and orchestrate realistic network services on such an emulator is of great significance.

This emulator's development is considered necessary to have a realistic environment on which test and demonstrate the design concepts for the future 6G systems. It is also considered beyond the state of the art, given that, to date, we are not aware of a similar asset. The implementation will be based on the well-known LXD technology [LXD], a system container and virtual machine manager that provides a unified environment for running and managing full Linux systems inside containers or virtual machines. It supports images for a large number of Linux distributions, from lightweight distributions (which are quite useful to be able to deploy a large number of Linux nodes on small-scale equipment) up to regular/complete Linux distributions (e.g., able to host and execute complex network services). LXD can scale from one instance on a single machine to a cluster emulating a full data center, making it suitable for simulating different scenarios. More details on this ILE component and the current status can be found on [GITILE].

c) Implementation of a predictive system to target the inherent volatility of the extreme-edge domain.

Rationale:

The approach described in (a) requires the M&O system to dynamically adapt the deployment status of the network service components to the available infrastructure resources, which can also dynamically change. As already commented, devices in this domain can be error-prone (they are not necessarily in well-controlled premises), they just could asynchronously connect/disconnect (because the owner may decide that), or suddenly change their available

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

resources (e.g., memory or computing resources). However, to ensure the network services continuity, the network M&O system should be fault-tolerant, being able to automatically (and quickly) manage these kinds of events, even if the network service components were deployed on failing nodes.

To manage this volatility, an AI/ML-based predictive approach is proposed, based on the integration of a component called ISPM (Infrastructure Status Prediction Module). This module would be used to complement/support the M&O system reactive decision-making behaviors, by providing information on possible infrastructure failures preemptively. To achieve this, the ISPM will be fed with historical data from different network domains and slices, which will be used to train the AI/ML models within it. Once trained, it will be used to predict the activation status of the different infrastructure devices in those slices and domains. These predictions would trigger the 6G DAWN M&O system to, e.g., drive the relocation of the service components that may be affected. The rationale behind this ISPM module is that, even so, the behavior of certain infrastructure devices can be indeed very random, other devices can behave following (pseudo)regular patterns (e.g., being connected/disconnected on weekends or workdays, or depending on certain usage patterns), which can be learned by AI/ML algorithms, even for non-obvious cases. The envisaged ISPM Main Features are:

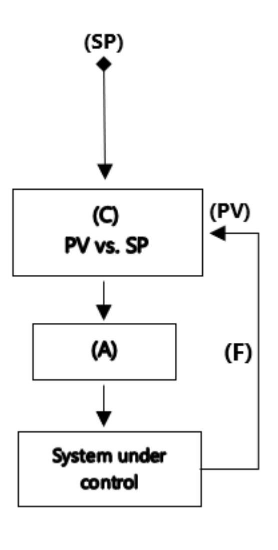
- This would be a Management Service integrated as part of the overall M&O system.
- It will correlate information from different network domains and slices (intra- and inter-domain and slice) to find out related events to produce reliable predictions about the availability of network resources (different instances of the ISPM module could be available, to produce predictions at different timescales, and for different network domains and slices).
- It would alert the M&O system in the event that a resource (or set of resources) in use could become unavailable within a short span of time.
- It could also be queried from the M&O System to more effectively deploy new network slices, or to scale/configure slices already in operation.

2.3 AI/ML agents in control loops

Control loops themselves are not a cutting-edge technology. On the contrary, they are a well-known concept in the control engineering field, used to maintain the desired state of a system under control by adjusting the system behavior based on a feedback signal. Since a long time ago, they have been widely used in multiple engineering disciplines, including process control, automation, and robotics. Mechanical implementations of control loops have been in scope for quite a long time ago. The first feedback control device on record is thought to be the ancient Ktesibios's water clock in Alexandria, Egypt, around the third century BCE. It kept time by regulating the water level in a vessel and, therefore, the water flow from that vessel [KTE300BC]. Already in the 17th century, another relevant example was the well-known fly-ball governor control loop, invented by Christiaan Huygens to regulate the distance and pressure between millstones in windmills, which later (in 1769) was adapted by James Watt to automatically control the speed of the steam engine [HAN22]. Later, already in the mid-20th century, the concept was further formalised and developed in a more general way in the field of Cybernetics by Norbert Wiener [WIE48], associated with topics such as time series processing, information and communication systems, and computing machines, among others.

The basic components of a closed control loop are summarised in Table 2.1, which also includes a high-level diagram.

TABLE 2.1. BASIC COMPONENTS OF A CONTROL LOOP.

Component	Meaning	Diagram
The Setpoint (SP)	*The desired value or target that the system is supposed to maintain (e.g., the desired speed in the steam engine or the temperature setting on a thermostat).*	 (SP) (C) PV vs. SP (PV) (A) (F) System under control
The Process Variable (PV)	*The actual measured value of the parameter being controlled (e.g., the current temperature in a room).*	
The Controller (C)	*A device or algorithm that compares PV vs. SP, and decides what action is needed to bring the PV closer to the SP.*	
The Actuator (A)	*A device or system that can influence the PV (e.g., a heater that can raise the temperature of a room).*	
The Feedback (F)	*Represents the information from the process that is used by the controller to make decisions. It typically involves measuring the PV and feeding it back to the controller.*	

As it could not be otherwise, this well-known technology has also been used in multiple ways in the field of telecommunications with multiple systems acting on different network parameters. For example, it can automatically control the power of the base stations in the mobile networks, the level of congestion in data networks, or synchronise the timing signals among different network nodes, among others.

With the advancements in 5G technology, there is a renewed interest in closed-loop technology. This is due to the potential to interact with virtualized service components that can be treated as managed objects. This technology has potential applications in the management and orchestration of network services. For example, relevant KPIs associated with network services such as QoS, continuity, and latency can be used to define specific Setpoints (SPs). These SPs can be automatically regulated by a closed-loop controller (C) which can dynamically act on the virtual network service components. This could involve triggering the scaling or relocation of the service components involved. Furthermore, the algorithms in the controller could be implemented using AI/ML techniques. This would enable the management of services in situations where dealing with large data sets is necessary, such as integrating a large amount of information from the extreme-edge domain.

It would also address problems where a regular algorithmic approach is not feasible, for example, triggering service component migration actions by processing data from the application layer using AI/ML-based image recognition techniques.

2.3.1 State of the Art

Nowadays, the concept of control loops is widely recognized and implemented across various telecommunications standards and frameworks. One of the most cited references when it comes to the application of control loops in the field of the management of telecommunications networks and services is the work being performed by the ETSI Zero-touch network and Service Management (ZSM) working group, and more specifically, the closed-loop related technical specifications [ZSM-009-1], [ZSM-009-2] and [ZSM-009-3]. This work is being also explored in the EC 6G flagship project Hexa-X-II [HEX2D63-24], which proposes a practical implementation of the ZSM concepts.

As a whole, the ETSI ZSM specifications define a framework for managing network services and resources with minimal human intervention, relying on the control-loops technology as a key concept, which are considered essential for automating and optimizing the management processes. Control loops consist of of several stages (data collection, analysis, decision-making, and action), and can operate at different levels (e.g., network slice, service, or resource). Although for full automation closed control loops are considered, open control loops that may involve some human interventions are also in scope. Control loops can be also designed with different levels of granularity, depending on the specific management task (e.g., a fine-grained loop might manage specific resources, while a coarse-grained loop might manage an entire network slice or service). They can also be designed to operate across multiple domains (e.g., different network segments or service types), making it possible to coordinate actions across these domains. The ZSM framework also emphasizes the integration of AI/ML into control loops, which is considered to enhance the decision-making process by providing predictive insights and enabling more sophisticated automation.

Besides the ETSI ZSM specifications, there are also other specifications in the telecommunications scope incorporating the control-loops concept. One of them is the 3GPP [28.867], which defines several control-loop mechanisms within its management architecture, and considers different use cases. These loops are designed to support dynamic network slicing, service assurance, and real-time management operations. Also, the 3GPP long-supported Self-Organizing Networks (SON) technology [28.861] relies heavily on control loops to automate network optimization, such as adjusting parameters for load balancing, coverage, and interference management. Also, the Open Digital Architecture (ODA) proposed by TM Forum includes concepts related to control loops in its framework for managing digital services. The ODA defines components like the Autonomous Service Management block, which leverages control-loops to automate service lifecycle management, including their application in automated fault management, performance optimization, and predictive maintenance, among others [IG1220I] [IG1219A1]. Another reference commonly cited is the IETF RFC7575, which claims that a fundamental concept towards the autonomic networking

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

concept involves eliminating external systems from a system's control loops and closing of control loops within the autonomic system itself, with the goal of providing the system with self-management capabilities, including self-configuration, self-optimization, self-healing, and self-protection [7575].

In what regards the application of AI/ML in control loops, they can play a transformative role in enhancing them in telecommunication networks: by applying AI/ML to control loops, networks can achieve higher levels of automation, efficiency, and adaptability [BSS+21] [TBA+23]. The ITU has provided the recommendation [Y.3173] on AI in Networks, where the ITU proposes the use of AI-driven control loops to optimize network operations, such as traffic management, fault detection, and dynamic resource allocation. AI/ML techniques can be applied to the data collection and the preprocessing processes, where AI/ML techniques can be used to collect, filter, and preprocess large volumes of data from various sources (e.g., different slices or network domains) within the network [SAS+23]. Also, AI/ML models can be trained on historical data to predict future network conditions, such as traffic surges, equipment failures, or quality of service degradations. This predictive capability allows control loops to proactively manage resources [CCR+23]. Also, in the scope of the data analysis and inference, AI/ML can identify patterns in the network data that might not be obvious through traditional analysis methods (e.g., clustering algorithms can group similar network events or behaviors, helping the control loops to identify underlying causes of issues) [VBM+21]. In other kinds of applications, reinforcement learning (which is based in implementing a reinforcement control loop itself) can be used within control loops where the AI agent can learn the optimal actions by interacting with the environment and receiving feedback [TOU22]. This is particularly useful in scenarios where the network conditions are highly dynamic and complex. Also, machine learning can enhance policy-based control loops by continuously refining and optimizing the policies based on real-time data [TSG+21]. For example, AI can dynamically adjust quality-of-service (QoS) policies to meet changing user demands. Besides, AI/ML can be used for the optimization of the control loops themselves [SP23], e.g., by adjusting the frequency of the data collection, the thresholds for triggering actions, or the balance between reactive and proactive management, among others.

In regard to the AI/ML algorithms, various approaches can be utilized to make predictions. For example,, [SJN22] describes how to perform forecasting on time series using various deep learning models. Long Short-Term Memories (LSTM) are also a well-known recurrent neural network (RNN) well-suited for sequence predictions tasks [HS97]. Also, Gated Recurrent Units (GRU) are another type of RNN that could be used to make predictions based on time-series data [WWW+21]. Besides, Gradient Boosting Machines (GBM) such as XGBoost [FYL+22] and LightGBM [CWY+23] have demonstrated a good performance on prediction tasks as well. Also, Networks based on transformers such as Bidirectional Encoder Representations for Transformers (BERT) [DIF21] or Transformers for Time Series (TFT) [ANT+23]. Besides, Convolutional Neural Networks (CNNs) and regular feed-forward networks can also be used for time-series predictions [BBO18] [GCW+22].

Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D · Gobierno de España · cttc Centre Tecnològic de Telecomunicacions de Catalunya

2.3.2 Beyond the State of the Art in 6G DAWN Context

The contribution of 6G DAWN beyond the state-of-the-art will be the application of AI/ML-based control loops to proactively minimise potential disruptions in running network services, even when they may be deployed over highly volatile extreme-edge infrastructure resources (see Section 2.2). The role of the AI/ML algorithms in the control loop will be to predict the activity status of the infrastructure devices on which the network service components could be deployed.

For a better understanding of this beyond the state of the art contribution, Table 2.2 represents an unfolding of Table 2.1 with the main closed loop components, but including two new columns: the central one, describing how these generic components would be aligned with the 6G DAWN architectural components with the target of minimizing disruptions in the deployed network services in a reactive way (i.e., triggering the actions right at the moment a problem were detected), and the right column, describing how such reactive model becomes proactive (which is the case targeted here), and which, as it can be seen, is based on a decision-making process based on the future state of the infrastructure devices on which the network service may be deployed. As it can be appreciated, such information of the possible future state of the infrastructure devices will be provided based on AI/ML algorithms part of the control loop, embedded in the already mentioned ISPM component (Section 2.2)[1].

TABLE 2.2. TABLE 1 EXTENDED WITH REACTIVE AND PROACTIVE CASES.

Component	**Meaning (from Table 2.1)**	**Reactive Control Loop**	**Proactive Control Loop**
The Setpoint (SP)	*The desired value or target that the system is supposed to maintain (e.g., the desired speed in the steam engine or the temperature setting on a thermostat).*	*Target: Minimise network service disruptions (keep the network service up and running).*	*Target: Minimise network service disruptions in a proactive way, i.e., anticipating the possible unavailability of the infrastructure nodes on which the network service components could be deployed.*
The Process Variable (PV)	*The actual measured value of the parameter being controlled (e.g., the current temperature in a room).*	*Number of network service components up and running. The PV value would be provided by the 6GDAWN M&O*	*Number of network service components at risk of not being up and running in the future (due to the unavailability of the infrastructure nodes*

[1] Please, note that the reactive control loop case is provided here just as an example, for a better understanding on how the control-loop will be implemented in the proactive case, which is the actual beyond the state-of-the-art case targeted in 6G DAWN.

Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D · Gobierno de España · cttc Centre Tecnològic de Telecomunicacions de Catalunya

		system (through the MS and the AE components).	*on which they could be deployed). The PV value would be provided by the 6GDAWN M&O system, through the MS and the AE components (the latter, including the ISPM component for the predictions).*
The Controller (C)	*A device or algorithm that compares PV vs. SP, and decides what action is needed to bring the PV closer to the SP.*	*6GDAWN M&O System (DE component).*	*6GDAWN M&O System (DE component).*
The Actuator (A)	*A device or system that can influence the PV (e.g., a heater that can raise the temperature of a room).*	*6GDAWN M&O System (ACT component).*	*6GDAWN M&O System (ACT component).*
The Feedback (F)	*Represents the information from the process that is used by the controller to make decisions. It typically involves measuring the PV and feeding it back to the controller.*	*PV value. Provided by the 6GDAWN M&O System (MS and AE components).*	*PV value. Provided by the 6GDAWN M&O System. Provided by the MS and the ISPM component (within the AE).*

As it can be seen, the approach is aligned with the 6G DAWN architectural components already described in the previous Deliverable E3 [6GDE3]: The MS (Monitoring System) contributes, together with the ISPM (part of the Analytic Engine), to provide the PV value, i.e., the number of network service components at risk of not being up and running in the future, based on the predictions provided by the ISPM. Besides, the control-loop Controller (C) would be implemented by the Decision Engine (DE), while the ACT component would play the role of the closed-loop Actuator (A). Figure 4 provides a high-level view of the proactive control-loop implementation for a single network service, in line with the general control-loop abstractions in Table 2.1.

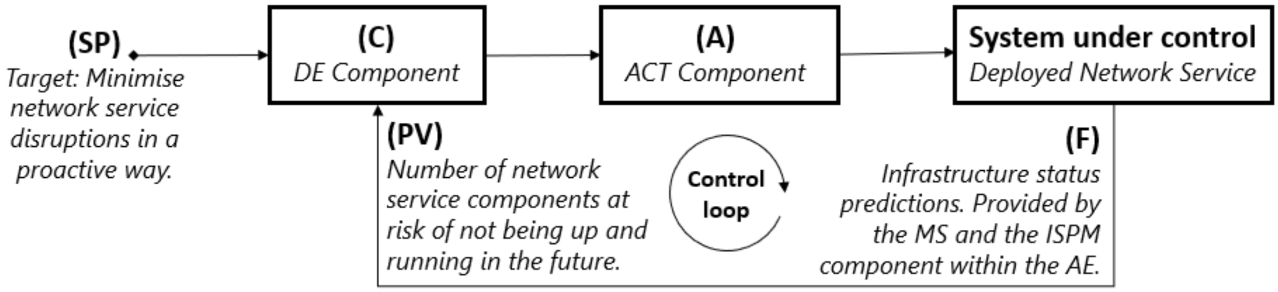

(SP)
Target: Minimise network service disruptions in a proactive way.
(C)
DE Component
(A)
ACT Component
System under control
Deployed Network Service
(PV)
Number of network service components at risk of not being up and running in the future.
Control loop
(F)
Infrastructure status predictions. Provided by the MS and the ISPM component within the AE.

FIGURE 4. PROACTIVE AI/ML-BASED CONTROL LOOP IMPLEMENTATION. HIGH-LEVEL VIEW.

Of course, for the proactive model to work, the AI/ML models in the ISPM should be properly trained to provide accurate predictions on the infrastructure status, i.e., the control-loop implementation is based on pre-trained AI/ML models. The pre-training of the model could be done initially off-line, and taken into production once the results of the predictions are deemed reliable enough. In a realistic environment, MLOps techniques could be used to keep and apply different versions of the models (e.g., to provide predictions on different network domains, slices, or timescales), and to apply automatic updates when necessary (e.g., in case of drift).

As it can be appreciated, no specific algorithms are specified here to implement the AI/ML models. This is intentional. As referred in Section 2.3.1 there are different algorithms in the state-of-the-art that could be used to make the infrastructure status predictions. So, for realistic implementations, it is considered better to keep the choice of the most suitable algorithm for each case to the discretion of the ISPM developer.

2.4 xApps in O-RAN

The role of xApps, short for eXtended Applications, within the Open Radio Access Network (O-RAN) architecture is important. These modular software applications operate on the Near-Real-Time RAN Intelligent Controller (Near-RT RIC). They are instrumental in optimizing and managing various aspects of the Radio Access Network (RAN) through advanced AI-driven algorithms and real-time data analytics. The O-RAN architecture is built on an open, disaggregated network system, allowing for seamless interoperability between software and hardware from different vendors through standardized interfaces. Within this flexible ecosystem, xApps are crucial facilitators of network intelligence and adaptability, empowering operators to efficiently manage network resources, enhance user experience, and reduce operational costs. Specifically designed to address RAN-specific tasks such as radio resource management (RRM), interference mitigation, beamforming, load balancing, and mobility management, xApps can be customized to achieve various performance objectives, including improved energy efficiency, network slicing, latency reduction, and throughput optimization. By dynamically responding to network conditions, xApps offer finer control over the RAN, paving the way for autonomous network operations and enhanced service quality. In the context of O-RAN, the openness and flexibility of xApps enable vendor-neutral innovation, allowing third-party developers to create and implement new functionalities that enhance overall network performance. This modular approach enables continuous upgrades and refinements, thus ensuring the network remains adaptable to evolving standards such as 6G.

2.4.1 State of the Art

In the current 5G and 6G deployments, xApps are highly valuable in addressing crucial network challenges. Within the O-RAN framework, the latest advancements in xApps revolve around their capacity to optimize near-real-time control decisions and deliver improved network performance across various use cases [REZAZA], [DRYJA] [KOUCH].

Resource Allocation: One of xApps' primary applications is real-time radio resource allocation. They monitor the RAN and dynamically adjust resources to ensure an optimal user experience. By utilizing machine learning algorithms, xApps can predict traffic patterns and proactively allocate resources to high-demand areas while optimizing spectrum usage.

Interference Management: xApps can reduce interference between adjacent cells, particularly in densely populated urban environments. AI-based xApps mitigate interference by managing power levels and adjusting beamforming parameters, leading to enhanced spectrum efficiency.

Network Slicing: Present xApps are being utilized in early 5G networks to manage network slicing, enabling different services (e.g., IoT, enhanced mobile broadband, ultra-reliable low-latency communication) to coexist on the same physical infrastructure while upholding specific quality-of-service (QoS) requirements.

Energy Efficiency: Many of today's xApps are focused on minimizing energy consumption by optimizing the operation of base stations under varying loads. For example, xApps can dynamically shut down or put certain hardware elements into low-power states during periods of low traffic, thereby reducing the network's overall power footprint.

However, while these applications are advancing, the xApps ecosystem is predominantly centered on 5G and has yet to fully meet the requirements that will emerge with the introduction of 6G. Current solutions are mostly centralized and often need more agility and scalability for large-scale 6G deployments, where real-time adaptability, decentralized control, and robust security are essential [ATAL],[QAZZ].

2.4.2 Beyond the State of the Art in 6G DAWN Context

In the 6G DAWN context, xApps will move far beyond the current state-of-the-art by addressing the need for decentralized, scalable, and resilient network management solutions, particularly in handling massive wireless network slicing. As 6G networks will need to support a much larger number of connected devices and services, along with highly diverse QoS requirements, the limitations of current centralized management systems will become apparent. The 6G DAWN RESILIENT subprojects will introduce new paradigms for xApp development and deployment to overcome these challenges.

6G DAWN RESILIENT: Security and Reliability

6G DAWN RESILIENT will focus on extending xApp functionalities to incorporate robust security and resilience mechanisms. As 6G networks are expected to support critical infrastructure and services, ensuring network slices' continuous, secure operation becomes paramount.

Advanced Security Schemes: xApps in 6G DAWN RESILIENT will be equipped with AI-driven security algorithms to detect and mitigate real-time threats. This includes defenses against distributed denial of service (DDoS) attacks, intrusion detection, and privacy-preserving mechanisms for sensitive data. These xApps will monitor the network and adapt security protocols based on emerging threats.

Secure Life-Cycle Management (LCM): As network slices evolve over their life cycle, from instantiation to termination, xApps will ensure that security measures are maintained consistently. The life-cycle management (LCM) of slices will be equipped with mechanisms to secure communication paths, protect data integrity, and guarantee service continuity even under attack scenarios.

Resilience to Failures: In the 6G DAWN framework, xApps will ensure the resilience of network slices by implementing self-healing mechanisms. These will automatically detect failures or performance degradations within a slice and trigger corrective actions, such as reallocating resources or rerouting traffic through alternative paths. This guarantees that mission-critical applications continue to function seamlessly, even in the face of hardware or software failures.

2.5 Relation of vertical KPIs with the network configuration

While the envisioned NPN Digital Twin system (key concept 2.1) provides visibility towards the performance of the network components (i.e., the Core and the RAN), it is also key to consider how these changes affect the performance of the end-applications and their impact on the quality of experience of end-users to achieve true end-to-end energy optimizations. This will enable CSP and third-parties to not only select the appropriate network configuration from the rank-ordered list of feasible configurations provided by the NDT, but also to monitor the impact that these changes have on end users and to re-adjust selected configurations based on the performance information obtained.

This customization, adjusted towards the optimization of energy consumption within the network to achieve better operational efficiency or to improve the network's availability and resiliency in situations where the availability of energy is limited, allows both close loop and open-ended approaches. Within the devised system, KPIs related to mission-critical services, such as MCPTT, MCVideo, and MCData, will be monitored and used to select the best configuration the NDT provides.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

2.5.1 State of the Art

Within the 3GPP MCX standards, namely TS 22.179 [TS22179] for MCPTT, TS 22.281 [TS22281] for MCVideo and TS 22.282 [TS22282] MCData, the industry and academia have established a set of performance goals and KPIs for mission-critical services. Some notable examples are the establishment of a maximum M2E (Mouth-to-Ear) latency for audio of 200 ms or a maximum E2E (End-to-End) access time, i.e., the delay between the push of button and the ability to start transmitting voice/speak.

In terms of voice quality and intelligibility, which is impacted by packet loss, jitter, quality of radio channels & codec impairments, the following metrics have been recommended as key indicators of the performance of voice-related mission-critical services:

- MOS (Mean Opinion Score) Scale: measured using quality models such as POLQA (Perceptual Objective Listening Quality Analysis) or PESQ (Perceptual Evaluation of Speech Quality) and quantified with the MOS scale (with scores between 1 and 5).
- Intelligibility: measures how the transmitted speech can be heard by the receiving party (e.g., takes into account background noise). Measured using an MRT—Mean Rhyme Test, such as ABC-MRT.
- Other: the ability to Establish/Retain Calls, E2E Access Time in group calls, Late Call Entry or the observed Grant Time, among others.

However, how the main set of network configurations impacts these metrics changes that optimization systems, such as those implemented by an NDT, may apply to achieve better energy efficiency still needs to be rigorously studied.

2.5.2 Beyond the State of the Art in 6G DAWN Context

By creating a system capable of using both application-specific QoE metrics and energy consumption information obtained via theoretical, empirical (from lab measurements) and field data, the system devised as part of this project will enable the creation of close loop and open-ended mechanisms that will help reducing the energy consumption of the network while minimizing the impact to end users ahead of standardization.

On the one side, the integration of the AE and the rest of components of the NDT with the Core, gaining privileged access to the other core NFs and the RAN, will enable the enforcement of network configuration changes to optimize the energy consumption and to gather other network-related KPIs. On the other side, the deployment and integration of the system with MS agents at the endpoints (at the UEs and external AFs executing the mission-critical services) will give access to the relevant QoE KPIs that will help in choosing the optimal configuration.

The application-specific QoE will be gathered within the client and application servers via the inspection of the control traffic (i.e., the exchanged SIP messages), the monitoring of the user plane traffic (e.g., to obtain metrics such as application-level latency and jitter) and the analysis of the quality of the end-data (e.g., to obtain voice quality-related metrics). By focusing on PNI-NPNs, and more concretely, within mission-critical scenarios with known traffic patterns, this will allow to use the learnt data from simulations and other domain knowledge sources to realize the potential of network digital twins in the realm of energy efficiency.

2.6 Inter/Intra-slice reconfiguration and massive slicing

Inter(a)-slice reconfiguration refers to the dynamic adaptation of resources and configurations between and within network slices to optimize performance in 6G networks. As 6G architecture will manage massive amounts of data and services, massive slicing is essential to accommodate a wide range of use cases, such as ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC). Inter-slice reconfiguration allows the flexible reallocation of resources across slices to adapt to changing traffic conditions, ensuring efficient use of available bandwidth and computing resources, reducing latency, and minimizing service disruption. Within-slice reconfiguration involves adjusting resources within a single slice to optimize performance based on user demand or network status.

2.6.1 State of the Art

Current 5G/6G deployments have introduced network slicing to allow mobile network operators to create virtual networks (slices) on shared physical infrastructure. Each slice is designed for a specific service and its requirements, such as latency, bandwidth, and reliability. However, these slices are often static or pre-configured, meaning they do not dynamically adjust to changing network conditions or traffic demands.

Static Allocation: Traditional network slicing allocates fixed resources (e.g., spectrum or computing capacity) to each slice, ensuring service isolation but leading to potential resource underutilization if a particular slice does not fully use the allocated resources. Current wireless network learning approaches have focused on traditional machine learning (ML) algorithms, which centralize the training data and perform sequential model learning over a large data set. However, performing training on a large dataset is inefficient; it is time-consuming and not energy and resource-efficient. Transfer Learning (TL) effectively addresses some challenges by training based on a small data set using pre-trained models for similar problems without impacting neural network model performance. The work in [THAN] detailing adaptive strategies to allocate resources in real-time, emphasizing dynamic management to resolve issues inherent in static allocations

Inter-slice Conflict: With concurrent slices competing for limited network resources, resource contention or conflicts may arise, leading to suboptimal performance or service degradation. Static allocation mechanisms lack the flexibility to resolve these conflicts dynamically. For example, the

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

work in [WU22] explores AI's role in each lifecycle phase of slicing, covering adaptive mechanisms and how AI can address real-time demands and conflicts (e.g., inter-slice resource contention) effectively in 6G networks. In [REZAZA], the work explores a novel explainable protocol framework for adaptive and conflict-free resource allocation in 6G O-RAN environments.

AI-Enhanced Slicing: Some advanced systems use AI and machine learning for predictive resource allocation, allowing slices to adjust to future demands. However, this approach remains limited in its ability to dynamically reconfigure resources in real-time based on actual, moment-to-moment network conditions. In [ABDE23], proposes an optimized, intelligent network slicing framework to maintain a high performance of network operation by supporting diverse and heterogeneous services, while meeting new KPIs, e.g., reliability, energy consumption, and data quality. Different from the existing works, which are mainly designed considering traditional metrics like throughput and latency, they present a novel methodology and resource allocation schemes that enable high-quality selection of radio points of access, VNF placement and data routing, as well as data compression ratios, from the end users to the cloud.

Despite progress, the current systems still struggle to meet the scalability and flexibility demands of 6G. With billions of devices interacting with numerous vertical industries, there is a critical need for advancements in 5G/6G deployments that can provide constant, dynamic resource adjustments across an unprecedented number of slices.

The work undertaken on scalable network slicing focuses on developing advanced algorithms for the Distributed Decision Engine, specifically designed to facilitate expansive slice environments. This research prioritizes efficient resource allocation and dynamic adaptability within the network, thereby enabling the seamless management and optimization of a large-scale, multi-slice architecture. Each Non-Public Network (NPN) is characterized by its own internal slices, which allows both Public Networks (PNs) and NPNs to be configured with a multitude of individual slices. This configuration supports granular and flexible management of network resources.

The Decision Engine utilizes the Network Data Analytics Function (NWDAF) to learn from diverse domains, capturing the ramifications of various network configurations on performance and energy consumption across differing operational contexts. This multi-domain learning framework empowers the system to implement localized decisions within the same NPN to achieve specific optimizations while facilitating global decision-making across the network to enhance energy efficiency. This aspect holds particular significance during critical emergency scenarios.

Furthermore, the NWDAF framework is instrumental in executing these complex operations across multiple deployment scenarios by offering specialized NWDAFs for targeted analytics, including but not limited to abnormal behavior, user equipment (UE) mobility, and data congestion. Area NWDAFs aim to provide analytics for specific geographical coverage within a Mobile Network Operator (MNO) network. In addition, Aggregation NWDAFs serve to compile analytics from various sources based on distinct Analytics IDs, thereby augmenting the system's ability to adapt decisions across interconnected network layers. This comprehensive, analytics-oriented approach to inter-slice reconfiguration and extensive slicing addresses the scalability challenges inherent in distributed networks while ensuring efficient and energy-aware network operations.

2.6.2 Beyond the State of the Art in 6G DAWN Context

In the context of 6G DAWN, inter(a)-slice reconfiguration will move beyond static and centralized resource management, introducing a distributed, AI-driven approach allowing massive slicing with real-time adaptability.

6G DAWN RESILIENT: Security and Reliability in Reconfiguration

6G DAWN RESILIENT focuses on ensuring that inter(a)-slice reconfiguration happens securely and reliably:

Real-Time Conflict Resolution: AI-based algorithms is developed to resolve conflicts between slices in real time, preventing issues such as resource contention and underutilization. Unlike current static systems, the AI continuously monitors the resource usage across slices and adjust allocations dynamically to avoid conflicts.

Secure Reconfiguration: Given the critical nature of many 6G applications (e.g., remote surgery, autonomous driving), reconfiguration processes must be protected from cyber threats. 6G DAWN RESILIENT introduces security mechanisms that protect against malicious attacks during reconfiguration, ensuring that resource allocations are not tampered with and that slices maintain their integrity.

Resilient Operations: The project ensures that, even in the case of hardware failures or network issues, reconfiguration can continue without disruption to critical services. This resilience is achieved through decentralized control and AI-driven failover mechanisms.

The current architectures being discussed have only been analyzed theoretically. In 6G DAWN we partially implement these features and design an architecture that can also incorporate the decision engine across multiple domains.

More specifically, if we focus on NPN Network Digital twin modelling, predictions and recommendations are based on blending three complementary assets: domain knowledge, broad range of empirical data and Machine Learning. These models evolve and adapt in parallel with the 5G NPN Physical NPN assets they are twinned to, being able to scale appropriately.

Furthermore, the vertical service QoE metric collector (MS) is being designed to be able to collect data from multiple domains (e.g. multiple agencies) and extract conclusions from all of these (of how they are affected by the network configuration changes) as an aggregated result.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

2.7 NEF instance for KPI data and configuration capabilities exposures of NPNs

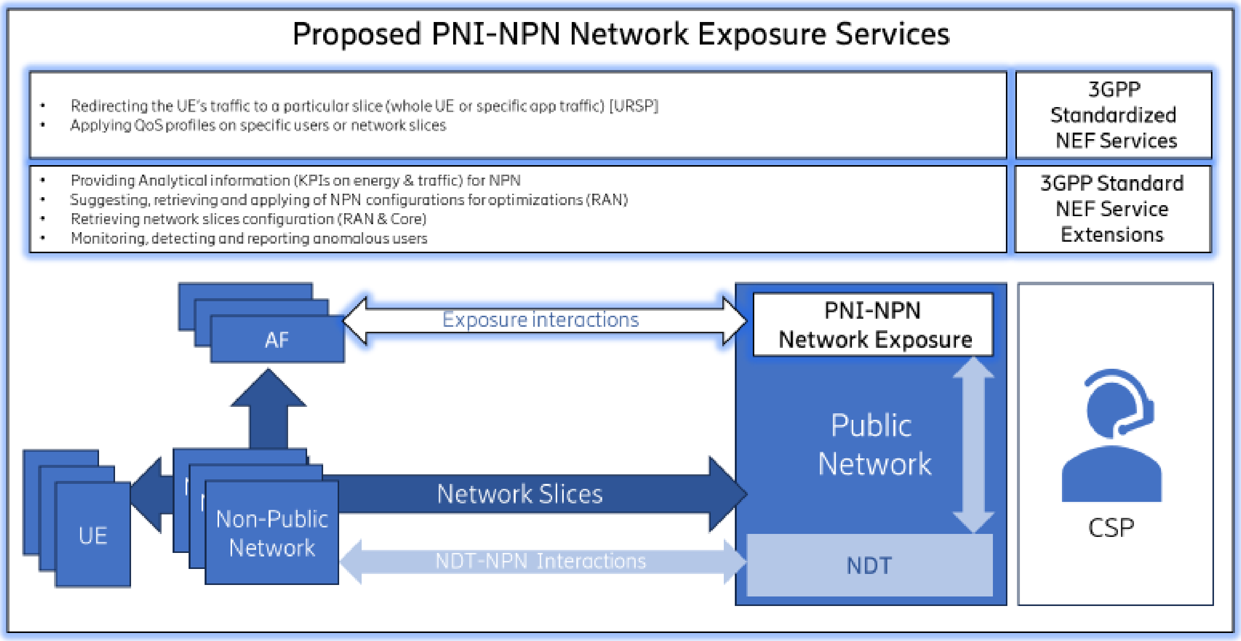
The deployment of 5G Non-Public Networks (NPNs) is rapidly transforming industries by providing dedicated, secure, and ultra-reliable connectivity. These networks cater to diverse use cases across various sectors, but managing their complexity requires advanced tools.

Network Digital Twins (NDTs) emerge as a game-changer, creating real-time digital replicas of physical NPNs. These replicas continuously capture data on various aspects of the network, including performance metrics like latency and jitter, resource allocation, energy consumption, and potential issues like detection of anomalies. This comprehensive data empowers NPN owners/operators in several ways. Firstly, NDTs enable proactive network optimization. By simulating changes within the NDT environment, NPN owners can test new configurations and optimize network parameters before implementing them in the real network. This minimizes the risk of disruptions and ensures optimal network performance. Secondly, NDTs offer real-time visualization of network analytics that will help NPN operators identify potential problems and act on them quickly.

However, unlocking the full potential of NDTs requires a standardized and secure way to access and interact with them. This is where the 3GPP Network Exposure Function (NEF) comes into play. NEF acts as a standardized interface that provides secure access to network information and functionalities. Integrating NDTs with NEF offers several advantages. Firstly, NEF enhances NPN management for owners/operators. They gain a secure way to interact with the NDT, allowing them to retrieve real-time and historical data, adjust configurations within the NDT environment for testing and optimization purposes, and gain valuable insights through user-friendly applications that leverage NDT data. Secondly, NEF facilitates the development of third-party applications that can leverage NDT data and functionalities. These applications can offer advanced network management tools, specialized analytics for specific industries, or even automated optimization services.

NEF integration ensures secure and controlled access to NDT data, preventing unauthorized modifications and maintaining network integrity. Additionally, using standardized NEF APIs fosters a diverse ecosystem of NDT-powered applications. However, to fulfill the diverse requirements of use cases that NDT provides, standardized NEF APIs are not adequate. Our concept in this project is to leverage and extend 3GPP standard NEF APIs to allow implementation of fundamental services like the capability to configure NPN components, to expose analytical information of NPN instances, and to influence steering of user or application traffic to desired network slices which as a result enables seamless integration with various NDT implementations.

Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D · Gobierno de España · cttc Centre Tecnològic de Telecomunicacions de Catalunya



- Redirecting the UE's traffic to a particular slice (whole UE or specific app traffic) [URSP]
- Applying QoS profiles on specific users or network slices

- Providing Analytical information (KPIs on energy & traffic) for NPN
- Suggesting, retrieving and applying of NPN configurations for optimizations (RAN)
- Retrieving network slices configuration (RAN & Core)
- Monitoring, detecting and reporting anomalous users

FIGURE 5: PNI-NPN NEF SERVICES

In conclusion, the combination of Network Digital Twins and 3GPP NEF integration represents a significant leap in NPN management. NDTs empower NPN owners/operators with proactive optimization, faster troubleshooting, and futureproofing capabilities. NEF integration unlocks this potential by providing standardized access and fostering innovation through third-party applications. This convergence paves the way for a future of intelligent, self-optimizing, and future-proof 5G NPNs that can fully support the evolving needs of various industries. As the field evolves, future research can explore areas like standardization of NDT interfaces, and tailoring NDT functionalities to address specific industry needs and use cases.

2.7.1 State of the Art

NEF enables exposing the Core Network capabilities internally in the operator or to an external partner with commercial agreements. The exposable core capabilities add value to internal or external users, for example, connectivity, optimization, identity, security, data, and analytics. Exposure also ensures that the internal or external developer can access the exposed capabilities in a secure, predictable, and reliable manner. NEF provides 3GPP standardized RESTFul APIs to expose network data (ex. User equipment status, network conditions, etc.) to external applications to enable programmability of the network so that external applications can influence network behavior and perform optimizations. Using standardized APIs with open specifications on NEF not only fosters innovation that creates a diverse ecosystem of applications and emerging services but also ensures interoperability.

Network exposure unlocks the value of the Core Network, enabling the collaboration of operators with the ecosystem of partners or developers that generates new revenue opportunities through unique use cases. With those aspects, NEF presents a promising approach for managing 5G Non-

Public Networks (NPNs) with greater efficiency and ease. This section explores the current state of NEF adoption in NPNs and highlights potential limitations in addressing NPN-related use cases.

NEF and its relevant services are defined across multiple technical specification documents however the ones given below include the fundamentals:

- **TS 29.522 - 5G System Network Exposure Function Northbound APIs Stage 3**: This document details the northbound APIs exposed by the NEF, which allow external entities to interact with the 5G Core Network (5GC).
- **TS 23.502 - Procedures for the 5G System (5GS); Stage 2:** 3GPP TS 23.502 defines system procedures for the 5G system. It includes standard services and operations supported by NEF.

These specifications offer a comprehensive understanding of NEF's capabilities for exposing network data and functionalities to external applications and services. They cover aspects like the definition of information models, northbound APIs, and security mechanisms for interacting with the NEF.

On the other hand, using NDTs to operate and manage NPNs might revolutionize the management of 5G Non-Public Networks (NPNs) by offering the services mentioned in the Network Digital Twin – Concept section. By exposing these services to external entities like NPN operators/3rd party applications via NEF, NPNs can be managed more easily in a secure manner.

While a NEF approach holds significant promise to manage and operate NPNs together with NDT, it’s important to acknowledge that the current 3GPP specifications might not encompass the full spectrum of NPN service and management requirements. Specific use cases within NPNs may necessitate functionalities beyond the current NEF capabilities. Specific examples of this scenario are:

- **Analytics information exposure service:** It would be needed to extend the currently defined 3GPP standard service with the KPIs (performance or energy-related) defined for NPNs.
- **Network configuration parameter provisioning service:** It would be needed to extend this service to cover NPN configurations.
- **Optimization service for energy consumption and performance of NPNs:** There is no standard NEF service to cover optimization scenarios of NPN (actual optimization service is provided by NDT and proposed to be exposed by NEF).

It's also crucial to recognize that the Network Digital Twin itself is not a standardized element within 3GPP specifications which makes integration activities with NEF more difficult. However, this integration has a great potential to increase the serviceability aspect of NPN thus allowing wide adoption of the usage of NPNs in multiple industries.

This state-of-the-art analysis highlights the potential of using NEF-NDT integration for NPN management and operation activities while acknowledging limitations in the current standardization activities. Further research and potential standardization efforts are essential to fully leverage NEF's capabilities within the NPN ecosystem, ensuring broader adoption and smoother integration with NDT functionalities.

2.7.2 Beyond the State of the Art in 6G DAWN Context

As already discussed in the SoTA section, current 3GPP standards does not cover the specific scenarios of PoCs that we are implementing for this project. We aim to reuse existing 3GPP standard

NEF services as a base and define customized extensions that cover the needs of NDT-managed NPN scenarios. In this project, specifically, we will be addressing the gaps in these NEF services to make them usable in NDT-managed NPNs:

- **Optimization service for energy consumption and performance of NPNs:** There is no standard NEF service to cover optimization scenarios of NPNs. This service allows AFs to request optimized NPN configurations based on the desired context, such as prioritizing network performance or maximizing energy efficiency. By leveraging this service, AFs can dynamically adjust network configurations to meet changing demands within the NPN.
- **Network configuration parameter provisioning service:** Standard NEF Network Configuration Parameter Provisioning service will be enhanced to encompass the suggested optimized NPN configurations. This empowers AFs to not only receive suggestions for optimal configurations but also directly implement them within the NPN, streamlining the optimization process.
- **Analytic information exposure service:** Expand standard Analytic Information Exposure service to encompass a wider range of Key Performance Indicators (KPIs) specifically relevant to NPNs. Mentioned NPN KPIs can be found in the previous deliverables. By exposing these NPN-centric KPIs, AFs gain deeper insights into network performance, enabling proactive problem identification and optimization strategies.

The proposed NEF extensions aim to improve the management landscape of NPNs. By providing a more comprehensive and dynamic approach to network configuration and optimization, these enhancements empower NPN operators to achieve superior performance, resource & energy efficiency, and resilience within their NPNs. Furthermore, these advancements pave the way for the development of even more sophisticated NPN management tools and applications, fostering further innovation in this rapidly evolving field.

2.8 AI/ML methods for reducing energy consumption

The basic objective of this key concept is to use advanced ML techniques to obtain reduction of the energy consumption while maintaining the QoS required by UEs. Thanks to the benefits that O-RAN introduces, mostly the support of advanced network intelligence and automation, done through the RIC, we will be able to introduce such ML techniques in order to optimize the network, because an over-utilization of resources can lead to unnecessary costs and energy consumption, while under-provisioning can impact performance and user experience.

2.8.1 State of the Art

Energy efficiency in mobile networks has attracted the attention of telecom industry stakeholders, including vendors and Mobile Network Operators (MNO), academia and standardization bodies. Significant efforts have recently been made by the 3rd Generation Partnership Project (3GPP) to develop unified mechanisms that offer MNO substantial energy savings in their networks. These efforts are further supported by the O-RAN Alliance, which considers 3 potential use cases [ORAN] i) Carrier and cell switch off; ii) RF channel switch off/on; iii) Advanced sleep mode. We will focus on the first use case. Cell switch off feature in O-RAN needs to be enabled by the non-real-time RIC, due to timescale of the operation and aims to reduce RAN (i.e., O-CU/DU/RU) power consumption by switching cells on/off. AI/ML-assisted solutions in the Non-RT RIC manage traffic load and automate the decision to switch carriers or cells on/off using O1 and/or Open fronthaul M-plane parameters.

By dynamically adjusting the active cells based on real-time traffic demands, network resources can be utilized more efficiently. During low-traffic periods, certain cells can be switched off, reducing unnecessary resource usage and providing significant energy savings. References [LOPEZ] and [LARSEN] are excellent surveys on sustainable technologies and energy efficient approaches for Radio Access Networks, including key enablers such as massive MIMO, lean carrier design or advanced idle modes. All the techniques can be classified into three categories: time domain, antenna domain or carrier-domain. Base station on/off switching fall into the first category. Several cell on/off switching algorithms have been proposed in the literature, e.g., [HAN], based on traditional optimization methods, which consider just the traffic load or the number of user equipments attached. Other approaches, such as [HOFFM], rely on more advanced techniques and consider ML algorithms, such as reinforcement learning for base station deactivation in massive-MIMO networks, hence, not relying in an O-RAN architecture.

Furthermore, it is often challenging to determine how improvements in the prediction of network traffic and/or resources could help in minimizing the power consumption in the RAN. Nevertheless, it has been recently shown that energy/power consumption is closely related to the requested radio resources. In particular, recent research in [BEGA] has provided a realistic characterization of 5G multi-carrier Base Stations (BSs), offering an analytical energy consumption model based on extensive data collection campaigns. In line with this study, references [PIOVESAN] and [BEGA] have shown that, in modern BSs, power consumption increases linearly with the PRB load —defined as the ratio of average used PRBs in a BS to the maximum number of PRBs available at the remote unit. Furthermore, as highlighted in [LOPEZ2], the Downlink (DL) PRB load is the parameter that holds the highest significance in the radio unit energy consumption. Generally, BSs are designed to handle high traffic volumes during peak hours, which often results in significant underutilized bandwidth during most of the day.

The concept of probabilistic forecasting has been extensively researched in various areas such as energy production [ZHANG] or supply chain management [SPILIOT]. However, its use in telecommunications, particularly in the context of O-RAN, is still relatively little researched. Previous studies have mainly focused on single-point forecasting models that provide deterministic predictions without taking into account the uncertainties inherent in the dynamics of the network [WANG]. Recent research has begun to fill this gap by investigating probabilistic methods such as Bayesian networks [BENRH], Gaussian processes [CHEN22], and deep learning-based models [LI] for network parameter forecasting. Above studies have shown that probabilistic forecasting can improve the performance and reliability of networks. However, there is a need for further research and validation of these techniques specifically in the context of sustainable O-RAN, as its unique architectural and operational characteristics need to be taken into account when integrating probabilistic forecasting techniques into its framework.

2.8.2 Beyond the State of the Art in 6G DAWN Context

Conventional strategies for resource allocation and network management often rely on deterministic models that fail to account for the inherent uncertainties and dynamic nature of modern mobile networks. Such approaches can lead to suboptimal performance, inefficient use of resources and increased operating costs. In contrast, probabilistic forecasting techniques offer a promising solution as they provide a more comprehensive understanding of network conditions and resource requirements. By incorporating uncertainty and variability into forecasting models, these techniques can significantly improve the accuracy and reliability of network analysis, enabling more informed decision making and optimized resource allocation. This KC aims to contribute to the growing body of knowledge on multivariate probabilistic forecasting techniques in telecommunications by focusing on its application within the O-RAN architecture. The primary contributions 6G DAWN of this are the following:

- We propose a novel framework that integrates probabilistic forecasting techniques to enhance the power efficiency of O-RAN deployments. This framework utilizes advanced state-of-the-art AI-based forecasting models to predict network conditions and resource requirements and enables dynamic and adaptive resource allocation.
- The accurate prediction of the PRB utilization and to improve the sustainability of O-RAN operations by achieving power-saving.

3 RESILIENT Key Concepts

3.1 Develop standard-compliant network interfaces to support AI driven network fault management systems at NPN

Non-Public Networks (NPNs) hold the potential to revolutionize industrial automation and innovation. However, their energy consumption, particularly for networks with constant high bandwidth demands, presents a significant challenge. Identifying and addressing high-energy users (defined as "anomalous" users in the rest of the text) within an NPN offers a solution with environmental and economic benefits, while also bolstering network resiliency.

Early detection of energy-intensive devices or processes allows NPN operators to implement targeted optimization strategies. This not only minimizes the network's environmental impact but also aligns with sustainability goals. Reduced energy consumption translates directly to cost savings for NPN owners, freeing up resources for reinvestment in network optimization, process improvements, or further industrial advancements. Ultimately, addressing anomalous users leads to greater efficiency and profitability.

Furthermore, isolating anomalous users plays a crucial role in enhancing network resiliency. Unidentified and unaddressed anomalous users can strain limited network resources, potentially exceeding the network's capacity and causing disruptions. By proactively detecting and isolating such users, NPN operators prevent resource depletion and ensure overall network efficiency and resiliency. This proactive approach creates a more stable and reliable network environment, critical for industrial applications that rely on consistent performance.

In essence, addressing high-energy anomalous users translates to a more sustainable, cost-effective, and resilient network. This paves the way for the vast adoption and deployment of NPNs not only in the present but also as we enter the 6G era, laying a strong foundation for the future of industrial automation and innovation.

In this project, we develop an E2E solution that covers all aspects of detecting, reporting, and isolating high energy-consuming users in an NPN by harmonizing Network Digital Twin, Network Exposure & Policy Functions together with the capabilities of external AFs. By using effective ML algorithms, NDT can detect "potential" anomalous users and report them to external AFs, and then external AFs can use additional information coming from other domains to conclude whether the "potential" anomalous user is a real threat to network resiliency and take further action to isolate it in a special "quarantine" network slice.

NEF plays a crucial role in allowing communication between external AF and NDT just replicating the 3GPP standard way of exposing network services to external entities. There are two different services to be exposed by NEF in our concept. One of them is the monitoring event service which is related to detecting and reporting the anomalous users whereas the other service is the application guidance for URSP (UE Route Selection Policy) determination, and it provides the capability of changing the existing network slice of the anomalous user to the quarantine network slice for isolation.

3.1.1 State of the Art

Traditional fault management systems rely heavily on predefined thresholds and rule-based algorithms to identify abnormalities in network user behavior. While effective to a certain extent, these systems lack the adaptability and precision required to address the nuanced challenges posed by anomalous user behaviors in the current mobile networks. Moreover, traditional network monitoring / fault management systems are not equipped with powerful AI/ML techniques for network monitoring, anomaly detection and isolation of the detected anomalies.

When it comes to standardization related to mobile networks, there are no standards defined by 3GPP related to use of Network Digital Twin concept for either monitoring or fault management.

As mentioned in the chapter 2.1.1 Digital Twinning technology and 3GPP Mobile networks, specifically NPNs, are two separate domains which are evolving in their own tracks; however, by creating a Network Digital Twin platform for NPNs to bridge these two separate worlds holds a great deal of potential to implement several use cases in network management space, including anomaly detection and isolation. Moreover, the accuracy and effectiveness of anomaly detection methods can be further strengthened by using the power of AI/ML techniques.

3.1.2 Beyond the State of the Art in 6G DAWN Context

In this project, NDT and NEF are the two main pillars to take the role of a network monitoring and reporting system particularly for the use case of anomalous (high-energy consuming) user detection and reporting. AI/ML powered NDT platform which is tightly integrated with NPN system is performing real-time monitoring of energy consumption patterns and detects anomalous usage patterns and reports them to the external application functions as “potential” anomalies. Such NDT anomaly detection service is exposed to external entities outside of the NPN network, (external AFs) through 3GPP standardized network APIs which fosters interoperability and innovation. 3GPP defined network service APIs will be extended

3.2 Decentralized Intrusion Detection to integrate novel trust-based evaluation mechanisms

Container-based microservices management platforms such as Kubernetes offer highly dynamic and automated computing clusters’ creation and management, multiple computing clusters co-existence and operation, and require interactions between all parts (containers) of the system to enable the final business logic.

Given the increasing dynamicity and complexity of computing clusters nowadays:

- It becomes challenging to identify erroneous behaviour inside clusters,
- It increases vector of attacks and room for malicious activities to take place undetected, and
- It becomes essential to monitor *all* components of the environment to ensure the entire cluster is healthy.

Decentralised intrusion detection in the context of 6G DAWN is defined on the capacity to detect intrusions on each and every working node in a computing cluster, and across different computing clusters and even across name spaces within clusters. It is essential to clarify that this translates to visibility on the behaviour of each and every worker node independently of the other worker nodes. In other words, the ability to take decisions on the *healthy* status of each worker node in a distributed computing cluster environment independently of other working nodes but towards a consolidated normal behavioural baseline.

The concept is centred on *anomaly detection in computing clusters* as the more solid approach to ensure the healthy status of computing clusters with no (or acceptable) deviation from normal behaviour. In contrast, the concept of signature-based intrusion detection focuses on ensuring that the computing cluster environment is free from *known* attack signatures but is unclear if unknown ("zero-day" exploit based) intrusions take place.

Furthermore, anomaly detection gives a solid baseline for *trust-based evaluation* mechanisms where trust in the underlying network infrastructure comes from its *conformity to the normal behaviour* observed during a training period rather than on the absence of known attack signatures (but with potential abnormal behaviour).

However, we explicitly acknowledge that concept of decentralised intrusion detection based on anomaly detection does not limit the concept to be further extended to a hybrid approach where anomaly detection is combined with a signature-based detection for additional assurance.

3.2.1 State of the Art

The state of the art focuses on approaches to cybersecurity anomaly detection in computing clusters. The detection of anomalies in computing clusters has recently gained attention in the literature.

In [KOS] the authors proposed a Kubernetes Anomaly Detector (KAD) that uses ML techniques. The proposed solution uses various ML models that facilitate the detection of different types of anomalies. The user selects a model to use in a given situation or context, or the system automatically selects the appropriate model based on scoring procedures. Among the different ML algorithms considered in KAD, one can find LSTM and Autoencoders, both are interesting and promising in the domain of anomaly detection. While the otherwise a very interesting and related to us work, the main focus of this work is on validation of different ML algorithms on some public datasets with no focus on scalability to decentralized detection within a cluster, and what level of intrusion (type of attacks) can be detected. In their conclusions, Autoencoder shown the most promising results of detection, well aligned with our decision of adoption of Autoencoder for anomaly detection.

An anomaly detector for Docker orchestration platforms, called KubAnomaly, was proposed in [KUBAB]. It offers a *container monitoring module* for Kubernetes and uses ML algorithms (neural networks) to produce classification models to support detection of abnormal behavior. Their focus is on monitoring logs from Docker container operations and selected system calls. Their work is complementary to our work on network anomaly detection and illustrates the pressing need to address decentralized network behavior monitoring and detection within a Kubernetes cluster – aligned with our endeavor in 6G DAWN.

An inspiring work was proposed in [DCID] on anomaly detection for container applications and Kubernetes environments based on *system calls* monitoring. It adopts Autoencoder algorithm for learning and detection of anomalies. This work has certain similarity to the approach in [KUBAB] and is seen as complementary to our approach in 6G DAWN on decentralised *network* anomaly detection in computing clusters. A relevant conclusion of this work is on the high performance of Autoencoder on detection of anomalies in system calls of container applications.

In [KUB], the authors proposed a relevant to 6G DAWN dataset for Kubernetes misuse detection. Their work addresses a research gap on *limited availability* of publicly available datasets for container security. The dataset consists of network flows, collected from a Kubernetes cluster. The limited factor of the dataset is on the offering of CSV files only of already processed traffic captures (PCAP files) using the community version of CICFlowMeter. In our case, we need PCAP files to extract network flow features based on our in-house feature extractor specifically tailored for anomaly detection.

In summary, we can conclude that the strong point of related work is advancing the knowhow on:

- What and how AI/ML algorithms facilitate learning and detection of anomalous behaviour in Kubernetes and Docker environments.
- How to use Kubernetes and Docker environments for deployment and data processing for anomaly detection.

3.2.2 Beyond the State of the Art in 6G DAWN Context

The vision of a decentralised intrusion detection is based on a combination of a local per worker node anomaly detection consolidated per computing cluster training of legitimate (normal) behaviour.

6G DAWN builds on two following main pillars – value proposition – that place the project over state of the art:

- *Multi-cluster monitoring and anomaly detection*: monitoring and anomaly detection per computing cluster and between cooperating clusters,

- *Zero-touch (automated) deployment, configuration, and ML model training*: Unsupervised learning on baseline traffic behaviour of all pods and services per worker nod (local) and per computing cluster (consolidated) modality with a specific visibility and conformity level.
- *Extended behavioural telemetry* of a computing cluster for increased detection rate of attacks and intrusions,
- *Reference architecture and deployment* for Kubernetes' clusters to achieve necessary traffic visibility (with respect to an adversary) and data processing for anomaly detection,
- *Enriched explainability* of what cluster's network behaviours aspects cause anomalies to improve utility in higher-level security management and processing.

3.3 Blockchain in 5G

The basic objective of this key concept is to use Blockchain technology for trustworthy administrative relationships. Blockchain offers significant potential to enhance trust and transparency in 5G networks through its decentralized architecture and immutable ledger. By incorporating smart contracts, blockchain enables automatic execution of predefined conditions, ensuring data integrity and compliance without the need for central oversight. This use of a tamper-proof ledger to store data enables transparency. Each transaction and its associated data are verifiable by all network participants, leading to increased trust among stakeholders. This trust is crucial in 5G ecosystem where trust is crucial in multi-administrative domains.

3.3.1 State of the Art

Blockchain technology has effectively addressed issues of security and integrity in federated learning (FL). In particular, Chen *et al.* [CHEN24] utilized blockchain to mitigate malicious activities and manage reward distribution, thereby enhancing transparency within FL systems. Similarly, Chakraborty and Boudguiga [CHAKRA] integrated cryptographic methods with reputation metrics to refine leader selection, encouraging a decentralized structure in FL networks. Moreover, the works of Yang *et al.* [YANG] and Tang *et al.* [TANG] emphasized blockchain's role in bolstering node privacy and ensuring reliability in IoT applications. Blockchain's impact extends into RAN, where it enhances security and network management, particularly within O-RAN, enabling mobile operators to exchange resources and improve network adaptability [GIUPPO]. Xu *et al.* [XU] and Heider-Aviet *et al.* [HEID] further investigated decentralized RAN frameworks, focusing on privacy-preserving designs and cross-border blockchain-based connectivity.

Despite these advances, the integration of blockchain-based reputation systems in collaborative RAN environments remains underexplored, especially where O-RAN and FL intersect. This gap represents a promising area for improving network optimization, predictive maintenance, and user experience without compromising privacy [ZEYDAN]. The convergence of these technologies could significantly improve network responsiveness and adaptability.

Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D · Gobierno de España · cttc Centre Tecnològic de Telecomunicacions de Catalunya

3.3.2 Beyond the State of the Art in 6G DAWN Context

The 6G DAWN project explores the role of blockchain in building trust. Specifically, this project develops a trust layer that combines four components: the Monitoring System (MS) for data collection, the Analytical Engine (AE) for data analysis and behavior prediction, the Decision Engine (DE) for performance optimization, and the Actuator (ACT) for real-time network adjustments. Together, these components manage network slices with high efficiency and reliability [KUKL].

Framework:

Blockchain and its integration in 6G DAWN framework, illustrated in Figure 6 below, enhances smart contract operations in blockchain networks by distinguishing between off-chain and on-chain computations [BARRAC]. Off-chain computations handle resource-intensive tasks, such as local and global model training in FL, which reduces costs and latency on the blockchain. These processes are managed by FL clients and servers, where clients handle local updates and servers perform global aggregation, with both employing robust authentication mechanisms. On-chain operations ensure transparency through smart contracts that log final model updates and client reputations, maintaining an immutable record.

Oracle smart contracts facilitate communication between off-chain servers and the blockchain by transferring key metrics, such as normalized mean square error, which minimizes blockchain bloat and enhances security verification. By leveraging both off-chain and on-chain environments, the framework offers a scalable and efficient solution for FL systems. Offloading heavy computations to off-chain servers and service providers accelerates processing, alleviating blockchain's computational load and reducing transaction costs and network congestion.

A major contribution of this framework is the use of on-chain digital identities for servers and clients. These identities link off-chain processes with the blockchain's digital ledger, improving transparency, security, and traceability. This system supports better auditability while maintaining robust data processing and blockchain security.

Another key feature is the reputation system, where performance-based scores are assigned to servers and clients in the FL model. Although model training occurs off-chain, reputation scores and relevant metrics are managed on-chain. Clients submit local model updates, which are aggregated off-chain by servers. The ability to effectively learn from the global model influences their reputation score—positive results enhance their score, while poor outcomes decrease it. This feedback loop ensures quality contributions and reliable data integration.

All interactions and transactions are securely logged on the blockchain, providing a tamper-proof record of activities. This transparent documentation informs real-world decisions and strategies, creating a dynamic, bidirectional data flow between the blockchain and physical components. By

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

combining blockchain's digital ledger with the computational capabilities of off-chain servers and clients, our framework ensures accountability and verifiability throughout the FL environment.

In 6G DAWN, Analytical Engines (AEs) operate as FL clients, establishing communication with a central FL server. The server sends training instructions to the AEs for local model training, after which the AEs transmit model weights back to the server for aggregation. These weights are combined to produce an updated model, which is redistributed to AEs for further refinement. This iterative process enhances model accuracy and adapts to changing network conditions [JAVED]. Within O-RAN, a modular network architecture allows operators to integrate diverse equipment. A critical component in this setup is the Radio Intelligent Controller (RIC), which manages radio resources. The integration of FL into the RIC enables localized model training at network nodes or base stations, while adhering to strict data privacy guidelines. This distributed data processing across network cells allows for rapid model updates, which are essential for managing dynamic network conditions, such as radio resource allocation and traffic alignment.

Building upon these concepts, this work introduces a trust layer into the FL framework, particularly focusing on B2B administrative relationships between the FL server and its AEs. By automating these relationships and embedding trust mechanisms via smart contracts on the blockchain, the framework improves the reliability of these interactions. This improvement is crucial for fostering a collaborative and trustworthy environment across multiple RAN providers, ensuring privacy and decentralized cooperation.

We outline the framework for trust in collaborative FL within O-RAN environments. Figure 6 below depicts the system architecture that combines FL with blockchain across multiple RAN service providers. In this system, FL clients independently train local models at each RAN provider, submitting their updates (local weights) to a federated aggregation server. These updates are authenticated and aggregated into a global model, ensuring data confidentiality and integrity. Smart contracts on a blockchain network manage the aggregation of weights, securely facilitate transactions, and enforce compliance with protocols. Additionally, Chainlink supports data exchanges between the blockchain and external sources, increasing system responsiveness and accuracy. This design decentralizes the learning process and utilizes blockchain's immutability to create a trustworthy, transparent environment for collaborative model training and evaluation across multiple entities.

However, critical questions remain about the reliability of FL clients within this framework. The ZK-ML method offers a solution to these concerns by ensuring that the system's algorithms run correctly. ZK-ML principles complement FL, allowing decentralized and collaborative model training without sharing data. This method ensures consistent execution and protection against threats. An example of this approach is the ezkl library, which transforms TensorFlow or PyTorch computational graphs into zero-knowledge proofs (ZK-SNARK circuits), enabling verification of FL client operations on private data. Through this contribution of Blockchain integration in 6G DAWN, this work briefly touches upon ZK-ML's potential, its future applications warrant deeper exploration.

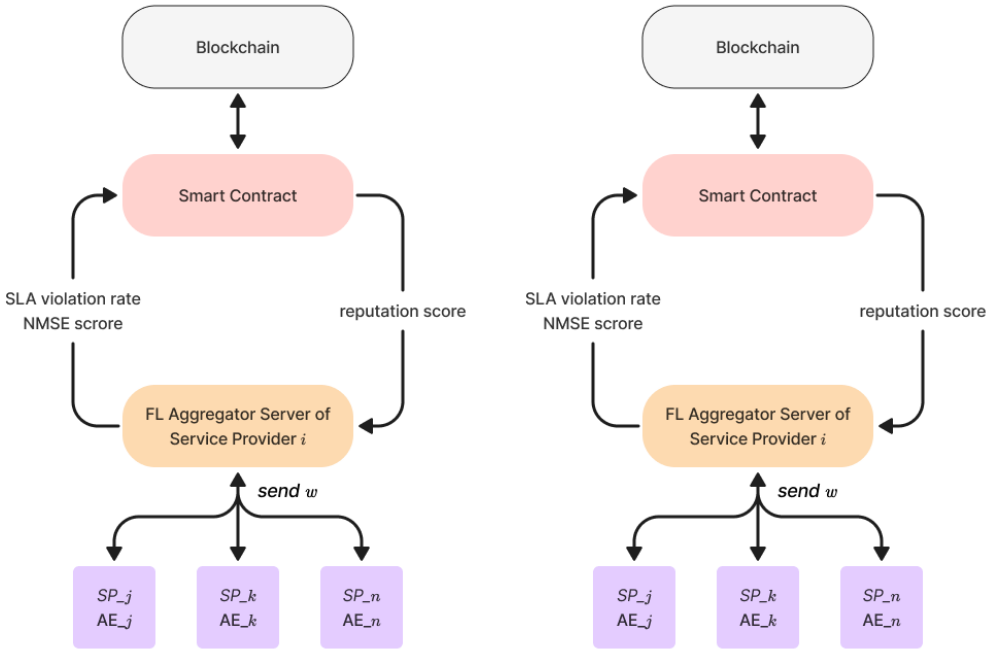


FIGURE 6: BLOCKCHAIN-ENABLED REPUTATION USING SMART CONTRACTS

3.4 Identification and handling of abnormally heavy energy components

As seen in the key concept 3.1 (Develop standard-compliant network interfaces to support AI driven network fault management systems at NPN), anomaly detection plays a pivotal role in addressing the challenge of energy efficiency in 6G networks by identifying irregularities, inefficiencies, and deviations from expected behavior. The identification and handling of anomalous high-energy users can both reduce the operational expenses and improve the network resiliency of NPNs. More concretely, the proposed detection mechanism within the RESILIENT subproject relies on the integration of the three sources of knowledge within the proposed network digital twin (key concept 2.1 – NPN Digital Twin System) and on the use of domain knowledge at the application level. Furthermore, the inclusion of an actuation mechanism based on the isolation of heavy energy components via their isolation of the main network, into a quarantine slice, is key to effectively improve the resiliency of the network.

3.4.1 State of the Art

The importance of identifying anomalous events within cellular networks has been partly recognized in Release 17 and 18, which have included, as part of the Network Data Analytics Function (NWDAF)

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

and 5GS architecture, specific triggers for requesting abnormal behavior analytics when unusual events such as frequent mobility re-registration of UEs or a high density of PDU session establishment in a given area are detected (see Table 3.1 for the complete list). However, these efforts have only put their focus towards the detection of strictly mobility or communication related abnormal behavior, overlooking the detection of events that, although individually may not be considered anomalous from a mobility and/or communications point of view, can jointly deviate energy consumption enough to be considered an anomaly.

TABLE 3.1 EXCEPTION IDS. BASED ON TS 29.520 [TS29520, TBL. 5.1.6.3.6–1]

Exception ID value	Description
UNEXPECTED_UE_LOCATION	Detection of a UE in an atypical or unexpected location (e.g. due to robbery). A possible response is to extend the Service Area Restrictions.
UNEXPECTED_LONG_LIVE_FLOW	UE establishing abnormally long traffic flows compared to historical data.
UNEXPECTED_LARGE_RATE_FLOW	UE establishing abnormally throughput-heavy traffic flows compared to historical data. Possible response is to decrease the MBR.
UNEXPECTED_WAKEUP	UE configured to send data periodically wakes up at an unexpected time.
SUSPICION_OF_DDOS_ATTACK	UE is suspected to be part or the target of a DDoS attack.
WRONG_DESTINATION_ADDRESS	UE is trying to send data to an incorrect/invalid address (e.g. due to outdated SRI).
TOO_FREQUENT_SERVICE_ACCESS	UE is establishing PDU sessions with an abnormal frequency or data volume.
UNEXPECTED_RADIO_LINK_FAILURES	UE link radio connectivity is being disrupted abnormally (e.g. too frequently).
PING_PONG_ACROSS_CELLS	UE frequently re-registering in the AMF from different cells

Within the literature, there is existing work on how to exploit and implement the NWDAF network function [LEE] [KIM], analyzing its potential use cases and possible architecture. More concretely, some of this work has centered around the use of this NF for anomaly detection, such as for detecting suddenly high traffic in a particular cell [SEVG]. However, work from an energy efficiency perspective is very limited in this regard.

3.4.2 Beyond the State of the Art in 6G DAWN Context

Due to the existing gap in the NWDAF specification, which does not encompass the detection of energy-related exceptions, and the lack of research in this regard, the RESILIENT architecture is

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

proposing the extension of the NWDAF analytics type values to also include energy abnormal analytics/events (on top of mobility and communication), exploring how anomalous energy-related events may be detected within the network and handled to reduce their negative impact on the network performance and availability. More concretely, a semi close loop detection algorithm, based on the three sources of knowledge within the designed network digital twin, together with an actuation policy-based algorithm is proposed, as the mechanism to detect these events. This system will be comprised of the following elements:

- **Monitoring**: for real-time energy consumption and performance data
- **Analytics**: for performing anomaly detection (for heavy energy users) and alerting the relevant services
- **Isolation**: leveraging URSP procedures for user allocation to quarantine slices
- **Policy**: creation of rules for managing the inspected behavior

This scenario, which shares some similarities with the use case 5.6 outlined in 3GPP TR 22.882 [TR22882], introduces multiple key differences and enhancements with respect to the latter one. More concretely, in terms of the detection responsibility, this enhanced architecture shifts the responsibility of detecting that a component (e.g. the UPF) is experiencing an abnormal energy consumption from the operator of the NPN to the 6G DAWN framework, which detects the event via the AE using the metrics collected by the MS and notifies the interested parties. This greatly reduces the effort required by NPN operators/external AF developers to handle these situations (who only need to worry about business domain). Additionally, while TR 22.882 [TR22882] only considers an application-specific fix (outside the scope of the network Core and RAN) to anomalous behaviors, the 6G DAWN framework introduces an interface that allows the enforcement of preventive network-specific actions with the intention to reduce its negative impact on the network. Via the DE component, it enables the controlled modification of PCF policies with the purpose of isolating the anomalous device via USRP rules.

Overall, the proposed system has the potential to simplify the work required by NPN operators to handle these events, achieves better interoperability with 3rd party systems (i.e. external AFs) and improves the resiliency of the network by automatically actuating upon the detection of the anomalous behavior.

4 Mapping of Key Concepts with Use Cases Proof of Concepts

6G DAWN RESILIENT contributions on technological key concepts are mapped in this section to the proof of concepts when they are dealt with, in some cases theoretically, which is indicated on Table 4.1, or in implementation and results.

TABLE 4.1: KEY CONCEPTS MAPPING VERSUS POCS

Key Concepts	**E UC1 PoC1**	**E UC1 PoC2**	**E UC2 PoC1**	**R UC1 PoC1**	**R UC1 PoC2**	**R UC2 PoC1**
NPN Digital Twin System		X				X
Extreme Edge			X	X		
AI/ML agent for control loops			X	X		X
xApps in O-RAN	X			X(theoretical)		
Relation of vertical KPIs with the network configuration		X		X		X
Inter(a)-slice reconfiguration and massive slicing		X				X
NEF instance for KPI data and configuration capabilities exposures of NPNs		X				X
AI/ML methods for reducing energy consumption	X	X				X
Develop standard compliant network interfaces to support AI driven network fault management systems at NPN						X
Decentralized Intrusion Detection to integrate novel trust-based evaluation mechanisms				X		
Blockchain in 5G					X	
Identification and handling of abnormally heavy energy components						X

5 Conclusions

The document presents the contributions of 6G DAWN RESILIENT subproject to the development of some key technical concepts beyond the state of the art. For each of the topics a state of the art is presented, and 6G DAWN RESILIENT contribution towards its enhancement. Many of the topics have been dealt with in collaboration of 6G DAWN ELASTIC subproject. In the last section the proofs of concepts where the key concepts have been implemented, or treated in theoretical way are identified.

This analysis resulted in the following architectural contributions that go beyond state of the art:

- The proposed Network Digital Twin (NDT) concept offers a comprehensive platform that covers everything from performance monitoring to optimization, focusing on PNI-NPN scenarios and utilizing theoretical, empirical, and machine learning models to enhance performance. Additionally, it integrates with 5G network architecture and supports both autonomous and collaborative strategies, providing a modular and open solution for intelligent network optimization.
- 6G DAWN introduces the extreme-edge domain as part of a multi-stakeholder, multi-domain network continuum, which enables unified orchestration of diverse resources across different technological and administrative domains. Additionally, 6G DAWN will develop a realistic emulator to simulate the complexity and volatility of the extreme-edge and implement an AI/ML-based predictive system to manage resource availability, enhancing fault tolerance and service continuity in 6G networks.
- The 6G DAWN architecture is designed to consider xApps that enable decentralized, scalable, and resilient network management, particularly for massive network slicing in 6G. The 6G DAWN ELASTIC sub-project focuses on distributed intelligence, allowing xApps to dynamically scale and adapt to real-time network demands, leveraging AI-driven adaptability for efficient resource management and proactive fault handling across multiple domains.
- The project focuses on reducing network energy consumption without compromising user experience by utilizing application-specific QoE metrics and energy data from theoretical models, lab measurements, and field data. Through integration with core network functions and monitoring agents at endpoints, the system will enable real-time optimization of energy usage and performance, particularly in mission-critical PNI-NPN scenarios, using insights from network digital twins for enhanced efficiency.
- In 6G DAWN, inter(a)-slice reconfiguration will adopt a distributed, AI-driven approach, enabling dynamic resource sharing between network slices based on real-time demand. The 6G DAWN ELASTIC project introduces scalable slicing, cross-domain coordination, and dynamic resource reallocation, enhancing efficiency and ensuring that multiple slices are simultaneously optimized without compromising service quality, with insights derived from digital twin modeling and machine learning.
- 6G-DAWN enhances 3GPP NEF services to address specific gaps for Non-Public Networks (NPNs), focusing on optimization for energy and performance, dynamic configuration

provisioning, and expanded analytic KPIs. These NEF extensions will empower Application Functions (AFs) to manage NPNs more effectively, improving network performance, resource efficiency, and resilience while enabling innovation in NPN management tools.

- The project uses probabilistic forecasting techniques to enhance O-RAN power efficiency deployments thanks to the use of AI to predict network conditions and resource requirements.
- 6G DAWN RESILIENT uses NDT and NEF as core components for a real-time, AI/ML-powered network monitoring system that detects and reports high-energy consumption anomalies to external applications via extended 3GPP APIs, promoting interoperability and innovation..
- The 6G DAWN-RESILIENT subproject envisions decentralized intrusion detection by combining local anomaly detection on each worker node with consolidated cluster-level monitoring and training for normal behavior patterns. Key innovations include multi-cluster monitoring, zero-touch deployment and ML training for unsupervised baseline detection, extended behavioral telemetry for higher intrusion detection rates, and enhanced explainability for network behavior analysis within Kubernetes environments.
- The 6G DAWN project incorporates blockchain to create a robust trust layer within a federated learning (FL) framework, integrating components like Monitoring Systems, Analytical Engines, Decision Engines, and Actuators to enhance secure, efficient management of network slices. By combining off-chain computations for resource-heavy tasks with on-chain smart contracts, digital identities, and a reputation system, the framework ensures transparency, security, and scalability, enabling collaborative, decentralized model training across multiple RAN providers with rigorous privacy and verification safeguards.
- The RESILIENT architecture within 6G DAWN enhances NWDAF by introducing energy anomaly detection and automated responses, addressing gaps in handling abnormal energy consumption events that impact network performance. By incorporating monitoring, anomaly detection, isolation, and policy management, this framework automates energy-related anomaly handling, reduces operator workload, and allows preventive, network-level actions to mitigate negative impacts on network resilience and availability.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

6 References

[6GDE3] 6G DAWN. Deliverable E3. First release of the use case requirements, KPIs and 6G DAWN architecture. 31 January 2024. [Online] Available at: https://6gdawn.cttc.es/images/6GDAWN_E3_RESILIENT_v01.pdf. Accessed: August 2023.

[5GACIA] 5G-ACIA "5G-ACIA 5G Non-Public Networks for Industrial Scenarios" https://5g-acia.org/wp-content/uploads/5G-ACIA_5G_Non-Public_Networks_for_Industrial_Scenarios_09-2021.pdf (accessed on December 12th, 2023)

[NDTERI] "Network digital twins – outlook and opportunities." retrieved from https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/network-digital-twins-outlook-and-opportunities.

[GRIEVE] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds., Cham: Springer International Publishing, 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7_4. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems | SpringerLink

[OLCO] Digital Twin Consortium, Digital Twin Consortium Defines Digital Twin, December 3, 2020, Olcott, S; Mullen, C

[BARR] "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications" IEEE Access, vol. 7, pp. 167653-167671, A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications, 2019, Barricelli, B.R.; Casiraghi, E; and Fogli, D (accessed on December 11th, 2023)

[HEX] Hexa-X. European level 6G Flagship projec A flagship for 6G vision and intelligent fabric of technology enablers connecting human, physical, and digital worlds. [Online] Available at: https://hexa-x.eu. Accessed: Aug. 2024.

[HEX2] Hexa-X-II. European level 6G Flagship project. [Online] Available at: https://hexa-x-ii.eu. Accessed: Aug. 2024.

[23.558] 3GPP TS 23.558 V19.1.0 (2024-03). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architecture for enabling Edge Applications. Release 19.

[NW-15] NetWorld. White Paper for Research Beyond 5G. Version: 1.0; (20-Oct-2015). Available at: https://www.sce.carleton.ca/faculty/yanikomeroglu/Pub/B5G-Vision-for-Researchv-1.0-for-public-consultation.pdf. Accessed: August 2024. (the same White

Paper can be also found at: https://5g-ppp.eu/wp-content/uploads/2015/12/Beyond5G.pdf, signed by the 5G Infrastructure Association).

[PWB16] Liu Peng, Dale Willis, and Suman Banerjee. "Paradrop: Enabling lightweight multi-tenancy at the network's extreme edge." 2016 IEEE/ACM Symposium on Edge Computing (SEC). IEEE, 2016. [Online] Available at: https://ieeexplore.ieee.org/abstract/document/7774349. Accessed: August 2024.

[PMJ+19] Portilla Jorge, Mujica Gabriel, Lee Jin-Shyan, Riesgo Teresa. (2019). The Extreme Edge at the Bottom of the Internet of Things: A Review. IEEE Sensors Journal. PP. 1-1. 10.1109/JSEN.2019.2891911.

[FUT21] Futuriom. 5G Catalysts: Telco Cloud and Edge Trends 2021. [Online] Available at: https://www.futuriom.com/articles/news/5g-catalysts-telco-cloud-and-edge-trends/2021/03. Accessed: August 2024.

[HV19] Cheol-Ho Hong and Blesson Varghese. "Resource Management in Fog/Edge Computing: A Survey on Architectures, Infrastructure, and Algorithms". In: ACM Comput. Surv. 52.5 (Sept. 2019). ISSN: 0360- 0300. DOI: 10.1145/3326066.

[ACC-19] Accedian. White Paper. The Extreme Edge. Managing performance at the network edge. 2019. [Online] Available at: https://tomfishercontent.wordpress.com/wp-content/uploads/2020/06/accedian_the-extreme-edge_whitepaper-2.pdf. Accessed: August 2024.

[MMS+20] Merino P, Mujica G, Señor J, Portilla J. A Modular IoT Hardware Platform for Distributed and Secured Extreme Edge Computing. Electronics. 2020; 9(3):538. https://doi.org/10.3390/electronics9030538

[RKN23] Visal Rajapakse, Ishan Karunanayake, and Nadeem Ahmed. 2023. Intelligence at the Extreme Edge: A Survey on Reformable TinyML. ACM Comput. Surv. 55, 13s, Article 282 (December 2023), 30 pages. https://doi.org/10.1145/3583683

[MKZ+17] Evangelos K. Markakis; Kimon Karras; Nikolaos Zotos; Anargyros Sideris; Theoharris Moysiadis; Angelo Corsaro et al., "EXEGESIS: Extreme Edge Resource Harvesting for a Virtualized Fog Environment," in IEEE Communications Magazine, vol. 55, no. 7, pp. 173-179, July 2017, doi: 10.1109/MCOM.2017.1600730.

[MEC003] ETSI GS MEC 003 V3.2.1 (2024-04). Multi-access Edge Computing (MEC); Framework and Reference Architecture. Reference RGS/MEC-0003v321Arch. [Online] Available at: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/03.02.01_60/gs_MEC003v030201p.pdf. Accessed: August 2024.

[HEXD61-21] Hexa-X. Deliverable D6.1 Gaps, features and enablers for B5G/6G service management and orchestration. June 2021. [Online] Available at: https://hexa-x.eu/wp-content/uploads/2021/06/Hexa-X_D6.1.pdf. Accessed: August 2024.

[HEX2D63-24] Hexa-X-II. Deliverable D6.3. Deliverable D6.3. Initial Design of 6G Smart Network Management Framework. June 2024. [Online] Available at: https://hexa-x-ii.eu/wp-content/uploads/2024/07/Hexa-X-II_D6-3_v1.0.pdf. Accessed: August 2024.

[5GC16] 5GCity. [Online] Available at: https://www.5gcity.eu/a-homepage-section. Accessed: August 2024.

[HEXD62] Hexa-X. Deliverable D6.2. Design of service management and orchestration functionalities. April 2022. [Online] Available at: https://hexa-x.eu/wp-content/uploads/2022/05/Hexa-X_D6.2_V1.1.pdf. Accessed: August 2024.

[ACC-19] Accedian. White Paper. The Extreme Edge. Managing performance at the network edge. 2019. [Online] Available at: https://tomfishercontent.wordpress.com/wp-content/uploads/2020/06/accedian_the-extreme-edge_whitepaper-2.pdf. Accessed: August 2024.

[ECP+02] D. Estrin, D. Culler, K. Pister and G. Sukhatme, “Connecting the physical world with pervasive networks,” in IEEE Pervasive Computing, vol. 1, no. 1, pp. 59-69, Jan.-March 2002, doi: 10.1109/MPRV.2002.993145.

[HEXD63] Hexa-X. Deliverable D6.3. Final evaluation of service management and orchestration mechanisms. 30/04/2023. [Online] Available at: https://hexa-x.eu/wp-content/uploads/2023/05/Hexa-X_D6.3_v.1.1.pdf. Accessed: August 2024.

[HEXPOC51] Hexa-X Demo 5 - Scenario 1- "Continuum orchestration of AI/ML driven traffic lights control service". [Online] Available at: https://youtu.be/daOKXkUAF60?list=PLkCf5oI9dmv0QpQjGqxVOzCV9djSSePAH. Accessed: August 2024. Described also in [HEXD63].

[HEX2DEM1] Hexa-X-II. Advanced M&O, Flexible topologies and Network beyond communications enablers in Cobot-powered WIM. [Online] Available at: https://youtu.be/Tmg6ihERzoM?list=PLVofa534-OwFuAx2AQeGlFXuGUbaXCkaR. Accessed: August 2024.

[HEXPOC52] Hexa-X: Towards the empowerment of manufacturing with cooperative robots & resilience through 6G. [Online] Available at: https://youtu.be/uHPtTyHX-O8. Accessed: August 2024. Described also in [HEXD63].

[HEX2DEM2] Hexa-X-II. E2E Extended Reality: Scalable Extended Reality (XR) testbed. [Online] Available at: https://youtu.be/bqyiN8AlBmA?list=PLVofa534-OwFuAx2AQeGlFXuGUbaXCkaR. Accessed: August 2024.

[LXD] LXD. [Online] Available at: https://documentation.ubuntu.com/lxd/en/latest. Accessed: August 2024.

[GITILE] https://gitlab.com/decentralized-continuum-orchestration/infrastructure-layer-emulator

[KTE300BC] Wikipedia. Control engineering. [Online] Available ar: https://en.wikipedia.org/wiki/Control_engineering. Accessed: Aug.2024.

[HAN22] John Hannavy. The Governor: Controlling the Power of Steam Machines. Pen and Sword Transport, 2022. ISBN 1399090895, 9781399090896.

[WIE48] Norbert Wiener. Cybernetics: Or Control and Communication in the Animal and the Machine. Paris, (Hermann & Cie) & Camb. Mass. (MIT Press) ISBN 978-0-262-73009-9; 1948.

[ZSM-009-1] ETSI GS ZSM 009-01, "Zero-touch network and Service Management (ZSM); Closed loop Automation; Part 1: Enablers", v1.1.1. June 2021.

[ZSM-009-2] ETSI GS ZSM 009-02, "Zero-touch network and Service Management (ZSM); Closed loop Automation; Part 2: Solutions for automation of E2E service and network management use cases", v1.1.1. June 2022.

[ZSM-009-3] ETSI GS ZSM 009-03, "Zero-touch network and Service Management (ZSM); Closed loop Automation; Part 3: Advanced topics", v1.1.1. August 2023.

[HEX2D63-24] Hexa-X-II. Deliverable D6.3. Deliverable D6.3. Initial Design of 6G Smart Network Management Framework. June 2024. [Online] Available at: https://hexa-x-ii.eu/wp-content/uploads/2024/07/Hexa-X-II_D6-3_v1.0.pdf. Accessed: August 2024.

[28.867] 3GPP Technical Report 28.867 V0.3.0 (2024-06). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; Closed control loop management (Release 19). [Online] Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4270. Accessed: Aug. 2024.

[28.861] 3GPP Technical Report 28.861 V16.0.0 (2019-12). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Study on the Self-Organizing Networks (SON) for 5G networks (Release 16). [Online] Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3556. Accessed: Aug. 2024.

[IG1220I] TM Forum. IG1220I Closed Loop Management Concept and Implications. v1.0.0. 24-Oct-2023. [Online] Available at: https://www.tmforum.org/resources/introductory-guide/ig1220i-closed-loop-management-concept-and-implications-v1-0-0. Accessed: Aug. 2024.

[IG1219A1] TM Forum Closed Loop Implementation and Management Levels. v1.0.0. 22-Dec-2023. [Online] Available at: https://www.tmforum.org/resources/introductory-guide/ig1219a1-closed-loop-implementation-and-management-levels-v1-0-0. Accessed: Aug. 2024.

[7575] Internet Research Task Force (IRTF). RFC 7575. Autonomic Networking: Definitions and Design Goals. June 2015. [Online] Available at: https://datatracker.ietf.org/doc/rfc7575. Accessed: Aug. 2024.

[BSS+21] Raouf Boutaba, Nashid Shahriar, Mohammad A. Salahuddin, Shihabur R. Chowdhury, Niloy Saha, and Alexander James. 2021. AI-driven Closed-loop Automation in 5G and beyond Mobile Networks. In Proceedings of the 4th FlexNets Workshop on Flexible Networks Artificial Intelligence Supported Network Flexibility and Agility (FlexNets '21). Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3472735.3474458.

[TBA+23] Tarik Taleb, Chafika Benzaïd, Rami Akrem Addad, Konstantinos Samdanis, AI/ML for beyond 5G systems: Concepts, technology enablers & solutions, Computer Networks, Volume 237, 2023, 110044, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2023.110044 . (https://www.sciencedirect.com/science/article/pii/S1389128623004899)

[Y.3173] International Telecommunication Union. ITU-T. Y.3172. (02/2020). Framework for evaluating intelligence levels of future networks including IMT-2020. [Online] Available at: file:///C:/Users/s234741/Downloads/T-REC-Y.3173-202002-I!!PDF-E.pdf. Accessed: Aug. 2024.

[SAS+23] K. Samdanis, A. N. Abbou, J. Song and T. Taleb, "AI/ML Service Enablers and Model Maintenance for Beyond 5G Networks," in IEEE Network, vol. 37, no. 5, pp. 162-172, Sept. 2023, doi: 10.1109/MNET.129.2200417.

[CCR+23] Juan Sebastian Camargo, Estefanía Coronado, Wilson Ramirez, et al, Dynamic slicing reconfiguration for virtualized 5G networks using ML forecasting of computing capacity, omputer Networks, Volume 236, 2023, 110001, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2023.110001

[VBM+21] D. de Vleeschauwer, Jorge Baranda, Josep Mangues-Bafalluy, et al., "5Growth Data-Driven AI-Based Scaling," 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Porto, Portugal, 2021, pp. 383-388, doi: 10.1109/EuCNC/6GSummit51104.2021.9482476.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

[TOU22] Thomas Tournaire. Model-based reinforcement learning for dynamic resource allocation in cloud environments. Computer science. Institut Polytechnique de Paris, 2022.

[TSG+21] Trakadas P, Sarakis L, Giannopoulos A, Spantideas S, Capsalis N, Gkonis P, Karkazis P, Rigazzi G, Antonopoulos A, Cambeiro MA, et al. A Cost-Efficient 5G Non-Public Network Architectural Approach: Key Concepts and Enablers, Building Blocks and Potential Use Cases. Sensors. 2021; 21(16):5578. https://doi.org/10.3390/s21165578.

[SP23] Schöning J, Pfisterer H-J. Safe and Trustful AI for Closed-Loop Control Systems. Electronics. 2023; 12(16):3489. https://doi.org/10.3390/electronics12163489

[SJN22] J. Shi, M. Jain, G. Narasimhan, "Time Series Forecasting (TSF) Using Various Deep Learning Models", arXiv:2204.11115v1 [cs.LG], https://arxiv.org/abs/2204.11115v1.

[HS97] Hochreiter, S. and Schmidhuber, J. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735. [Online]. Available at: https://www.researchgate.net/publication/13853244_Long_Short-term_Memory.

[WWW+21] P .B. Weerakody, K. W. Wong, G. Wang, W. Ela. A review of irregular time series data handling with gated recurrent neural networks, June 2021, DOI https://doi.org/10.1016/j.neucom.2021.02.046 [Online]. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0925231221003003?via%3Dihub [Accessed: 23 January 2024].

[FYL+22] Z. Fang, S. Yang, C. Lv, S.i An, W. Wu. Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study, July 2022, DOI: 10.1136/bmjopen-2021-056685 [Online]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9251895

[CWY+23] Q. Cao, Y. Wu,J. Yang, J. Yin, Greenhouse Temperature Prediction Based on Time-Series Features and LightGBM, January 2023, DOI https://doi.org/10.3390/app13031610 [Online]. Available at: https://www.mdpi.com/2076-3417/13/3/1610

[DIF21] I. Deznabi, M. Iyyer, M. Fiterau, Predicting in-hospital mortality by combining clinical notes with time-series data, January 2021 DOI 10.18653/v1/2021.findings-acl.352 [Online]. Available at: https://aclanthology.org/2021.findings-acl.352.pdf

[ANT+23] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, G. Rasool, Transformers in Time-Series Analysis: A Tutorial, July 2023, DOI https://doi.org/10.1007/s00034-023-02454-8 [Online]. Available at: https://link.springer.com/article/10.1007/s00034-023-02454-8.

[BBO18] A. Borovykh, S. Bohte, C. W. Oosterlee, Conditional Time Series Forecasting with Convolutional Neural Networks, September 2018, DOI https://doi.org/10.48550/arXiv.1703.04691 [Online]. Available at: https://arxiv.org/abs/1703.04691.

[GCW+22] Geva, M., Caciularu, A., Wang, K. R., & Goldberg, Y. (2022). Transformer feed-forward layers build predictions by promoting concepts in the vocabulary space. arXiv preprint arXiv:2203.14680.

[REZAZA] F. Rezazadeh et al., "Intelligible protocol learning for resource allocation in 6G O-RAN slicing,"IEEE Wireless Communications, vol. 31, no. 5, pp. 192-199, 2024.

[DRYJA] Dryjański, M., Kułacz, Ł., Kliks, A., 2021. Toward modular and flexible open RAN implementations in 6G networks: Traffic steering use case and O-RAN xapps. Sensors 21 (24), 8173.

[KOUCH] M. Kouchaki and V. Marojevic, "Actor-critic network for o-ran resource allocation: xapp design, deployment, and analysis," in 2022 IEEE Globecom Workshops (GC Wkshps). IEEE, 2022, pp. 968–973

[ATAL] T. O. Atalay, S. Maitra, D. Stojadinovic, A. Stavrou, and H. Wang, "Securing 5g openran with a scalable authorization framework for xapps," arXiv preprint arXiv:2212.11465, 2022.

[QAZZ] M.M. Qazzaz, Ł. Kułacz, A. Kliks, S.A. Zaidi, M. Dryjanskim, D. McLernon, Machine learning-based xApp for dynamic resource allocation in O-RAN networks, in: 2024 IEEE International Conference on Machine Learning for Communication and Networking, ICMLCN, IEEE, 2024, pp. 1–6.

[TS22179] "TS 22.179 Mission Critical Push to Talk (MCPTT). Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=623

[TS22281] "TS 22.281 Mission Critical (MC) video." [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3018

[TS22282] "TS 22.282 Mission Critical (MC) data." [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3019

[THAN] A. Thantharate, C. Beard, ADAPTIVE6g: Adaptive resource management for network slicing architectures in current 5G and future 6G systems, J. Netw. Syst. Manage. 31 (1) (2023) 9

[WU22] W. Wu, C. Zhou, M. Li, H. Wu, H. Zhou, N. Zhang, X. S. Shen, and W. Zhuang, "AI-native network slicing for 6G networks," IEEE Wireless Communications, vol. 29, no. 1, pp. 96–103, Apr. 2022.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

[ABDE23] A. A. Abdellatif, A. Abo-Eleneen, A. Mohamed, A. Erbad, N. V. Navkar et al., “Intelligent-slicing: An AI-assisted network slicing framework for 5G-and-beyond networks,” IEEE Trans. Network Serv. Manage., vol. 20, no. 2, pp. 1024–1039, May 2023.

[KOS] J. Kosińska and M. Tobiasz, "Detection of Cluster Anomalies With ML Techniques," in IEEE Access, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3216080

[KUBAB] “KubAnomaly: Anomaly detection for the Docker orchestration platform with neural network approaches”, Engineering Reports, Vol 1, Issue 5, Wiley, 2019.

[DCID] “DCIDS—Distributed Container IDS” Applied Sciences 13, no. 16: 9301. 2023

[KUB] “A Kubernetes dataset for misuse detection” ITU Journal on Future and Evolving Technologies, Vol. 4, Issue 2, 2023

[CHEN24] L. Chen, D. Zhao, L. Tao, K. Wang, S. Qiao, X. Zeng, and C. W. Tan, “A credible and fair federated learning framework based on blockchain,” IEEE Transactions on Artificial Intelligence, 2024.

[CHAKRA] O. Chakraborty and A. Boudguiga, “A decentralized federated learning using reputation,” Cryptology ePrint Archive, 2024.

[YANG] R. Yang, T. Zhao, F. R. Yu, M. Li, D. Zhang, and X. Zhao, “Blockchain-based federated learning with enhanced privacy and security using homomorphic encryption and reputation,” IEEE Internet of Things Journal, 2024.

[TANG] Y. Tang, Y. Liang, Y. Liu, J. Zhang, L. Ni, and L. Qi, “Reliable federated learning based on dual-reputation reverse auction mechanism in internet of things,” Future Generation Computer Systems, vol. 156, pp. 269–284, 2024.

[GIUPPO] L. Giupponi and F. Wilhelmi, “Blockchain-enabled network sharing for o-ran in 5g and beyond,” IEEE Network, vol. 36, no. 4, pp. 218–225, 2022.

[XU] H. Xu, L. Zhang, Y. Sun et al., “Be-ran: Blockchain-enabled open ran with decentralized identity management and privacy-preserving communication,” arXiv preprint arXiv:2101.10856, 2021.

[HEID] A. Heider-Aviet, D. R. Ollik, S. Berlato, S. Ranise, R. Carbone, N. El Ioini, C. Pahl, H. R. Barzegar et al., “Blockchain based ran data sharing,” in 2021 IEEE International Conference on Smart Data Services (SMDS), IEEE, 2021, pp. 152–161.

[ZEYDAN] E. Zeydan, L. Blanco, J. Mangues, S. Arslan, and Y. Turk, “Decentralized identity management for secure resource sharing in o-ran,” in 2023 IEEE 12th International Conference on Cloud Networking (CloudNet), IEEE, 2023, pp. 483–488.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D
GOBIERNO DE ESPAÑA
cttc Centre Tecnològic de Telecomunicacions de Catalunya

[KUKL] S. Kukliński, R. Kołakowski, L. Tomaszewski, L. Sanabria-Russo, C. Verikoukis, C.-T. Phan, L. Zanzi, F. Devoti, A. Ksentini, C. Tselios et al., "Monb5g: Ai/ml-capable distributed orchestration and management framework for network slices," in 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), IEEE, 2021, pp. 29–34.

[BARRAC] S. Barrachina-Muñoz, E. Zeydan, L. Blanco, L. Vettori, F. Rezazadeh, and J. Mangues-Bafalluy, "Cloud native federated learning for streaming: An experimental demonstrator," in 2023 IEEE 24th International Conference on High Performance Switching and Routing (HPSR), IEEE, 2023, pp. 1–3.

[JAVED] F. Javed and J. Mangues-Bafalluy, "Blockchain-based sla management for 6g networks," Internet Technology Letters, vol. 7, no. 3, p. e472, 2024.

[LEE] S. Lee et al., "Design and Implementation of Network Data Analytics Function in 5G," in 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), Oct. 2022, pp. 757–759. doi: 10.1109/ICTC55196.2022.9952559

[KIM] T. Kim et al., "An Implementation Study of Network Data Analytic Function in 5G," in 2022 IEEE International Conference on Consumer Electronics (ICCE), Jan. 2022, pp. 1–3. doi: 10.1109/ICCE53296.2022.9730290.

[SEVG] S. Sevgican, M. Turan, K. Gökarslan, H. B. Yilmaz, and T. Tugcu, "Intelligent network data analytics function in 5G cellular networks using machine learning," J. Commun. Netw., vol. 22, no. 3, pp. 269–280, Jun. 2020, doi: 10.1109/JCN.2020.000019.

Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D · GOBIERNO DE ESPAÑA · cttc Centre Tecnològic de Telecomunicacions de Catalunya