DETERMINISTIC6G

First report on DETERMINISTIC6G architecture

D1.2

The DETERMINISTIC6G project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no 1010965604.

First report on DETERMINISTIC6G architecture

Grant agreement number: Project title: Project acronym: Project website: Programme:	101096504 First report on DETERMINISTIC6G architecture DETERMINISTIC6G Deterministic6g.eu EU JU SNS Phase 1
Deliverable type: Deliverable reference number: Contributing workpackages: Dissemination level: Due date: Actual submission date:	Report D12 WP1 Public M16 30-04-2024
Responsible organization: Editor(s): Version number: Status:	EAB/EDD Joachim Sachs V1.0 Final
Short abstract:	The report elaborates on the system architecture for E2E deterministic communication with 6G.
Keywords:	Network Architecture, system architecture, deterministic communication, TSN, 6G, DetNet, Industry 5.0, Time Sensitive Communication, Extended Reality, Exoskeleton, Adaptive Manufacturing, Smart Farming, Cyber-Physical Systems

Contributor(s):	Joachim Sachs (EDD) János Harmatos (ETH), Dávid Jocha (ETH) Jose Costa Requena (CMC) James Gross (KTH), Gourav Prateek Sharma (KTH) Oliver Hoeftberger (B&R), Franz Profelt (B&R) Ognjen Dobrijevic (ABB) Giulia BIGONI (IUVO), Francesco Giovacchini (IUVO) Filippo Dell'Agnello (SSSA), Emilio Trigili (SSSA) Marilet De Andrade Jardim (EAB) Edgardo Montes de Oca (MI), Huu Nghia Nguyen (MI) Frank Dürr (USTUTT)

Disclaimer

Executive summary

Dependable, time-critical communication is poised to become a key technology enabler for future 6G networks. This necessity stems from the demand to facilitate a wide range of indoor and outdoor applications with high availability for time-critical services, across various domains, such as adaptive manufacturing, smart farming, extended reality (XR) and occupational exoskeleton. While the 5G system architecture already includes functional components and solutions to support use cases with low latency requirements, it also reveals certain shortcomings. Consequently, there is a need for additional enablers in 6G to efficiently support these emerging, visionary use-cases.

This deliverable provides a comprehensive overview about the initial DETERMINISTIC6G system architecture tailored to dependable, time-critical communication, as well as offers a detailed description of the architecture's deployment for realizing several time-critical use cases.

To clearly identify the capabilities and the limitations of existing 5G system components, the deliverable begins with an exhaustive state-of-the-art overview of solutions relevant for 6G related to time-aware communication and compute. This includes analyzing 5G support of non-public networks, 5G support for TSN and DetNet, network exposure, integration with edge computing, data analytics and machine learning, as well as security features. Furthermore, pivotal directions towards the 6G architecture, along with essential architecture design principles for 6G, considering the data-driven and AI-native system design are also investigated.

To pave the way towards a comprehensive architectural view, several technology concepts are proposed as components of the 6G architecture in the relevance of dependable communication. These concepts encompass time synchronization, packet delay correction, time-aware edge compute services, data-driven latency prediction and security-by-design principles. The detailed, architecture-related analysis of all the developed concepts is also presented.

Leveraging the integration of these functionalities, the initial DETERMINISTIC6G functional architecture is described in end-to-end scope focusing on how the proposed functionalities can be mapped in a 6G architecture scope. This also encompasses a unified overview of application requirements and their service specifications, as well as the coordinated usage of functional system components for dependable, time-critical communication and compute services.

The deliverable also offers deployment descriptions and aspects of the functional architecture intended to realize the various use cases. It includes a detailed analysis of how the capabilities of the developed system components can be leveraged to realize specific use cases, such as industrial exoskeleton, XR, adaptive manufacturing and smart farming.

Contents

1. Introduction

Digital transformation of industries and society is resulting in the emergence of a larger family of time-critical services with needs for high availability which present unique requirements distinct from traditional Internet applications like video streaming or web browsing. Time-critical services are already known in industrial automation; for example, an industrial control application that might require an end-to-end "over the loop" (i.e., from the sensor to the controller back to the actuator) latency of 2 ms and with a communication service requirement of 99.9999% [3GPP19-22261]. In the same way, with the increasing digitalization similar requirements are appearing in a growing number of new application domains, such as extended reality, autonomous vehicles, and adaptive manufacturing [DET23-D11]. The general long-term trend of digitalization leads towards a Cyber-Physical Continuum where the monitoring, control and maintenance functionality is moved from physical objects (like a robot, a machine or a tablet device) to a compute platform at some other location, where a digital representation – or digital twin – of the object is operated. Such Cyber Physical System (CPS) applications need a frequent and consistent information exchange between the digital and physical twins. Several technological developments in the Information and Communications Technology (ICT) sector drive this transition. The proliferation of (edge-) cloud compute paradigms provides new cost-efficient and scalable computing capabilities that are often more efficient to maintain and evolve compared to embedded compute solutions integrated into the physical objects. It also enables the creation of digital twins as a tool for advanced monitoring, prediction, automation of system components, and improved coordination of systems of systems. New techniques based on Machine Learning (ML) can be applied in application design that can operate over large data sets and profit from scalable compute infrastructure. Offloading compute functionality can also reduce spatial footprint, weight, cost, and energy consumption of physical objects, which is particularly important for mobile components, like vehicles, mobile robots, or wearable devices. This approach leads to an increasing need for communication between physical and digital objects, and this communication can span over multiple communication and computational domains. Communication in this cyber-physical world often includes closed-loop control interactions which can have stringent end-to-end Key Performance Indicators (KPI) (e.g., maximum packet delay and packet delay variation) requirements over the entire loop. In addition, many operations may have high criticality, such as business-critical tasks or even safety relevant operations. Therefore, it is necessary to provide dependable time-critical communications which provide service-assurance to achieve the agreed service requirements.

1.1. DETERMINISTIC6G approach

In the past, time-critical communication has mainly been prevalent in industrial automation scenarios with special compute hardware like Programmable Logic Controller (PLC), and is based on a wired communication system, such as Powerlink and EtherCat, which is limited to local and isolated network domains configured according to the specific purpose of the local applications [ECAT] [PLNK]. With the standardization of Time-Sensitive Networking (TSN) and Deterministic Networking (DetNet), similar capabilities are being introduced into the Ethernet and IP networking technologies, which thereby provide a converged multi-service network allowing time critical applications in a managed network infrastructure aiming for consistent performance with zero packet loss and guaranteed low and bounded latency [IEEE-TSN] [IETF-DETNET]. The underlying principles are that the network elements (i.e., bridges or routers) and the PLCs can provide a consistent and known performance with negligible

stochastic variation, which allows to manage the network configuration according to the needs of time-critical applications with known traffic characteristics and requirements.

It turns out that several elements in the digitalization journey introduce characteristics that deviate from the assumptions that are considered as baseline in the planning of deterministic networks. There is often an assumption for compute and communication elements, and applications, that any stochastic behavior can be minimized such that the time characteristics of the element can be clearly associated with tight minimum/maximum bounds. Cloud computing offers efficient and scalable computing resources, but introduces uncertainty in execution times. Wireless communications provide flexibility and simplicity, however they contain inherently stochastic components that lead to significant packet delay variations compared to those found in wired counterparts. Additionally, emerging applications incorporate novel technologies (e.g., ML-based or machine-vision-based control) where the traffic characteristics deviate from the strictly deterministic behavior of old-school control [SPS+23]. In addition, it is expected that there will be an increase in dynamic behavior, where characteristics of applications and network or compute elements may change over time in contrast to a static behavior that does not change during runtime. It turns out that these deviations of stochastic characteristics make traditional approaches to planning and configuration of end-to-end time-critical communication networks such as TSN or DetNet fall short regarding service performance, scalability, and efficiency. Instead, a revolutionary approach to the design, planning, and operation of time-critical networks is needed which fully embraces the variability but also dynamic changes that come at the side of introducing wireless connectivity, cloud compute and application innovation. The objective of DETERMINISTIC6G is to address these challenges: including the planning of communications and compute resource allocation for diverse time-critical services end-to-end over multiple domains, while providing efficient resource usage and a scalable solution [SPS+23].

DETERMINISTIC6G takes a novel approach towards converged future infrastructures for scalable cyber-physical systems deployment. With respect to networked infrastructures, DETERMINISTIC6G advocates (I) the acceptance and integration of stochastic elements (like wireless links and computational elements) with respect to their stochastic behavior captured through either short-term or longer-term envelopes. Monitoring and prediction of KPIs, for instance latency or reliability, can be leveraged to make individual elements plannable despite a remaining stochastic variance. Nevertheless, system enhancements to mitigate stochastic variances in communication and compute elements are also developed. (II) Next, DETERMINISTIC6G attempts to manage the entire end-to-end interaction loop (e.g., the control loop from the sensor to the controller to the actuator) with the underlying stochastic characteristics, especially while embracing the integration of compute elements. (III) Finally, due to unavoidable stochastic degradations of individual elements, DETERMINISTIC6G advocates allowing for adaptation between applications running on top of such converged and managed network infrastructures. The idea is to introduce flexibility in the application operation such that its requirements can be adjusted at runtime based on prevailing system conditions. This encompasses a larger set of application requirements that (a) can also accept stochastic end-to-end KPIs, and (b) that possibly can adapt end-to-end KPI requirements at run-time in harmonization with the networked infrastructure. DETERMINISTIC6G builds on a notion of time-awareness, by ensuring accurate and reliable time synchronicity while also ensuring security-by-design for such dependable time-critical communications. Generally, we extend a notion of deterministic communication, where all behavior of network and compute nodes and applications are pre-determined, towards dependable time-critical communication, where the focus is on ensuring that the communication (and

compute) characteristics are managed in order to provide the KPIs and reliability levels that are required by the application. DETERMINISTIC6G facilitates architectures and algorithms for scalable and converged future network infrastructures that enable dependable time-critical communication end-to-end, across domains, including 6G.

1.2. Objective of the document

DETERMINISTIC6G has developed several use cases and their requirements [DET23-D11]. Functionality for time-awareness based on time synchronization [DET23-D22], 6G capabilities for dependable communication [DET23-D21] and approaches for end-to-end dependable time-critical communication including fixed and wireless domains [DET23-D31] including security solutions [DET23-D32] have been proposed. The objective of this document is to describe a first architecture proposal that integrates the functionality listed above in a 6G network architecture and provides dependable time critical communication end-to-end.

An architecture can be described with different purposes [RÖT+23]. A *functional architecture* describes functional blocks and their relationships and interactions. It is often the baseline for standardization. An implementation architecture describes how functionality is realized in a real system. Often different design choices exist, on how functional blocks are grouped and how they are implemented. The functional architecture should provide sufficient freedom for implementation choices and optimizations. A central part of the functional architecture is to define interfaces where system components of several different vendors are integrated, and where open standardized interfaces enable commercially relevant system realizations. A deployment architecture describes how a network is practically deployed in a specific environment. The functional and implementation architectures shall allow for flexible deployments, so that it can efficiently realize the use cases envisioned in the deployment area. The focus of this document is on a functional architecture description, but also addresses some deployment aspects.

1.3. Relation to other work packages and deliverables

This deliverable is part of work package (WP) 1 and has relation to other technical work packages, as shown in Figure 1.1. To analyse the architectural impacts as well as being able to define a new dependable E2E architecture based on 6G for time-critical applications, WP1 used as input the outcome of WP2 “6G Centric Enablers for Deterministic Communication” and WP3 “Enablers of 6G Convergence for Deterministic Communication”. In particular, these work packages have provided results that are documented in the following deliverables:

- D2.1: 6G Centric Enablers [DET23-D21]
- D2.2: Time Synchronization for E2E Time Awareness [DET23-D22]
- D3.1: 6G Convergence Enablers Towards Deterministic Communication Standards [DET23-D31]
- D3.2: Security solutions [DET23-D32]

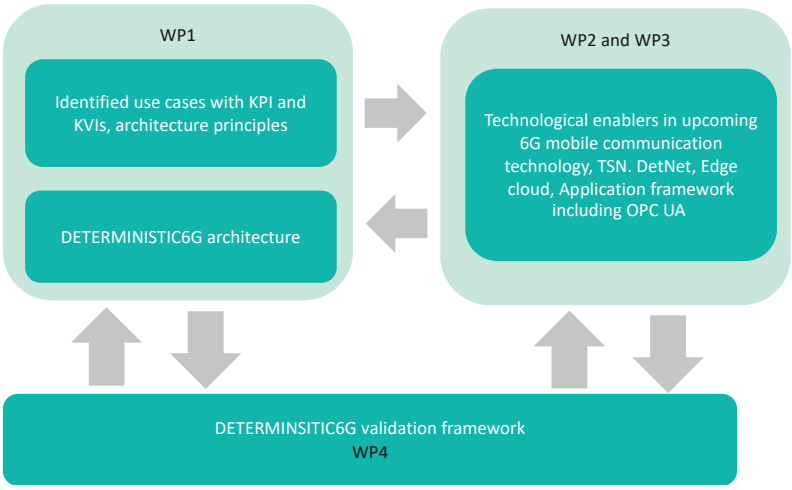

Figure 1.1. Relation between work packages in DETERMINISTIC6G

As input WP2 and WP3 used the results of WP1 in D1.1 entitled “DETERMINISTIC6G Use Cases and Architecture Principles” [DET23-D11] including the visionary use cases, and their KPIs and KVIs. The use cases and architectural principles outlined in [DET23-D11] are also used as basis and reference in this report. In the future, WP4 will then evaluate the new architecture in this deliverable via the validation framework that was published in D4.1, entitled “DETERMINISTIC6G DetCom simulator framework release 1” [DET23-D41]. WP2 and WP3 will as well consider the defined architecture in this deliverable D1.2 to complete the study on enablers for dependable E2E time-critical communications involving 6G converged with deterministic technologies.

1.4. Structure of the document

The document is structured as follows. In chapter 2, the state-of-the-art (derived from 5G) and external proposals for the 6G architecture are reviewed. Chapter 3 introduces the functionality that has been developed in DETERMINISTIC6G. Chapter 4 describes how the functionality of chapter 3 can be addressed in a 6G architecture. This is followed by chapter 5 which shows how the architecture can be deployed to address some of the DETERMINSTIC6G use cases.

2. Review of system architecture concepts

The system architecture of DETERMINISTIC6G proposed later in chapter 4 builds on earlier architecture work, which is reviewed in this chapter. First, 5G system architecture already comprises components that are relevant for DETERMINISTIC6G, as described in section 2.1, and can be expected to be reused in the migration from 5G towards 6G as explained in section 2.1.6. Furthermore, outside DETERMINISTIC6G, significant efforts are put into the exploration of 6G technology components and their integration into a 6G system architecture, which is summarized in section 2.2.

2.1. System architecture components in 5G relevant for DETERMINISTIC6G

The concepts of end-to-end dependable communication with 6G build on earlier work with 5G. In this section we review those architecture concepts of 5G that are relevant.

2.1.1. Support for non-public networks

In order to address use cases where a closed group of mobile devices obtain access to a private network, solutions for non-public networks (NPN) have been standardized for 5G. These approaches have been motivated largely by the interest of reliable private 5G networks for industrial use cases, and have been described in e.g. [5GAC21b] [GLS+22] [5GS20-D52] [5GAC24] [3GPP18-23501]. They are particularly motivated with the regulatory decisions in several regions to make spectrum available for local licenses in order to deploy local on-site private networks [FHB+24]. There are different approaches on how a private 5G network can be deployed depending on a possible coordination or integration of the private 5G network with a public 5G network, as shown in Figure 2.1. In so-called *standalone non-public networks* (SNPN), an isolated private network is locally deployed and operated, for example, within the premises of a factory by making use of a local spectrum license. All network equipment is deployed on-site and there is no need for a connection to a public mobile network. As one option, the local private 5G network may apply *Radio Access Network (RAN) sharing* with a public network, which means that the same radio network equipment can serve public 5G devices and private 5G devices. RAN sharing is known from public networks when multiple network operators share the costs for deploying a common RAN, while the common RAN appears to each of the mobile networks as their own RAN. In the context of local NPNs with local spectrum licenses, the motivation is less a sharing of RAN costs, but the benefit of spectrum coordination, which also provides better and more spectrally efficient connectivity for both public and private UEs. With locally licensed spectrum, the same spectrum may be used by neighbouring networks with own local spectrum licenses in vicinity. By the nature of radio propagation, simultaneous usage of the same spectrum in vicinity can cause mutual interference and requires efficient coexistence, which is relevant not only for co-channels but also adjacent channels [5GS20-D14] [5GS21-D15] [CAS+22] [CAS+23]. RAN sharing can harmonize the spectrum usage and provide benefits in coexistence.

As alternative to SNPNs, *public-network integrated NPNs* (PNI-NPN) have been standardized. In a PNI-NPN, the NPN is typically provided by a mobile network operator that is sharing parts of its existing network infrastructure to deliver the NPN to a customer. For example, for providing a non-public network at a local confined area (e.g., a factory), a mobile network operator may provide dedicated core network user-plane nodes locally in addition to providing a local RAN at the premises. This ensures low latency performance and good integration with available local network infrastructure; also, sensitive data is not leaving the customer premises. In contrast, some of the control-plane functionality of the network may be using the public network instances, which may allow some of the NPN devices to also roam into the public instances. The realization of a PNI-NPN is flexible depending on how much public network resources are shared with the NPN. In the extreme, the PNI-NPN is realized as a network slice (i.e., a virtual set of private network functions) that is provided by the public mobile network operator to the NPN customer (and which can still be complemented with local RAN nodes to guarantee sufficient capacity and availability). Note, that a PNI-NPN can be configured to use the spectrum assets licensed by the mobile network operator but also other spectrum resources like those obtained by the factory via a local spectrum license.

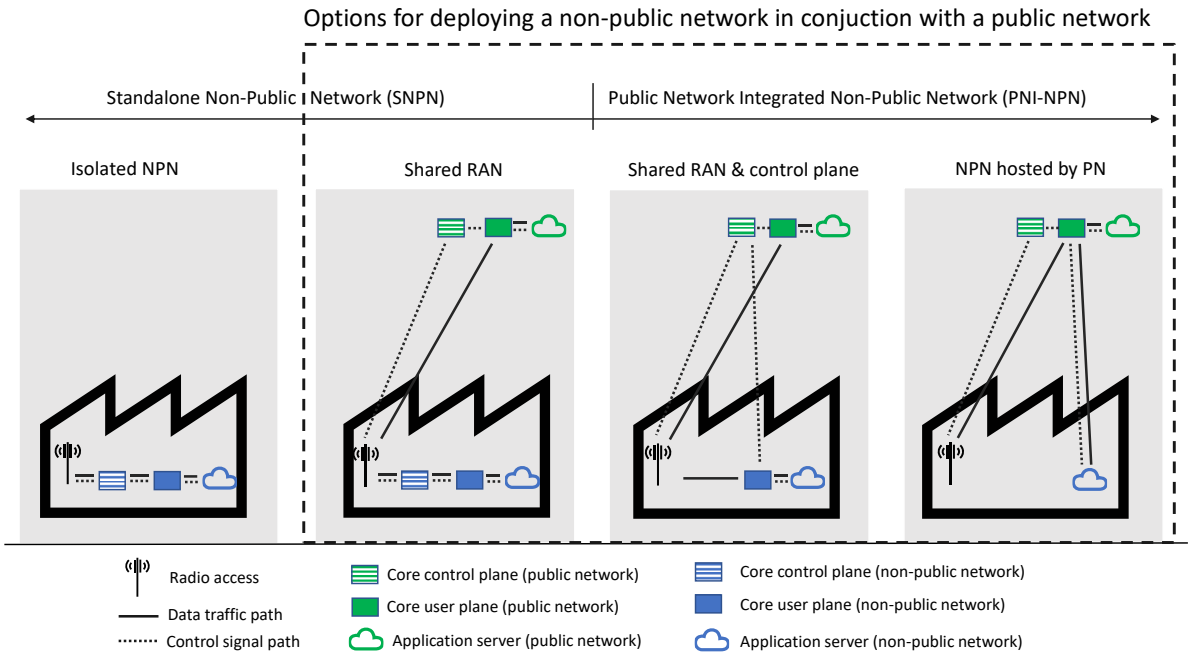

Figure 2.1. Non-public networks (based on [FHB+24])

2.1.2. Support for TSN and DetNet in 5G System

3GPP introduced the time-sensitive communications (TSC) framework in 5G in order to support deterministic type of communications that have been standardized in IEEE, namely Time-Sensitive Networking (TSN), and in IETF, namely Deterministic Networking (DetNet). The main enabler to support these technologies in the RAN is Ultra Reliable Low Latency Communications (URLLC) which guarantees a bounded delay, delay-critical type of quality of service (QoS), and reliable communication.

With TSC the 5G system is modelled as a node which mimics the behavior of a regular fixed node that supports time-sensitive communication [5GS20-D51] [3GPP18-23501] [5GAC21c]. Figure 2.2 shows an example how a 5G logical TSN bridge is placed in a TSN network. Specifically, the 5G logical TSN bridge is connected to other TSN bridges or TSN end stations (such as Talker or Listener). The 5G logical bridge also interfaces an external control plane management entity, in this case the TSN controller (aka Centralized Network Configuration (CNC)). Note that in the case of a fully centralized architecture in TSN, applications in the Talker or Listener (aka TSN end stations) communicate with a Centralized User Configuration (CUC) which oversees user configurations, which then is passed in terms of communication requirements to the CNC, the network controller. After the CNC has collected all TSN flows' requirements from the CUC and all bridge components' capabilities, the CNC sets the configuration for all the bridge components (TSN bridges and TSN end station Network Card Interface (NIC)) [IEEEQcc] [IEEEQdj] using a configuration protocol such as NETCONF [RFC 6241] or RESTCONF [RFC 8040].

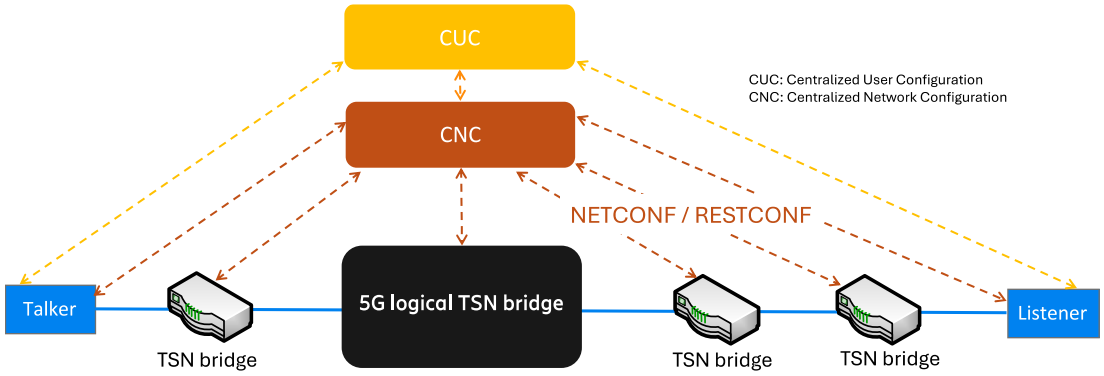

Figure 2.2. 5G logical TSN bridge in a TSN network

To interface 5G system with the other TSN nodes, 3GPP introduced TSN translators (TT) at the device side (DS-TT) and at the network side (NW-TT), as shown in Figure 2.3. The translators provide the TSN Ethernet interface and mimic the expected behavior of a number of TSN features, without the need to radically modify the functionalities of the 5G nodes and functions. To interface with the CNC, a TSN application function (TSN AF) was defined. The TSN AF obtains the 5G bridge capabilities and provides them to the CNC, while the CNC configures the 5G bridge via the TSN AF. The TSN AF translates the capabilities to the parameters that the CNC handles and translates the configuration from the CNC into a flow set up in the 5G system. Finally, the TSN AF also handles the typical time synchronization data sets that can be configured by the CNC.

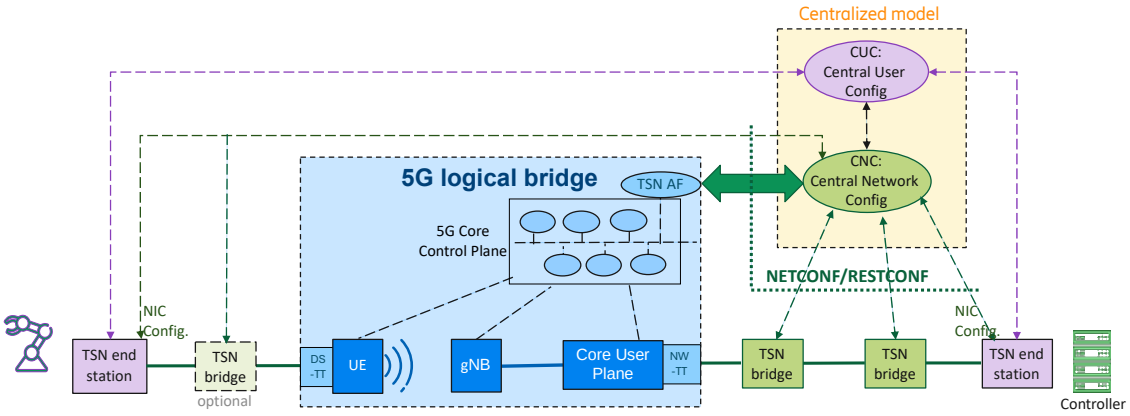

Figure 2.3. 5G system acting as a TSN bridge.

Note that TSN is a specific case of TSC. In general, TSC would involve the use of a similar entity to the TSN AF, known as the TSC and Time Synchronization Function (TSCTSF). TSCTSF performs similar tasks as the TSN AF but with a generalized exposure interface that is proxied via the Network Exposure Function (NEF) as shown in Figure 2.4, unless the external Application Function (AF) is a trusted entity for the 5G system, in which case the NEF is not needed. In the general case, any AF can request a deterministic service via 5G and the 5G system is considered a node.

Figure 2.4. TSN as a special case of the 5G Time-Sensitive Communications (TSC)

A similar technology to TSN has been standardized in IETF, namely Deterministic Networking (DetNet), which is implemented in layer 3 of the OSI reference model supporting IP-based communications. Similarly, 3GPP has modeled 5G system as a logical DetNet node, just as illustrated in Figure 2.5.

In the case of DetNet, the general TSC framework was reused where the TSCTSF interacts with the DetNet controller (instead of AF) without the need for NEF since the controller is considered a trusted entity. The information used in this interface TSCTSF-DetNet controller uses the defined YANG models for DetNet. TSCTSF translates the configuration from the DetNet controller into requirements towards the 5G system, similarly as in the case of TSN AF. With DetNet there is no requirement for a DS-TT since the interface is IP-based which is already supported in 5G, and no additional translation is needed. If a time synchronization service is required, then the DS-TT will be required.

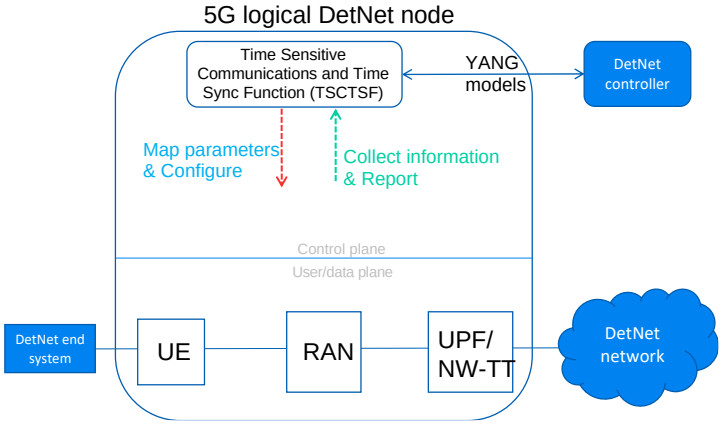

Figure 2.5. 5G logical DetNet node

2.1.3. Network Exposure

Network exposure provides a means for configuring and monitoring the network and the communication services. It is based on Application Programming Interfaces (API) to provide a level of network programmability that allows to customize the network services to the desired use cases. [5GAC21a] has defined the requirements on network exposure for 5G non-public networks to be able to address industrial use cases. Network exposure requirements are grouped into capabilities. One group is related to device management, which includes the device connectivity management and connectivity monitoring that allow to establish communication services to different applications that then provide the required QoS [5GAC21d]. Another group of exposure requirements is focused on network management. [5GS21-D55] [SK23] [GSD+22] [KDS+23] describe how network exposure can be applied in industrial control use cases. It is also worthwhile mentioning ongoing 3GPP efforts on 5G network exposure APIs, namely 5G NEF [3GPP23-29522] and 5G Service Enabler Architecture Layer for Verticals (SEAL) [3GPP23-23434], to address some specific aspects of TSC. At the time of preparing this deliverable, the latter exposure features relate to, e.g., providing application QoS requirements to the network and establishing a TSC session in 5G networks.

Network programmability via APIs is not only relevant for non-public networks, but also plays a role in wide-area and public mobile networks. These APIs allow to make network capabilities consumable by end users [SKM+21]. Requesting differentiated capabilities from the network can include commercial agreements to be initiated via network exposure [FMK23]. In recent times, new public network exposure initiatives have started that specify industry-aligned APIs to provide application developers easy ways to make public network usage flexible and fit for the purpose of applications. For example, *the Telco Global API Alliance* (CAMARA) is working on standardized network APIs that allow to configure the network with a dynamic network slice at a defined geographic area and time period. Via such an API it is possible to obtain for devices connected to this network slice a guaranteed service level performance [CAM24a] or alternatively a specific QoS support for an application can be requested [CAM24b].

2.1.4. Support for edge computing in 5G

Edge computing leverages the distributed computing paradigm and provides an ecosystem where the execution environment (e.g., compute and resource storages) is closer to the location where the task is invoked compared to the traditional cloud computing paradigm. The proximity of edge premise results in reduced latency between a client and the server application, so edge computing is able to support use cases where low latency is a crucial requirement. Hence, edge computing enables to realize use cases where time-critical applications are moved to the virtualized environment, instead of using dedicated, specialized hardware. Furthermore, the cloudification makes possible the further evolution of the applications by leveraging the cloud-native design paradigm.

2.1.4.1. 3GPP edge computing support architecture

3GPP edge computing is a new paradigm introduced with 5G to bring the data processing of applications closer to their physical location, either end user or data source. This reduces the traffic load that reaches the mobile operator infrastructure and outsources the computing process to edge nodes closer to the user.

3GPP SA2 group introduced features to support edge computing defined in TS 23.548 [3GPP23-23548]. This specification outlines three connectivity models supported by the 5G core to enable edge computing. It defines various functionalities for traffic steering and User Plane Function (UPF) selection for realizing these different connectivity models. TS 23.548 provides a detailed description of how the 5G Core supports the Edge Application Server (EAS) discovery/re-discovery in the case of the various connectivity models.

The TS23.588 [3GPP23-23558], specified by the 3GPP SA6 group described Application Layer Architecture for enabling edge applications, is illustrated in Figure 2.6.

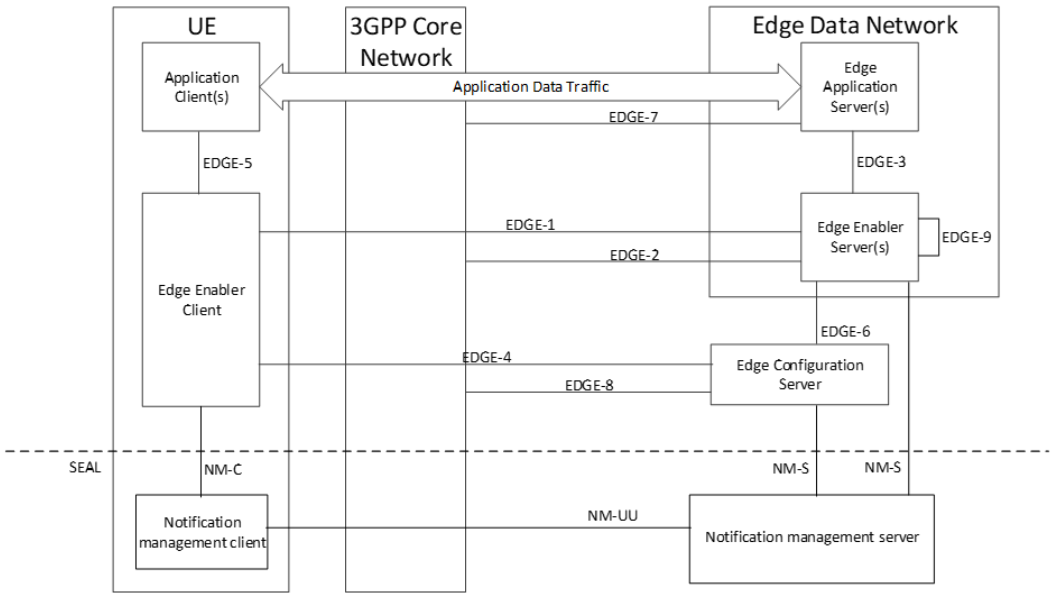

Figure 2.6. 3GPP TS23.558 defined edge computing architecture

This architecture makes the user equipment (UE) edge-aware, which means that all the devices include an Edge Enabler Client (EEC). The EEC communicates with the Edge Configuration Server (ECS), which provides the required configuration and supporting functions to setup a data session from the application client (hosted by the UE) to the Edge Application Server (EAS).

TS 23.548 and TS 23.558 3GPP defines a set of Network Functions (NFs), listed in Table 2.1, to provide all the default functionality to register, authenticate and provide data sessions to the devices to support edge computing and interact with the rest of the NFs in the 5G system (5GS).

Table 2.1. 3GPP-defined edge computing support Network Functions

NF Acronym	NF Definition	NF Functionality
AC	Application Context (Client)	A set of data about the Application Client that resides in the Edge Application server
ACT	Application Context Transfer	Refers to the transfer of the Application Context between the source Edge Application Server and the target Edge Application Server
ACR	Application Context Relocation	Refers to the end-to-end service continuity procedure of relocating the Edge Application Server due to some reason (e.g., user plane change, AF request, etc.)
AF	Application Function	Control plane function interacting with NEF within 5G core network, to provide application services to the subscriber
EAS	Edge Application Server	Application software resident in the edge performing the server function
EASDF	Edge Application Server Discovery Function	A DNS resolver/server locally deployed by the 5GC operator within the local data network, responsible

		for resolving UE DNS queries into suitable EAS IP address(es)
ECS	Edge Configuration Server	Provides configurations to the EEC to connect with an EAS.
EDN	Edge Data Network	A local data network that supports the architecture for enabling edge applications.
EEC	Edge Enabler Client	Provides support functions, such as EAS discovery to the ACs in the UE (DNS client)
EES	Edge Enabler Server	Formerly responsible for enabling discovery of the EAS

2.1.4.2. ETSI MEC Architecture

Multi-access edge computing (MEC) is an application scenario of edge computing used for mobile networks, defined by the European Telecommunications Standards Institute (ETSI). The objective of the MEC as defined in ETSI is to create a continuum between the telco and IT-cloud worlds. The ETSI architecture provides a generic architecture to facilitate the integration of IT and cloud computing capabilities with mobile networks. Figure 2.7 shows the components defined in the ETSI architecture that is structured in three layers. The lower layer consists of the network infrastructure that provides the basic connectivity i.e., local network and 3GPP mobile network, between the devices and the MEC platform. The middle layer provides the platform that is hosting the edge computing infrastructure, including the virtualization components required to run the edge applications and the management system that handles the available resources on the host where the MEC platform is deployed. The higher layer consists of a system level management that provides overall visibility of the devices and edge computing platform.

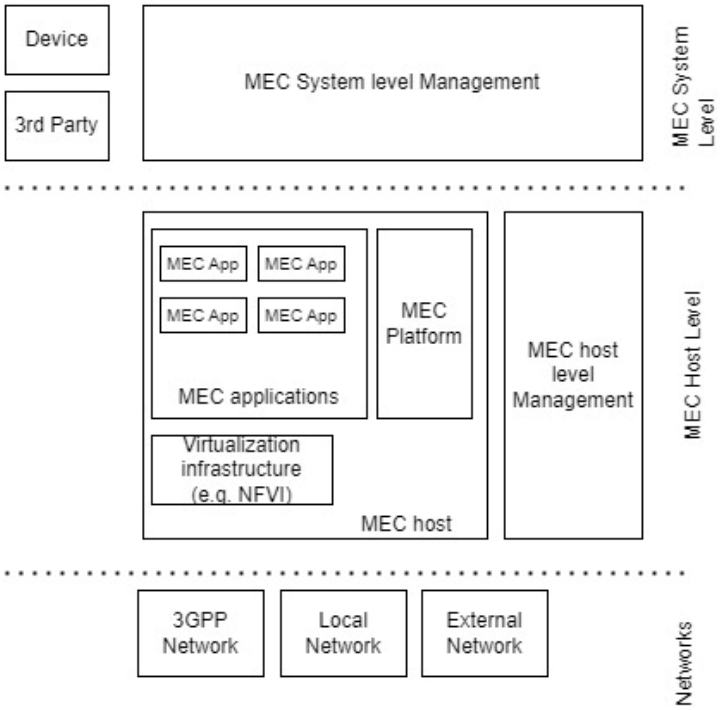

Figure 2.7. Multi-access edge computing framework (Ref ETSI GS MEC 003 V3.1.1)

2.1.4.3. Integration aspects of edge computing with 5G-TSN

The architecture integration aspects of edge computing and various NPN deployment options are extensively analyzed in [5GS21-D54] in the context of manufacturing use cases. Similar topic is overviewed by a 5G-ACIA white paper [5GAC23] to cover use cases, architecture, and different deployment options of industrial 5G edge computing. However, the aspects of dependability were discussed in a very limited way, and the functional integration of the edge computing deployment with the TSN communication domain was not covered. One major objective of DETERMINISTIC6G is to fill this gap, concentrating first on the seamless IEEE 802.1Qbv end-to-end traffic handling support for the cloudified applications.

The integration of the TSN 802.1CB Frame Replication and Elimination for Reliability (FRER) with the cloud reliability features is also discussed in [5GS21-D54]. The essence of the proposed architecture options is to leverage the capability of the cloud to host multiple application instances to provide redundancy, while hiding these instances and emulating them as a single instance towards the end device to ensure backward compatibility, as shown in Figure 2.8.

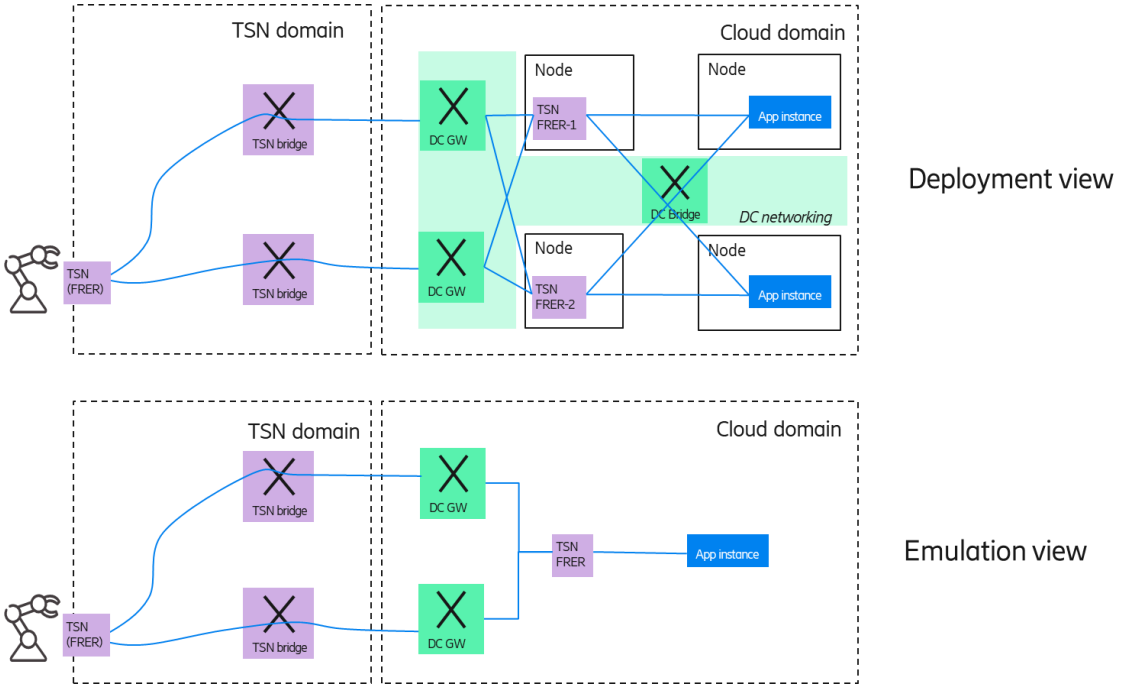

Figure 2.8. Emulated, cloudified application handling scenario for seamless FRER support

These studies are focused on scenarios in which multiple application instances are deployed in the edge domain, but only one instance communicates with the end device at any given time. If a failure occurs, another application instance can be invoked to continue the serving of the end device. A downside of this solution is that it cannot ensure continuous communication, since the failure needs to be detected and the other application instance needs to be invoked, which takes some time. To overcome this limitation, in DETERMINISTIC6G the intention is to develop a concept, which enables to guarantee seamless operation with zero-failover time, in the case of any single failure, but still ensures backward compatibility.

2.1.5. Review of AI/ML in 5G

The operations of mobile networks have been transitioning from manual processes towards sophisticated automated process flows as the networks become much larger and more complex. Machine Learning (ML) and Artificial Intelligence (AI) have been envisioned to be crucial technologies for enabling automation of network operations (e.g., resource optimization, fault prediction, security policies) in future mobile networks. In recent years, techniques based on AI/ML are being proposed to be used across various domains of the 5G system, including operations, administration, and management (OAM) (e.g., Management and Orchestration), 5G core network, or RAN (e.g., AI/ML-enabled RAN intelligence) [LIN23].

The Network Data Analytics Function (NWDAF) is a key component of the 5G core network, introduced in Release 15 by the 3GPP to enhance 5G core network capabilities with respect to AI/ML [3GPP17-23288]. The primary objective of NWDAF is comprehensive data collection and analytics functionalities. To this end, NWDAF is responsible for gathering vast amounts of data from various sources within the 5G system and for providing analytics to the consumer NFs. To this end, NWDAF can interact with different entities in the 5G system as shown in Figure 2.9.

The first interaction domain is the 5G core network itself where the NWDAF is located. Here, various NFs can be the producers of the data towards NWDAF and also the consumers of its generated analytics. For example, the Access and Mobility Management Function (AMF) and the Session Management Function (SMF) might produce data regarding user mobility and service usage, which can then be analyzed by NWDAF to optimize handover protocols and session resource allocation across different network cells [3GPP17-23288]. The second interaction domain for NWDAF is OAM. The OAM feeds the data to NWDAF which it has obtained from the measurement probes located in RAN and relevant 5G NFs. In the OAM domain, the Management Data Analytics Function (MDAF) has been specified which is also responsible for the interactions with NWDAF [3GPP16-28533]. Lastly, the third set of interactions for NWDAF is in the service domain. The functions residing outside the scope of 3GPP trust domain must provide data or consume analytics from NWDAF through AFs. For instance, the NWDAF might produce statistics and predictions about user service quality which is consumed in the service domain through an AF.

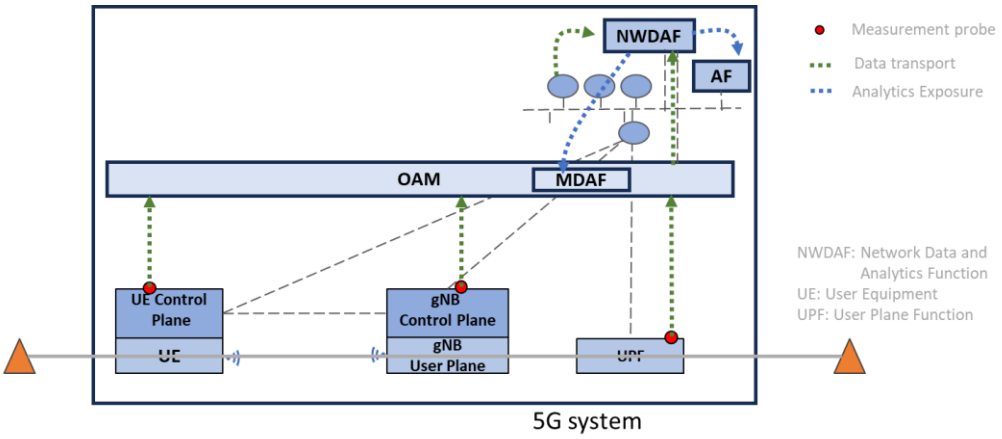

Figure 2.9 An illustration of data collection and exposure in 5G based on NWDAF.

It is important to point out that the Data Collection Coordination Function (DCCF) has been specified to avoid duplication of requests for data as well as analytics between various NFs and NWDAF. In other words, all requests for data and analytics are sent to DCCF which might further rely on a messaging framework to collect analytics and deliver it to the NFs.

To facilitate the deployment and operation of AI/ML capabilities in the 5G system, machine learning models need to be managed throughout their lifecycle. To this end, the functionality in the NWDAF is handled by two logical sub functions: (i) Analytics Logical Function (AnLF) and (ii) Model Training Logical Function (MTLF) [3GPP17-23288]. AnLF using DCCF can access the vast amounts of data collected and applies ML models to generate predictions, e.g., predicting network congestion based on UE mobility patterns. On the other hand, the MTLF is responsible for building and refining the ML models that AnLF utilizes. MTLF is tasked with training and updating models using the data collected.

A functional framework for AI-enabled RAN intelligence has been described in [3GPP22-37817]. This framework describes the key functional entities relevant for integrating intelligence into RAN for selected use cases. The framework serves as a good foundation to explore the architecture aspects of integrating data-driven (AI/ML) approaches for latency prediction into the 5G-Advanced (5G-Adv) / 6G architecture. We describe different components of this framework with respect to data-driven latency predictions that are envisioned to be a part of future 5G-Adv/6G networks in section 3.4.

2.1.6. Architecture directions from 5G to 6G

As a new 6G RAN and core network functionality will be deployed, it is important to understand how it relates to already deployed 5G networks and how a network migration can take place to integrate new functionality. A similar challenge has happened before, most recently with the introduction of 5G after finalization of the first 5G standard in 2018. For 5G introduction several options have been standardized, allowing different migration paths from a 4G deployment to 5G. The migration comprised several network domains, the existing 4G radio access network and a new 5G radio access network, the existing 4G core network and the new 5G core network, and the paths on how the new 5G RAN and 5G CN could be connected to existing 4G network domains, resulting also in different ways in which a mobile device i.e. UE can connect to a 5G network [GSMA18] [CRVW18]. This range of migration options required support in the 5G standardization, and a total of seven different migration options (plus some additional sub-variants) were identified and largely standardized. Despite the large standardization effort to define the different migration paths, not all of them have been applied in the network migration towards 5G. By January 2024, the vast majority of deployed public 5G networks are so-called 5G non-standalone networks (5G NSA) [GSA24], which means that a 5G radio access network is used to provide radio connectivity to the 4G core network. 5G NSA has been intended as an intermediate step towards a 5G standalone (5G SA) deployment, where a 5G UE is connected via the 5G RAN to the 5G Core Network. As a consequence of the prevalent 5G NSA deployments, a large part of the functionality that has been standardized for the 5G Core Network, such as Network Slicing, a cloud-native design, and also including support for industrial IoT – Ethernet LAN, TSN, time synchronization – cannot be fully used in most deployed 5G networks, even if the amount of UEs that support 5G SA has been steadily increasing and is now supported in the majority of UEs [GSA24]. It is expected that it will still take several years before 5G SA can be considered common in public 5G networks. In hindsight, the range of (partly complex) migration options has led to a fragmentation of the market and a need for multiple stepwise network upgrades, and this has contributed to a slow uptake of 5G network adoption (with full 5G SA capabilities).

For the introduction and migration towards 6G, it is proposed to limit the migration options and target directly a 6G standalone deployment providing the full 6G capabilities [RÖT23] [CMRV+23] [Hex2-D21], as shown in Figure 2.10. As the most efficient migration of the core network, it is proposed to evolve the 5G core network to support the 6G RAN and 6G UEs. Already the 5G core network has been defined as a flexible and extensible network platform, and the service-based architecture is well suited for a cloud-based network deployment. It can be flexibly extended with functionality to support a 6G RAN and 6G UEs [RÖT23] [CMRV+23] [HEX2-D21] while building on the investments that are being made for deploying the 5G core network. In order to provide 6G access to (and aggregation of) existing spectrum carriers that are used by e.g. 5G, it is suggested to apply dynamic multi-radio spectrum sharing, which allows for efficient and flexible radio capacity sharing (and eventual capacity migration towards 6G), which builds on the dynamic spectrum sharing defined for 4G and 5G.

The timeline of 6G standardization is described in [LGP+24]. 3GPP standardization work towards 6G will start in 2024, where the first phase will investigate use cases for and requirements on 6G – in alignment with the framework and objectives specified by ITU for 6G (denoted by ITU as IMT-2030) [ITU23]. Based on this first phase, a technical study phase will take place in 3GPP during approximately 2025-2027 and will be followed by a work item phase in which standard specifications will be developed until approximately the end of 2028, including a self-evaluation to be sent to ITU. This will allow commercial 6G networks to be applied around 2030.

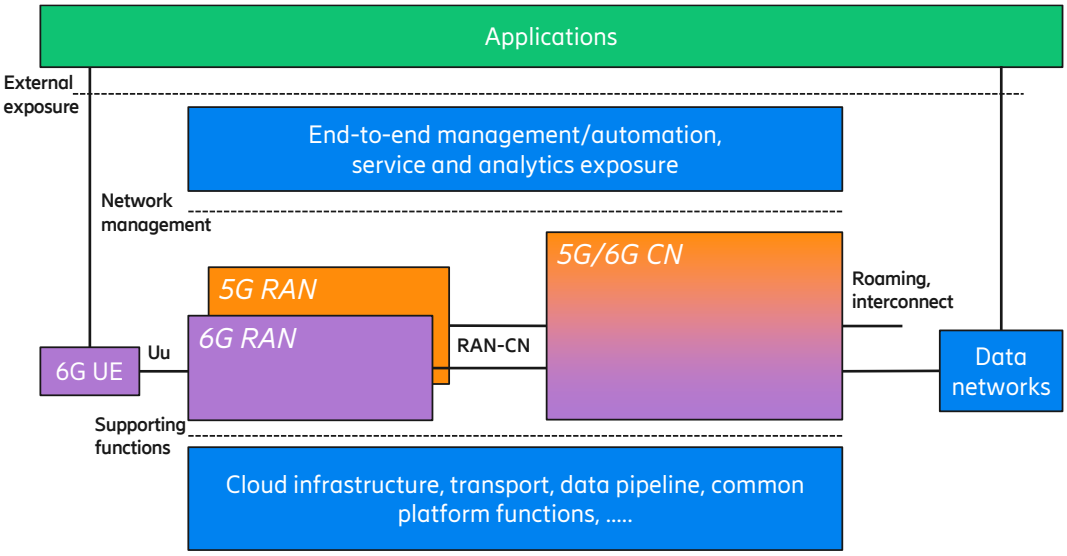

Figure 2.10. Proposal for 6G architecture from [CMRV+23]

2.2. Architecture principles investigated for 6G

A large number of technology components are currently investigated and researched as potential candidates for a future 6G standard. With the upcoming start of 6G standardization, there is a need to assess the potential benefits and level of maturity of those different technology components, to agree within the ecosystem on the technology foundation for 6G. Furthermore, agreed technology components need to be aligned within a consistent architecture view on the 6G system.

In the following we review several technology trends that find wider interest and support in the ecosystem, and which are relevant to the objectives of DETERMINISTIC6G [HEX2-D21] [HEX-D14]. A system blueprint developed in [HEX2-D21] is shown in Figure 2.11.

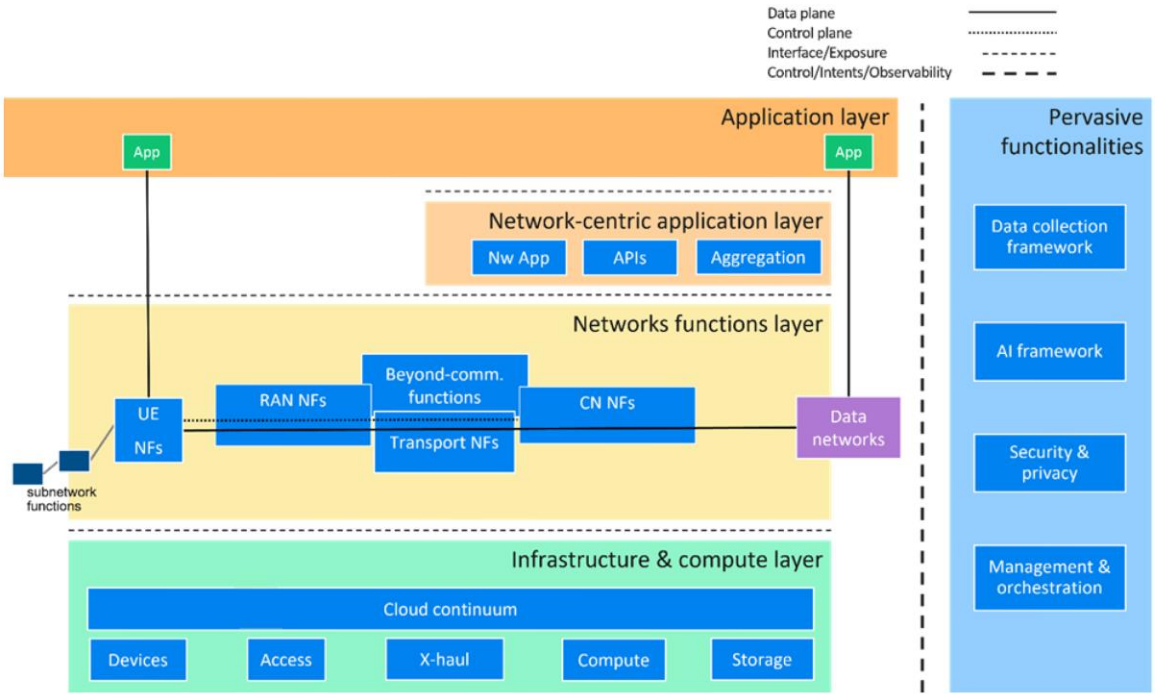

Figure 2.11. Initial 6G E2E system blueprint as proposed in [HEX2-D21]

[Eri23] presents a future (mobile) network architecture, which is depicted in Figure 2.12. It shows the network architecture in a horizontalized way for different functional domains. For some time, networks have evolved to become softwarized, which is based on the separation of hardware and software. Network functionality is realized as software functionality (network functions) and can be executed on a suitable (and ideally interchangeable) compute platform. Regarding compute platforms in particular, (hyper-scale) cloud compute platforms have attracted significant interest, as they provide flexible, scalable, and cost-efficient compute infrastructure. This means that at the bottom of the network architecture functional domain (see Figure 2.12), there are the implementation and infrastructure layers which comprise a compute infrastructure, and a transport network that interconnects the sites at which the compute infrastructure is located. Softwarization of the network already started in the 4G timeframe, and has been the foundation for 5G, and is sometimes referred to as a cloudified mobile network. It shall be noted that compute-intensive operations, like for radio access functionality, are still mostly based on dedicated and highly optimized compute platforms. With the increasing integration of hardware accelerators into the compute domain, also for the RAN, in future more cloudified solutions are envisaged. A cloud-native approach is considered one foundation for the 6G network [HEX2-D21] [Eri23]. With the increasing adoption of AI in the optimization and automation of network functions, there is a need for data collection and transport towards ML functions (as explained in more detail in section 2.2.1) [HEX2-D21] [Eri23]. As a consequence, data collection / data pipeline functionality [HEX2-D21] [Eri23] is also needed in the infrastructure layer. Based on these two lower layers of transport and compute, network functionality can be realized, providing wireless access, mobility, and including further network applications. In this softwarized network paradigm, the network design can be configured flexibly. This is, network functionality can be flexibly instantiated and distributed on a number of distributed sites with compute infrastructure, that are interconnected via a transport network. Virtual networks can be configured on this infrastructure, and several of them can coexist as network slices. The design of a virtual network shall

be optimized to provide the best value to the applications it is targeting to serve. Networks can provide optimized performance according to use case needs, in local network deployments, or over wide area. A management and orchestration layer is needed to configure, manage, and orchestrate the network operation in an automated fashion. To this end, intent-based management is investigated [Eri23] [HEX2-D21], where requirements, goals and constraints are formally specified to be realized via an autonomous cognitive intent-based management framework. A value-based delivery of network services to use cases and applications is provided by monetization functionality, which allows to define e.g., service level agreements (SLAs) or operational agreements between the application domain and the network domain. With an inherent distributed compute platform as basis for the network realization, the 6G network platform can also provide to the application domain compute-as-a-service [HEX2-D21]. This is in particular beneficial for offloading compute workloads of mobile devices, in order to provide advanced compute resources to the device applications and reduced device footprint, power consumption and complexity of the device. Compute offloading [RJS+23] [BBW+23] can bring significant benefits to the use cases investigated in DETERMINISTIC6G [DET23-D11], like the exoskeleton and the XR devices [AKP+21] [ADF+23]. Network and service exposure functionality enables the interaction of the application / user domain with the network via standardized APIs, to configure service requests, configure and use network service capabilities [Eri23] [HEX2-D21] which turns the network into a programmable service platform.

From a standardization perspective, it is important to identify the commercially relevant multi-vendor interfaces in the architecture that provide commercial value, see [Eri23] [CMRV+23]. With this focus, the 6G network architecture shall allow flexible service innovation for a large variety of use cases, in an interoperable and industry-aligned way that is commercially viable.

At large, the horizontalized network architecture of [Eri23] (Figure 2.12) resembles the network architecture of [HEX2-D21] (Figure 2.11), where the pervasive functionalities of [HEX2-D21] are integrated into the horizontal structure.

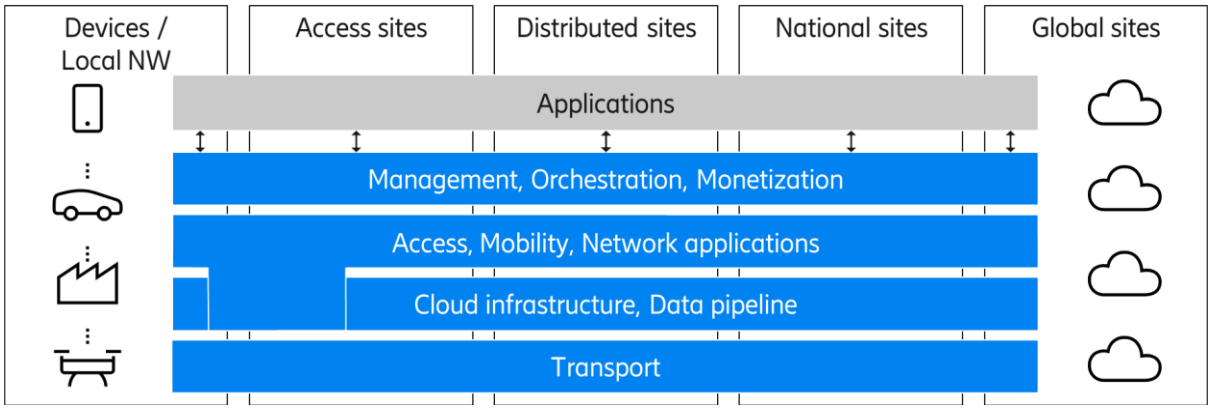

Figure 2.12. A horizontal 6G architecture [Eri23]

The 6G network platform is further foreseen to provide services beyond pure communication [HEX2-D21], which can include time synchronization, compute-as-a-service, positioning, sensing, and AI-as-a-service of which the former ones are also relevant within DETERMINSTIC6G.

2.2.1. Data-driven and AI-native system design

There is a general understanding that artificial intelligence (AI), and in particular machine learning (ML), is going to play an increasing role in future networks, but also in the application domains and

systems that use networks [HEX2-D21] [HEX-D14]. AI technology has matured and is considered beneficial to address problems that are highly complex, comprise inherent randomness and non-determinism [IJR+23]. While AI can be introduced by enhancing system components with AI, or introducing new AI-based components, a next level system design is to become AI native [HEX2-D21], which is in [IJR+23] defined as: "the concept of having intrinsic trustworthy AI capabilities, where AI is a natural part of the functionality, in terms of design, deployment, operation, and maintenance. An AI native implementation leverages a data-driven and knowledge-based ecosystem, where data/knowledge is consumed and produced to realize new AI-based functionality or augment and replace static, rule-based mechanisms with learning and adaptive AI when needed." This leads to the following four aspects [IJR+23]:

- Intelligence everywhere: AI workloads should be executable where it is beneficial from a cost-benefit perspective. As the number of AI models grows, automation is needed, including automated model lifecycle management.
- A distributed data infrastructure is needed that enables data collection and data transport in alignment with possible constraints related to data handling. This allows for data-driven intelligence leveraging AI capabilities. More detailed discussion on a data-driven network architecture is provided in [Roe20].
- It is increasingly complex for human operators to manage the network and data infrastructure, which calls for zero-touch management with autonomous networks based on intent-based design.
- AI-as-a-service: as the network architecture becomes AI native, it integrates AI-related capabilities, such as AI model lifecycle management or data handling. Such capabilities could be exposed to external (network) users as platform services provided by the network infrastructure. More details on a possible AI-aaS service capability can be found in [SAR+23] and also in [HEX2-D21].

Capabilities of AI-native and data-driven design can be described differently from an architecture perspective. In Figure 2.11 they are depicted as "Pervasive functionalities" that is available throughout the architecture; in Figure 2.10 and Figure 2.12 they are part of the *infrastructure layer* (together with compute and transport) that is the foundation of the architecture and available to the other architecture components.

2.2.2. Security-related prior work

In general, security-by-design requires embedding security in the design of systems via the adoption of both *software security assurance processes* and *trusted hardware*. Software assurance processes include, for instance, carrying out a comprehensive threat analysis, include the design of countermeasures against existing threats as part of the system architecture, adopting repeated reviews and audits, and executing rigorous security testing [CD13]. We review in this section the security-by-design architecture for beyond 5G and 6G networks. We focus especially on the security aspects concerning deterministic networks.

6G is projected to be secure-by-design by integrating security at the heart of the infrastructure, adopting end-to-end and defense-in-depth concepts, and with new security control assurance and privacy-protection mechanisms [Sol21].

Creating a security-by-design architecture for 6G, but also the Industrial Internet of Things (IIoT), massive IoT and Industry 4.0 networks which may extensively utilize wireless communications, involves considering several critical components. Security measures need to be embedded at every stage of the network's design and operation, rather than being added as an afterthought. Making sure that the security measures do not overly perturb the quality, or the performance of the networks is of primordial importance. This is especially true when industrial vertical applications require dependable communications, as it is the focus of the DETERMINISTIC6G project.

[GSS+21] presents a layered functional architecture (Figure 2.13) that integrates 5G, TSN and DetNet. DetNet provides end-to-end deterministic networking across TSNs industrial automation systems, 5G/6G mobile networks, and large-scale sensor networks. 5G/6G acts as a bridge between TSNs industrial automation systems. The main layers described by [GSS+21] with regard to Figure 2.13 are the *application layer*, the *management and orchestration layer* (overseeing network resources, orchestrating deployment and management of network functions), the *control layer*, the *data layer* (with the user-plane data packet handling). The figure also shows the *Security Layer*. This layer spans across the entire architecture and provides the functions and measures to protect data integrity, confidentiality, and availability. It encompasses network security monitoring, encryption, authentication, authorization, and other security functions to safeguard the network from malicious activities and ensure trustworthiness, safety, security, resilience, reliability, privacy and scalability of communications. From the deterministic perspective, it provides end-to-end visibility for the detection of anomalies in the packet deliveries and time synchronization. It also provides prevention, mitigation, and reaction capabilities to respond to detected security breaches. The figure indicates the wide scope of security by design; also in Figure 2.11 security is indicated as a *pervasive functionality* over the entire architecture scope. In DETERMINISTIC6G we mainly focus on the aspects related to dependable communication for time-critical services.

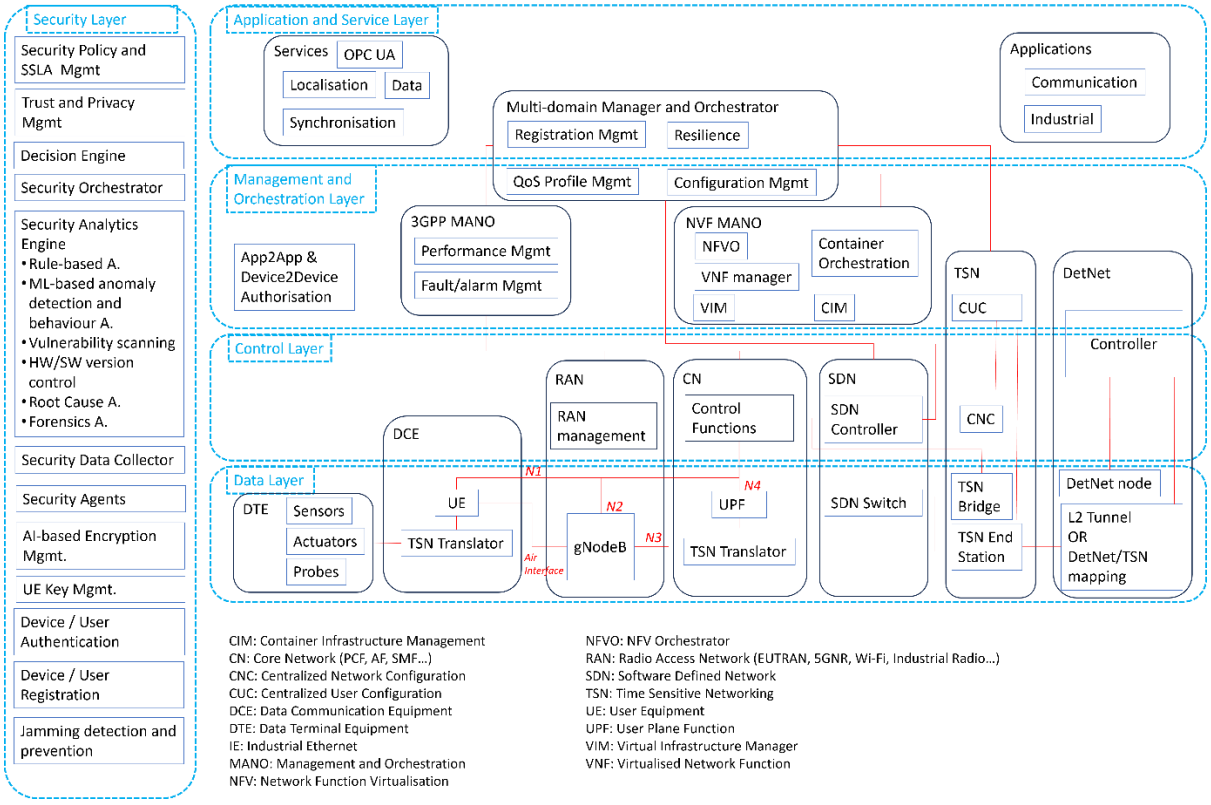

Figure 2.13. Integrated 5G/TSN/DetNet layered functional architecture according to [GSS+21]

It has to be noted that, for the sake of clarity not all functions and concepts that are relevant in the context of DETERMINISTIC6G are shown in Figure 2.13 as, for instance, MEC for optimization and privacy protection, IIoT networks, integration with legacy networks (e.g., industrial Ethernet), cyber resilience mechanisms, etc.

The main concepts and components to consider that have strong impact on determinism include end-to-end encryption, threat detection and response, and frequency spectrum management to counter jamming attacks. These are described in the following. Then, other important aspects are presented, including the need for: trade-offs between costs, risks, and security; security service level agreements for specifying requirements; detecting attacks on the Precision Time Protocol (PTP); security enablers that are needed to implement a zero-touch network and Service Management (ZSM) system; and other aspects.

End-to-End Encryption

End-to-end encryption concerns ensuring that data is encrypted from the point of origin to the point of destination of each communication. Encryption prevents eavesdropping and leakage of sensitive information. But encryption has a cost, so its protocols and parameters need to be carefully selected (e.g., by specifying and enforcing security policies) to get the best compromise between the risks that need to be prevented and the quality of the communications (e.g., latency, bandwidth, reliability). The time required for encoding and decoding can be accurately predicted, thus, its impact on latency can be accounted for when planning traffic to ensure that encryption will not disrupt the required deterministic properties. In 5G networks, encryption is used for user data that is transmitted over the radio interface, as well as on the transport interfaces.

IIoT networks can involve a massive number of interconnected devices. These devices range from high-powered machinery to low-resource sensors, and the criticality of the data being transmitted can vary greatly. Advanced Encryption Standard (AES) can be used for strong security of sensors and actuators with limited processing power and battery life. AES can operate in different modes such as Cipher Block Chaining (CBC) mode and Galois/Counter Mode (GCM). For resource-constrained devices, Elliptic Curve Cryptography (ECC) can also be used for secure key exchange and digital signatures. Transport Layer Security (TLS) and Datagram TLS (DTLS) can also be used to secure communications over IP networks. Pre-Shared Key (PSK) Ciphers allow configuring a shared secret key used for encryption and authentication, reducing the complexity, and overhead associated with certificate management. Lightweight cryptography is being standardized by National Institute of Standards and Technology (NIST) for securing devices with lower computational capabilities. It is also necessary to ensure firmware updates using techniques like code signing using private and public keys to verify firmware modifications.

Threat detection and response is based on continuous monitoring of the network exchanges at the different OSI layers. Several techniques can be used, such as rule-based detections, but also using AI/ML algorithms to detect and respond to unusual behavior and potential threats in real-time and adapting the security enforcement to new and evolving security threats.

Security Service Level Agreements (SSLAs)

Security Service Level Agreements (SSLAs) can play an important role in the management of network security, establishing clear specifications for the level of protection required and the performance expected from security controls. These agreements are tailored according to various security levels and domains, ensuring that different parts of the network receive the appropriate type of security measures based on their criticality and the data they handle.

An effective SSLA can clearly specify the security requirements for each domain (e.g., RAN, Core, IoT, M2M communications, edge, cloud, TSN, DetNet) in a multi-tenant environment. They allow detailing the functioning of security controls, their expected performance, and how these controls are monitored and validated. It outlines the scope of the security services, defines responsibilities, and sets out the procedures for responding to security incidents. Additionally, these agreements cover the detailed description of services provided, along with guarantees on service availability, integrity, and confidentiality. Fundamental to SSLAs is the understanding and negotiation of the trade-off between security risk, cost, and performance.

Security enablers for implementing a Zero-touch network and Service Management (ZSM) system

Zero-touch network and Service Management (ZSM) [ETSI-ZSM] aims to minimize human intervention by automating various processes such as risk analysis, security assessment, and responses to security events. This automation is made possible by sophisticated algorithms (e.g., AI/ML-based) that can rapidly analyze vast amounts of data to identify potential risks and anomalies. Once detected, the system can autonomously implement pre-defined security strategies to mitigate threats, ensuring an effective response.

Such a security architecture (see Figure 2.14) is composed of many different security functions or enablers that work together to offer zero-touch network security management both at the domain level and cross-domain. The enablers that are most relevant to deterministic networking are discussed in the following.

First and most important are the *security data collectors*. These are probes that can be passive or active, meaning that they can observe the traffic flows with or without impacting them. They can be beneficial for monitoring and detecting anomalies in both the data and control planes. They have been developed and are provided as open-source [MMT-Probe] and allow capturing data from many different protocols and performing an initial local analysis based on rules or ML-based algorithms. Agents are deployed to collect information from different observation points (virtual or physical), and to provide the information to a *Data Collector* for further and more global analysis and visualization. The data can also be stored to perform historical or forensics analysis. The aggregated data can subsequently undergo a more comprehensive and holistic analysis by the *Security Analytics Engine*, employing methodologies similar to those used by the *Security Agents*, but in a more exhaustive manner, and with a global network and system perspective.

Once anomalies have been detected, the decision engine needs to determine their cause and if they need to be acted on within a particular domain. The determined reaction, mitigation, or prevention action can be carried out by a *security orchestrator* that is an extension of the Network Function Virtualization (NFV) orchestrator (e.g., Open Network Automation Platform (ONAP), Open Source Management and Orchestration (OSM)). The following section describes how the ZSM process works. Other enablers, such as, the policy and trust managements are also pertinent but are less critical for assuring the determinism and low latency focus of the DETERMINISTIC6G project. Cyber Threat Intelligence (CTI) can also be useful to determine the reputation of endpoints and hosts at a wider network scale. This includes the use of adapted honeypots, related to the system to be protected, to attract malicious behavior that can then be prevented.

Precision monitoring is an integral to the ZSM concept, as it provides the detailed, real-time data necessary for the automated systems to make informed decisions. This involves collecting and analyzing network metrics to produce actionable insights that facilitate timely responses to any network anomaly or security breach. Furthermore, ZSM incorporates traffic steering to navigate data flows through secure paths automatically, enhancing the security posture and resilience of the network. Several techniques can be used for this, including network slicing, moving target defense, redundancy, etc.

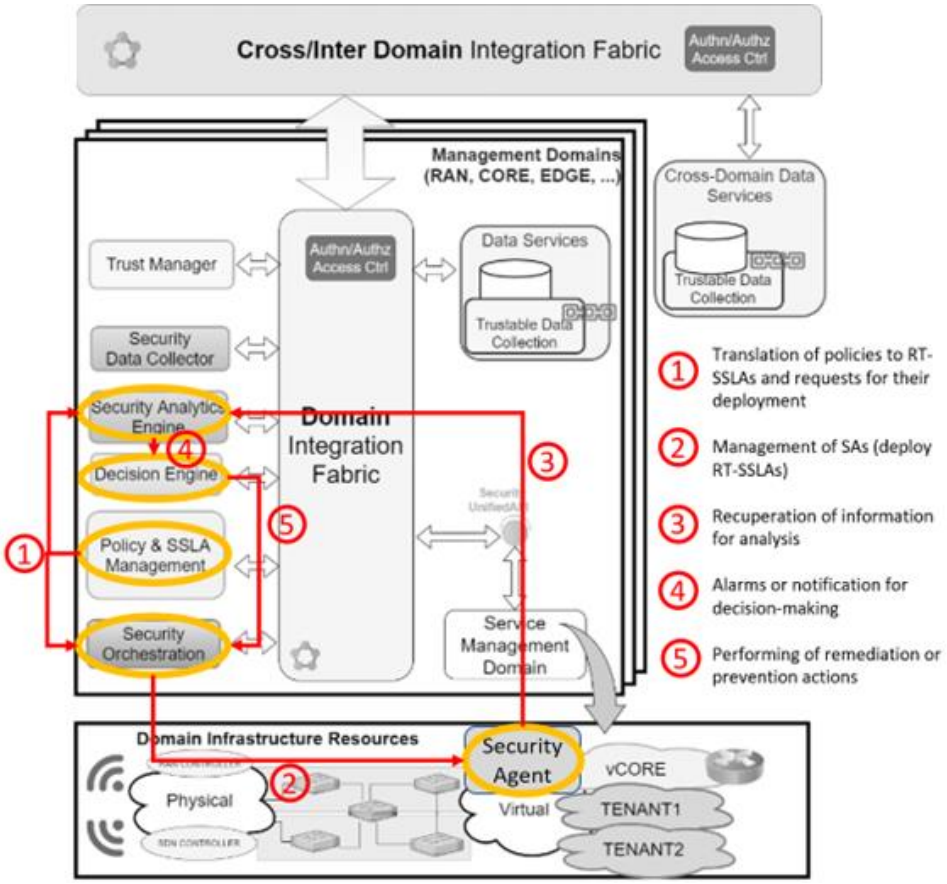

Figure 2.14. Closed loop zero-touch security management as described in [Per22]

As described in [Per22] and shown in Figure 2.14, several different steps are involved in zero-touch security management that illustrates the closed loop interactions between the main enablers to achieve the zero-touch security management. This architecture has been developed and demonstrated [Per22] and was selected as a Proof of Concept (PoC) related to ZSM called "Security SLA assurance in 5G network slices" [ZSM PoC]: Steps 1 and 2 involve converting the high-level security policies to real time security rules and algorithms (referred to as RT-SSLAs), and deploying them so that they are considered by the Security Analytics Engine (SAE) and the Security Orchestrator that in turn deploys them in the Security Agents that will extract the data required for verifying the SSLAs and make it available for analysis (Step 3) by the SAE. In Step 4, the SAE will communicate the results to the Decision Engine in the form of alarms or notifications. Finally Step 5 involves informing the Security Orchestrator of the actions that need to be taken to mitigate or respond to the detected security breaches. SSLAs are essential since they allow specifying the security requirements from the applications' perspective, enable E2E security across domains, and most important find the right trade-off between the actual risks and the costs of detection and response.

Techniques that can be used to prevent, react, or mitigate anomalies or attacks include:

- Trusted execution environments that ensure that devices boot using only trusted software and hardware components, preventing unauthorized firmware and hardware from compromising the network.

- Network slicing that allows creating virtually isolated networks within the same physical infrastructure, so that security policies can be customized and prevent propagation of threats across the network. Network segmentation (e.g., network slicing) and Virtual Private Networks (VPNs) don't provide encryption by default but can be combined with encryption to enhance security. These techniques limit the exposure of the devices and communications to potential threats by isolating them within the network.
- Software-Defined Networking (SDN) and NFV that enhance network flexibility and scalability while allowing for centralized security policy management and rapid deployment of security functions.
- Resilience mechanisms to ensure the ability to withstand attacks and quickly recover from security incidents, minimizing downtime and data loss.

The ZSM mechanisms are pertinent for deterministic networking, but they do not specifically address the deterministic and low latency requirements. For this it is necessary that the Security agents act as active probes that will swiftly react to mitigate any security breaches related to attacks on the latency, performance, determinism, and privacy. In-band Network Telemetry (INT) is an emerging technique that can be suitable for implementing the fast reactions needed.

Frequency spectrum management to prevent jamming attacks

With the adoption of 5G/6G wireless communications in industrial and safety-critical applications, preventing or avoiding the impact of jamming attacks is essential. The use of frequencies higher than 30 GHz make jamming less likely since it would require high levels of power [LRM+18]. Nevertheless, jamming is still possible (and easier in lower frequencies) and several techniques exist to ensure resilience, adaptability, and security of radio communications, some being described in [AF20] [SKT22] [KLY+13]. For protecting communications, other techniques are proposed such as strong encryption of the transmitted data [NLC+21], and wiretap channels [MCP+23] that introduce noise to make it harder for attackers to interpret the signals.

Other aspects to consider

Other aspects that need to be considered that will be important in 6G are MEC, Zero-Trust Networking (ZTN), data privacy, and secure transactions. In DETERMINISTIC6G we will consider them if needed, but at least any related requirements should be defined, and the different solutions investigated.

With respect to ZTN, the goal is to make sure that the services continue to function even under attack. With ZTN, no entity, whether inside or outside the network perimeter, is trusted by default. This implies that everything needs to be verified, and that access is granted based on strict identity verification and need-to-know basis. Since absolute security is not possible (i.e., too expensive) the networks need to consider that some or all elements are untrustworthy. E2E encryption, monitoring, rigorous and continuous verification for every request, integrating strict access control, multifaceted authentication, and precise user and device identification are the main tools for this, but in the deterministic and low latency context they need to be correctly managed and adapted. Furthermore, isolation strategies that effectively segment network resources, minimizing the risk of lateral movement by potential attackers, can enhance network security.

Management of MEC is pivotal for industries that require real-time processing capabilities both within and beyond their operational perimeters. By deploying edge nodes strategically within these boundaries, MEC management ensures that data processing occurs closer to where it is generated, significantly reducing latency. Ciphering data at its source before transmission is a critical security

measure, as it safeguards against interception and unauthorized access. Consequently, only encrypted data traverses to external MEC nodes, ensuring that confidentiality is preserved during transmission across various network segments. This approach not only bolsters security but also optimizes network performance by leveraging the proximity of edge computing resources.

In the field of machine learning (ML), adversarial attacks present a significant challenge, where attackers craft inputs specifically designed to fool ML models into making incorrect predictions or classifications. These attacks exploit weaknesses in the models that may not be apparent. As a result, the robustness of ML models against manipulations has become an important issue. Furthermore, the transparency and explainability in ML has become an important challenge. It's not enough for a model to be accurate. The decision-making process needs to be made understandable by humans to obtain user trust, but also for diagnosing and defending against adversarial attacks.

Concerning Data Privacy, it is necessary to ensure that data is handled, stored, and processed in compliance with privacy laws and regulations. This can involve the use of different techniques, such as data minimization, data anonymization, homomorphic encryption, federated learning (FL), and consent management.

Finally, Secure Transactions can be guaranteed using distributed ledger technologies to secure data exchanges across the network, ensuring integrity and non-repudiation, and providing, for instance, the means to establish smart contracts, immutable records for supply chain management, managing digital identities, record keeping, accountability and liability management, device management, etc.

From the broad scope of design and functionality needed for a security-by-design as described above, we focus in DETERMINISTIC6G on INT-based data extraction and security analytics.

3. Review of proposed DETERMINISTIC6G functionality

Within the work of DETERMINISTIC6G, several technology concepts are being developed and proposed as technology components for 6G, and would thus need to be integrated into a 6G architecture. In this chapter we review these concepts.

3.1. Time synchronization

A time synchronization service is available in 5G since Release 16, and other improvements have been added in Releases 17 and 18 [GLR+20] [MAG+19] [PDR+21] [3GPP18-23501]. However, time synchronization reliability remains an important issue. Indeed, there is a need to support resilient time synchronization mechanisms in time critical 6G-TSN applications to meet the performance requirements. The current TSN time synchronization (generic Precision Time Protocol (gPTP), IEEE 802.1AS) relies on the Best timeTransmitter Clock Algorithm (BTCA) to find the next grandmaster (GM) clock, when the current GM clock has failed or degraded in performance. However, BTCA may take time to find the next GM. Additionally, BTCA is unable to detect transient faults in a GM, hence could lead to a ping-pong effect between two potential GMs. For the emerging 6G-TSN use cases, this waiting time without a synchronization source is not desirable. With the objective to provide a continuous timing, a new amendment of the IEEE 802.1AS standard is ongoing (802.1ASdm) which modifies the hot standby mechanism. In the new setting, the hot standby mechanism is modified, where BTCA is not used to select the next GM, and instead a static configuration is sent down from the network management system (e.g., the CNC in the case of TSN). Here a hot standby GM is always transmitting timing messages along with a primary GM. In order to keep the two GMs in sync, the hot

standby GM synchronizes itself to the primary GM before it starts sending timing messages. Hence each end station has access to two time domains at any given time. If there is a failure or performance degradation in the primary GM, the hot standby GM takes over immediately.

From an architecture point of view the deliverable D2.2 [DET23-D22] analyses the different implications on the location options of the GM, the redundancy design, and the 3GPP support for such a scenario. Indeed, the location of the primary and hot standby GM will determine the coverage of the synchronization redundancy. Hence, a careful design to optimize the location of such GM clocks is required. More importantly is the aspect of 3GPP support of the 802.1ASdm amendment, and consequences based on whether the 6G GM becomes the primary GM or the hot standby GM.

3GPP support for 802.1ASdm has not been specified since this standard amendment is still under development. There are different possibilities for how the 6G GM behaves in a hot standby enabled 6G-TSN network. In the case 6G GM is neither the primary nor the hot standby GM, then it is enough that the 6G system maintains the primary and secondary time domains as independent time domains. This is already possible with available 3GPP 5G support for multiple time domains. If the 6G GM becomes the hot standby GM, this would not be acceptable from a 3GPP viewpoint since the 6G GM would need to synchronize with the external primary GM. Such an option is not feasible because the 6G GM must be guaranteed for the operation of the base stations and the general 6G network availability, and it is also not supported in the current 3GPP standards. If the 6G GM becomes the primary GM for the external system (e.g., TSN network), this would be possible as the 3GPP standards already support this case where 5G GM is the external GM. More work is required for the upcoming 6G to investigate the support for the hot standby amendment and the above description forms the ground for such study.

3.2. Packet delay correction

As described in the deliverable D2.1 [DET23-D21], it is possible for 5G to provide low bounded latencies through the URLLC features, where a fewer number of HARQ retransmissions are possible. In this case, packet transmissions which perceive the worst transmission conditions are improved by boosting the reliability of their transmission. Such transmission conditions are not known beforehand, so all transmissions must be generally protected to a higher level. This implies a high resource cost. On the other hand, a bounded latency is not sufficient to provide time-critical dependable communications. The delay should also be stable, that is the packet delay variation (PDV) experienced should be very low[1]. It is difficult to guarantee such low packet delay variation only in the actual radio transmission, which is inherently subject to stochastic variations. An alternative approach is to compensate for the incurred PDV at the edges of the 6G systems. We propose a packet delay correction (PDC) mechanism for such purpose.

PDC is a mechanism by which every packet is forced to remain in the 6G system approximately for the same amount of time. For example, if all the packets spend the same delay within the 6G system before it is forwarded to the next node, then the packet delay variation in 6G system is zero. This mechanism is intended to be applied at the egress of the 6G system, e.g., at the DS-TT in downlink or at the NW-TT in uplink. The egress entity or the TT will hold the packet for some time, until it reaches

[1]An alternative approach would be if the application would be robust to latency variations, e.g. if the application messages would contain timestamps a receiver could compensate for packet delay variations and apply a time-aware application logic. This could however not compensate for the sensitivity to packet delay variation of an intermediate network layer like TSN.

the desired time, e.g, maxDelay, and then it is forwarded to the next node in the external network, e.g., the TSN network. Note that to decide how long the packet should wait, additional information is required which is transported within every packet as metadata. For example, if a timestamp-based PDC is applied, then a timestamp metadata is carried in every packet which is added at the ingress of the 6G system. Then the egress TT can calculate the time that a packet already spent within the 6G system and then the TT can calculate the time it needs to wait to reach maxDelay. This is illustrated in Figure 3.1.

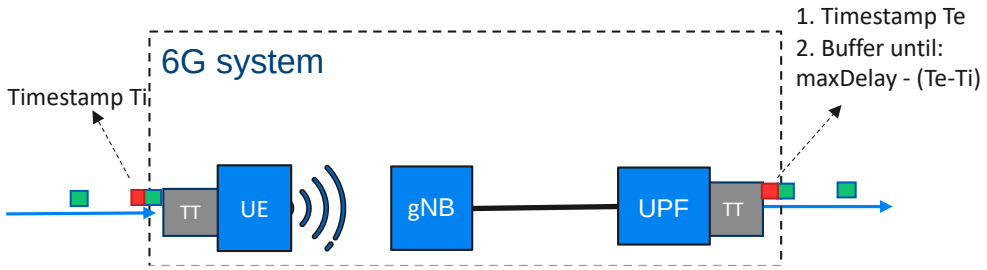

Figure 3.1. Timestamp-based PDC

A broader description of PDC can be found in D2.1[DET23-D21].

PDC basically compensates (or corrects) the packet delay variation, which is basically introduced by stochastic variations in different parts of the 6G system. Since the main component of PDV intrinsically comes from the radio transmission, PDC could also be applied in the RAN segment of the 6G network (i.e. between gNB and UE in Figure 3.1).

3.3. Edge computing for time-critical applications

As it was mentioned in Section 2.1.4, edge computing enables us to realize time-critical use cases. The typical use cases are discussed in DETERMISISTIC6G, including extended reality (XR), occupational exoskeleton with function offloading to the edge and various industry automation use cases, where the controller application is deployed on an edge computing host. However, if we intend to utilize the TSN functionalities for the application hosted by the edge computing, the integration of the edge computing and TSN domains is required, considering the architecture and conceptual (real-time compute, time-aware traffic handling in the virtualized domain) aspects.

3.3.1. Integration of 3GPP edge computing support and TSN support architectures

If we consider a use case scenario, in which the TSN Talker/Listener (e.g., cloudified application instance) is moved to the edge computing domain, and 3GPP-enabled TSN support is utilized for communication, then we encounter the challenge of integrating the 3GPP-defined TSN support and edge computing support architectures. Two basic deployment options are identified. In one of the options the *edge computing service is enabled by the 5G system* and connected to the 5G/6G virtual TSN bridge that includes all TSN network functions defined by the 3GPP as shown in Figure 3.2. In the other option, the edge computing service is deployed by using a standalone, dedicated infrastructure to host the TSN Talkers/Listeners as shown in Figure 3.3.

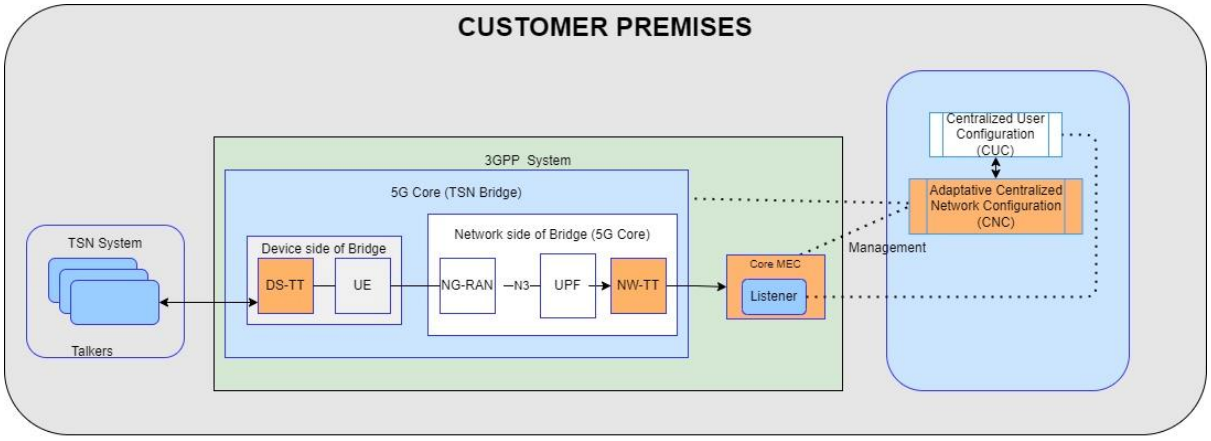

Figure 3.2. Edge computing platform is integrated into 6G system.

This solution shows a scenario, where the edge computing is tightly integrated with the 3GPP system by leveraging the distributed cloud infrastructure resource principle. One major characteristics of this approach is its ability to utilize the 3GPP SA6 specified edge computing support features (details are outlined in TS23.558). Depending on the deployment case, various levels of integration are possible. In the case of an Enterprise deployment, for instance, where the 3GPP domain may act as a Standalone Non-public Network (SNPN), a shared computing infrastructure could host the 3GPP and edge services. Certainly, even in this case, dedicated resources can still be reserved for the time-critical applications within the Enterprise domain. Another option is when the edge computing is enabled by the Mobile Network Operator (MNO). The geographical density of RAN sites enables the deployment of edge resources within approximately 10km range from the end device, thereby supporting time-critical services. Using MNO premises, 3rd party edge service providers can offer full commercial Infrastructure as a Service (IaaS) or Platform as a Services (PaaS) for the customers. Alternatively, the MNO may act as edge computing service provider, offering its own platform services directly to the customers.

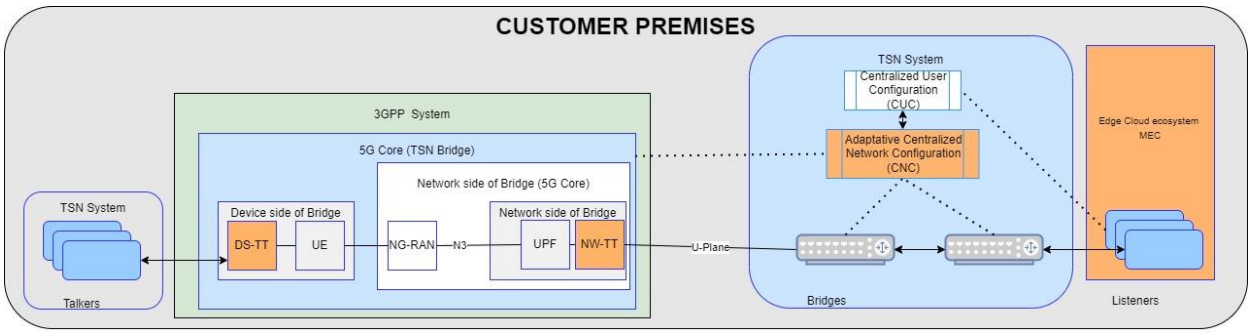

Figure 3.3. Standalone edge computing deployment

In this option the edge computing deployment is isolated from the mobile infrastructure, and a single, standalone datacenter infrastructure is deployed to host the cloudified applications (aka TSN Talkers/Listeners). This type of edge deployment enables lots of flexibility, such as distributed edge infrastructure for ensuring high reliability, traffic handling, and networking solutions in the virtualized domain tailored to tight interworking with TSN domain. However, the edge owner is responsible for

managing the full edge computing stack, including both the hardware and software components, which requires solid knowledge in the cloud execution and management ecosystem. Since the edge infrastructure is separated from the communication infrastructure in this option, the TSN system could be used to handle both the wired and wireless end devices. Leveraging the 3GPP specified TSN support, the wireless end devices are handled by the 3GPP virtual TSN bridges, while the wired end devices can be handled by the legacy (wired) TSN segment. Since this solution ensures high-level of customization, it fits very well for realizing such use cases, with precise timing requirements and where TSN communication is crucial.

3.3.2. Enablers for hosting virtualized, time-critical applications in the edge domain

Although – due to the geographical proximity – edge computing provides the potential to reduce latency, it alone cannot guarantee a solution for dependable, time-critical use cases. On one hand, the shared resource paradigm of cloud computing applies to edge-enabled application deployment as well, resulting in uncertainties in resource scheduling and execution times for applications, which may lead to unacceptable jitter. On the other hand, these uncertainties in the compute domain do not allow the utilization of various TSN features, such as 802.1Qbv scheduled traffic, and 802.1CB (Frame Replication and Elimination for Reliability - FRER) for cloudified applications.

Hence two main challenges for enabling TSN/DetNet usage for cloudified applications can be identified. The first is related to how time-bounded compute execution is guaranteed for the time-critical applications. The second pertains to the timely and coordinated delivery of messages generated by the various applications, in accordance with the configured 802.1Qbv schedule on the physical NIC.

To address the above challenges in the DETERMINISTIC6G design, our approach is twofold. On one hand, the goal is to mitigate the harmful uncertainties of the compute components, by leveraging the resource allocation, task scheduling, isolation, etc. toolset of cloud computing in a coordinated way. On the other hand, basic design principles for the seamless integration of the TSN and edge computing domain are defined to ensure the TSN-aware coordination and handling of the traffic generated by different application instances deployed on the same host.

A cloud system is ideal to run applications designed for a virtualized, distributed environment. For applications not designed for a cloud environment the goal is to provide similar behavior to native execution. Virtualization techniques typically focus on providing spatial isolation meaning that resources of an application cannot be accessed by other applications, while trying to reduce virtualization overhead as much as possible. Time critical application need temporal isolation meaning that timing related constraints of an application does not depend on or interact with other applications. For proper allocation of computational resources and quick reaction to external or internal events, low latency and appropriate scheduling[2] is needed. To provide low latency and appropriate resource allocation, a Real-Time Operating System (RTOS), like Real-time Linux may be needed. A CPU scheduler based on the Earliest Deadline First (EDF) and Constant Bandwidth Server (CBS) algorithms, called SCHED_DEADLINE is suitable for real-time applications. This way it can be guaranteed that a process is executed by the required deadline, however executing multiple

[2] The term scheduling is used here for the low-level scheduling, i.e., CPU timeslot allocation to processes.

applications on the same host and in addition to the deadlines requiring additional synchronization between them is still a challenging task.

In Kubernetes when a workload (Pod) is specified, resource (like CPU) requests and limits can be specified, to be taken into account at orchestration[3] and scheduling. The requested amount will be always provided, and when there are additional free resources, it may consume up to the limit specified. It's important to note, that having allocated a given CPU to a Pod does not give guarantees when it will be scheduled, therefore EDF application scheduling must be applied together with workload resource specification. Workloads with real time requirements can benefit from obtaining dedicated resources, like CPU, or even whole bare-metal computers. However, assigning dedicated compute resources goes against one of the main benefits of the cloud approach: the gain from multiplexing, so this option should be considered as an extreme.

Furthermore, orchestration solutions should take care of placing components of a complex real-time application close to each other, like the same machine or close-proximity blades. Requirements can be described with affinity or anti-affinity rules, or with explicit (time) distance rules.

Even in the case of a perfect scheduler or dedicated un-shared resources, there are challenges. The cloud environment may re-orchestrate time-sensitive applications or components resulting in severe or at least transient failures to fulfill time related requirements. Although optimization related re-orchestrations can be disabled, failure handling related re-orchestration cannot be avoided.

Consequently, the above-mentioned real-time cloud support features alone are not enough to guarantee the full utilization of the available TSN communication features. This could only be ensured by the tight integration of communication network (e.g., TSN) and the host networking of an edge computing server node, covering both the user and control planes.

3.3.3. 802.1Qbv-aware handling in the virtualized networking

Regarding the support of 802.1Qbv for cloudified applications the main challenge is the virtualized execution environment, where the application instances (aka TSN end points) are deployed in the edge ecosystem. In this architecture, the deployment and management details of the application instances (including resource reservation, orchestration, etc.) are hidden from the TSN control plane. This means that although the host NIC can be configured with 802.1Qbv traffic schedule plan by the CNC, it cannot be guaranteed that the different applications will generate the frames according to the 802.1Qbv configuration. If the frames arrive at the physical queue of the NIC in an improper order, there is no way to correct this, and the incorrect order of frames may cause unpredictable traffic disturbances in the other TSN bridges. To address this issue, an 802.1Qbv-aware traffic handling scheme is proposed to be applied in the virtualized networking of the host. This scheme ensures that frames originated from different application instances arrive at the NIC according to the configured traffic schedule, regardless of the scheduling order of the applications.

The essence of the method is to define a time-gating mechanisms, which is able to mimic a traffic scheduling, which is compatible with the configured 802.1Qbv traffic schedule on the host's NIC interface. Considering a containerized application deployment, the following two options for the time-gating mechanism are proposed:

[3] Orchestration is the process of allocating the workloads to hosts (referred to as "scheduling" in Kubernetes, but we try to avoid using the overloaded term for this).

1. Define a coordinated time-gating mechanism for the container interfaces: This solution leverages the TAPRIO (Time-Aware Priority Shaper) queuing discipline[4], which can be configured for the virtual Ethernet interfaces of the containers. The TAPRIO qdisc is configured on the different container interfaces in a coordinated manner, ensuring that at any given time, only a one containerized application is allowed to send traffic towards the physical NIC. The order of the application instances and the timing configuration of the TAPRIO qdisc are derived from the 802.1Qbv schedule on the NIC. Since the TAPRIO qdisc utilizes shared CPU resources as well, it may introduce timing uncertainties. To account for these uncertainties and ensure proper, timely packet forwarding between the container's virtual Ethernet interfaces and the NIC, a guard time must be configured on the container interfaces between the opening times for the traffic of the consecutive containers. A crucial consideration of this option is that the guard times must take into account, when the 802.1Qbv schedule is planned for the NIC by the TSN control plane.
2. The other option is to deploy the proposed time-gating mechanism using an Open vSwitch (OVS) in the Kubernetes container network interface (CNI) plugin. In this case, the time-gating mechanism is realized on the egress interface of the OVS, which is connected to the physical NIC of the host. In order to ensure the proper timing and ordering of the frames a hierarchical scheduling concept is proposed, as it can be seen in Figure 3.4.

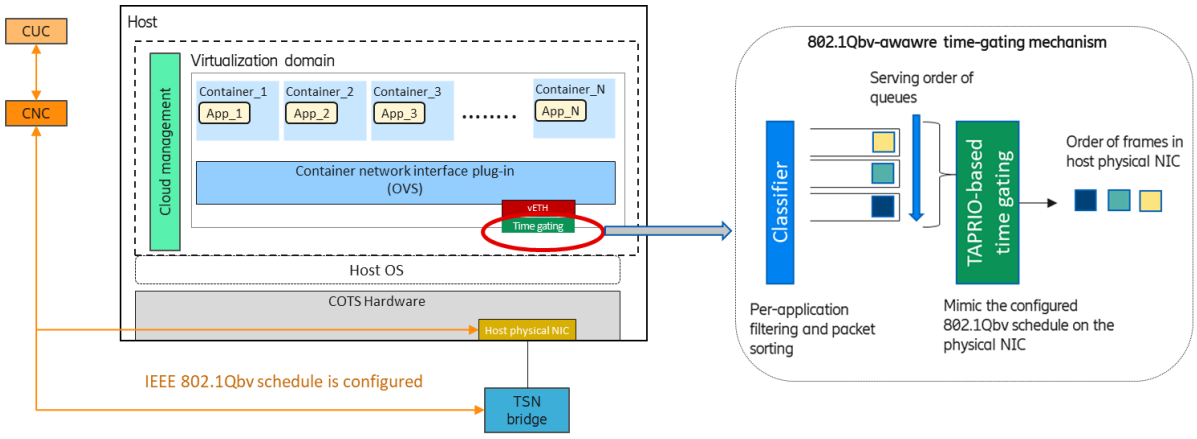

Figure 3.4. 802.1Qbv-aware time-gating mechanism

A priority queuing or Cyclic-Queuing and Forwarding (CQF) is used to ensure the proper ordering of the frames, regardless of their incoming order. Using the classifier component, the first frame to be forwarded according to the 802.1Qbv schedule is handled by the first queue to be served. The TAPRIO qdisc is used to timely forward the frames to the NIC's physical queue, where the traffic handling can be performed according to the configured 802.1Qbv schedule by the CNC. One advantage of this deployment option is the centralized handling of the traffic belonging to a specific NIC. Therefore, the guard times mentioned in the previous case are not required, resulting in better utilization of the network resources.

In both above cases, the cornerstone is the timely generation of the packets by the applications, therefore the enablers (e.g., deadline scheduling) described in section 3.3.3 has to be applied for these time-critical application in order to ensure a time bound for the execution of the applications.

[4] https://manpages.ubuntu.com/manpages/focal/en/man8/tc-taprio.8.html

3.4. Architectural aspects of data-driven latency predictions

We propose latency prediction as a mechanism for providing dependable communication in 6G networks. By utilizing such predictions, we can characterize the variations in packet delay that occur across the network. The obtained latency characterization can further be used, e.g., to optimize end-to-end TSN schedules. Analogous to the functional framework for integrating AI/ML in RAN, latency predictors in the 6G architecture have three key components [3GPP17-37817]: (i) data collection, (ii) model training and (iii) model inference. We will elaborate on each of these aspects in detail next.

3.4.1. Data Collection

Data collection acts as the initial step for collecting training data necessary for building data-driven latency predictors. Its primary responsibility is gathering relevant measurement data from different network entities (e.g., UE, gNB, UPF) corresponding to the traffic flowing in both uplink (UL) and downlink (DL) directions. Comprehensive data collection for latency prediction may entail deploying multiple monitoring probes spanning the entire 6G network in the RAN, in transport network and the core network. These monitoring probes are tasked with timestamping packets along their path from the sender to the receiver. The monitoring can be realized either in passive or active mode. In the passive mode, these probes can initiate such monitoring by performing measurements based on the already existing data plane traffic (QoS) flows. However, a drawback of passive monitoring is that the measurements can only be conducted on those specified QoS flows. Active monitoring tackles this issue by conducting measurements on traffic specifically injected for monitoring purposes, thus resulting in higher flexibility. Nevertheless, the active mode incurs additional overhead on the network, which can be intrusive. It also does not guarantee that the dedicated monitoring traffic obtains the same packet treatment (and QoS handling) as the real traffic and may thus not be representative.

Merely timestamping the arrival and departure of packets at a monitoring probe is insufficient for advanced data-driven latency prediction schemes. It is important for such schemes to measure both network state (e.g., Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ), # of retransmission) as well as traffic profile (e.g., packet arrival pattern, number of flows) in addition to the timestamps. It is important to note the potential need for the collection of metadata from various points within the 5G/6G system, not just at the endpoints. For example, for downlink latency queue states may be monitored at the gNB, while packet arrival timestamps at the UE are required. This highlights the necessity for careful consideration regarding where and how metadata is collected. This metadata is required to construct a dataset that can be used for building supervised learning models for latency predictions. A crucial assumption in data collection at multiple monitoring probes, in particular for latency measurements, is that relevant entities within the 6G system (e.g., UE, gNB and UPF) are sufficiently synchronized and the timestamping capabilities in the monitoring probes have sufficient precision. Accurate timestamps corresponding to this metadata are important to correlate conditions and events in the network that result in the measured latencies. End-to-end time synchronization mechanisms for future 5G-Adv/6G architectures have been described in [DET23-D22] [3GPP18-23501].

QoS monitoring mechanisms have been defined by 3GPP in Release 16 that allow for delay measurements on QoS flows between the UE and the UPF [3GPP18-23501] including the delay measurements in RAN. The definitions for packet delay measurements in the RAN part have been

specified by 3GPP in [3GPP18-28552][3GPP17-38314]. However, it is worth noting that these measurements mechanism might not suffice for data collection for data-driven latency predictors. On the one hand, these measurements can be imprecise, e.g., measurements like average RAN delays measurements or measurement distributions using bins of certain sizes (e.g., 0.1 ms) may not provide enough granularity to accurately capture tail distribution of latency. This might not be adequate for the cases when reliability levels of 99.999% or more are required at a given delay target. On the other hand, if a more fine-grained breakdown of end-to-end delay in the network and its components is desired a wider set of QoS monitoring capabilities (along with the collection of corresponding metadata) may be desirable than what is defined within the proposed 3GPP mechanisms.

The data collection is tasked with configuring the monitoring probes for timestamping packets. The configuration could, for example, include the parameters related to sampling (i.e., identifiers for QoS flows whose packets need to be timestamped, time interval between each measurement, precision of each timestamp, etc.) and the type of measurements to be performed (i.e., which layers should timestamp data units, metadata to be collected) at a given network entity. For instance, monitoring probes in the gNB could be configured to measure delay contributions finer grained, see e.g. [MTS+23]. Generally, a framework for data collection is desirable as an integral part of the network architecture. On the selection of what probes or data elements are to be selected, a careful trade-off needs to be made between the complexity and overhead of data collection, versus the benefit that each data type can provide. DETERMINISTIC6G will explore this tradeoff for packet latency and reliability prediction.

The actual transport of measurement data from monitoring probes to the data collection function (or NWDAF) is accomplished using two broad approaches: (i) out-of-band telemetry and (ii) in-band telemetry (INT). The conventional out-of-band telemetry approach is based on having separate dedicated flows/channels for data transport which are distinct from the QoS flows for regular data plane traffic. This approach allows collecting monitoring data avoiding potential interference on QoS flows and impact on network performance. Furthermore, tracking timestamps and metadata for a packet along the network path (e.g., at UE, gNB and UPF) could be challenging with this approach. A potential solution could be to use unique identifiers for each data unit to track them through different layers of the protocol stack. For example, a data unit in the RLC layer can be associated to the data unit in the PDCP layer using the unique PDCP sequence number.

In contrast to the out-of-band telemetry, there exists concept of In-band Telemetry (INT) for collecting data for latency predictors. This involves embedding packet timestamps within the payload of QoS flows rather than transporting them through separate channels. The INT approach allows for data collection without requiring additional infrastructure or resources. However, considering the number of timestamps per packet could become quite large depending on the number of measurement points, care should be taken to optimize the overhead due to INT fields in each packet and select an appropriate Maximum Transmission Unit (MTU) size that can accommodate all the data.

When the data is collected, optional data filtering can be used to optimize data volume and retain only useful information from the data. For instance, if the newly collected latency data and the corresponding network conditions exhibit no significant statistical differences with respect to the historical data, then it can be filtered as it does improve the model generalizability. Furthermore, preprocessing steps like normalization and standardization of features are typically beneficial to scale

their values appropriately such that performance of the predictors is not degraded due to varying value ranges of the features.

In general, the selection of the ML algorithm will have an impact on the nature of latency predictions that can be obtained. For example, ML models used for forecasting point estimates like average latency may not provide any insight into delays at higher quantiles but do not require huge datasets for training. Under specific conditions, even if an average delay of 5 ms is predicted, it cannot be guaranteed that the network will deliver a latency of 10 ms with 99.999% reliability. However, Mixture Density Networks approach, particularly, based on Gaussian mixture Models combined with Generalized Pareto Distribution can be leveraged for accurately characterizing the tail of latency distribution as described in [DET23-D21]. Furthermore, architectures like Recurrent Neural Networks could be used to capture the temporal dependencies in latency predictions which in turn require large amount of training data to learn temporal patterns. In summary, the goal of latency prediction dictates the choice of machine learning models and the scope of data collection needed.

3.4.2. Model Training

The model training function is responsible for coordinating the entire learning process of ML models that includes training, validation and testing phases. The datasets gathered by the data collection function will typically consist of a substantial number of samples, ranging from tens of thousands of samples to millions of samples collected over different durations. These datasets facilitate a rich representation of network and traffic conditions and corresponding latencies under representative set of network and traffic conditions. Training ML models with such a huge data set requires running complex training algorithms for long durations which in turn requires significant computational resources (e.g., GPUs or specialized hardware), which can, for example, be available at an edge cloud. In contrast to this centralized learning approach, the learning process of an ML model can also be approached in a decentralized way. On the other hand, Federated Learning (FL) presents a decentralized approach where ML models are trained locally at each client (e.g., a base station), rather than communicating data to a centralized server for training with the entire dataset. In FL, clients periodically transmit only model weights (or its updates) to a centralized location where these models are amalgamated into one model which is redistributed back to all clients [3GPP18-22874] [3GPP19-22876]. In FL, model training occurs on distributed data sources avoiding the need to centralize data from all sources, thus addressing privacy concerns and minimizing data transmission overheads. The selection of the learning scheme to train the ML models has to be determined by analyzing the two schemes in terms of their tradeoffs.

The trained latency prediction ML models should be deployed, for example to each base station, to allow for optimizing radio resource allocation to meet application latency requirements. Furthermore, latency predictions can be exposed to the domains (e.g., TSN CNC) external to the 5G/6G system using, e.g., TSN Application Function (AF), in order to make the mobile network characteristics plannable and allow to compute efficient end-to-end schedules. Post training, the ML models are validated and tested prior to deployment. However, the performance of these deployed models can still deviate from expectation to the differences in the distribution of the data used for training and the current network data distribution, necessitating effective management of these performance discrepancies.

The approach of offline training involves performing the training, validation, and testing procedures using the datasets collected in the past. As mentioned, the changes in the network conditions (e.g., mobility) or traffic (e.g., new interference source) can result in a drift in the performance of latency

predictors. Hence, additional fine tuning of the trained models post deployment based on local data is required to adapt the latency predictors and learn from new data over time without discarding previous knowledge. To this end, the models can be either trained periodically or model performance (e.g., in terms of accuracy, precision, recall or F1) can be continuously monitored and retraining is only initiated if the performance metrics fall below the set threshold. In contrast to these two approaches, a predictive approach can be used to predict when to trigger training using an unsupervised classifier [GCC+23].

3.4.3. Model Inference

Once the latency prediction models are trained, the Model Inference function can take charge. This function utilizes the trained models and the live data collected from the 5G-Adv/6G systems to predict the latency distribution.

The placement of latency inference function is crucial and is dependent on where and how the latency predictions are utilized, corresponding to the associated actor function's role. For example, the inference function for RAN latencies should be placed on each base station to not only allow collection of real-time data for latency inference but also for optimization of radio resource allocation to meet tight QoS requirements.

Ultimately, the Actor function utilizes the latency predictions from model inference to accomplish specific objectives, such as optimizing resource allocation within the network while at the same time fulfilling the application requirements. The Actor also provides valuable feedback to the OAM / data collection function, ensuring that the data gathering strategies are refined and optimized to enhance the quality and relevance of the training dataset.

3.5. Security architecture, monitoring and analysis, remediation

In this section, we first examine the potential threats on deterministic properties, we then present our remediation strategies to prevent, detect and mitigate the threats.
Table 3.1 summarizes different threats identified from [RFC-7384] [DetNet-Sec] [WMs+21] [BCAMM+18] [DCJ+21] [FBZ+21] for TSN and DetNet. For each threat the table provides the type of threat, i.e., if it is an internal or external threat, if it involves a man in the middle (MITM), or if it is an Injection attack.) Also the impact of the threat and possible mitigation strategies are listed. For example, MITM Delay attacks will certainly impact the determinism and can be detected using performance analytics and can be mitigated using path redundancy. In the threat model presented in Table 3.1, an internal attacker either has access to a trusted segment of the network or possesses the encryption or authentication keys. An external attacker, on the other hand, does not have the keys and has access only to encrypted or authenticated traffic. An MITM attack is in a position that allows intercept, modify, or drop in-flight protocol packets, whereas an inject attack can only inject protocol packets.

Table 3.1. List of potential threats on deterministic properties

Threat	**Description**	**Internal**		**External**		**Impact**	**Mitigation**
		MITM	**Inject**	**MITM**	**Inject**		
Delay attack	Attacker maliciously delays data flow traffic.	X		X		Non-deterministic delay, Data disruption, Increased resource consumption	Path redundancy, Performance analytics
Flow modification / spoofing	Modification of headers of enroute packets, or spoofs packets → Manipulate the resource consumption.	X	X			Increased resource consumption, Data disruption	Path redundancy, Integrity protection, Node authentication For deterministic traffic, traffic policing could be applied, as e.g. per-stream filtering (PSFP) in TSN according to IEEE 802.1Qci.
Inter-segment attack	Attacker injects traffic from one segment, affecting the	X	X	X	X	Increased resource consumption, Data disruption	Path redundancy, Performance analytics For deterministic traffic, traffic policing could be applied, as e.g. per-

	performance of other segments.						stream filtering (PSFP) in TSN according to IEEE 802.1Qci.
Replication: Increased Attack Surface	Multiple paths → more points in the network that can potentially be attacked.	X	X	X	X	All impacts of other attacks	Integrity protection, Node authentication For deterministic traffic, traffic policing could be applied, as e.g. per-stream filtering (PSFP) in TSN according to IEEE 802.1Qci.
Replication Header Manipulation	Attacker modifies replication header → Forward both replicas / eliminate both replicas / flow hijacking.	X				Non-deterministic delay, Data disruption	
Path Manipulation	Attack control plane → Manipulate the paths being used.	X	X			Enabler for other attacks	Control or signaling message protection
Path Choice: Increase Attack Surface	Attack control plane → Increase number of points that can potentially be attacked.	X	X	X	X	All impacts of other attacks	
Control /Signalling Packet Modif.	Modify control / signalling packets → Manipulate path / resource allocation.	X				Increased resource consumption, Non-deterministic delay, Data disruption	
Control/ Signalling Packet Injection	Inject control / signalling packets → Manipulate path / resource allocation.		X				

Reconnaissance	Passive eavesdropping → Gather information about flows, bandwidths, schedules.	X		X		Steal confidential info., Enabler for other attacks	Encryption
Attacks on Time Sync Mechanisms	Attack time sync mechanism. → Disrupt flow forwarding.	X	X	X	X	Non-deterministic delay, Increased resource consumption, Data disruption	Path redundancy, Control message protection, Performance analytics

Even though E2E encryption is a solution to protect against the threats concerning packet modifications in the Table above, it is not effective against delaying attacks, particularly Denial of Services attacks or against an internal attack who has access to a trusted segment of the network or possesses the encryption or authentication keys. Detecting and preventing these involve setting up robust systems that can swiftly detect any deviations from expected behaviors. Security measures, including Intrusion Detection Systems (IDS), must be tailored to analyze specific characteristics of deterministic networks, where consistent timing and delivery of packets are critical. For instance, an adapted IDS in such an environment would not only look for malicious activities but also monitor the timing and order of packet arrivals to detect anomalies.

The DETERMINISTIC6G deliverable D3.2 [DET23-D32] provides a description of the security monitoring framework developed in the project. It first presents a state-of-the-art analysis and different high-precision techniques for security monitoring. It then describes a multi-domain security architecture and its enablers, that are taking into account the security aspects previously presented in Section 2.2.2. We briefly present the investigated architecture components subsequently, please refer to the DETERMINISTIC6G deliverable D3.2 [DET23-D32] for a detailed description. For the architecture evaluation we use the In-band Network Telemetry (INT) techniques [P4-INT] implemented in DETERMINISTIC6G using P4 language for fine grained analysis of flow and packet level end-to-end performance metrics to detect anomalies. INT is a technique that allows network monitoring and – with an appropriate analytics function – detecting when an anomaly occurs Currently, the INT research is led by two working groups, the Internet Engineering Task Force (IETF) IP Performance Measurement (IPPM) and the P4 Application (p4.org). IETF conducts research and standardization on the architecture and protocol of INT. It promotes insitu Operation Administration and Maintenance (OAM) which complements current out-of-band OAM mechanisms, based on Internet Control Message Protocol (ICMP) or other types of probe packet, to compose the set of tools used for OAM. The p4.org focuses on the implementation of INT using the programable data-plane and proposes basic implementation ideas using the P4 language to leverage network softwarization techniques in a flexible manner. INT has three modes of operation that go from “no”, “limited”, to “considerable” packet modification. The first mode configures a node to extract a given set of metrics (called Watchlist). The second mode adds a header that indicates what metrics need to be extracted from the packets and sent to the Collector for analysis. The last mode also carries the data captured by the

previous nodes. It is demonstrated in the Figure 3.5 which depicts various network nodes that add INT headers and captured data, enabling an INT Collector to retrieve and examine this information to identify irregularities in packet transmissions.

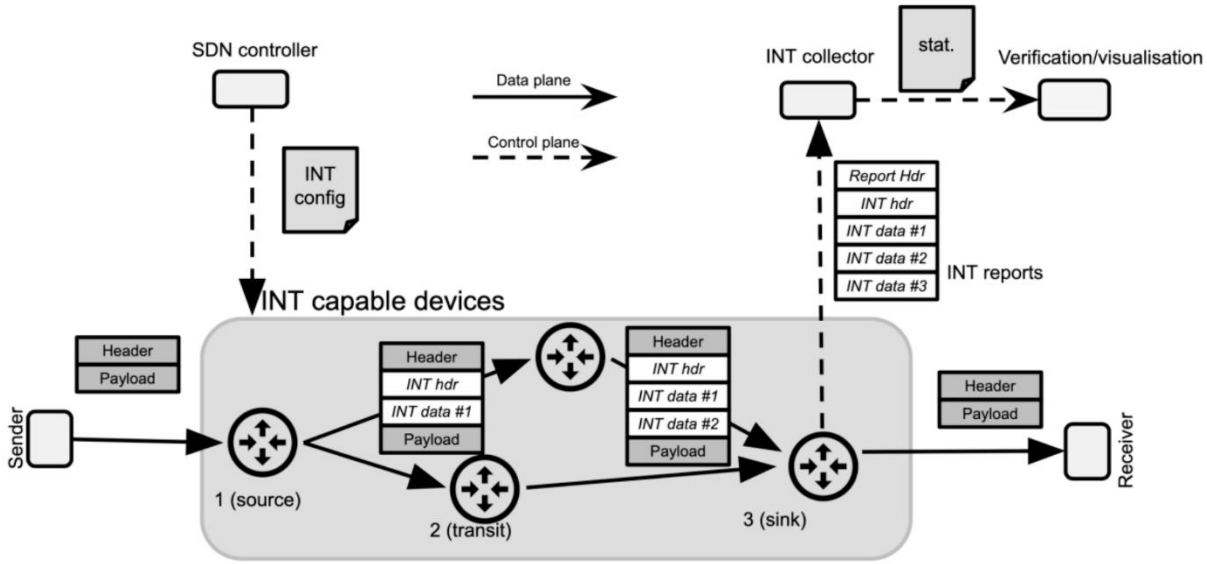

Figure 3.5. In-band Network Telemetry (INT) for fine-grained analysis of packet transmissions.

It must be noted that the overhead of INT over the latency and data payload size is inevitable but it can be minimized by limiting the data extracted (e.g., 10 values) and preferably using the first mode. We will in future work in DETERMINISTIC6G evaluate the impact of this overhead. Possible optimizations could involve selective processing of specific flows or employing sampling techniques, such as analyzing a subset of packets chosen based on defined criteria or at random, to enhance the likelihood of identifying irregularities.

In our framework, we also implement passive probes to collect data. These probes can observe the traffic flows with or without impacting them. These probes do not suffer from the overhead as being caused by INT. Deploying probes at the end points allow E2E detection of anomalies and deploying them at several strategic observation points would allow detecting where the problem occurs. Thus, the deployment ideally needs to be throughout the TSN, DetNet and 6G systems.

To mitigate and prevent attacks it is also necessary that the probe provide the captured information to the Security Analytics, that in turn interacts with the Orchestrators or Controllers via the Decision Engine to change the configuration or deploy new paths, security services, network slices, change network topology, deploy new probes, etc. Exactly where these probes need to be deployed is an open question and depends greatly on the network infrastructure and use cases.

Multiple security management domains (SMD) can be employed at the local level to analyze the behavior of deterministic network protocols (e.g., Precision Time Protocol (PTP), TSN and DetNet protocols) ensuring that the timing and reliability requirements are met. The observation points are placed within the network to detect any deviations from expected performance and trigger local corrective actions. It can be beneficial to complement this local surveillance by cross-domain analytics, which could provide end-to-end (E2E) network awareness, and may allow to coordinate a network-

wide response to any issues, guaranteeing that the entire system maintains the required deterministic properties. It is subject to further work to investigate how such security analysis and mitigation can be coordinated across multiple network domains.

When deviations in network traffic behavior are observed, it is essential to promptly pinpoint the root causes. This could range from configuration errors, software bugs, lack of resources, hardware failures, or malicious attacks. Quick identification allows network administrators or automated reaction mechanisms to take decisive action to mitigate any issue. Additionally, network traffic must be constantly scrutinized for irregular patterns or non-credible sources. Non-reputable traffic, which might indicate a security threat or network malfunction, must be identified and either rerouted (to be further analyzed) or blocked to maintain network integrity and performance. By blocking, isolating, or diverting suspicious traffic, the impact on the network's deterministic performance is minimized, preserving the crucial predictability and reliability required by applications that depend on deterministic networking.

One strategy to achieve minimal latency impact involves deploying monitoring techniques that can implement machine learning-based analysis near the data source, e.g., in a local SMD, or in a federated way, and eventually use reduced monitoring techniques, such as random sampling, possibly incorporating statistical analysis methods for predictive modelling. These approaches aim to intelligently divide the analysis task and filter the volume of traffic to be inspected, thus reducing the processing burden and data transmitted, and conserving network resources.

Moreover, network slicing facilitates segregating DetNet/TSN traffic from non-deterministic flows, thereby allowing each slice to maintain its deterministic nature while being shielded from potential disruptions caused by less critical traffic. This segregation also simplifies the process of identifying and isolating anomalous or malicious traffic, as it is easier to monitor and control within the confined boundaries of a network slice. Furthermore, slicing can be used to redirect network traffic to ensure the deterministic requirements of the communications flows with high priority and isolate them from DoS attacks.

Thus, an integral part of the security architecture is the careful calibration of risk, cost, and performance trade-offs. Decisions regarding security policies and the level of security measures to be implemented must align with the value of the assets being protected, the potential impact on network performance, the safety requirements, the likelihood of threats, and, at the same time, not overly affecting the deterministic properties of the communications.

4. Initial architecture for E2E Deterministic Communication with 6G

We base our architecture on the work in [Eri23] as depicted in Figure 2.12 and extend it to environment in focus in DETERMINISTIC6G. The resulting draft architecture is shown in Figure 4.1. In contrast to Figure 2.12 a *deterministic network* domain has been added. A 6G system will provide to 6G devices access to an external data network via the 6G network, following the same principles as 5G. In our context we expect this to be a specific data network with support for deterministic communication and it is expected to be either a deterministic LAN network, based on Ethernet TSN, or a deterministic IP network, based on IP DetNet. This deterministic network may extend to the device side, where a local deterministic network on device side is connected via a 6G UE gateway and the 6G network connectivity to the deterministic data network. The deterministic data network can provide access server running time-critical application, or those can be hosted in an edge computing infrastructure. In some cases, the applications may use a middleware framework to realize their time-critical services, which in turn uses the end-to-end deterministic networks available. An example is OPC UA with its extension to (dependable time-critical) field level communication (denoted as field exchange), which is being specified for a wider range of industrial use cases.

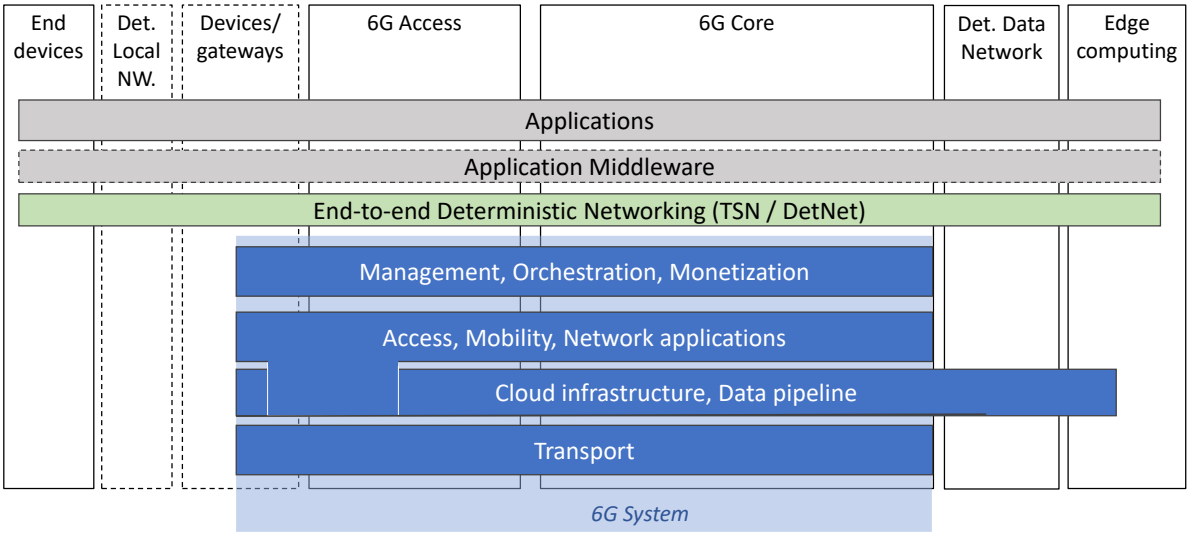

Figure 4.1. Draft DETERMINISTIC6G architecture (derived from Figure 2.12 [Eri23])

Figure 4.2 shows how the concepts developed in DETERMINISTIC6G integrate into the end-to-end architecture with 6G based, where the numerals in the list correspond to the corresponding numerals in Figure 4.2:

- 1: In [DET23-D11] we have described future time-critical use cases and applications and analyzed their requirements in terms of *key performance indicators* (KPI) and value creation in terms of *key value indicators* (KVI). In the architecture the applications are operating in the application domain end-to-end, and may make use of some application middleware.
- 2: To invoke dependable communication for time-critical services, the applications need to provide their service specification. i.e., to specify their traffic characteristics and performance requirements, in order to request a dependable communication service from the network. By

enhancing the information exchange and provide situational awareness between the application domain and the network domain, better service provisioning is envisaged. To some extent this is described in [DET23-D11], but is generally planned for future work.

- 3: The network needs to provide a dependable communication service. This means that it must be able to comply with and deliver the performance that is requested from the applications. To this end, the network needs to be able to monitor the KPIs that characterize the delivered service performance. Furthermore, by data-driven (latency) performance prediction, the 6G network shall be able to specify which (latency) performance levels it can promise to what reliability level [DET23-D21] [DET23-D42]. One important characteristic is to be able to control also the packet delay variation as explained in [DET23-D31}, for which mechanisms like packet delay corrections are proposed [DET23-D21]. These functions are part of the access and network applications in Figure 4.2 and build on the time awareness described below. The data-driven latency prediction further builds on the availability of a data pipeline for data collection and distribution to feed machine learning models, as described in [DET23-D21] [DET23-D41].
- 4: Dependable time-critical communication builds on time-awareness throughout the system. This is achieved by robust time synchronization which should also include hot standby support for time synchronization. Time-awareness is provided I the transport and infrastructure layers, and may be used in the network function layer. To provide robust and secure 6G network services a paradigm of security by design shall be applied. To this end, data monitoring at the transport layer shall be possible, in combination with the data pipeline that allows for smart security assessment based on observed network behavior (see [DET23-D32].
- 5: With the increasing interest to apply cloud compute capabilities, a cloudification of application / control functionality towards an edge cloud is of primary interest. This is in particular of interest for applications where functionality is offloaded to network-side (edge) compute capabilities to improve device performance. The integration of edge cloud with deterministic networking, and providing dependable compute to the application domain ensure timely and effective integration of compute with time-sensitive communication in an end-to-end manner.
- 6: When considering latency variations of sub-components in an end-to-end system, gains can be provided for end-to-end deterministic networking by making the end-to-end traffic handling aware of the latency characteristics of sub-components (see [DET23-D31] To this end, optimizations will be proposed and evaluated, which are applied in the end-to-end deterministic network domain (i.e., TSN and DetNet). They will provide more robust and optimized end-to-end deterministic network configurations that take the characteristics of the 6G network into consideration. Such information is provided to the end-to-end deterministic networking controller (TSN or DetNet) from the 6G management layer and is based on network insights as described above in item 3.

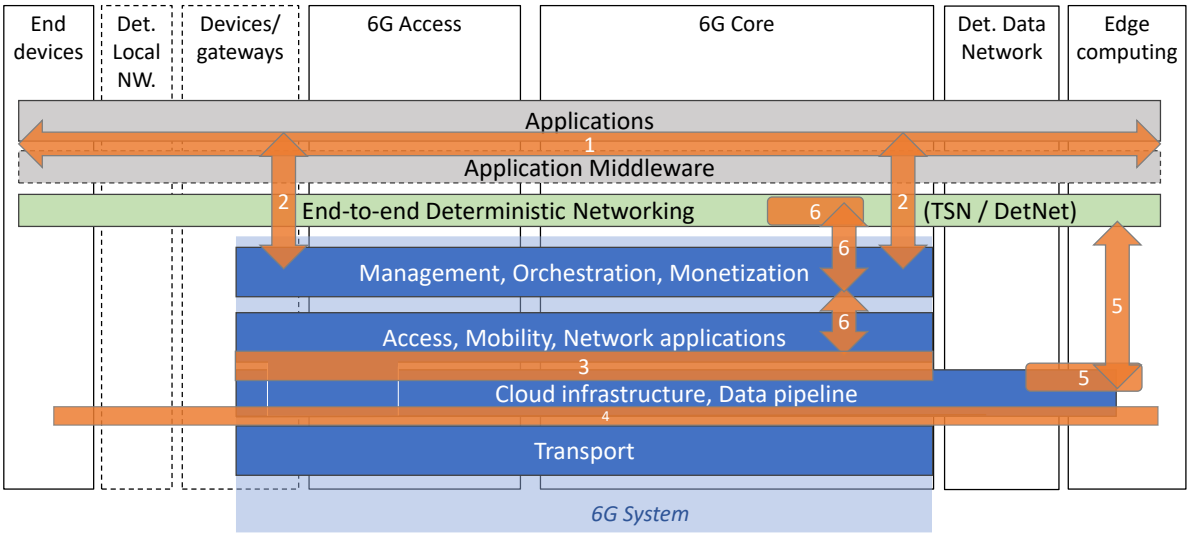

Figure 4.2. DETERMINISTIC6G enhancements in the end-to-end architecture

5. Realization of different use cases

Innovative use cases have been described in [DET23-D11], which all require dependable time critical communication that can profit from, or rely on the solutions developed in DETERMINSTIC6G. In the previous chapters, DETERMINISTIC6G technology components are described, architectural design choices are analyzed and a functional architecture for 6G is proposed in chapter 4 that integrates the DETERMINISTIC6G technology components. In this chapter, we want to explore how this functional architecture can be applied to the different use cases in [DET23-D11] with a corresponding network deployment architecture and a discussion on how the technology components benefit the DETERMINISTIC6G use cases.

5.1. Review of DETERMINISTIC6G use cases

When imagining a physical deployment of the use cases, we can see that the DETERMINSTIC6G use cases can be grouped into two categories:

- Shopfloor-based use cases
- Outdoor confined area use cases

In fact, both of these use case categories are focused on confined areas: a well-specified geographic region within which the use case takes place. In the shopfloor-based use cases, this is the shopfloor of an industrial site like a factory; typically, it is an indoor location within one or more buildings. Other use cases exist in, e.g., other industrial environments, like a mine, a port, a construction area, which can include outdoor areas. Such use cases have similarities with the shopfloor-based and the outdoor confined area use cases.

5.1.1. Shopfloor-based use cases

The shopfloor-based use cases described in [DET23-D11] are the following:

- Extended Reality (XR) for industrial workers
- Occupational exoskeletons
- Adaptive manufacturing

In Section 5.2.1, a detailed analysis of a use case realization for occupational exoskeletons on a factory shopfloor is made and the corresponding 6G network deployment architecture is discussed. For the other shopfloor-based use cases a similar approach could be chosen, and it is briefly discussed how the illustrated 6G network deployment can be extended to also include adaptive manufacturing and XR for industrial workers.

5.1.2. Outdoor confined area use case

The smart farming use case described in [DET23-D11] is focused on the automation of farming operations on the fields, like ploughing, sowing, and harvesting. These operations are concentrated at the physical locations of the fields, somewhat remote from inhabited areas and the farms owning the fields. But the use case may include transit of machines, agriculture equipment, and goods between the fields and the farms. Other smart farming scenarios located at the farms are not explicitly investigated here. Although the physical location of the smart farming automation is on the fields, the automation system for the planning, monitoring & supervision of farming operations and some remote control is expected to be distributed over a geographical area. One option is having an "automation server room" at the farm itself, but a more likely alternative is that some of the automation functions are hosted in a data center in vicinity of the farms and the fields.

5.2. Shopfloor-based use cases

5.2.1. Industrial Exoskeletons

5.2.1.1. Review of the use case, traffic and topology

As detailed in [DET23-D11], occupational exoskeletons (OEs) are wearable robots aimed at reducing the physical load of workers performing demanding activities [MAD20]. Active OEs (i.e., exoskeletons relying on powered actuators to generate the assistive action) have several advantages, such as:

- they can provide adaptive support based on the user's or environment's inputs,
- they may be fully integrated with the smart factory digital ecosystem, allowing the possibility of a real-time monitoring or tuning of the system.

On the other hand, they also have some disadvantages:

- they may be heavy and cumbersome to wear for long periods of time, due to the presence of actuators, electronic components, batteries,
- they have high-demand requirements in terms of power supply,
- they require real-time, deterministic networking of subsystems (e.g., on board sensors of OEs, external sensors for monitoring the user's status and environment) and elaboration through complex control strategies to deliver the correct amount of assistance depending on the user's needs.

The use case related to industrial exoskeletons taken into consideration during the project involves an active OE for lumbar support, connected through the 6G network to a cloud-based system, which collects information both from the exoskeleton and from different environmental sensors to identify, instant-by-instant, the optimal task-oriented assistive strategy.
The delocalized controller, real-time processing and monitoring of a huge amount of information gathered from multiple subsystems would boost the maturity of active exoskeletons, by enabling off-board feasibility of complex assistive strategies that can adapt to the different tasks and activities that

the workers have to perform, accounting for kinematics, physiological, and environmental information.
Furthermore, the offloading of the control would also allow to delocalize hardware components, thus reducing the exoskeleton's weight/bulkiness and its power consumption.

It is also expected that a virtual replica of the system is created into the edge cloud: this digital twin allows an external expert, such as an ergonomic specialist, to remotely monitor and analyse the worker's movements in order to provide real-time feedback and suggestions for posture correction.

Since two different scenarios have been identified, depending on the computing offloading strategy, a brief description of the exoskeleton's control levels is provided hereafter, for the sake of clarity.

- *Low-level control* refers to the basic control of exoskeleton's actuators; it acquires information from sensors embedded in the device, by implementing the on-board sensors reading protocols, receives input from the middle-level control and computes the precise instructions to transmit to the actuators.
- *Middle-level control* refers to the part of the control responsible to translate the high-level commands into reference commands for the low-level control.
- *High-level control* is responsible for the definition of the assistance provided by the exoskeleton: it oversees the task recognition and the detection of worker's movements, and it elaborates information from the exoskeleton and from external sensors, such as cameras, to make decisions about the exoskeleton's mode of action.

The two scenarios taken into considerations, which differ by the computing offloading strategy, are:

- **near-term scenario**, in which the low-level control is embedded in the exoskeleton, while middle and high-level control are delocalized in the cloud;
- **long-term scenario**, in which all the control levels (i.e., low, middle and high) are delocalized in the cloud.

Despite higher frequency of data packets exchange, the long-term scenario should rely on a simpler embedded architecture which offers benefits in terms of overall OE encumbrance and power consumption. The environment that is the background of this use case is therefore a shopfloor, whose components are connected via 6G network to the edge cloud, which is in charge of different functions. Both shopfloor and edge cloud elements are described below.

Shopfloor elements

- **Exoskeletons**, whose number inside the shopfloor may vary depending on the number of employees. Each exoskeleton comprises, for each powered joint, an actuator, its driver and several embedded sensors, such as joint position sensor, torque sensor, motor axis position sensor. On-board sensors and actuators/drivers are controlled by the exoskeleton's embedded unit, which oversees acquiring data from the on-board sensors and providing the proper commands to the actuator drivers. Data acquired from the on-board sensors is then transmitted by the exoskeleton's embedded unit to the control unit in the edge cloud:
    - every 10 ms in the near-term scenario, in which they are transmitted after having been processed by the low-level control,
    - every 1 ms, in the long-term scenario, in which they are processed by all the control levels in the cloud-based system.

The size of the data packets transmitted from the exoskeleton embedded unit is 100 bytes and it has been defined taking into consideration, as defined in [DET23-D11], an active lumbar exoskeleton, thus embedded with 2 powered joints (left and right hips), able to provide assistance, both during loads lifting and during walking.

- **Environmental sensors**, which are sensors located in the shopfloor and whose number depends mainly on the shopfloor dimension. They could be sensors of different types, such as cameras, providing information on the workers' movements and on the loads' shapes and locations, or sensors (such as force sensors, Inertial Measurement Units (IMUs) or switches) positioned on the loads to be lifted. They transmit data:
  - periodically, in case of cameras, for which transmission is foreseen every 16 ms (which corresponds to a typical frame rate of security cameras);
  - aperiodically, in case of sensors positioned on the loads, which transmit data only when the load is lifted/moved.

- **Dashboard**, which shows a virtual replica of the system, created via a digital twin of the exoskeleton, enabling the possibility to visualize and monitor the exoskeleton's action and the worker's movements. This allows ergonomics experts to detect any bad posture and send notifications to the workers for posture correction (e.g., haptic feedback provided through the device's actuators, or visual feedback provided through exoskeletons' embedded Light-Emitting Diodes (LEDs)). Moreover, it allows them to identify any risky movement or critical load to be lifted and, thus, to modify the assistance provided by the exoskeleton as needed, to enhance the support to the worker and improve the quality of the movements.

Edge cloud elements

- **Exoskeleton controller**, one for each exoskeleton in the shopfloor. Each controller elaborates (i) data coming from the sensors on board the exoskeletons, which provide information on the current exoskeleton status and on the workers' movements, (ii) environmental sensors data, which, as mentioned above, provide information such as locations/movements of both workers and loads inside the shopfloor, load dimensions, force applied on a load and (iii) information from the dashboard, which may lead to fine-tuning of the assistance or to change the actuators action to provide haptic feedback to the worker. The controller exploits all this data to identify, via a complex algorithm, the optimal assistive strategy for the movements and the task that the worker is accomplishing. The assistive strategy is translated into reference commands for the low-level control in the near-term scenario and in precise commands for the actuators in the long-term scenario and they are then transmitted from the edge cloud server to the exoskeleton:
  - every 10 ms in the near-term scenario, in which, when received by the exoskeleton embedded control unit, they are processed by the low-level control,
  - every 1 ms, in the long-term scenario, in which they are transmitted after having been processed by all the control levels in the cloud-based system.

  The data packets transmitted from the edge cloud controller contain, together with the information concerning the assistive strategy, information to be shown via Liquid Crystal Display (LCD) displays/LEDs integrated into the exoskeleton. The size of the data packets is 100 bytes.

- **Digital twin**, one for each exoskeleton. Exploiting the data coming both from the exoskeleton and from the environmental sensors, a virtual replica of the exoskeleton is created whose function is to allow to visualize, monitor and analyse the exoskeleton's action and the worker's movements, via the dashboard.

Figure 5.1 shows the topology envisioned for this use case, as described above:

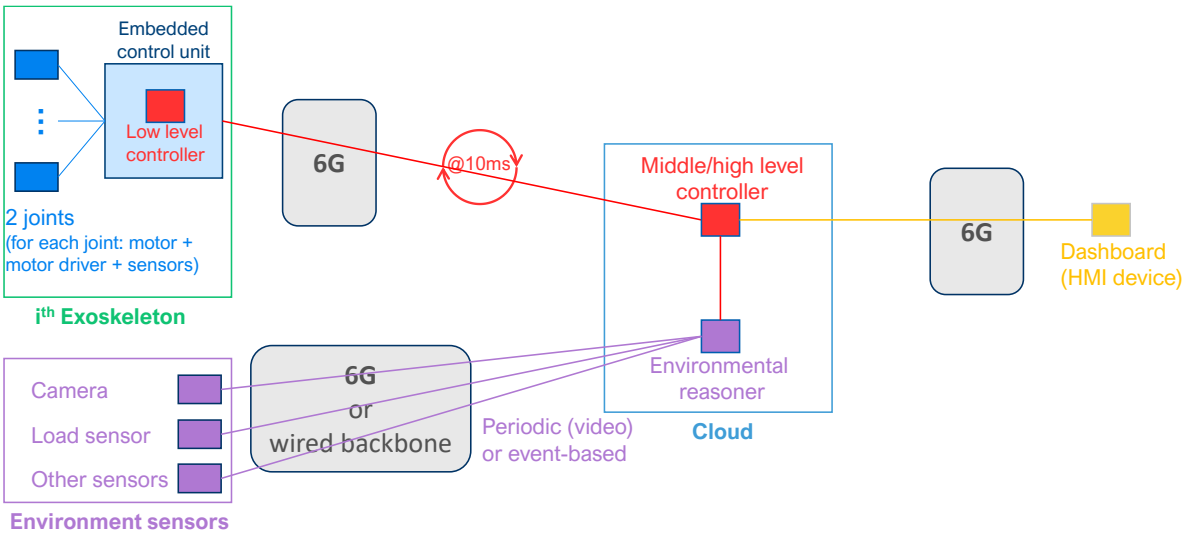

Figure 5.1: Near-term scenario for occupational exoskeleton.

An evolution of this use case is shown in Figure 5.2, where also the low-level control of the exoskeleton is located in the compute environment of an edge cloud.

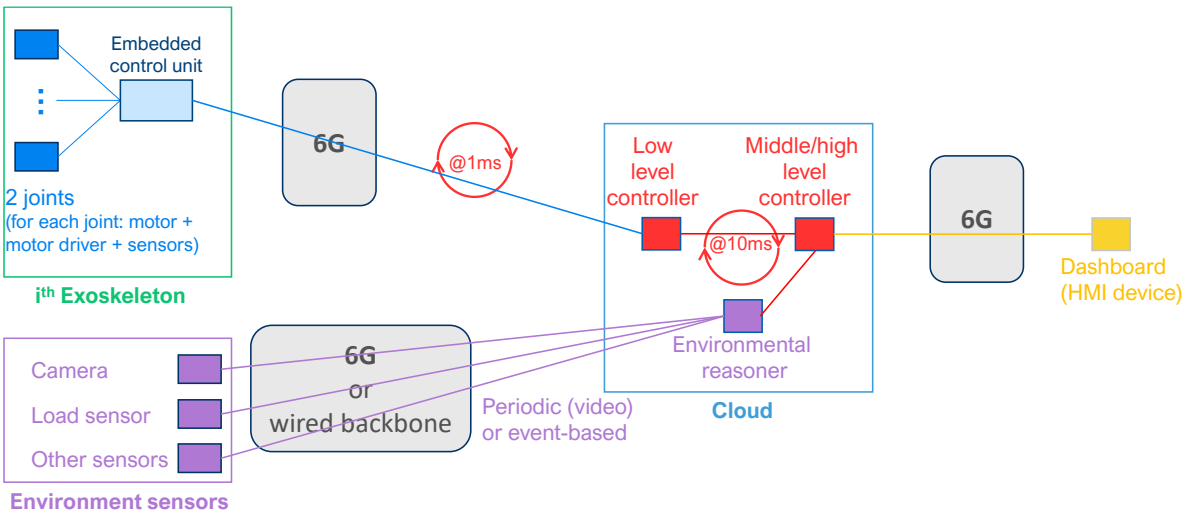

Figure 5.2: Long-term scenario for occupational exoskeleton.

5.2.1.2. 6G deployment architecture

An example for the topology of a factory shopfloor in this use case is shown in Figure 5.3. The factory comprises different zones, like an inventory with high-shelf storage close to the supply management.

At the center of the shopfloor there are production lines with industrial robots and assembly stations. The intra-logistics is realized via automated vehicles like automated guided vehicles (AGVs) that transport goods, workpieces, and materials on the shopfloor where needed. The shopfloor has an operation center, where the digital automation platforms are located, and which also comprises a corresponding on-site server room which offers edge-cloud computing capabilities. In this document it is assumed that a part of the automation functionality is realized as cloud-hosted applications on this compute infrastructure, but the operation center could also host special-purpose hardware, like PLCs. There, the controllers for the exoskeleton and the environmental reasoner are located. Fixed installed equipment, such as the production lines and some cameras, are connected via a TSN backbone with the server room. A private 6G network is deployed in the factory and covers the entire shopfloor. The workers wearing the occupational exoskeletons are located on the shopfloor, e.g., at the production lines or at assembly stations. In addition, environmental sensors and cameras are distributed over the shopfloor. AGVs, exoskeletons, sensors and (some) cameras are connected via 6G to the server room.

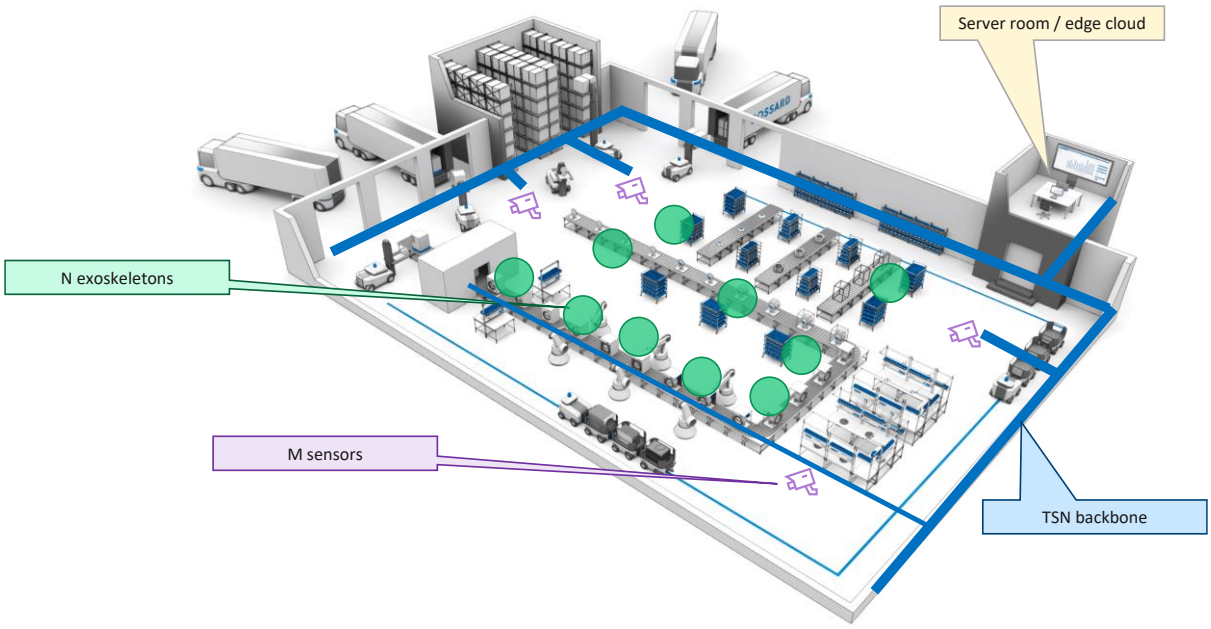

Figure 5.3: Topology of the shopfloor for the industrial exoskeleton use case.

A logical 6G network architecture for the shopfloor is depicted in Figure 5.4. The figure shows also the communication relations between main shopfloor elements. Different functions located in the edge cloud are communicating among themselves via the TSN backbone. Mobile devices, like the exoskeletons, some Human Machine Interface (HMI) devices, and sensors, are all connected to the 6G virtual TSN bridge, and then via the TSN backbone to the edge cloud.

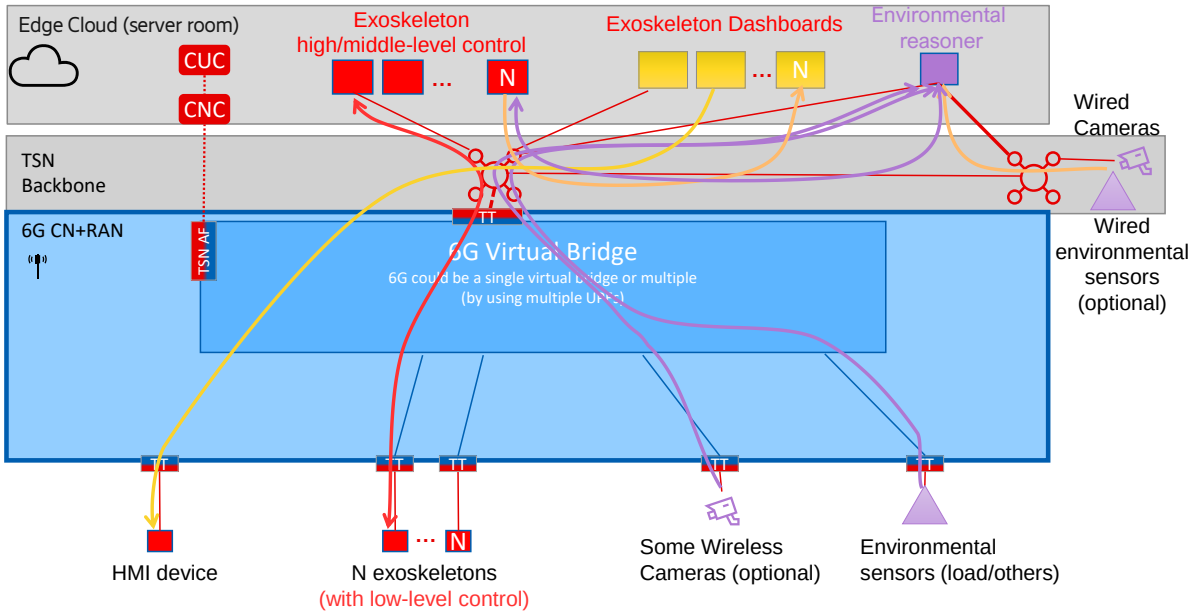

Figure 5.4: Logical 6G network architecture for the occupational exoskeleton (near-term scenario) on the shopfloor.

5.2.2. Additional shopfloor-based use cases

5.2.2.1. Extended Reality for connected workers

The use case on XR for connected workers is in detail described in [DET23-D11]. Workers on the shopfloor can wear XR glasses, which provide augmented reality where digital information is blended into the visually observed physical surrounding. As discussed in [DET23-D11], this comprises several steps. The recording of the physical environment and seen by the worker can be augmented with information from the automation system or the digital twins of the assets on the shopfloor. This allows for the following use cases:

- The worker interacts with the shopfloor system, machines and equipment present there, to e.g.
  - Manipulate, configure, or maintain assets. Via augmented reality the workers are provided with guidelines, instructions and other support information that are overlaid via the glasses to the view of the worker.
  - Visualization of shopfloor information, like next planned steps of machines.
- Collaborative design among multiple workers, working on a shared digital model.

For XR, use cases comprise two compute intensive tasks: spatial compute (to obtain a spatial understanding of the local environment) and rendering of the scene with the multiple integrated (digital and physical) objects. Great benefits are obtained if these functions are offloaded from the device (i.e., the XR glasses) to the edge cloud [DET23-D11] as shown in Figure 5.5. In addition, information about the digital objects related to the shopfloor that are to be immersed into the scene needs to be provided to the rendering engine; this information is provided e.g., by digital twins of the shopfloor assets. XR use cases introduce communication between the XR device and the server to which functionality has been offloaded, which is expected to be located in the server room of the factory.

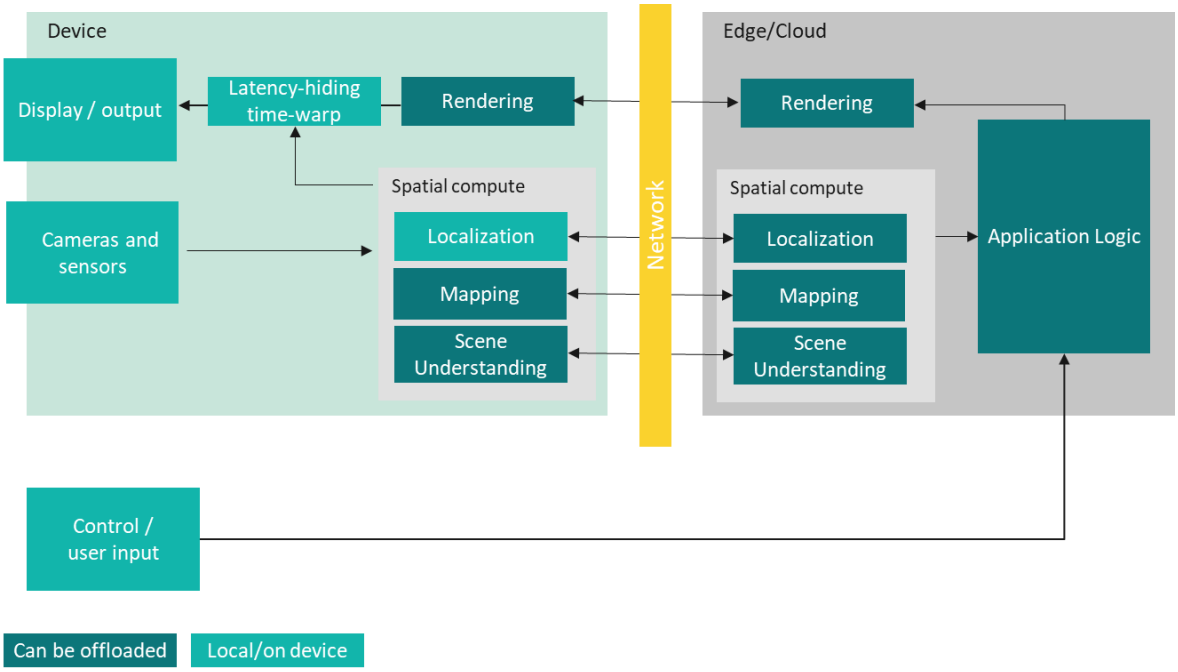

Figure 5.5: XR functionality with offloading options, see [DET23-D11].

5.2.2.2. *Adaptive Manufacturing*

The Adaptive Manufacturing Use Case as described in detail in [DET23-D11] brings in multiple different scenarios that show the inclusion of mobile components and actors within an automated factory. These mobile components include Automated Guided Vehicles (AGVs) and Mobile Processing Modules (MPMs). Both types of devices are based on the idea to have a mobile device that can move freely within the factory floor and either transport parts or more complex machinery (like tools or robot arms) which can be combined and used in cooperation with stationary components like a processing cell.

The mobility of these components and the capability of moving automatically introduces many possibilities for flexible and adaptive applications, but it also introduces a higher level of complexity and the necessity of avoiding obstacles.

The motivation behind this use case is twofold. Firstly, it aims to accommodate the faster adoption of cutting-edge technologies into manufacturing use. This includes the integration of mobile robots (MRs), machine learning, and advanced sensing technologies. These technologies can significantly enhance the efficiency and effectiveness of manufacturing processes. Secondly, the use case leverages 6G to provide a reliable wireless communication infrastructure. This infrastructure is designed to match the diverse requirements put forward by use case enablers, such as ensuring functional safety over wireless and facilitating MR-MR interaction.

The benefits of implementing this use case are substantial. It leads to a higher degree of production flexibility and manufacturing adaptivity. This means that the manufacturing process can easily adapt to changes in demand or production requirements, thereby improving efficiency. Furthermore, the use case can decrease the downtime or changeover time of production lines, leading to an increase in

overall productivity. This can significantly enhance the factory's output and potentially lead to increased profitability.

The following lists show the different components and functions that are either physically present on the factory-floor or are executed in the Edge Cloud. For readability reasons, it is left to the reader to identify appropriate locations of these components in Figure 5.3 and Figure 5.4.

Factory-floor elements

- **Automated Guided Vehicle (AGV)**, which is a generic device that can move automatically without the need of a human operator. Such devices can be used for various purposes. In the context of this use case, AGVs are used for moving parts or goods within the factory. This movement can have the simple purpose of transporting parts between static components, or the movement is performed in coordination with the surrounding environment. Such a coordinated movement can be part of the processing application. AGVs need different types of communication with their surrounding environment strongly depending on the application and usage of the AGV.
- **AGV swarm**, is a logical combination of multiple Automated Guided Vehicles (AGVs) moving in unison. They work together to transport manufacturing parts. The swarm coordinates its movement through the periodic exchange of status and control information with an exemplary interval of 5 ms. It also communicates with the safety system and processing cell for registration and deregistration through aperiodic exchange of device information. Functional safety-related messaging within the swarm usually involves the periodic exchange of information with typical intervals of 20 ms.
- **Mobile Processing Module (MPM)**, which is an AGV equipped with a robotic arm or another tool. It communicates with the production line for registration and deregistration through aperiodic exchange of device information. During the combined operation, together with the processing line, the communication changes to a low latency periodic form with typical intervals of 1 ms. MPMs also inter-report during their movement, but they do not form a swarm. They also use aperiodic communication with the MPM base station for registration and deregistration.
- **Safety System**, which controls an AGV swarm for functional safety purposes. It communicates with the swarm through the periodic exchange of control information, such as every 20 ms for a functional safety stop.
- **Surveillance Camera / Object Detection System**, which identifies potential other actors in a factory region shared by AGVs/MPMs, such as human personnel or human-controlled vehicles. It performs obstacle detection through the periodic exchange of object information every 10 ms, but with different possible levels of operation. These levels of operation can be adapted to the available communication resources and achievable capacities.
- **Processing Cell**, which performs different operations on the part transported by an AGV swarm. It coordinates its movement with the swarm through the periodic exchange of status and control information every 5 ms.
- **Production Line**, which performs cooperative processing operations with an MPM. It operates in conjunction with the MPM through the periodic exchange of status and control information every 1 ms.

- **MPM Base Station**, which charges the battery of an MPM while it's docked, changes the MPM tool, etc. It communicates with the MPM through both, periodic and aperiodic exchange of status and/or control information, depending on the implemented logic.

Edge cloud elements

- **AGV Swarm Coordination**, which is a crucial function within the edge cloud on a factory floor. This function is responsible for coordinating the activities of the AGVs that operate as a swarm. The coordination could be driven by any of the swarm AGVs or a separate process running on an edge cloud. This ensures that the AGVs work in harmony and effectively perform their intended operations.
- **Safety System**, which is another vital function within the edge cloud. This function is responsible for executing safety-related controls to ensure the safe operation of the factory floor. The system can be designed to provide redundant computing and communication components, ensuring that safety measures are always in place and operational, thereby minimizing the risk of accidents or mishaps on the factory floor.
- **Object Detection System**, which is a function within the edge cloud that uses image processing to detect objects (i.e., static, or dynamic ones) on the factory floor. By processing images from cameras or sensors around the factory, the system can identify and track objects, and indicate potential obstacles in real-time. This allows, for instance, to take immediate actions to avoid any disruptions to the factory operations, and thus, to ensure a smooth and efficient workflow. Running such a system on an edge cloud can enable the usage of resource intensive technologies and algorithms for improved object detection.

5.2.2.3. 6G deployment architecture

An exemplary shopfloor topology and a corresponding logical 6G architecture are depicted in Figure 5.3 and Figure 5.4, respectively, for the exoskeleton use case. In a realistic scenario all of the above use cases would be realized on the shopfloor at the same time and the elements of the XR and adaptive manufacturing use cases should accordingly be added to Figure 5.3 and Figure 5.4. This would make the figures rather unreadable, and thus, we abstain from it. Instead, we highlight here the envisaged updates to Figure 5.3 and Figure 5.4. Each XR user would add an additional device on the shopfloor that is connected via the 6G network to the edge cloud. In addition, there would be active communication between the offloaded XR functionality (i.e., spatial compute and rendering) in the edge cloud and digital twins of shopfloor assets (also expected to be hosted in the edge cloud) via the TSN backbone network.

To include adaptive manufacturing, the figures would need to be extended as follows:

- The shopfloor of a factory is structured into production lines and processing cells, with specified safety-critical zones and AGV-/MPM-shared regions, and also MPM base stations present in different locations,
- AGVs and MPMs are distributed over the shopfloor in accordance with their task assignments,
- Safety Systems are associated to the safety-critical zones,
- Surveillance Cameras are placed in a way to overlook the AGV-/MPM-shared regions,
- Edge computing in the server room hosts machine vision and other algorithms for the object detection system to process camera images and track objects.

Regarding the logical 6G architecture, the following additions would need to be depicted in the corresponding figure. Many shopfloor elements (e.g., exoskeletons, XR devices, AGVs, and MPMs) would contain a 6G UE and communicate via the 6G network to their corresponding applications or offloaded functionalities in the edge cloud. Communication between different applications in the edge cloud would be realized via the TSN backbone network. In order to achieve the AGV/MPM swarm operation, the different mobile devices communicate with each other via the 6G network.

5.3. Smart Farming

5.3.1.1. Review of the use case, traffic and topology

Smart Farming addresses a societally important question of global food production and supply, which require increased levels of automation to efficiently use, e.g., available land and water. Distinguishing features of this use case [DET23-D11] entail: scalable field monitoring and exploitation of the collected data in managing farming operations, a timely identification of crops affected by pests and bad weather as well as planning ways of crop treatment, ground and/or aerial vehicles which inter-share information and cooperate to execute farming tasks, etc.

As compared to the other DETERMINISTIC6G use cases, Smart Farming targets mobile automation and outdoor communication over possibly large areas. A remote-control center delivers work plans to Unmanned Ground Vehicles (UGVs) and Unmanned Aerial Vehicles (UAVs), whose navigation paths are, in turn, planned by motion controllers geographically closer to the farming vehicles. The global motion planning is based on tasks which are provided by a specific farming application, which, in turn, uses reporting on both task status and vehicle status from the UGVs and/or the UAVs to decide on the execution of the next task. For the decision making, farming applications may also take advantage of sensor data collected from the vehicles by one or more monitoring applications. All farming tasks are supported by safety applications, which are responsible for, e.g., avoiding collisions among the farming vehicles and with field personnel and animals.

Edge computing will offer capabilities to offload computationally intensive processing of data such as video from the farming vehicles to reduce power consumption and increase their battery life. That way, a total footprint of specialized hardware in the UGVs and UAVs may be decreased, leading to more cost- and energy-efficient solutions. Furthermore, edge computing can host AI/ML algorithms, which allow to carry out an advanced fusion of data from different sensors (e.g., temperature, humidity, and air pressure) and, thus, adapt decisions for both high-level work planning and motion control of the farming vehicles. At a core of the whole system infrastructure is 6G, which is expected to provide a dependable wireless communication for a diverse set of applications.

Ultimate benefits of using the technological “pillars” for Smart Farming include a more efficient use of critical resources, such as land and water, improvement in crop yield from existing fields, as well as a reduction of production waste.

The two lists below a) summarize main components and functions (“actors”) that are involved in communication scenarios of this use case and b) list those of them which may specifically be executed in an edge cloud.

Main actors

- **Unmanned Ground Vehicle** is an autonomous or remote-controlled farming vehicle, which is used for different tasks such as ploughing, sowing, harvesting, etc. Examples of an UGV

encompass planters, (combine) harvesters, trolleys, and rollers. Being a central field device for the overall farming process, UGVs put forth different communication requirements. For motion planning of a single UGV, a periodic exchange of control data every 2-20 ms is established by its associated motion controller. To (locally) coordinate movements among different UGVs, a complementary, periodic transmission of vehicle status information is carried out every 2-20 ms. In addition, when the UGVs need to coordinate their actions, for instance, a harvester emptying yield to different trolleys, two types of information flow are used among the vehicles: an aperiodic transmission of, e.g., connect/disconnect commands and a periodic exchange of application status information, every 50-100 ms.

- **Unmanned Aerial Vehicle** is an autonomous or remote-controlled farming vehicle used for crop inspection, transportation of light objects, etc. This can, for instance, be a drone or light sport aircraft. Motion planning for each UAV relies on a periodic exchange of control data with its controller, every 2-20 ms. Real-time communication among the UAVs and the UGVs is employed to coordinate their inter-movement while collaborating on a farming task. That communication is a periodic transmission of vehicle status information every 2-20 ms.
- **Farming application** refers to the use of specific technologies and procedures to, e.g., manage growing crops. One example of such an application is Variable Rate Treatment, to enforce a variable rate of fertilizers or pesticides.
- **Motion planning application** comes in one or more instances, to calculate navigation paths for the UGVs and the UAVs.
- **Monitoring application** also comes in one or more instances, to collect data from the farming vehicles, sensors, etc., predict possible problems regarding crop damage and waste, and determine countermeasures. For the purposes of monitoring different sensors, a periodic exchange of monitoring data is run every >= 100 ms.
- **Safety application** processes images/videos, detects possible obstacles from them, and aims to help avoiding collisions, e.g., between the UGVs and field personnel. To keep avoiding path obstacles, a periodic exchange of images/videos every 2-100 ms is performed with one or more of the involved UAVs as well as an aperiodic transmission of notifications on obstacles to the UGVs.
- **Remote control center** is used for overall work planning and monitoring of farming operations, but also, in specific cases, to remotely control motion of the farming vehicles. For instance, to monitor different sensors, a periodic exchange of monitoring data is carried out every >= 100 ms.

Edge cloud elements

- **Farming applications** benefit from data shared by motion planning and monitoring applications, to plan specific tasks.
- **Motion planning applications** take, as inputs, scheduled tasks from farming applications, vehicle status information, as well as possible obstacle notifications from safety applications, to plan navigation paths for the UGVs and the UAVs.
- **Monitoring applications** collect information from the farming vehicles and different sensors and may also make that information available to the other applications.
- **Safety applications** support execution of all farming tasks which include field personnel and the farming vehicles.

5.3.1.2. 6G deployment architecture

A logical 6G network architecture for this use case is depicted in Figure 5.6. The smart farming area is a confined area that comprises, e.g., the agricultural fields. It seems unlikely to build a dedicated private 6G network for providing communication coverage and connectivity on the field. Public mobile network operators already offer national mobile network coverage, for which a future 6G support is assumed in this document. An efficient way to offer communication services for the use case is to request a dedicated network slice (or a so-called public network integrated non-public network) from one of the mobile network operators. This creates a virtual network, including the coverage at the pre-defined area with guaranteed wireless access for the authorized devices according to the agreed and configured slice capabilities. For the purpose of cost effectiveness, the smart farming applications are desired to be hosted on an efficient compute platform. A viable solution is to run these functions on an edge cloud that is provided by the mobile network operator, or the local data center of a hyper-scale cloud provider. The area covered by the use case spans a comparatively large area and longer distances, so DetNet can be considered a well-suited IP-based communication system for deterministic networking.

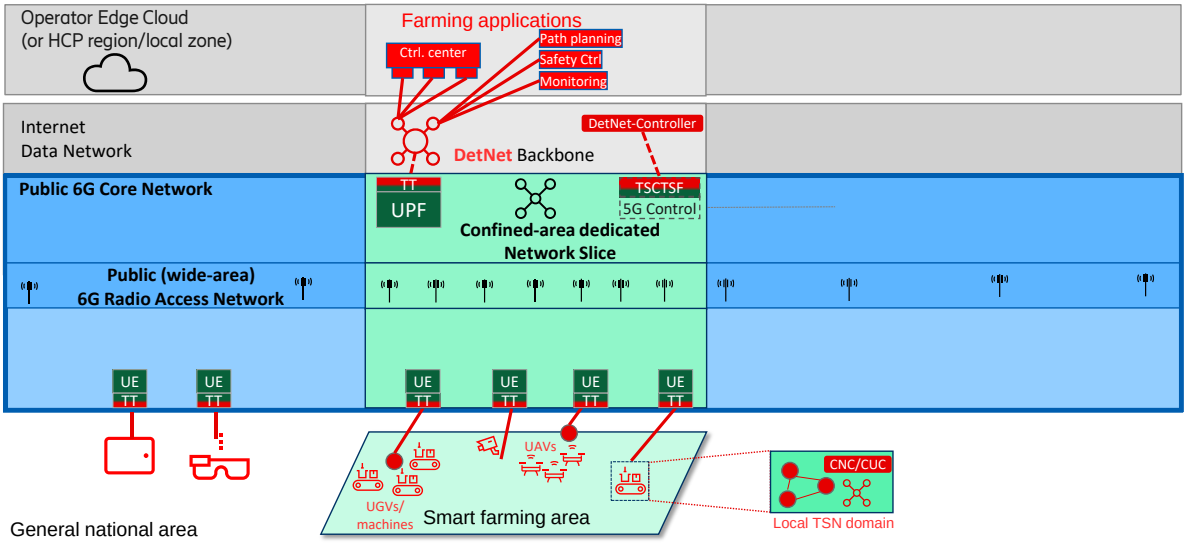

Figure 5.6: Logical 6G network architecture for the smart farming use case.

The communication interactions in the use case are shown in Figure 5.7. A large part of communication happens between local devices in the field and their applications hosted in the edge cloud. It is important to note, that devices can even be part of the dedicated network slice, if they are outside of the confined area of the confined area for which the network slice was set up (see Figure 5.6). This allows, e.g., the farmer to monitor from distance farming operations on the field via a dashboard in the remote-control center, where the progress of the operations is tracked. It is also possible to connect to one of the machines on the field from the same control center, for example to teleoperate a machine.

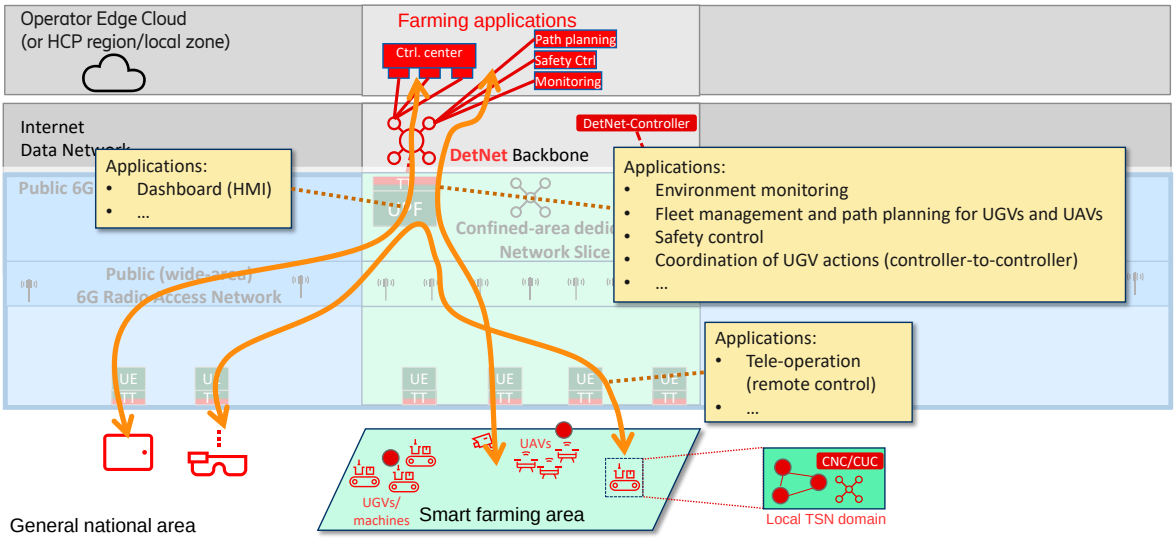

Figure 5.7: Communication relations for the smart farming use case.

Larger machines, like UGVs, have a set of functions integrated, like controllers, cameras, drives, sensors, and actuators that are inter-connected for the autonomous operation of the vehicle. To this end, a local TSN network within the vehicle can be assumed. However, the UGVs will also communicate with the applications located in the edge cloud. In this case, a function on the UGV communicates via DetNet with an application in the edge cloud, whereas the local communication within the UGV is based on TSN. The DetNet capability can be applied, where DetNet uses TSN as a subnet on a part of the end-to-end path [5GS21-D53], as shown in Figure 5.8.

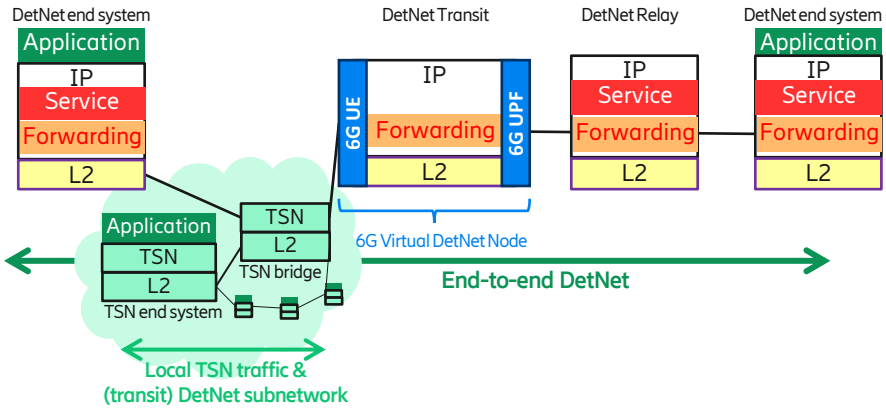

Figure 5.8: DetNet communication from a machine with a local TSN network towards the edge cloud.

In some scenarios multiple vehicles are acting as a synchronized swarm, e.g., when a UAV is used for the purpose of environment surveillance ahead of a UGV, or when a harvester is loading a trolley during the harvesting process. In such a case, communication between, for instance, controllers of the different vehicles pass through multiple TSN configuration domains, as shown in Figure 5.9. In this case also the controller-to-controller coordination within the swarm happens largely within the swarm domain (in contrast to Figure 5.7 where the controllers are in the edge cloud). It is for further work to investigate the most appropriate end-to-end communication setup for such scenarios.

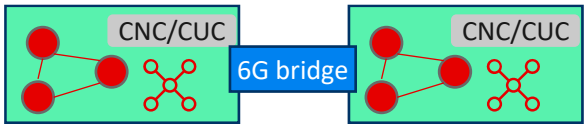

Figure 5.9: Communication across multiple TSN configuration domains, e.g., for interconnecting two vehicles within a swarm.

5.4. Use case to DETERMINISTIC6G capabilities mapping

This section indicates the benefits of the availability of specific DETERMINISTIC6G capabilities in different application scenarios at a high level. A quantitative benefit depends on multiple parameters of the individual use cases, like overall communication system load, number of involved communication endpoints, etc., as well as the specific configuration of 6G network (e.g., used spectrum and assigned radio bandwidth). The listing is not exhaustive, while the reader may find additional beneficiary usage scenarios.

Proposed DETERMINISTIC6G capabilities include:

1) Dynamic 6G network behavior based on latency monitoring and predictions
   a) Description: This capability allows to quantify the latency performance that can be provided by the 6G network to an application. It also enables to dynamically react to changes in the system and its environment that lead to degraded or improved predictions. Communication and computation resources can then be reassigned based on updated predictions.
   b) Beneficiary application scenarios
      i) **Exoskeleton functional safety:** the consequences of a 6G network degradation can be mitigated (e.g., by zeroing the interaction torque between the exoskeleton and the worker), thus avoiding potential accidents which could result from unexpected latency values.
      ii) In-swarm AGV functional safety: Timeout margins for safety-relevant communication can be reduced, which allows increasing operational performance (e.g., maximum allowed speed of AGVs).
      iii) AGV swarm-to-system functional safety: same as ii).
      iv) Combined operation between production line and MPM: Control loop cycle times can be reduced due to improved reliability of 6G communication service.
      v) Coordinated operation of UGVs and UAVs: same as iv).
2) Time synchronization reliability
   a) Description: This capability reduces the probability of time synchronization service to fail completely. Thus, also the probability of performance degradation of applications, which require the synchronized time for operation, can also be reduced.
   b) Beneficiary application scenarios
      i) **Coordination between worker's movements and exoskeleton's action:** the precise functioning of the closed-loop controller can only be guaranteed with reliable time synchronization.
      ii) Exoskeleton functional safety: same as i).
      iii) Coordinated operation of UGVs and UAVs: the validity of data usually decays with the progression of time. This is especially true for moving objects (i.e., humans, animals) that are detected by the drone. Timestamp information exchanged between the UAV (e.g., a

drone) and the UGV (e.g., a harvester) is only valuable if the time is synchronized and the synchronization is reliable.

iv) **Video based object detection on factory floor:** similar as iii).

3) Packet delay correction
   a) Description: With this capability, the packet delay variation of the wireless transmission is reduced to provide a deterministic communication latency performance. This allows to support use cases where applications are very sensitive to delay variations. It also helps to improve the TSN/DetNet traffic management, e.g., for time-scheduled transmission of different traffic streams through the network.
   b) Beneficiary application scenarios
      i) **Coordinated operation between processing cell and AGVs:** communication within high performance processing cells requires tight schedules with often extremely short cycle times. Control information exchanged between an AGV and such processing cells additionally needs to fit into the schedule. A high packet delay variation may lead to missed timeslots and reordering of packets in the queues of the network switches, which subsequently may result in schedule violations (i.e., packets are not transmitted in the scheduled time window).
      ii) **Coordination between worker's movements and exoskeleton's action:** similarly to what stated in i), also information exchanged between the exoskeleton and the edge cloud needs to fit into the expected time frame, to guarantee the proper synchronization between the assistance provided by the device and the movements performed by the worker.
      iii) **General TSN/DetNet-based use cases with tight delay bounds and high availability:** Packet delay correction allows better plannability of the latency characteristics of the 6G system which enables improved and more robust TSN traffic management. All applications that rely on robust TSN configurations benefit from this.
4) TSN/DetNet capable edge computing
   a) Description: This capability enables computing resources to be part of end-to-end and time-aware traffic scheduling, and thus provides benefits for time-critical applications.
   b) Beneficiary application scenarios
      i) **Coordination between worker's movements and exoskeleton's action:** complex assistive strategies, requiring the elaboration of huge amount of data acquired from multiple subsystems, are made feasible.
      ii) **Exoskeleton's weight/bulkiness reduction:** offloading computing functionalities allows to delocalize part of the exoskeleton's hardware components.
      iii) **Precise task recognition:** the collection and elaboration of data coming from all the environmental sensors allow to precisely recognize the characteristics of the task performed by the worker.
      iv) **AGV swarm coordination**: using edge computing technology allows to flexibly create multiple instances of swarm controllers on demand. Thereby, timely control of all AGVs in the swarm is key to prevent hazards and potential damages caused by the swarm (e.g., collisions between AGVs or with their surrounding).
      v) **Coordination of UAVs and UGVs**: similar as iv).
5) Wireless-friendly end-to-end scheduling

a) Description: This capability allows for scheduling time-sensitive real-time traffic with dependable end-to-end timing guarantees. On the one hand, this includes the capability to cope with a stochastic packet delay and high packet delay variation. Furthermore, this includes the dynamic adaptation of end-to-end schedules if the quality of the wireless medium changes significantly based on predictions of the packet delay.

b) Beneficiary application scenarios:

i) In general, any of the described scenarios since they all include end-to-end paths with wireless links and time-critical network traffic.

6) Security

a) Description: The security capabilities ensure that the DETERMINISTIC6G architectural building blocks do not introduce additional security threats.

b) Beneficiary application scenarios

i) All of the described scenarios benefit from an elaborated security architecture.

With the review of the use cases additional desirable capabilities may be identified. For example, some use cases would benefit from multicast data delivery, like when the location of identified obstacles can be sent to multiple AGV, or movement changes of one vehicle can be communicated to all other vehicles within a swarm. Support for multicast will be investigated in future work.

6. Conclusions and Future Work

Digitalization is continuing to drive use cases to increasing levels of adaptivity, embracing of enablers like cloud computing and data-driven design with ML and leading towards a cyber-physical design. This journey is already progressing and will continue well into the time frame of 6G. There will be a wide range of time-critical services with the need for high availability. In many cases, such services will have to be supported end-to-end by deterministic networking technologies, such as Ethernet TSN or IP DetNet, which will need to work seamlessly also for (sub-)systems connected wirelessly with 6G. DETERMINISTIC6G is studying several use cases with time-critical applications to explore and define corresponding 6G capabilities.

In this report we review architectural design directions and propose a first 6G architecture design that integrates functionality for dependable time-critical services as developed in DETEMRINISTIC6G. The proposed architecture integrates robust time synchronization, packet-delay control reducing large packet delay variations, supports data-driven latency prediction, integrated time-aware edge computing, and considers security-by-design principles for dependable time-critical services. It shall also enable improved interactions between applications to invoke dependable communication services. Furthermore, improved end-to-end dependable networking shall be provided, where the end-to-end traffic management applied in TSN and DetNet is optimized in conjunction with 6G wireless communication. The proposed architecture framework is reviewed with regard to DETERMINISTIC6G use cases and includes local deployments on an industrial shopfloor, but also deployments over wider areas.

In future work we will investigate the functionality for dependable time-critical services in depth and evaluate the performance of the proposed functionality. Those insights will be applied to refine the proposed 6G architecture and provide more design details. A concept validation of the architecture design will be made by investigation of the DETERMINISTIC6G use cases.

References

[3GPP16-28533] 3GPP TS 28.533, “Management and Orchestration of Networks and Network Slicing; Management and Orchestration Architecture (Release 16),” v16.0.0, Jun. 2019.

[3GPP17-23288] 3GPP TS 23.288, “Architecture enhancements for 5G system (5GS) to support network data analytics services,” v17.9.0, June 2023.

[3GPP17-37817] 3GPP TR 37.817, “Study on enhancement for Data Collection for NR and EN-DC,” v17.0.0

[3GPP17-38314] 3GPP TS 38.314, “Layer 2 measurements,” v17.4.0, January 2024.

[3GPP18-22874] 3GPP TR 22.874, “5G System (5GS); Study on traffic characteristics and performance requirements for AI/ML model transfer,” v18.2.0, Jun. 2021

[3GPP18-23501] 3GPP TS 23.501, “System Architecture for the 5G system,” v18.4.0, Dec. 2023, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[3GPP18-28552] 3GPP TS 28.552, “5G performance measurements,” v18.5.0

[3GPP19-22261] 3GPP TS 22.261, “Service requirements for the 5G system,” v19.5.0, March 2024.

[3GPP19-22876] 3GPP TR 22.876, “Study on AI/ML Model Transfer Phase2”, 3GPP TR 22.876 v19.0.0, Jun. 2023.

[3GPP23-23434] 3GPP TS 23.434, ”Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows”, December 2023, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[3GPP23-23548] 3GPP TS 23.548, ”5G System Enhancements for Edge Computing; Stage 2”, technical specification, April 2023, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3856

[3GPP23-23558] 3GPP TS 23.558, ”Architecture for enabling Edge Applications”, technical specification, March 2023, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3723

[3GPP23-29522] 3GPP TS 29.522, ”5G System; Network Exposure Function Northbound APIs; Stage 3”, technical specification, December 2023, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3437

[5GAC23] 5G-ACIA, “Industrial 5G Edge Computing – Use Cases, Architecture and Deployment”, white paper, February 2023, https://5g-acia.org/whitepapers/industrial-5g-edge-computing-use-cases-architecture-and-deployment/

[5GAC21a] 5G-ACIA, "Exposure of 5G Capabilities for Connected Industries and Automation Applications," white paper, Feb. 2021, https://5g-

acia.org/whitepapers/exposure-of-5g-capabilities-for-connected-industries-and-automation-applications-2/

[5GAC21b] 5G-ACIA, “5G non-public networks for industrial scnearios,” white paper, September 2021, https://5g-acia.org/download/18734/?tmstv=1706108824

[5GAC24] 5G-ACIA, “NPNs for Industrial Scenarios,” white paper, March 2024, https://5g-acia.org/download/18718/?version=a4

[5GAC21c] 5G-ACIA, “Integration of 5G with Time-Sensitive Networking for Industrial Communications”, white paper, February 2021, https://5g-acia.org/whitepapers/exposure-of-5g-capabilities-for-connected-industries-and-automation-applications-2/

[5GAC21d] 5G-ACIA, “5G QoS for Industrial Automation”, white paper, November 2021, https://5g-acia.org/whitepapers/5g-quality-of-service-for-industrial-automation-2/

[5GS20-D14] 5G-SMART Deliverable 1.4, ”Report describing the framework for 5G system and network management functions”, November 2020, https://5gsmart.eu/deliverables/

[5GS21-D15] 5G-SMART Deliverable 1.5, ”Evaluation of radio network deployment options”, November 2021, https://5gsmart.eu/deliverables/

[5GS20-D51] 5G-SMART Deliverable 5.1, ”First report on new technological features to be supported by 5g standardization and their Implementation impact”, May 2020, https://5gsmart.eu/deliverables/

[5GS20-D52] 5G-SMART Deliverable 5.2, ”First Report on 5g network architecture options and assessments”, November 2020, https://5gsmart.eu/deliverables/

[5GS21-D54] 5G-SMART Deliverable 5.4, ”Second report on 5g network architecture options and assessments”, November 2021, https://5gsmart.eu/deliverables/

[5GS21-D53] 5G-SMART Deliverable 5.3, ”Second report on new technological features to be supported by 5g standardization,” November 2021, https://5gsmart.eu/deliverables/

[5GS21-D55] 5G-SMART Deliverable 5.5, ” Framework for 5g system and network management functions,” November 2021, https://5gsmart.eu/deliverables/

[ADF+23] F. Alriksson, E. Drugge, A. Furuskär, D.H. Kang, J. Kronander, J.L.Pradas, Y.Sun, “Future network requirements for extended reality use cases”, Ericsson Technology review, April 2023, https://www.ericsson.com/496150/assets/local/reports-papers/ericsson-technology-review/docs/2023/future-network-requirements-for-xr-apps.pdf

[AF20] Arjoune, Y. and Faruque, S., 2020, January. Smart jamming attacks in 5G new radio: A review. In 2020 10th annual computing and communication workshop and conference (CCWC) (pp. 1010-1015). IEEE.

[AKP+21] F. Alriksson, D.H. Kang, C. Phillips, J.L. Pradas, A. Zaidi, ”XR and 5G: Extended reality at scale with time-critical communication,” n *Ericsson Technology Review*, vol. 2021, no. 8, pp. 2-13, August 2021, doi: 10.23919/ETR.2021.9904681, https://ieeexplore.ieee.org/document/9904681

[BBW+23] C.G. Blázquez, A. Balador, A. Williams, A, Hata, “Offloading for the future: current use cases and scenarios,” Ericsson blog, August 24, 2023, https://www.ericsson.com/en/blog/2023/8/computational-offloading-for-future-innovations

[BCAMM+18] B. Moussa, C. Robillard, A. Zugenmaier, M. Kassouf, M. Debbabi, and C. Assi, “Securing the Precision Time Protocol (PTP) Against Fake Timestamps,” IEEE Commun. Lett., vol. 23, no. 2, pp. 278–281, 2018.

[CAM24a] CAMARA, “Network Slice Boooking,” draft API, accessed April 17, 2024, https://github.com/camaraproject/NetworkSliceBooking/tree/main/documentation

[CAM24b] CAMARA, “Quality on Demand,” draft API, accessed April 17, 2024, https://github.com/camaraproject/QualityOnDemand_PI1/tree/main

[CAS+22] J.B. Caro, J. Ansari, J. Sachs, P. de Bruin, S. Sivri, L. Grosjean, N. König, R. H. Schmitt, "Empirical Study on 5G NR Cochannel Coexistence" Electronics 11, no. 11: 1676. May 2022, https://doi.org/10.3390/electronics11111676

[CAS+23] J. B. Caro, J. Ansari, A.R. Sayyed, P. de Bruin, J. Sachs, N. König, R.H. Schmitt, "Empirical study on 5G NR Adjacent Channel Coexistence," 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom, 2023, pp. 1-6, doi: 10.1109/WCNC55385.2023.10119074.

[CD13] Cavoukian, A. and Dixon, M., 2013. *Privacy and security by design: An enterprise architecture approach*. Information and Privacy Commissioner of Ontario, Canada.

[CMRV+23] T. Cagenius, G. Mildh, G. Rune, J. Vikberg, M. Wahlqvist, P. Willars, “6G network architecture – a proposal for early alignment”, in *Ericsson Technology Review*, vol. 2023, no. 11, pp. 2-7, October 2023, doi: 10.23919/ETR.2023.10313589, https://ieeexplore.ieee.org/document/10313589

[CRVW18] T. Cagenius, A. Ryde, J. Vikberg, P. Willars, “Simplifying the 5G ecosystem by reducing architecture options”, in *Ericsson Technology Review*, November 2018, https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/simplifying-the-5g-ecosystem-by-reducing-architecture-options

[DCJ+21] D. Ergenc, C. Brulhart, J. Neumann, L. Kruger, and M. Fischer, “On the Security of IEEE 802.1 Time-Sensitive Networking,” 2021 IEEE Int. Conf. Commun. Work. ICC Work. 2021 - Proc., 2021.

[DET23-D11] DETERMINISTIC6G, Deliverable 1.1, “DETERMINISTIC6G use cases and architecture principles,” Jun. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D21] DETERMINISTIC6G, Deliverable 2.1, “First report on 6G centric enablers”, Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D22] DETERMINISTIC6G, Deliverable 2.2, “First Report on the time synchronization for E2E time awareness,” Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D31] DETERMINISTIC6G, Deliverable 3.1, “Report on 6G convergence enablers towards deterministic communication standards,” Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D32] DETERMINISTIC6G, Deliverable 3.2, “Report on the Security solutions,” Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D41] DETERMINISTIC6G, Deliverable 4.1, “DETERMINISTIC6G DetCom simulator framework release 1,” Dec. 2023, https://deterministic6g.eu/index.php/library-m/deliverables

[DET23-D42] DETERMINISTIC6G, Deliverable 4.2, “Latency measurement framework,” March 2024, https://deterministic6g.eu/index.php/library-m/deliverables

[DetNet-Sec] E. E. Grossman, T. Mizrahi, and A. Hacker, RFC 9055 - Deterministic Networking (DetNet) Security Considerations, 2021.

[ECAT] EtherCAT Technology Group, EtherCAT Specification, https://www.ethercat.org/en/downloads/downloads_A02E436C7A97479F9261FDFA8A6D71E5.htm

[Eri23] Ericsson, “Future Network Architecture”, Ericsson White Paper, April 2023, available at https://wcm.ericsson.net/en/future-technologies/architecture

[FBZ+21] F. Luo, B. Wang, Z. Fang, Z. Yang, and Y. Jiang, "Security Analysis of the TSN Backbone Architecture and Anomaly Detection System Design Based on IEEE 802.1Qci", Secur. Commun. Networks, vol. 2021, 2021.

[FHB+24] F. Pedersen, R. Högman, M. Buchmayer, A. Zaidi, “5G spectrum for local industrial networks,” Ericsson white paper, April 2024, https://www.ericsson.com/en/reports-and-papers/white-papers/5g-spectrum-for-local-industrial-networks

[FMK23] J. Friman, E. Mueller and B. van Kaathoven, "Monetizing API exposure for enterprises with evolved BSS," in Ericsson Technology Review, vol. 2023, no. 1, pp. 2-10, January 2023, doi: 10.23919/ETR.2023.10035875. https://ieeexplore.ieee.org/document/10035875

[GLS+22] L. Grosjean, K. Landernäs, B. Sayrac, O. Dobrijevic, N. König, D. Harutyunyan, D. Patel, J.F. Monserrat, J. Sachs, “5G-enabled smart manufacturing – a booklet of 5G-SMART,” technical report, September 2022, https://arxiv.org/abs/2209.10300

[GSA24] Global mobile Suppliers Association (GSA), “5G - 5G-Standalone January 2024 Member Report,” January 2024, https://gsacom.com/paper/5g-standalone-january-2024-member-report/

[GSD+22] G. Seres, D. Schulz, O. Dobrijevic, A. Karaağaç, H. Przybysz, A. Nazari, P. Chen, M.L. Mikecz, Á.D. Szabó, "Creating programmable 5G systems for the Industrial

IoT," Ericsson Technology Review, vol. 2022, no. 10, pp. 2-12, October 2022, doi: 10.23919/ETR.2022.9934828. https://ieeexplore.ieee.org/document/9934828

[GSS+21] Michael Gundall et al. [Gundall, M., Strufe, M., Schotten, H.D., Rost, P., Markwart, C., Blunk, R., Neumann, A., Grießbach, J., Aleksy, M. and Wübben, D., 2021. Introduction of a 5G-enabled architecture for the realization of industry 4.0 use cases. *IEEE access*, *9*, pp.25508-25521.

[GSMA18] GSMA, "Road to 5G: Introduction and Migration", report, April 2018, https://www.gsma.com/futurenetworks/wp-content/uploads/2018/04/Road-to-5G-Introduction-and-Migration_FINAL.pdf

[ETSI-ZSM] ETSI, Zero touch network & Service Management (ZSM), https://www.etsi.org/technologies/zero-touch-network-service-management (accessed April 2024)

[GCC+23] V. Gudepu, V. R. Chintapalli, P. Castoldi, L. Valcarenghi, B. R. Tamma and K. Kondepu, "Adaptive Retraining of AI/ML Model for Beyond 5G Networks: A Predictive Approach," 2023 IEEE 9th International Conference on Network Softwarization (NetSoft), Madrid, Spain, 2023, pp. 282-286, doi: 10.1109/NetSoft57336.2023.10175451.

[GLR+20] I. Godor, M. Luvisotto, S. Ruffini, K. Wang, D. Patel, J. Sachs, O. Dobrijevic, D. P. Venmani, O. Le Moult, J. Costa-Requena, A. Poutanen, C. Marshall, J. Farkas, "A Look Inside 5G Standards to Support Time Synchronization for Smart Manufacturing," in *IEEE Communications Standards Magazine*, vol. 4, no. 3, pp. 14-21, September 2020, doi: 10.1109/MCOMSTD.001.2000010. https://ieeexplore.ieee.org/document/9204594

[HEX2-D21] HEXA-X-II, "Initial Architectural enablers," deliverable D3.2, October 2023, , https://hexa-x-ii.eu/results/

[HEX-D14] HEXA-X, "Hexa-X architecture for B5G/6G networks – final release," deliverable D1.4, July 2023, https://hexa-x.eu/deliverables/

[IEEEQcc] "IEEE Standard for Local and Metropolitan Area Networks--Bridges and Bridged Networks -- Amendment 31: Stream Reservation Protocol (SRP) Enhancements and Performance Improvements," in IEEE Std 802.1Qcc-2018 (Amendment to IEEE Std 802.1Q-2018 as amended by IEEE Std 802.1Qcp-2018) , vol., no., pp.1-208, 31 Oct. 2018, doi: 10.1109/IEEESTD.2018.8514112. https://ieeexplore.ieee.org/document/8514112

[IEEEQdj] "IEEE Draft Standard for Local and Metropolitan Area Networks--Bridges and Bridged Networks -- Amendment 38: Configuration Enhancements for Time-Sensitive Networking," in *IEEE P802.1Qdj/D2.0, November 2023* , vol., no., pp.1-49, 20 Nov. 2023.. https://ieeexplore.ieee.org/document/10326153

[IEEE-TSN] IEEE 802.1 Time-Sensitive Networking (TSN) Task Group, https://1.ieee802.org/tsn/ (accessed April 2024)

[IETF-DETNET] IETF Deterministic Networking Working Group, https://datatracker.ietf.org/wg/detnet/about/ (accessed April 2024)

[IJR+23] M. Iovene, L. Jonsson, D. Roeland, M. D'Angelo, G. Hall, M. Erol-Kantarci, "Defning AI native: A key enabler for advanced intelligent telecom networks," Ericsson white paper, February 2023, https://www.ericsson.com/en/reports-and-papers/white-papers/ai-native

[ITU23] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond," recommendation ITU-R M.2160-0, November 2023, https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2030/Pages/default.aspx

[KDS+23] A. Karaagac, O. Dobrijevic, D. Schulz, G. Seres, A. Nazari, H. Przybysz, J. Sachs, "Managing 5G Non-Public Networks from Industrial Automation Systems", 2023 IEEE 19th International Conference on Factory Communication Systems (WFCS), Pavia, Italy, 2023, pp. 1-8, doi: 10.1109/WFCS57264.2023.10144248. https://ieeexplore.ieee.org/document/10144248

[KLY+13] Kang, T., Li, X., Yu, C. and Kim, J., 2013. A survey of security mechanisms with direct sequence spread spectrum signals. Journal of Computing Science and Engineering, 7(3), pp.187-197.

[LGP+24] D.C. Larsson, A. Grövlen, S. Parkvall, O. Liberg, "6G standardization – an overview of timeline and high-level technology principles," Ericsson blog, March 22, 2024, https://www.ericsson.com/en/blog/2024/3/6g-standardization-timeline-and-technology-principles

[LRM+18] Lichtman, M., Rao, R., Marojevic, V., Reed, J. and Jover, R.P., 2018, May. 5G NR jamming, spoofing, and sniffing: Threat assessment and mitigation. In 2018 IEEE international conference on communications workshops (ICC Workshops) (pp. 1-6). IEEE.

[MAG+19] A. Mahmood, M. I. Ashraf, M. Gidlund, J. Torsner and J. Sachs, "Time Synchronization in 5G Wireless Edge: Requirements and Solutions for Critical-MTC," in IEEE Communications Magazine, vol. 57, no. 12, pp. 45-51, December 2019, doi: 10.1109/MCOM.001.1900379. https://ieeexplore.ieee.org/document/8930825

[MAD20] Monica L, Anastasi S and Draicchio F, "Occupational exoskeletons: Wearable robotic devices and preventing work-related musculoskeletal disorders in the workplace of the future", pp. 1–12., September 2020, https://osha.europa.eu/en/publications/occupational-exoskeletons-wearable-robotic-devices-and-preventing-work-related

[MCP+23] Mitev, M., Chorti, A., Poor, H.V. and Fettweis, G.P., 2023. What physical layer security can do for 6G security. IEEE Open Journal of Vehicular Technology, 4, pp.375-388.

[MMT-Probe] https://github.com/Montimage/mmt-probe

[MTS+23] S. Mostafavi, M. Tillner, G.P. Sharma and J. Gross. "EDAF: An End-to-End Delay Analytics Framework for 5G-and-Beyond Networks.", Jan. 2024, arXiv preprint arXiv:2401.09856.

[NLC+21] Nguyen, V.L., Lin, P.C., Cheng, B.C., Hwang, R.H. and Lin, Y.D., 2021. Security and privacy for 6G: A survey on prospective technologies and challenges. IEEE Communications Surveys & Tutorials, 23(4), pp.2384-2428.

[P4-INT] P4.org Applications Working Group. Contributions from Alibaba, Arista, CableLabs, Cisco Systems, Dell, Intel, Marvell, Netronome, VMware. In-band Network Telemetry (INT) Dataplane Specification. Version 2.1. 2020-11-11 https://p4.org/p4-spec/docs/INT_v2_1.pdf

[PDR+21] D. Patel, J. Diachina, S. Ruffini, M. De Andrade, J. Sachs and D. P. Venmani, "Time error analysis of 5G time synchronization solutions for time aware industrial networks", 2021 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS), NA, FL, USA, 2021, pp. 1-6, doi: 10.1109/ISPCS49990.2021.9615318. https://ieeexplore.ieee.org/document/9615318

[Per22] A.P. Perales, "INSPIRE-5Gplus: Vision on Security Beyond 5G," presentation, ETSI Security Conference 2022, 3-5 October 2022, https://www.etsi.org/events/2068-etsi-security-conference#pane-1/

[PLNK] EPSG Draft Standard 301, Ethernet POWERLINK Communication Profile Specification Version 1.5.1, 2023,https://www.br-automation.com/downloads_br_productcatalogue/assets/EPSG_301_V-1-5-1_DS-c710608e.pdf

[RFC 6241] IETF, Network Configuration Protocol (NETCONF), Request for Comments 6241, June 2011, https://datatracker.ietf.org/doc/html/rfc6241

[RFC 8040] IETF, RESTCONF Protocol, Request for Comments 8040, January 2017, https://datatracker.ietf.org/doc/html/rfc8040

[RFC 7384] T. Mizrahi, "RFC 7384 - Security Requirements of Time Protocols in Packet Switched Networks", 2014.

[RJS+23] R Robert, W. John, J. Sjöberg, J. Halén, "Opportunities with dynamic device offloading as a 6G service," Ericsson blog, September 07, 2023. https://www.ericsson.com/en/blog/2023/9/dynamic-device-offloading-as-a-6g-service

[Roe20] Dinand Roeland, "An introduction to data-driven network architecture," Ericsson blog, October 06, 2020. https://www.ericsson.com/en/blog/2020/10/data-driven-network-architecture

[Roe20] D. Roeland, "An introduction to data-driven network architecture", October 2020, https://www.ericsson.com/en/blog/2020/10/data-driven-network-architecture

[RÖT23] G. Rune, P. Öhlén, Z. Turányi, G. Mildh "Six talking points for architecting the next wireless generation", Ericsson blog, September 13 2023, https://wcm.ericsson.net/en/blog/2023/9/six-talking-points-future-network-architecture

[Sol21] Soldani, D., 2021. 6G fundamentals: Vision and enabling technologies. *Journal of Telecommunications and the Digital Economy*, *9*(3), pp.58-86. and Soldani, D. "From security-enhanced 5G networks to security-by-design 6G systems"; https://pubhtml5.com/uifs/ztur/CYBER_DEFENSE_EMAGAZINE_FOR_AUGUST_2021/83]

[SAR+23] M. Saimler, M. D'Angelo, D. Roeland, A. Ahmed, A. Kattepur, "AI as a service: How AI applications can benefit from the network," Ericsson blog, December 14, 2023. https://www.ericsson.com/en/blog/2023/12/ai-as-a-service

[SK23] D. Schulz, A. Karaagac, "Cutting the cables," ABB review, January 2023, https://new.abb.com/news/detail/99141/cutting-the-cables

[SKM+21] M. Svensson, B. Kovács, E. Mueller, M. Maggiari and R. Szabó, "Service exposure and automated life-cycle management: The Key Enablers for 5G Services," in Ericsson Technology Review, vol. 2021, no. 13, pp. 2-12, December 2021, doi: 10.23919/ETR.2021.9904697 https://ieeexplore.ieee.org/document/9904697

[SKT22] Sharma, H., Kumar, N. and Tekchandani, R., 2022. Physical layer security using beamforming techniques for 5G and beyond networks: A systematic review. Physical Communication, 54, p.101791.

[SPS+23] G. P. Sharma, D. Patel, J. Sachs, M. De Andrade, J. Farkas, J. Harmatos, B. Varga, H. -P., Bernhard, R. Muzaffar, M. Ahmed, F. Duerr, D. Bruckner, E.M. De Oca, D. Houatra, H. Zhang and J. Gross, "Toward Deterministic Communications in 6G Networks: State of the Art, Open Challenges and the Way Forward," in IEEE Access, vol. 11, pp. 106898-106923, 2023, doi: 10.1109/ACCESS.2023.3316605.

[WMs+21] W. Alghamdi and M. Schukat, "Precision time protocol attack strategies and their resistance to existing security extensions," Cybersecurity, vol. 4, no. 1, 2021.

[ZSM PoC] "Security SLA assurance in 5G network slices," https://zsmwiki.etsi.org/index.php?title=PoC_6_Security_SLA_assurance_in_5G_network_slices (accesses April 2024)

List of abbreviations

5G NSA	5G Non-Standalone
5G SA	5G Standalone
3GPP	3rd Generation Partnership Project
5G-ACIA	5G Alliance for Connected Industries and Automation
5G-Adv	5G-Advanced
5GS	5G System
ABE	Attribute-Based Encryption
AC	Application Context [Client]
ACR	Application Context Relocation
ACT	Application Context Transfer
AES	Advanced Encryption Standard
AF	Application Function
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AI-aaS	Artificial Intelligence as a Service
AMF	Access & Mobility Management Function
AnLF	Analytics Logical Function
API	Application Programming Interface
AR	Augmented Reality
AV	Aerial Vehicle
BTCA	Best timeTransmitter Clock Algorithm
CAMARA	*the Telco Global API Alliance*
CBC	Cipher Block Chaining
CBS	Constant Bandwidth Server
CN	Core Network
CNC	Centralized Network Configuration
CNI	Container Network Interface
CPS	Cyber-Physical System
CPU	Central Processing Unit
CQF	Cyclic Queuing and Forwarding
CQI	Channel Quality Indicator
CTI	Cyber Threat Intelligence
CUC	Centralized User Configuration
DCCF	Data Collection Coordination Function
DetNet	Deterministic Networking
DL	Downlink
DoS	Denial of Service

DS-TT	Device-Side Time Sensitive Networking Translator
DTLS	Datagram Transport Layer Security
E2E	End-to-End
EAS	Edge Application Server
EASDF	Edge Application Server Discovery Function
ECC	Elliptic Curve Cryptography
ECS	Edge Configuration Server
EDF	Earliest Deadline First
EDN	Edge Data Network
EEC	Edge Enabler Client
EES	Edge Enabler Server
ERP	Enterprise Resource Planning
ETSI	European Telecommunications Standards Institute
EV	Electric Vehicle
FL	Federated Learning
FPGA	Field Programmable Gate Array
FRER	Frame Replication and Elimination for Reliability
FSM	Finite State Machine
FX	Field eXchange
GCM	Galois/Counter Mode
GM	Grand Master
gNB	Next Generation Node B
GPP	General Purpose Processor
gPTP	Generic Precision Time Protocol
GPU	Graphics Processing Unit
GV	Ground Vehicle
HARQ	Hybrid Automatic Repeat Request
HMD	Head Mounted Device
HMI	Human Machine Interface
HVAC	Heating Ventilation and Air Conditioning
IaaS	Infrastructure as a Service
ICT	Information and Communication Technology
ICMT	Internet Control Message Protocol
IDS	Intrusion Detection Systems
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMT-2030	International Mobile Telecommunications towards 2030 and beyond
IMU	Inertial Measurement Unit
INT	In-band Network Telemetry

IoT	Internet of Things
IIoT	Industrial Internet of Things
IP	Internet Protocol
ITU	International Telecommunication Union
KPI	Key Performance Indicator
KV	Key Value
KVI	Key Societal Value Indicator
LAN	Local Area Network
LCD	Liquid Crystal Display
LED	Light-Emitting Diode
LIDAR	Light Detection and Ranging
M2M	Machine to Machine
MDAF	Management Data Analytics Function
MEC	Multi-Access Edge Computing
MITM	Man-in-the-middle
ML	Machine Learning
MMT	Montimage Monitoring Tool
MNO	Mobile Network Operator
MPM	Mobile Processing Module
MR	Mobile Robot
MTBF	Mean Time Between Failures
MTLF	Model Training Logical Function
MTP	Motion-to-Photon
MTU	Maximum Transmission Unit
NEF	Network Exposure Function
NETCONF	Network Configuration Protocol
NF	Network Function
NFV	Network Function Virtualization
NIC	Network Interface Card
NIST	National Institute of Standards and Technology
NPN	Non-public Networks
NSA	Non-Standalone
NWDAF	Network Data Analytics Function
NW-TT	Network-side Time Sensitive Networking Translator
OAM	Operations, Administration, and Maintenance
OE	Occupational Exoskeleton
ONAP	Open Network Automation Platform
OPC UA	Open Platform Communications Unified Architecture
OS	Operating System

OSI	Open Systems Interconnection
OSM	Open Source Management and Orchestration
OVS	Open vSwitch
PaaS	Platform as a Service
PDC	Packet Delay Correction
PDCP	Packet Data Convergence Protocol
PDV	Packet Delay Variation
PLC	Programmable Logic Controller
PN	Public Network
PNI-NPN	Public network integrated NPN
PoC	Proof of Concept
PSK	Pre-Shared Key
PTP	Precision Time Protocol
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RESTCONF	Representational State Transfer Configuration Protocol
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTOS	Real-time Operating System
RT-SSLA	Real-Time Security Service Level Agreement
SA	Stand Alone
SAE	Security Analytics Engine
SDG	Sustainability Development Goal
SDN	Software-Defined Networking
SEAL	Service Enabler Architecture Layer for Verticals
SLA	Service Level Agreement
SMD	Security Management Domain
SMF	Session Management Function
SNPN	Standalone NPN
SSLA	Security Service Level Agreement
TAPRIO	TSN Time-aware Priority Shaper
TLS	Transport Layer Security
TRL	Technology Readiness Level
TSC	Time-Sensitive Communication
TSCTSF	Time-Sensitive Communication and Time Synchronization Function
TSN	Time-Sensitive Networking
TSN AF	Time-Sensitive Networking Application Function

UAV	Unmanned Aerial Vehicle
UCS	Use Case Scenario
UE	User Equipment
UGV	Unmanned Ground Vehicle
UL	Uplink
UPF	Uer Plane Function
URLLC	Ultra Reliable and Low Latency Communications
VM	Virtual Machine
VNF	Virtual Network Function
VNP	Virtual Private Network
VR	Virtual Reality
WP	Work Package
VRT	Variable Rate Treatment
WRMD	Work-related Musculoskeletal Disorders
XR	Extended Reality
YANG	Yet Another Next Generation data modelling language
ZSM	Zero-touch network and Service Management
ZTN	Zero-Trust Networking