NEXT G ALLIANCE
An ATIS Initiative

6G

Next G Alliance Report:
Management and Orchestration

FOREWORD

As a leading technology and solutions development organization, the Alliance for Telecommunications Industry Solutions (ATIS) brings together the top global ICT companies to advance the industry's business priorities. ATIS' 150 member companies are currently working to address network reliability, 5G, robocall mitigation, smart cities, artificial intelligence (AI)-enabled networks, distributed ledger/blockchain technology, cybersecurity, IoT, emergency services, quality of service, billing support, operations and much more. These priorities follow a fast-track development lifecycle from design and innovation through standards, specifications, requirements, business use cases, software toolkits, open-source solutions, and interoperability testing.

ATIS is accredited by the American National Standards Institute (ANSI). ATIS is the North American Organizational Partner for the 3rd Generation Partnership Project (3GPP), a founding Partner of the oneM2M global initiative, a member of the International Telecommunication Union (ITU), as well as a member of the Inter-American Telecommunication Commission (CITEL).

For more information, visit www.atis.org. Follow ATIS on Twitter and on LinkedIn.

The ATIS 'Next G Alliance' is an initiative to advance North American wireless technology leadership over the next decade through private-sector-led efforts. With a strong emphasis on technology commercialization, the work will encompass the full lifecycle of research and development, manufacturing, standardization, and market readiness.

TABLE OF CONTENTS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

6G systems are expected to address a wider range of use cases and markets than their predecessors. This flexibility and capability come at the cost of increasing the number of features, greater programmability, and more dynamic configurability. Accordingly, the concepts of Management and Orchestration (M&O) must expand to ensure that the system operates efficiently and reliably, while at the same time meeting the performance and regulatory requirements of a varied and dynamic environment.

This report identifies nine areas that are key to ensuring that 6G M&O will meet the challenges of managing 6G systems in all its flavors. The report explores the issues and challenges in each of these areas. The areas are:

- Intelligent and Autonomous System Framework
- AI/ML Operations and Enablement
- System Robustness and Resilience
- Knowledge and Semantics Management
- Data Management
- Cloud M&O
- 6G System Digital Twin
- Energy Efficiency
- Security and Privacy

1 INTRODUCTION

Next G Alliance's *6G Technologies Report* [1] includes a high-level overview of 6G network Operations, Administration, and Management (OAM) and service enablement (see Section 5). It covers the areas of service M&O, data management and AI/ML-based intelligent network controllers for automation, public safety in emergencies and disaster scenarios, technology enablers for business services transformation, and energy-efficient green networks.

This paper gives an in-depth analysis of the key aspects and challenges to enable the intelligent and autonomous 6G system M&O to achieve the cognitive, energy-efficient, secure, robust, and resilient converged communication-computing 6G system. This is elaborated through a service framework that embraces AI/ML technologies, data management, cloud system M&O, system digitalization, and knowledge management for 6G systems.

2 KEY TECHNOLOGIES FOR INTELLIGENT AND AUTONOMOUS 6G SYSTEM M&O

The increasing levels of system-wide complexity, which follows a continuing evolution of next-generation systems and networks, demands a congruent evolution of system-wide embedded intelligence to automatically and self-adaptively manage complexity, while sustaining an intended suite of performance objectives for both connectivity and service. Along these directions, the emergence of complexity is propelled by a diversity of stakeholders in the evolving mobile ecosystem, with distributed and decentralized heterogeneous networks.

The elements of growing system-wide complexity reflect increasing levels of collaboration and interdependence. These occur across the protocol stack layers to support the demands of service sophistication and evolution offered by emerging mobile communication networking, computing, and storage systems.

An effective management of system-wide complexity optimizes various attributes including performance, operations, capital expenditures, converged access, flexibility, scalability, sustainability, privacy, security, customizable service experience, etc. This requires a system-wide adoption of autonomic principles, realized through embedded and distributed intelligence. The nature of this intelligence is characterized by an appropriate application of Artificial Intelligence (AI) and Machine Learning (ML) to enhance and optimize system-wide network operations in a cognitive and self-adaptive manner, where the performance expectations exceed human response limits. The AI/ML capabilities are exhibited through a variety of algorithms in concert with feedback control loops. They constitute an autonomous system, which has self-adaptive or self-Configuring, Healing, Optimizing, and Protecting (self-CHOP) characteristics [2]. These are intrinsic aspects of an emerging M&O framework, with system-wide scope. Optimization of system-wide attributes classified in terms of quantitative and qualitative measures is an integral aspect of delineating the benefits of emerging system-wide M&O capabilities.

An intelligent M&O framework imbued with the autonomic principles of self-CHOP behaviors intrinsically contains the requisite AI/ML modalities for complexity management. It also satisfies the relevant system-wide key performance indicators (KPIs), Key Value Indicators (KVIs) [3], and attractive human interfaces. Anomaly detection and prevision, together with intent-based networking, are among the capabilities for intelligent M&O for augmenting system-wide stability while effectively managing scalability.

Key considerations and technologies for the realization of an intelligent and autonomous M&O subsystem need to be examined to manage system-wide complexity accrual in 6G without operational human intervention.

2.1 Intelligent and Autonomous System Framework

Autonomous 6G M&O will provide the capabilities to automate the management of system-wide complexity across an end-to-end, heterogeneous, mobile network. It also will optimize operations to dynamically adapt to the system and service KPIs (quantitative measures) and KVIs (qualitative measures) in terms of safety, safety, sustainability, anonymity, scalability, reliability, and resilience.

2.1.1 Overview

6G M&O is expected to continue reaching higher levels of intelligence and autonomy, which are leveraged for self-CHOP of M&O of operations. Although AI/ML is a key tool for intelligent and autonomous systems, it is a prominent component of building intelligent and autonomous systems, which must also consider where and how AI/ML is used. This section explores the framework for such a system.

The operational efficiency of M&O in 6G is expected to be enhanced through the use of AI/ML-enabled management capabilities for resource and network function management in an end-to-end and multi-access mobile and fixed wireless network system. A system-wide embedding of AI/ML technologies, together with fast and slow feedback control loops, between a multi-access mobile and fixed wireless network system and its environment, constitutes an autonomous system. The 6G M&O subsystem needs to be capable of facilitating, enabling, and managing the 6G system's AI/ML operations. It must control and ensure the effect of the AI/ML is to embrace responsible AI for 6G.

In cloud-native and compute-native 6G systems, an autonomous M&O subsystem will provide an automated and cognitive Life Cycle Management (LCM) of distributed Network Functions (NFs). These NFs need to be portable and may be dynamically deployed at suitable locations in the system to suit the appropriate system performance requirements, load or fault conditions, availability of resources, anomaly detection, anomaly prevision, and other system integrity or intent conditions.

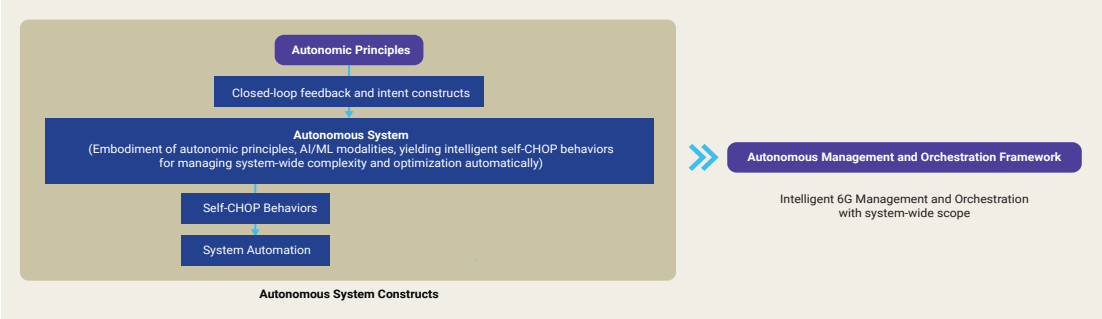


Figure 1 - Intelligent M&O framework context

The system resource allocation and scaling, such as for a Network Slice (NS) or an NF, is automated by an autonomous M&O process.

In 6G, data is essential for enabling intelligence for the whole system, including M&O, and Core Network and Radio Access Network (RAN). The management data Performance Metrics (PMs)/KPIs, alarms, configuration data, MDT measurements, Quality of Experience (QoE) data, logs, etc. and events (Trace) will be used to support the intelligence features not for M&O and for the networks. Therefore the M&O subsystem needs to support efficient data management, which in turn can provide on-demand data services for the various consumers.

With the development of digital twin technologies, it is envisioned that more and more management capabilities (e.g., emulation) can be actualized through the system's digital twin. Digital twin will be a new enabler for M&O of the 6G system.

To achieve the intelligence and autonomy of the mobile system, it is significant to not only have the data but also to know how to use it. The knowledge is the semantics abstracted from data, or the methodology summarized from experience. It is generated by one producer could be valuable and reusable for other consumers. Knowledge management and sharing will be a new area that can be investigated to lift the 6G system's level of intelligence and autonomy.

2.1.2 Key Machine Learning Paradigms and Techniques for Network Management

In the realm of autonomous M&O in the context of 6G networks, AI/ML will play an indispensable role. There are various categories of ML models, and each fits some specific types of use cases. The two distinct and fundamental approaches, generative AI and discriminative AI, are suitable for different use cases. Generative AI is used to generate new data based on the input data, while discriminative AI is suitable to classify the existing data.

Furthermore, different learning methods have been developed under each AI category, including supervised learning, unsupervised learning, and reinforcement learning. These learning methods can be applied to specific use cases. The ML has further evolved to achieve better performance and applicability with the paradigms including Federated Learning (FL), Transfer Learning (TL), and Automated ML (AutoML). These AI/ML techniques can be utilized for their strategic significance in addressing critical challenges and enhancing operational efficiency in autonomous M&O.

2.1.3 Architectural Framework Considerations

The 6G system will employ a foundational framework for sophisticated realizations of self-adaptive behavior through autonomous M&O. System-wide automation will allow continuous system improvement and situational operational optimization, with minimized human involvement.

It is anticipated that the traditional Telecommunications Management Network (TMN) layered compartments of functionality and interconnection will be transformed into cross-domain collaborative management functions/ capabilities that interact via management services. This transformation paradigm will also be reflected in the architectural considerations for an autonomous M&O subsystem to satisfy the emerging long-tailed distribution of demands and requirements.

The various functional benefits are offered by an autonomous M&O subsystem, which embraces self-CHOP processes built upon AI/ML operations, data management, cloud M&O, and Digital Twin, and are leveraged with the knowledge management.

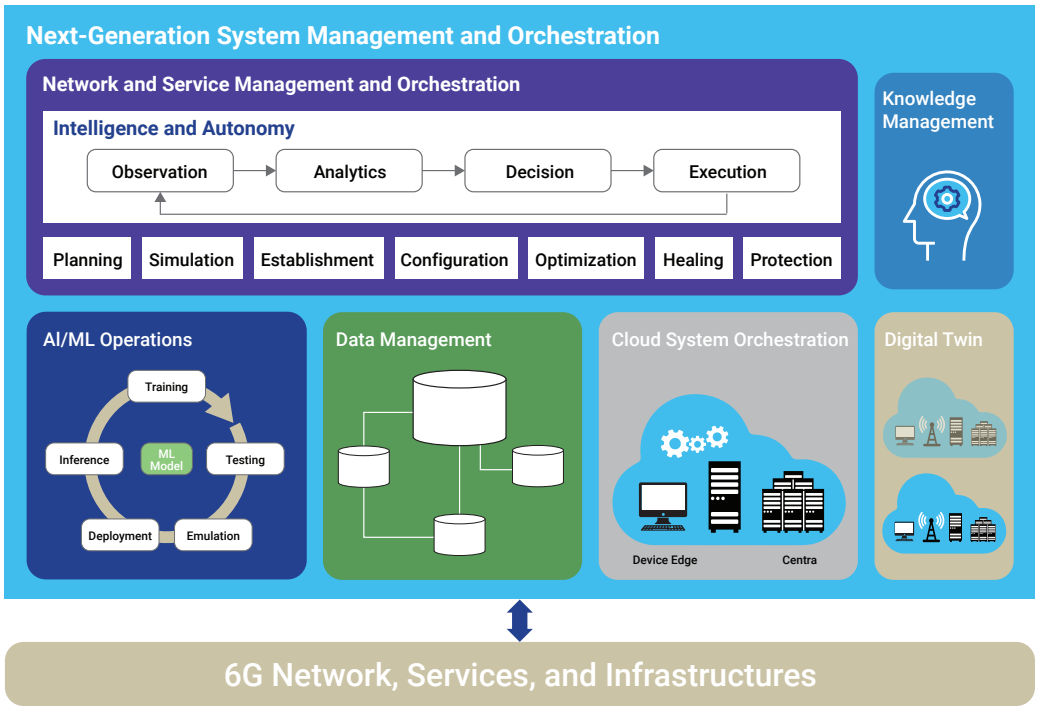


Figure 2 - 6G system M&O framework

A summary of the broad functional benefits of an autonomous M&O includes the following:

- Automation of data-oriented tasks in the network for zero-touch automation.
- Dynamic self-CHOP behaviors for management operations throughout the entire lifecycle of the networks, including planning, simulation/emulation, establishment, configuration, optimization, healing, and protection.
- Service-based architecture providing basic or atomic management services to support diverse management capabilities.
- Enabling and establishing responsible AI/ML in the 6G system, including management of AI/ML capabilities in devices, capabilities in the autonomous operations for the cloud-native, and compute-native 6G system by the evolved cloud M&O.
- Providing fundamental data management service to enable various management and network capabilities.
- Extending the management capabilities bolstered by Digital Twin and Knowledge management for 6G systems.
- Establishment of sustainable policies to optimize energy efficiency, operational efficiency, and service efficiency in an end-to-end manner across the mobile network system.
- Facilitation of intelligent intent-based networking, management, and orchestration, based on related logic, with respect to service objectives, system capabilities, and system environment conditions.

From a broad architectural perspective, the various types of NFs in a Service Based Architecture (SBA) framework are appropriately grouped on a functional basis, such as in terms of ML functions, core network functions, management functions, and monitoring functions. Among these functions, some are grouped to realize the functionality of an autonomous M&O subsystem. For example, the ML models required for this subsystem may be arranged at the design layer. The design layer of an intelligent M&O subsystem includes ML embedded autonomous functions, with feedback loops that may engage one or more classical layers of the protocol stack. This is used within an end-to-end, next-generation system to interpret and make decisions using relevant information gathered from the various segments of the end-to-end system (e.g., radio interface, transport, edge, application etc.). The interaction between the design layer and other layers in the system is facilitated by an appropriate Application Programming Interface (API) management exposure layer, which follows the SBA framework and Machine Learning Operations (MLOps) processes for a deployment of the related ML models and associated resources.

The ML models within an autonomous M&O subsystem can be utilized for the management of any artifact in the system, service fulfillment, and service assurance. The ML models can be also leveraged for system optimization within, as well as across distributed and cooperative autonomous domains (network service, network slice (NS), network function (NF), etc.), including across the core, edge, transport, and radio networks, and the user equipment (UE).

The trend toward. decentralization and distribution continues. As a result, edge networks that are pivotal for enhanced and customizable services, tuned to each user's preferences, the monitoring functions within an autonomous M&O subsystem are leveraged by the management function to perform the necessary orchestration of resources and functions for NS and NF scaling. The ML model-oriented functions may be arranged for either specific or cross-layer scope, through slow and fast feedback control loops. The fast control loops, with lower latency bounds, would be located closer to the network edge, while the slow feedback control loops in the core of the system would contain system-wide knowledge and more complex logic.

The management complexity, while rendering automatic system-wide optimization of resource utilization, is a pivotal characteristic of an autonomous M&O subsystem that leverages assorted ML models. These models leverage large and diverse data sets for decision-making and operational optimization, with high availability and reliability.

2.1.4 Key Value Indicators

The 6G ecosystem is anticipated to include a variety of enhanced capabilities, such as native ML capabilities, sensing, and energy-efficient networks and devices. These capabilities will exist across a decentralized and distributed network arrangement, together with cooperation across autonomous domains for connectivity and services. This would imply that there would be a need for different types of performance indicators, such as the ML model convergence, granularity of sensing, latency, fidelity of an ML model, etc.

A requisite assessment of these new and diverse KPIs would require an understanding of KVIs, which would facilitate a quantification of the value of 6G for society, beyond the performance metrics of energy networks that are delineated in terms of KPIs.

For an autonomous M&O subsystem, the automation and optimization of system-wide resources should consider KVIs associated with human, physical and digital environments, across cyber-physical interfaces, in terms of sustainability, trustworthiness, and digital inclusion. Predictive M&O are attributes of an autonomous M&O subsystem. This enables end-to-end network self-adaptability, as well as the management of the end-to-end system (cloud-edge-device segments) in a seamless manner.

Use cases include sustainable development, digital twinning, collaborative robots (cobots), immersive service experience, and localized trust zones. These are examples where the corresponding system design paradigm is transformed from one that is limited to performance measures to a complement of both KPIs and KVIs.

2.1.5 Standardization

Work on topics relating to intelligent and autonomous systems are ongoing in several Standards Development Organizations (SDOs). These initiatives cover separate aspects of automation, including use of autonomous technologies such as AI, intent-driven management, digital twins, data-driven management, and ML-Ops. Many SDOs are now in the phase where definitions of interfaces and models based on previously established use cases are specified. For example:

- **TM Forum's Autonomous Networks project** [4][5][6][7][8][9][10][11]: Common base models for intents and guide for the domain-specific extensions and common interfaces/APIs used for management of intents.
- **ETSI ZSM**: The "Intent-driven autonomous networks" study [12] is ongoing and planned to be followed by a work item for a normative specification, targeting reuse of TM Forum solutions.
- **ETSI NFV** [13][14][15]: Autonomous networks feature. Normative work on interfaces and models for intent specifications in and management data analytics interface definitions.
- **3GPP SA5** [16][17]: Studies on intent-based network slicing management and intent management enhancements for mobile networks. UML was selected for the 3GPP intents model.
- **O-RAN** [22][23][24]: The R1 interface is extended with intent support and RAN and autonomous functions are realized as one or more rApp.

2.1.6 Challenges and Research Directions

The heterogeneous, decentralized, and distributed nature of mobile wireless networks is likely to have multiple stakeholders with corresponding autonomous domains. The underlying ML models must have an inherent resilience towards malicious data streams, adversely impacting autonomous M&O. Therefore, sufficient levels of privacy and security are of paramount significance. These aspects must be addressed by utilizing FL models and homomorphic encryption techniques, which are considerations for ongoing research, associated with collaborative and connected intelligence across autonomous domains. The coordination of ML models with multiple feedback control loops with potentially different KVIs and related KPIs is pivotal for a consistent behavior of an autonomous M&O subsystem throughout the end-to-end system lifecycle. Further research is necessary to preserve a proper coordination of the underlying ML models. This includes research areas such as temporal shifts and priorities across different feedback control loops to avoid contention across different ML models that leverage the same environment.

The M&O subsystem needs to be built with a framework that can facilitate the multifold management capabilities by the SBA, which allows one service (a.k.a. management service) to be consumed by various management functions. These management capabilities for data management, AI/ML operations, cloud system orchestration, system digital twin, knowledge management, energy efficiency, and security and privacy management are envisioned to be the key components to interwork jointly to enable the 6G system's intelligent and autonomous M&O. These capabilities are discussed in detail later in this paper.

2.1.8 Conclusions

The ideas and themes that have been examined in Section 2.1, with respect to an autonomous M&O subsystem, in the context of heterogeneous mobile wireless network systems, are envisioned to serve as considerations for further study and elaboration in emerging, intelligent, and distributed connectivity and service paradigms in 6G.

2.2 AI/ML Operations and Enablement

2.2.1 Overview

With the objective of pursuing higher levels of intelligence and autonomy for 6G systems, AI/ML is a powerful technology to accelerate this journey. AI/ML is a key component in the M&O subsystem to enhance the autonomy as described in Section 2.1. It also is a significant catalyst to promote the intelligence of the network across the UEs, RAN, and transport and core networks.

In addition to the benefits of applying the relevant AI/ML models in the entire 6G system, it is imperative that AI/ML capabilities can be enabled and managed throughout the lifecycle of the AI/ML models. This should be done together with the related iterative processes, including data collection, data preparation, ML model training, validation, testing, emulation, deployment, and performance monitoring.

This section addresses the key considerations, technologies and challenges in AI/ML operations and management to enable consistent and responsible AI/ML model behaviors in the entire 6G system.

2.2.2 AI/ML Technologies Development Trends

In recent years, the rapid advancement of AI/ML technologies has left an indelible mark on various sectors, reshaping industries and redefining our relationship with technology. In the realm of computer vision, CNNs have proven to be highly effective. These neural networks have been employed in applications ranging from autonomous vehicles to Augmented Reality (AR). In the domain of natural language processing, Large Language Models (LLMs) within the generative AI field have emerged as game-changers. Their increasing potential hints at the possibility of intent-driven networks, revolutionizing how we interact with and manage networked systems. However, the AI/ML landscape is far from uniform; it adapts to the specific needs of different sectors. While some areas witness the proliferation of increasingly intricate machine learning models, others necessitate a more tailored approach. Lightweight tasks, particularly those inherent to Internet of Things (IoT) and edge AI applications, demand nimble solutions. This is where Tiny Machine Learning (TinyML) comes into play. It is a rapidly growing field characterized by its compact models and efficient computing, ideally suited to address the

challenges of resource-constrained environments and poised to empower a wide spectrum of IoT applications.

In this section, we will delve into how such advancements in AI/ML are poised to play a transformative role in the evolution of wireless networks, spanning the transition from 5G to 6G. We will highlight key aspects, ranging from data handling to the transparency of AI/ML models leveraged in these networks, shedding light on the significant impact these technologies are poised to have on the future of wireless communication.

2.2.3 AI/ML Capabilities and Operations in 5G

AI/ML capabilities are used in various 5G system domains, including M&O (e.g., Management Data Analytics (MDA)) [4], 5G Core (5GC) Networks, and Next Generation Radio Access Network (NG-RAN).

For M&O, AI/ML is used to enable MDA [4].

For 5GC Network, AI/ML is used to enable Network Data Analytics Function (NWDAF) [5].

For NG-RAN, AI/ML is used to support RAN intelligence functions, including Network Energy Saving, Load Balancing, and Mobility Optimization [6][7].

3GPP is studying AI/ML for NR air interface to address the beam management, Channel State Information (CSI) feedback, CSI prediction, and UE positioning [8].

3GPP has also defined the AI/ML operation and management capabilities [9][10] throughout the entire AI/ML operational workflow in the lifecycle of an ML model, as depicted in Figure 3.

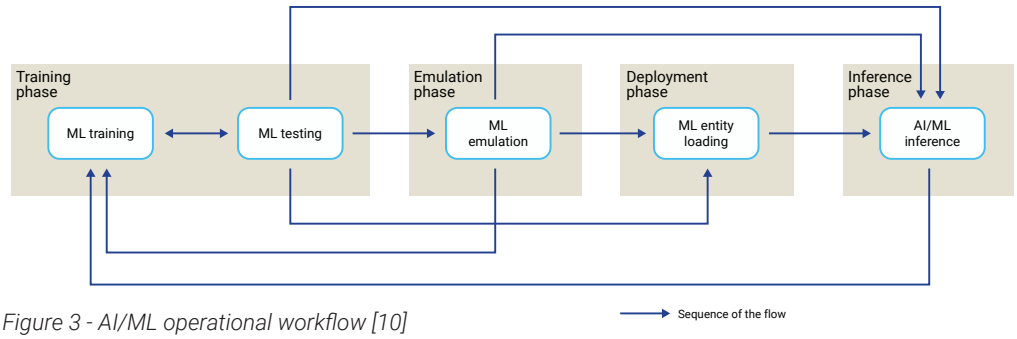


Figure 3 - AI/ML operational workflow [10]

The management capabilities and services are defined for each phase of the AI/ML operational workflow, including ML training, testing, emulation, model deployment, and inference.

O-RAN Alliance is developing support for AI/ML through design of AI/ML workflow services that support AI/ML model Lifecycle Management (LCM), including model training, inference, monitoring and model management, and data processing pipeline [22][23]. A variety of use cases have been defined that employ AI/ML algorithms in the solutions, such as QoE optimization and RAN slice service-level agreement (SLA) assurance [24].

2.2.4 Envisioned AI/ML Applications in 6G Systems

AI/ML will empower the 6G system to achieve greater intelligence and autonomy. AI/ML will be used at every layer and in every domain from the management system, network to the UE over air interface.

AI/ML will enable system intelligence and autonomy and minimize human intervention for the network operation and optimization. Based on AI/ML (e.g., generative AI), the management system will be able to automatically generate the detailed configurations from the operator's target or intent based on AI/ML and apply the configurations to the networks. This will significantly simplify the operations needed to manage the networks. Furthermore, the LLM has the potential to elevate the level of intelligence and autonomy of the 6G system to new heights by driving the evolution of management interfaces from ML-based to human language-oriented, which makes it much easier for both machine and human "consumers" to use.

As more 6G applications become AI-enabled, the 6G system needs to better support AI workloads. The networks need to be monitored closely in terms of the performance when supporting AI workloads and need to be optimized in an efficient manner when necessary.

2.2.5 Machine Learning Operations (MLOps)

MLOps is a set of workflow principles that enable simplification and automation of ML workflow, all the way from exploratory data analysis to production deployment. The principles facilitate features such as reproducibility, continuous integration and delivery, deployment, monitoring, and continuous training. MLOps is an extension of the software development and information technology (IT) operations (DevOps) process of continuous software engineering practices. It promotes the alignment and compliance of the associated ML models with the relevant policies and regulations. MLOPs also enables system operation drift checks, between expected and actual ML model behaviors, which is essential for sustaining a high system availability and reliability.

This approach facilitates the LCM of the ML models for a scalable, risk-reduced applicability of the autonomous M&O, which is critical for an automated and optimized operation of the 6G system.

2.2.6 AI/ML with Emulating Environment

AI/ML emulation is an essential step for making the AI trustable before deploying the ML models to the real systems. The ML model can be used for inference in the real systems only if its behavior and performance in the real networks or systems can be known or assessed before-hand (e.g., by emulation).

AI/ML emulation is studied in 5G, along with AI/ML that 6G will use. Even so, how to dynamically build and allow the consumer to choose the emulating environments remains challenging.

Digital twin could be a potential solution for emulating the system (see Section 2.7).

2.2.7 Data Handling for AI/ML

3GPP has defined the ML Training Management Service (MnS), through which the ML training capabilities are provided in the context of Service-Based Management Architecture (SBMA) to the authorized consumer(s) by ML training function playing the role of the MnS producer, as illustrated in Figure 4.

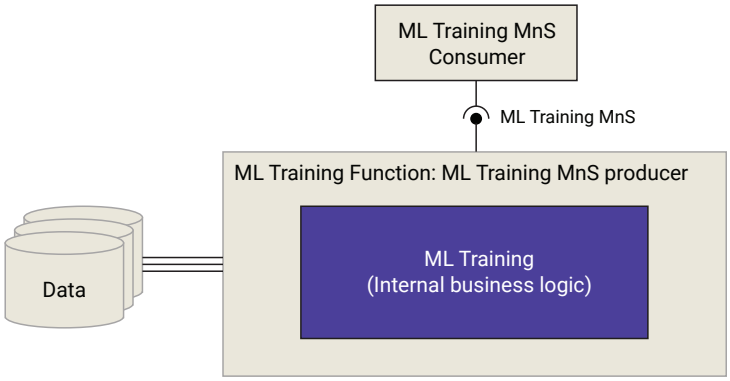


Figure 4 - Functional overview and service framework for ML training [9]

The ML training function leverages the current and/or historical data of the following types for training:

- Performance Measurements (PM) as per 3GPP TS 28.552 [11], 3GPP TS 32.425 [12] and KPIs as per 3GPP TS 28.554 [13].
- Trace/MDT/RLF/RCEF data as per 3GPP TS 32.422 [14] and 3GPP TS 32.423 [15].
- QoE and service experience data as per 3GPP TS 28.405 [16] and 3GPP TS 28.406 [17].
- Analytics data offered by NWDAF as per 3GPP TS 23.288 [5].
- Alarm information and notifications as per 3GPP TS 28.532 [18].
- Network Resource Model (NRM) configuration information and notifications 3GPP TS 28.541 [19].
- MDA reports from MDA MnS producers as per 3GPP TS 28.104 [4].
- Management data from non-3GPP systems.
- Other data that can be used for training.

For some AI/ML use cases, one type of issue (e.g., coverage) can be reflected by multiple types of data such as performance measurements, MDT data, and QoE data. To support ML training, different types of data may need to be correlated, integrated, and labelled in case of supervised learning. The data needs to be filtered, prepared, and potentially generalized for ML training to fit target network conditions. The data correlation, filtering, preparation, labelling, and generalization to support ML training need to be investigated.

Additionally, more and more AI/ML use cases need UE-level measurements in real time or near real time. How to collect and manage these UE level measurements is also a challenging issue.

2.2.8 Enabling AI as a Service

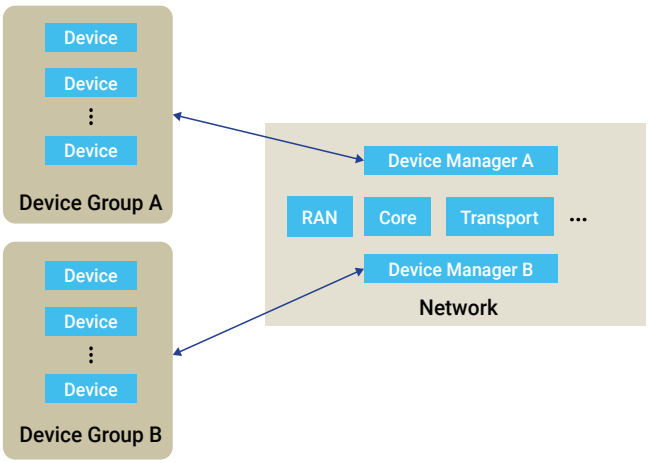
In 6G system, AI could be offered as a service for consumption by other system functions (NFs or management functions). In 5G, the management services for ML training, testing, emulation, deployment, and inference have been defined. How to streamline, automate, and operate these services requires additional research.

2.2.9 M&O of AI/ML Capabilities in Devices

Mobile devices can use AI/ML to improve performance when accessing the networks. ML models may be trained at different places by different parties, such as in an application server, in the management system, or in network nodes. The ML model on the device would impact both device and network performance. Therefore, deployment and management of ML models must follow current software release frameworks that employ rigorous testing, version control, and production deployment. It is important to research which entity or system would be responsible for deploying the ML model to the devices, and which entity or system manages (e.g., configure, activate/deactivate, monitor) the ML models.

Furthermore, 6G systems should allow management of each device vendor's specific AI/ML models. Different implementations require vendor-specific AI/ML models because of different hardware and software designs. Thus, the management entity needs to be device specific. This is shown in Figure 5, where two device groups are shown: A and B. The grouping of devices is based on device vendor, although this concept can be extended.

Figure 5 - Architecture for device management when device manager is hosted in the network



Each group has its own Device Manager that provides services tailored for that specific set of devices. It manages the device configurations including AI/ML model configurations and data collection from devices. Although, the Device Manager is shown as a separate network entity hosted within the network, it may be implemented as a part of another network entity or management system, or may be hosted outside the network.

2.2.10 Trustworthiness of AI/ML for Mobile Networks

AI/ML trustworthiness is critical. Their deployment demands technical robustness and safety, privacy and data governance, diversity, non-discrimination, fairness, and accountability. Addressing these aspects has been a subject of study for 5G systems, as exemplified by 3GPP's TR 28.908 report [20]. This section delves into key aspects of AI/ML trustworthiness, with a broader focus for 6G systems from a M&O perspective.

- **Availability, Reliability, and Resilience:** In telecommunications, network availability, reliability, and resilience are vital for end-to-end performance. Availability is the wireless network's consistent and sustained capacity to deliver long-term performance, immune to transient disruptions. Reliability is characterized by the system's steadfastness and ensures operational confidence with a high degree of certainty, mitigating the risk of failures. Resilience, on the other hand, defines the network's ability to swiftly recover from errors to guarantee operational continuity even in challenging scenarios. Mission-critical settings, like emergency response systems, emphasize the amplified importance of these attributes due to potential severe consequences.

  Integrating AI models into wireless networks, especially in mission-critical applications, necessitates alignment with stringent industry standards. These models must ensure sustained performance, reliability, and resilience, with heightened significance in time-sensitive missions.

- **Transparency:** In the dynamic landscape of AI/ML systems, transparency plays a pivotal role in enhancing both responsibility and security. This becomes particularly crucial when these systems are integrated into wireless networks supporting mission-critical functions.

  The responsibility of AI frameworks is augmented through transparency because it allows comprehensive examination of their decision-making processes. This scrutiny identifies potential vulnerabilities within AI frameworks and mitigates the risks of exploitation by adversaries in mission-critical scenarios. In this context, transparency becomes synonymous with accountability, ensuring that AI models adhere rigorously to ethical and operational standards.

  Moreover, the concept of explainability, as a subset of transparency, refers to the capability of certain AI systems to assist humans in interpreting the reasoning employed by AI workflows to reach an inference. In the realm of network management, it offers insights into AI-driven decisions, fosters trust, and ensures uncompromised safety. Network operators rely on understanding how AI models arrive at decisions that influence network performance. In alignment with this, the principles for explainable AI stress the need for providing explanations [26] that are comprehensible to individual users, especially within the intricate landscape of network management.

- **Privacy and Data Governance:** Privacy and data governance emerge as paramount considerations in network management, particularly given the sensitivity of the data involved. The integration of data privacy measures is vital within AI-driven network management practices. This encompasses controls such as pseudonymization, encryption, and user authentication to safeguard the integrity of sensitive network data.
- **Diversity, Non-discrimination, and Fairness:** Network management AI/ML systems are expected to be inherently fair and non-discriminatory, ensuring equitable treatment of all network users and services [20]. These systems are tasked with considering an array of network parameters and user profiles without introducing biases. Explainability plays a pivotal role in demonstrating and validating fairness within AI-driven decisions.
- **Human Agency and Oversight:** Human agency and oversight emerge as indispensable components in network management, particularly when AI systems are at the forefront. 3GPP's guidelines advocate for "humans in the loop" to intervene in AI-controlled network systems when necessary. Explainable AI systems furnish user-friendly interfaces that empower network operators to effectively intervene and make informed decisions [26].

Essentially, addressing the trustworthiness aspects of AI/ML in mobile networks presents both challenges and opportunities. Solving these issues is essential for unlocking the full potential of AI/ML in network management as the technology landscape evolves.

2.2.11 Sustainable AI/ML

Depending on the complexity of the ML model's task and structure, the energy consumption of the AI/ML solution could be very different. Some AI/ML solutions may consume high energy for both training and inference, while others may consume much more energy for training than inference. Therefore, it needs to be evaluated whether the energy cost is worth the benefits provided by a specific AI/ML capability. This evaluation must find a way to balance the energy cost and the benefits to make sustainable AI/ML for 6G.

2.2.12 Challenges and Research Directions

To improve the efficiency of AI/ML, the AI/ML operations are expected to be streamlined and automated.

The overhead associated with the online ML model training needs to be controlled and reduced in terms of the training time, network traffic, and computing resources, especially at the network edges, where latency budgets are low, while handling large datasets.

Data handling — especially data correlation, filtering, preparation, auto-labelling, and generalization — is essential and requires further research to support ML training.

The emulation of ML models is necessary to ensure consistent decision-making behaviors, under both normal and faulty conditions, for diverse usage scenarios that are anticipated as the service landscape continues to evolve.

ML model trustworthiness is complex because the decision-making process based on feedback control loops has system-wide scope and impact. This is an arena of ongoing research and enhancements. Device management needs to support device vendor specific AI/ML models. To achieve this functionality, support is needed for device identification, grouping, and routing of device traffic to the appropriate Device Manager. Functionality to coordinate across Device Managers also needs to be supported.

2.2.13 Conclusions

To make extensive use of AI/ML and unleash its power to promote 6G system intelligence and autonomy, efficient and streamlined AI/ML operations are required. So is management throughout the entire ML model lifecycle. In addition, in-depth research is needed to make AI/ML responsible, trustworthy, and sustainable for the wireless system.

2.3 System Robustness and Resilience

2.3.1 Overview

Cellular network systems are evolving with the merging of IT, Operational Technology (OT), and enterprise networks. Next generation technologies are expected to provide sustainable and intelligent solutions and applications while providing M&O tools to manage varying levels of SLAs, heterogenous network resources, and data sources [27]. However, there are a few objectives that are fundamental from a service provider's point of view: resilience, integration, and cost.

2.3.2 Resilience

A well-accepted definition of resilience is "the ability of the network to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation" [28]. Resilience includes disciplines of dependability (encompassing availability, reliability, etc.), security (encompassing confidentiality, integrity, availability, etc.), and performability. Some of the means for a resilient architecture include intelligent and adaptive networks, and redundant and diverse topological infrastructure [28]. Managing autonomous network operations is expected to be complex.

2.3.3 Integration

Integration involves the deployment of network assets in the most simplified and cost-efficient manner. Tradeoffs for resilience versus resources and resilience versus complexity need to be carefully studied [28].

2.3.3 Cost

A direct consequence of resilience and integration efforts is cost [28], which can be captured by Capital Expenditure (CapEx) and Operational Expenditure (OpEx). Orchestration and management tools should enable the cost-efficient deployment of heterogenous resources.

2.3.4 Challenges and Research Directions

Some research directions for resilient M&O are:

- A common and holistic architectural view of intelligence frameworks between SDOs, and even the working groups of a given SDO, is crucial to ensure that potential solutions are not fragmented. Alignment across SDOs — such as ETSI NFV-MANO, O-RAN, cloud orchestration frameworks (e.g., LF Nephio project), and AI/ML-related work from 3GPP working groups — becomes critical to have an explainable and trustworthy AI/ML system. A common framework would help operations personnel to understand interdependencies and complexities with the next-generation OAM platforms.
- The 3GPP 5G specifications for closed-loop control systems [29][30]. Control loop systems (either open with human in the loop or closed without human) will be essential for real-time operations of detection and recovery when the network is disturbed. These control loops should be improved to include advanced use cases with native-AI/ML systems [31].
- The orchestration and management plane should have the ability to process data from temporally and spatially different resources. Furthermore, the fast-paced data needs to be understandable among interoperable systems to achieve different SLA objectives for varying Enhanced Mobile Broadband (eMBB), Ultra Reliable and Low Latency Communications (URLLC), and Massive IoT (mIoT) application types. The AI/ML models need to provide *accurate* and *high-confidence* output with varying timescales to be able to address heterogenous data sources while assuring varying levels of SLAs. Moreover, identifying operational inputs and metrics for long-term adaptive and evolving resilient network state is another gap.
- Explainability of such AI/ML systems to manage and orchestrate network is crucial [32][33]. A system that is not explainable and not dependable may degrade network operations and increase network operating costs. New metrics to measure such explainaibility are necessary to assess and modify models, and if necessary for switching from a closed-loop to an open-loop operational state.
- Resilient autonomous operations are naturally expected to bring complexities (i.e., more resources, states, interdependencies). Tradeoffs between resilience and complexity, as well as resilience and energy consumption, need to be studied to reduce conflicting objectives. Tradeoffs between advanced AI/ML capabilities and reducing energy consumption to manage and operate sustainable networks should be studied in new algorithm development [34].

2.3.5 Conclusions

6G networks are expected to include heterogenous network resources to address varying SLA objectives for a multitude of applications. The M&O plane should have capabilities for resilient network operation via cost-efficient means. Important aspects to consider in this 6G journey include alignment across SDOs to avoid fragmenting solutions, a simplified integration

pathway to 6G, studying tradeoffs between resilient mechanisms and cost (as well as complexity, network resources, and energy efficiency), and explainability of AI/ML systems.

2.4 Knowledge and Semantics Management

In the dynamic landscape of network management, we find ourselves on the cusp of a profound transformation in the way we communicate. Our historical communication methods have predominantly operated within what we'll refer to as "Level A." This approach has prioritized the exchange of raw data among network entities while paying less attention to unraveling the deeper layers of meaning concealed within that data. However, as we step into the era of 6G and beyond, it becomes increasingly evident that communication must transcend the confines of Level A and embrace the concept of semantic communication.

2.4.1 Overview

In the ever-evolving realm of network management, a profound transformation is underway as we transition from traditional syntactic communication towards semantic communication. This shift goes beyond mere data exchange, delving into the nuances and depths of meaning within our communications, underpinned by the integration of AI. Essentially, the integration of various AI frameworks and their various foundational models plays an instrumental role in semantic inference and management [35]. It involves extracting meaning from information transmitted by a sender, leveraging associated Knowledge Bases (KBs) [36] linked to both the sender and the receiver. This semantic inference is pivotal in surpassing the convergence of computing and communication resources, enabling mobile wireless networks to operate at a level that transcends the conventional boundaries of bit/symbol rate.

Unlike conventional data streams over physical mediums such as wireless connectivity, semantic communication transmits only the vital and pertinent information, empowering the receiver with generative computing capabilities to faithfully recreate the intended message [37]. This intelligent approach can significantly reduce traffic volume and enhance energy efficiency. Processing latencies may also diminish because the information parsing is not dependent on a fixed syntax. For instance, in scenarios involving image or video transmission, AI technologies are central to minimizing the transmitted information while preserving fidelity. The synergy between system-wide KBs and semantic information, along with the autonomic and self-adaptive capabilities enabled by AI, establishes a robust foundation for cognitive and autonomous M&O within the 6G ecosystem.

The emphasis on high semantic fidelity remains unwavering, prioritizing it over bit-level accuracy, especially in wireless connectivity scenarios. In semantic communication, only indispensable features of multimedia content necessary for faithful reproduction at the receiving entity are transmitted. This approach offers a myriad of use cases in the 6G ecosystem, ranging from immersive communications, such as AR, Virtual Reality (VR), and Extended Reality (XR) services, all aimed at delivering engaging, interactive, and attractive user experiences. This marks a pivotal moment in the evolution of network management, ensuring that our communication is not just data transmission but a meaningful exchange that holds profound potential.

2.4.2 Architectural Considerations

The application of a network system representative KB with semantic reasoning is an intrinsic architectural consideration for cognitive and autonomous M&O in a 6G system. The underlying service-based framework of distributed and decentralized network functions, together with communications, computing, and storage resources, promotes the programmability of the network system as a whole. This includes an autonomous M&O subsystem, facilitating self-adaptive, zero-touch automation. The cognitive qualities imbued by a semantic M&O subsystem enable the organization and updates to the system-wide knowledge base, which collectively sustains self-CHOP system-wide behaviors, throughout the system's lifecycle. For semantic reasoning and decision-making within a M&O subsystem, the KB represents diverse system parameters, across the prominent system layers, consisting of service, network, and the underlying infrastructure. These aspects are cooperatively harnessed for decision-making strategies for an optimized and automated allocation of resources (e.g., NS). The programmability of the network in an SBA framework allows for the creation and exposure of relevant data. This data is leveraged for maintaining and updating the representative KB. This representative KB in turn serves as a foundation for a semantic and autonomous M&O subsystem.

Unlike a traditional network architecture, a cognitive and autonomous system architecture, which includes a M&O subsystem, is designed with the capability to dynamically self-adapt the network system to changes within the system and its operating environment while automatically sustaining the reliability, availability, and optimized behaviors. Such a system utilizes ontology modeling to create and update a KB, based on semantic entity relationships, within the system.

The system-wide components of the service layer — consisting of NSs supported over network layer resources such as NFs and operating over network elements at the infrastructure layer — provide the framework for a cognitive, autonomous M&O subsystem. Cross-layer APIs facilitate interactions across the service, network, and infrastructure layers. Cloud-native and AI-native capabilities that harness CI/CD/CT processes for relevant MLOps are a pivotal aspect of a cognitive, autonomous M&O subsystem.

Cognitive service architectural constructs infused into an SBA framework provide the semantic capabilities, in conjunction with a representative KB, to augment the user experience in innovative immersive services across human and machine interfaces. These directions further advance a customized service experience, tuned to the resolution of the subjective preferences of each end user, unveiling new value proposition opportunities, hinging on a coupling of innovative business models and system-wide cognitive capabilities.

Cooperation is pivotal for a cognitive and autonomous M&O subsystem across network system states described in terms of diverse attributes, such as bandwidth, communication, computing, and storage resources, latencies, packet loss, and load conditions. In emerging decentralized and distributed network system configurations, such as Multi-Access Edge Computing (MEC), a dynamic mapping of end user requirements and the network system resource conditions

is pivotal for choreographing semantic behaviors in an autonomous M&O of system-wide resources. This ensures that any gaps in meeting the end user service requirements are bridged gracefully as the system and its environment change dynamically over time.

2.4.3 Use Cases

A semantic and knowledge management paradigm is anticipated to usher an era of continuing service innovation, associated with a variety use cases [38] in the 6G ecosystem. Decentralized and distributed ML-enabled functions, within an end-to-end service based architectural framework, are anticipated to support optimized semantic learning and inferencing over diverse workloads, including at a network edge anywhere. Autonomous M&O, with anomaly detection and prevision supported by semantic inferencing and decision-making, are expected to underpin the system-wide availability, reliability, and sustainability to adequately and efficiently support the service experience associated with diverse use cases.

Flexible and novel deployment scenarios, while optimizing costs and performance through a relevant alignment with KVIs and KPIs, are among the foundational considerations for enabling innovative usage scenarios. In this context, flexible arrangements of terrestrial and non-terrestrial networks, diverse spectrum ranges (Terahertz, sub-Terahertz spectrum), cell-free Multiple-Input and Multiple-Output (MIMO), non-public networks, device-to-device communications, etc., that leverage ML models and semantic autonomous M&O procedures are envisioned to suit use cases associated with various forward-looking business and deployments models.

A few prominent use case examples include immersive communications (e.g., AR, VR, XR), ubiquitous connectivity (e.g., diverse spectrum utilization), intelligent health (e.g., curated and self-directed wellbeing), autonomous mobility (e.g., self-driving vehicles), holographic presence (e.g., three-dimensional virtual presence rendering), and sustainable development (e.g., energy efficiency, environmental conditions/sensing etc.). The use of a semantic approach for autonomous M&O is anticipated to enhance these categories of use cases while optimizing 6G system performance, operations, costs, and the associated utilization of system-wide resources.

2.4.4 Challenges and Research Directions

Among the challenges and research directions [39], 6G requires a shift away from superficial characteristics inferred through typical ML models that utilize related data patterns toward a contextual and semantic inference from observations of meaning. The inference of meaning promotes an efficient utilization of system resources for a proper inference of information within the system and between the system and its operating environment. This entails a requisite merging of ML models and a representative and system-wide KB. The associated semantic learning process is for further research to properly infer context, through knowledge sharing transfer learning and federated learning, to facilitate an inter-domain and intra-domain cognitive, autonomous M&O subsystem.

The nature of these research directions and challenges, associated with knowledge and semantics management in 6G, are interdisciplinary, spanning a variety of use cases, heterogeneous mobile wireless systems, and semantic communications.

2.4.5 Conclusions

Prominent facets of knowledge and semantics in a 6G autonomous M&O have been examined briefly. This approach reveals a distinct shift from a conventional syntactic inference of information to a semantic inference of information, where the latter is based on meaning and the former is based on a sequence of bits or symbols.

Semantic communications hold the potential for a realization of enormous system-wide benefits including enhancements associated with accurate inferencing of system information and knowledge, reduced channel bandwidth, improved resource utilization, and energy efficiency. Additionally, the use of semantics and associated knowledge advances the efficiency of autonomous M&O to support emerging use cases in a 6G ecosystem.

2.5 Data Management

2.5.1 Overview

6G is expected to support approximately 20 times the data capacity over 5G through a combination of increased channel bandwidth and spectral efficiency. New use cases including 3D XR, digital twinning, JCAS, and fiber-like FWA will generate enormous amounts of data. User data will be generated, stored, and shared. Knowledge (i.e., new data) will be extracted from real-time user data. In addition to providing high data throughput capacity, these future networks will enable low-complexity, energy-efficient end user devices by providing compute resources that may be distributed, at the edge, or centrally located.

Not only will 6G networks transport, process, and store data for end users and services, they will be data-driven, trained systems that natively incorporate AI/ML. Massive amounts of data must be transported, processed, and stored to manage the models that will enable a self-optimizing air interface. The result will be a 6G network that is perfectly adapted to any channel, any hardware, or any application running on top.

With 6G, the traditional data management will expand in scope and become even more critical. It will no longer be limited to traditional forms of management data (i.e., PMs, CMs, KPIs, alarms, and traces) but will include new types of data that 6G networks produce and require. Management of the vast amount of data will be required to support the applications and network operation. This is because many aspects of 6G networks will depend on the efficient discovery of the relevant data, data exposure, and access to various data while maintaining the privacy and security of all data (user and system).

2.5.2 Data Management in 5G Systems

Data management in 5G systems is specified differently depending on whether the data is categorized as management data.

5G specifies the management data as performance data and configuration data of NFs and systems, as well as alarms and traces. Additionally, any "external" data (i.e., data not specified by 3GPP) that is used for M&O of a 5G network is also considered management data [11][14]. 5G also specifies mechanisms for discovery and retrieval of management data. [56] The management of the management data is an integral part of OAM functionalities in 5G networks because the network operation depends on efficient access to the required data.

Other data required by NFs (and not categorized as management data) is specified and stored in the relevant NFs. Subscriber data, operational data, and policies are stored in the Unified Data Management (UDM). Application data, structured data for exposure, subscription data, and policy data are stored in a common database called the Unified Data Repository (UDR). Any NF can store unstructured data in the Unstructured Data Storage Function (UDSF). [41]

The primary focus of this section is the management of data used for operation and orchestration of the network (i.e., management of data as traditionally considered part of M&O functions).

2.5.3 Data Drivers in 6G

6G systems will see an immense increase in data volume and new types of data that need to be managed, potentially with new requirements. The sources of data in a 6G system include:

- Network data in 5G systems (e.g., management data related to performance, configuration, alarms, etc.) remains critical data that is produced by different NFs and entities. In 6G systems, the management data is enriched by data required by new applications and features, including additional data collected for automation. Additionally, some types of network data (e.g., performance management data) are consumed offline in 5G systems. But in 6G systems, they may need to be collected and managed in real time for network optimization.
- New and diverse access types, device types, and communication technologies introduce new types of data with different characteristics. For example, Ambient IoT and Joint Communication and Sensing (JCAS) produce new types of data specific to 6G systems.
- There are new 6G applications expected that would introduce vast amounts of data, including 3D XR, 3D digital twinning, and environment data (i.e., replication of aspects of the real world as digital twins).
- The analytics data is expected to increase compared to 5G systems as more aspects of the network are automated and operated by AI/ML systems. The analytics data includes the data from various sources mentioned in this list used for training and retraining of ML models, evaluation data of the ML models, the ML model itself, its related information, and the analytics output.

2.5.4 Consumers of Data

As with new sources of data, the number of consumers in 6G system also increases. Traditionally the network management system has been the main consumer of data in the network. In 6G, the AI-enabled and autonomous operations in network management system will expand the demand for the data.

With automation, observability of the network status at all domains, layers, and locations becomes critical. The data collected across the network is the input into AI/ML systems for real-time, near-real-time, or offline processing. AI/ML or data-driven trained systems are an integral part of 6G systems and will consume data for ML model training, federated learning, and knowledge transfer within ML systems. The volume of data exchanged among AI/ML systems within the network to enable these functions is massive, especially in large-scale distributed deployments.

Additionally, vertical services rely on data exposed by the network for the applications enabled and services provided to the users. The data needed by vertical services can vary from user-specific information to network-specific information and anything in between.

There are challenges associated with efficient management (i.e., discovery, transport, storage, and security) of the data for an increased number and diverse set of data consumers in the network. These consumers may be geographically close to the data sources (e.g., an edge application using user information at the edge) or far (e.g., an ML model training in the cloud). Moreover, the consumer may need real-time data (real-time performance management), near-real-time (configuration optimization), or historical data (model training). These challenges are discussed in the following sections.

2.5.5 Discovery of Data

As described in the previous section, there are diverse types of data from a diverse set of sources produced across the network. The produced data may be stored for future use, consumed instantly, or both. The data may be produced at the edge to be consumed by the network elements within the cloud or vice versa. The network data may also be used by entities that are not part of the 6G network (e.g., the vertical applications).

In all these scenarios, the consumers should be able to discover the data (i.e., identify the source of the data, learn whether the data is already produced, whether the data is stored, and in all cases the address of where the data is accessible).

Given the large number of producers of the data and types of data, and the large number of data consumers in a 6G system, it is critical that there is an efficient way for the consumers to discover the desired data. Consumers should be able to learn whether the data is produced and where it is stored with low overhead, regardless of their relative distance to the source of the data and where it is stored. Moreover, to ensure the efficiency of the discovery process, there should be a single mechanism to discover all data types, regardless of their source.

Another challenge with respect to the discovery of the data is the discovery and accessibility of the data across management domains and especially outside of the network management system (i.e., by vertical applications). In these scenarios, the data needs to be exposed based on pre-determined policies.

2.5.6 Data Storage

Traditionally, network data is stored according to the operator's policy for analysis of different aspects of network performance, as well as other policy-related usages. In 6G, there are additional demands for storage due to the expanded use of data by multiple consumers, including AI/ML systems, and use of data by various applications.

Data may be stored for short or long terms; raw or processed, and locally or in the cloud. In general, the farther from the cloud and the core, the more expensive the storage is and hence more appropriate for shorter duration storage. Storing raw data requires no or little processing, but it also requires more space and may be associated with longer latency when accessing the right data for a specific application. On the other hand, the processed data is quick to access and requires less space, but there is computation latency and overhead to process the data.

With storage distributed across the network, the data management system should determine whether a particular type of data should be stored locally at the far edge, at the edge, at a central location, or in the cloud. The choice depends on the specific use case and the latency and type of processing associated with the data. Data may also be moved across the network from storage at the edge to the cloud, for example, as it ages and its applicability for near-real-time or shorter time scale applications expires.

2.5.7 Data Transport

A 5G network may be thought of as having different domains: RAN, core, cloud, and transport. Historically, orchestration in mobile networks focused on the RAN and core domains. Less emphasis was placed on coordinating the M&O of the transport domain with the RAN and core, and cloud orchestration being relatively new. 6G networks will be closely integrated with applications, including performance requirements for specific traffic types. This requires closely coupled M&O of transport with all other domains, jointly with services.

6G networks can be characterized as heterogeneous. Data will be stored, processed, and transported among far edge, edge, core/central and public/private clouds. Moving the data between the physical resources (radio equipment, routers, switches, data centers, and cloud infrastructure) will present opportunities and challenges related to performance. The far edge can be characterized by a variety of devices using different, non-homogenous hardware and software technologies. A unified way for data retrieval, regardless of data type, will be necessary, along with efficient methods to retrieve data, regardless of the source or location. System performance measurements and KPIs (e.g., throughput, latency, packet loss) will depend on all domains, and well-performing networks will demand integrated M&O across all domains, including transport.

2.5.8 Security, Privacy, and Trust

One of the core values of 6G is trustworthiness. Security and privacy of data will be essential to earn and keep the trust of stakeholders. While the network services must be fit for purpose and accessible, security measures inevitably introduce some limitations (e.g., flexibility, performance), and a balance will be required. KVIs that measure trustworthiness must be developed and understood.

Specific to data management, 6G networks must provide resiliency against attacks (no single point of failure, theft of data/breach, denial of service, integrity) and preserve privacy. In the convergence of networking and computation, ensuring privacy is essential. Privacy-preserving technologies that balance minimal data exposure with legitimate analytics will be required. An example is using key aspects of the data for analysis while removing private aspects. Data security must be considered, along with AI/ML federation capabilities. Security and privacy are especially important when orchestrating services from third-party providers. Distributed ledger technologies have the potential to ensure trust of data.

2.5.9 Data as a Service (DaaS)

The previous sections highlight the challenges and the complexities associated with handling the huge amounts of data that will exist in 6G systems. These include data collection, storage, and efficient retrieval and transportation of the data, potentially to many consumers in different locations in the network (and applications outside) with various constraints and differing requirements.

DaaS is an efficient way to address these challenges in a 6G system. The provider of DaaS collects, stores, correlates the data and the requests for it, maintains the data, and makes it available to the consumers in a way that optimizes the use of network resources, thus reducing the cost associated with data management.

DaaS based on requests from various data consumers facilitates discovery, collection, and delivery of the data from many sources in 6G to the consumers. This is done in a way that minimizes the overhead on the data producers and optimizes the usage of resources required for data storage and transportation. In Figure 6, the DaaS provider receives requests for data from 6G consumers and coordinates the delivery of the data to them with the data sources and the storage via the control plane. The data plane is used for the transport and delivery of the requested data to the consumers.

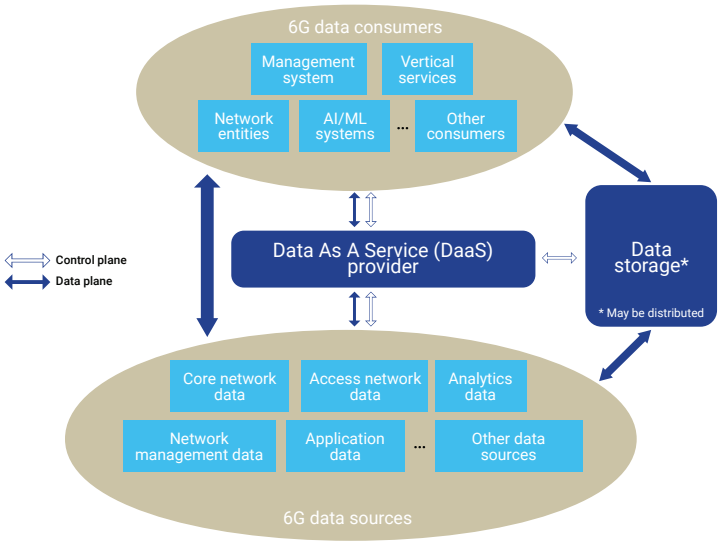


Figure 6 - DaaS and 6G sources and consumers of dat

2.5.10 Challenges and Research Directions

Some of the challenges related to data management in 6G systems are:

- Unified and efficient mechanisms for the discovery of different types of data across the network, including how data, irrespective of its type and producer, can be modeled for discovery.
- Low-overhead mechanisms for learning about availability of the data and its address by data consumers across the network and different management domains.
- Storage policies that increase efficiency of accessing the data for different use cases while reducing the cost associated with the storage and the overhead of transporting data across the network.
- Efficient mechanisms for transport and distribution of data across the network (i.e., to multiple consumers located anywhere between the edge to the core) while minimizing the duplication of the data and its transport path.
- Access, exposure, and delivery policies that maintain security and privacy rules associated with the data in a multi-domain and multi-consumer environment across the end-to-end network.

2.5.11 Conclusions

Extensive increase of data is expected in 6G systems. Network automation depends on analytics and observability across the network, which in turn requires the data to be efficiently collected and accessible at all domains, layers, and locations. Efficient solutions for data management (i.e., discovery, transport, storage, and security of the data) become even more critical in 6G systems.

2.6 Cloud System M&O

2.6.1 Overview

6G applications are envisioned to be relevant to virtually all aspects of life, society, and industries, to further improve and change the way people live and work. The new types of applications that will be enabled can be characterized into categories of everyday living, critical roles, experience, and societal goals [40]. Nearly all of these 6G applications are data-driven and need to be supported by both communication and computing, as illustrated in Figure 7. These applications have specific requirements for both communication and computing. Therefore the 6G cloud system is the integration of communication and computing, which can simultaneously take care of the requirements from both perspectives.

Figure 7 - Data-driven 6G applications

2.6.2 Cloud in the 5G Era

2.6.2.1 5G Telco Cloud

Enabled by Network Function Virtualization (NFV) and Open RAN (O-RAN) solutions, the 5GC network functions as defined in F [41] and some RAN units (e.g., O-RAN Central Unit (O-CU), O-RAN Distributed Unit (O-DU) as defined in O-RAN architecture [42]) can be virtualized and run in the telco cloud.

The NFV M&O as depicted in Figure 8 [43] are independent of the management of functionality of the 5G networks. This means the NFV M&O are agnostic about the functionalities of the virtualized NFs.

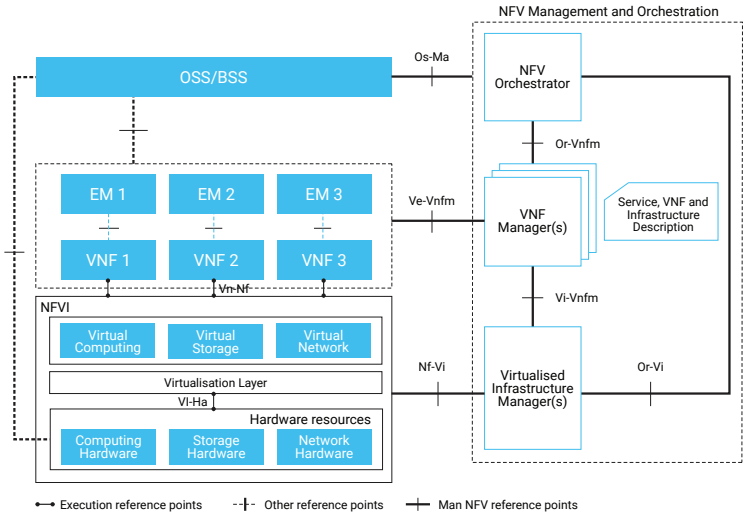


Figure 8 - NFV reference architectural framework

2.6.2.2 Edge Computing

In 5G era, MEC (formerly known as Mobile Edge Computing) offers cloud-computing capabilities and an IT service environment at the edge of the network to bring computing-intensive applications closer to users to reduce latency, based on the framework illustrated in Figure 9 [44]. MEC allows software applications to access real-time information about local-access network conditions to facilitate QoS assurance. However, from the management perspective, the edge data network and the mobile network are managed and orchestrated separately.

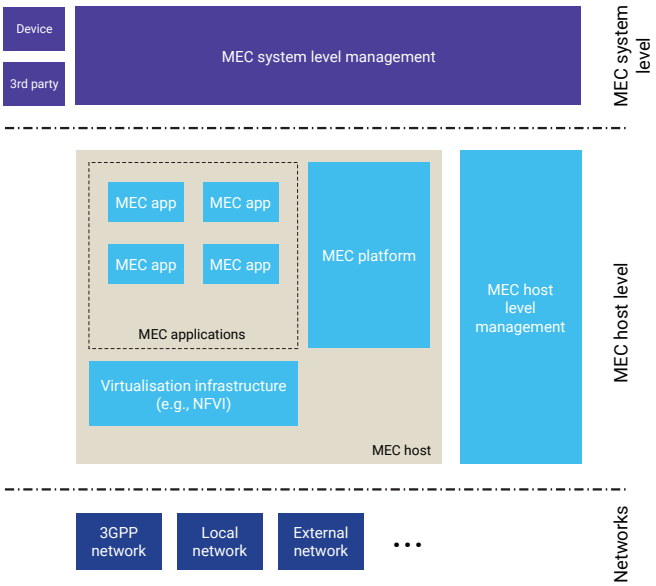


Figure 9 - ETSI ISG MEC Framework

3GPP has also enhanced the architecture for enabling edge computing. The corresponding management capabilities and services have been defined, as depicted in Figure 10 [45]. The alignment and mapping of 3GPP edge computing enabling architecture and ETSI ISG MEC architecture are underway.

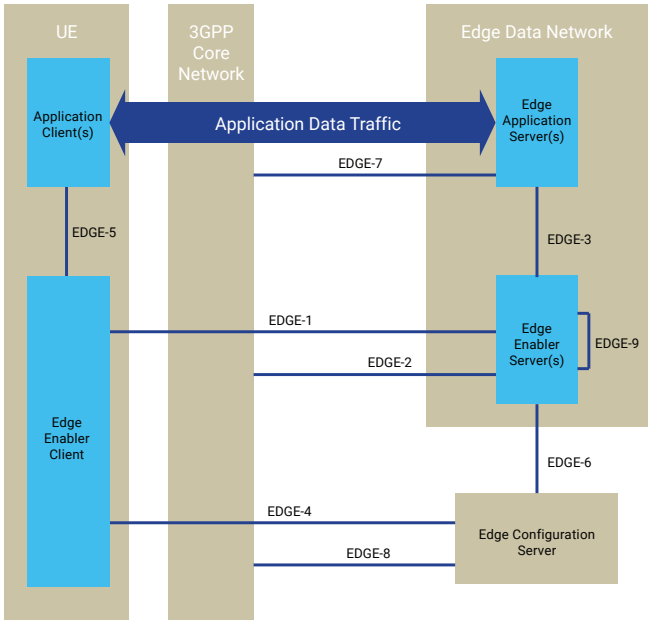


Figure 10 - Architecture for enabling edge applications

2.6.3 6G Wide-Area Cloud Evolution

The evolution of cloud in 6G is expected to be:

1) Distributed across devices, on premises, and in cell sites, edge sites, core sites, and central data centers, as depicted in Figure 11 [46]. This distributed and ubiquitous computing allows the processing of data closer to the source to achieve ultra-low latency, reduction of the end-to-end data transmission, and utilization of specialized compute resources for specific workloads.

2) Convergence of communication, computing, and data services [47]. Both the NFs and the applications can run on the cloud. They can be orchestrated and controlled jointly to offer the unique capability of meeting the requirements for both communication and computing perspectives, but without neglecting or sacrificing any one requirement.

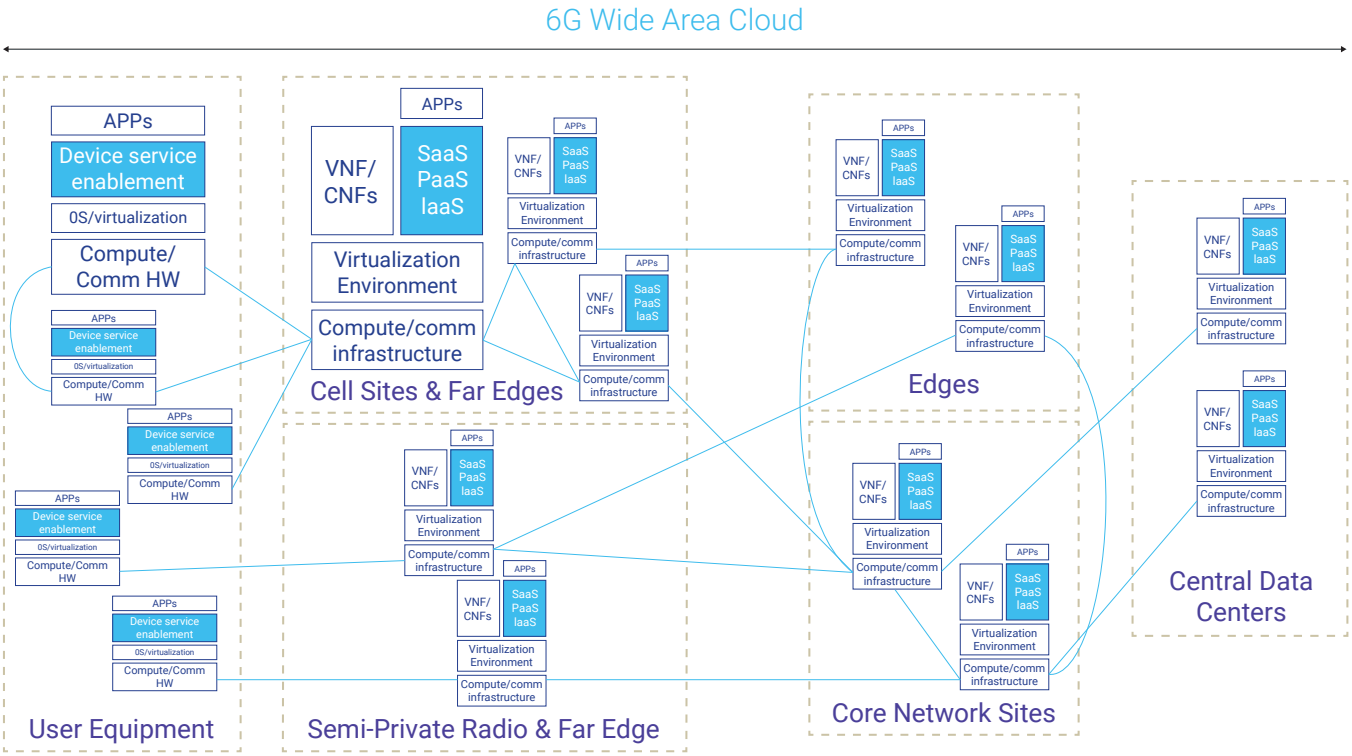


Figure 11 - 6G Wide-Area Cloud

2.6.4 M&O of 6G Cloud Systems

The performance of the 6G system offered to end users will depend on both communication and computing capabilities. The virtualization environment in the 6G cloud provides the virtualized resources to both NFs and applications. Due to the convergence of communication and computing, the cloud M&O system also needs to take care of both aspects jointly, especially for the following capabilities and aspects, which are essential and important areas for further research:

- Resource management
- Workload deployment
- Monitoring and analytics
- Healing
- Scaling
- Traffic routing management
- Management of cloud services
- APIs for management services
- System architecture considerations

2.6.4.1 Resource Management

Various types of compute resources are available today, and new types of compute resources are emerging to tackle increasingly complex or dedicated computing tasks. Each type of resource has its specific advantage for processing certain types of workloads. The compute resource may be virtualized resource or physical resource (i.e., bare-metal hardware).

To allow the cloud M&O system to allocate appropriate compute resources for the workloads and deploy them onto the appropriate compute resources, the variety of compute and storage resources need to be registered automatically. They must also be easily discovered and monitored by the cloud M&O system. The resource type, capability, location, and energy efficiency level/indicators are required for registration and discovery.

A workload requires a certain amount of computing capability (e.g., FLOPS) to meet the performance requirements. The computing capability that each compute resource could offer depends on various factors, such as the number of processing cores, clock rate, memory size, and temperature. To facilitate the resource allocation for the workload, the cloud M&O system needs to be able to calculate, measure, and verify the actual computing capability of each kind of the heterogenous resources when processing a specific type of workload. One challenge is how to quantify and measure the computing capability of diversified resources for different types of workloads.

2.6.4.2 Workload Deployment

One fundamental capability for cloud orchestration is workload deployment. The 6G cloud workload includes NFs and applications. The applications can be further categorized into applications provided by Application Service Provider (ASP) and consumed by end users, and applications provided by the end users (e.g., the end user offloads some workload to the cloud).

The cloud M&O system must be able to automatically deploy the workloads onto the appropriate compute resources.

That means the cloud M&O system needs to recognize the workload type and the communication and computing requirements, and then select the most appropriate resource to achieve the best performance for deployment. If both requirements cannot be met, a policy or instruction is needed to indicate which requirements prevail over the other.

For the NFs, it is desirable that the consumer (e.g., operator) requests to deploy a network or subnetwork that consists of multiple NFs (e.g., core NFs) or network units (e.g., O-CU, O-DU) with the connectivity and topology requirements. The cloud M&O system needs to be able to automatically deploy these workloads and establish the connectivity at one time.

For the ASP-provided applications, it is also very common that the consumer (ASP) requests to deploy multiple application instances in the distributed cloud at one time. One application may be composed of multiple microservices (MS). The cloud M&O system must be able to automatically deploy these workloads and establish required communication between the microservices at one time, as depicted in Figure 12.

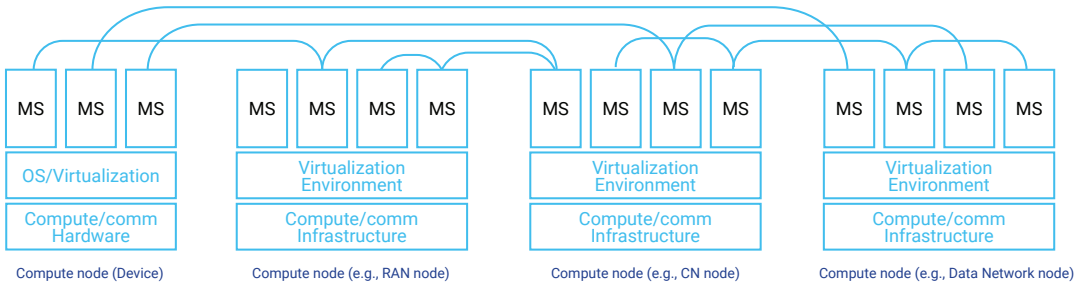


Figure 12 - Workload deployment in 6G wide-area cloud

Storage requirements also need to be considered. An example is whether permanent storage or ephemeral storage is needed for the workload, the required capacity, and data locality, etc.

Another factor that needs to be considered is whether the workload can share some resources with others (to a certain extent) or whether no sharing is allowed.

After the successful deployment or update of application(s), the cloud M&O system must make the application instances information available to session control. This is where the service chaining between those workload instances happens for each service instance.

2.6.4.3 Monitoring and Analytics

The communication system has very rigorous requirements for reliability and performance. With the convergence of communication and computing in the 6G cloud, the performance and the healthiness of the 6G cloud system need to be constantly monitored. To facilitate pinpointing the root cause of performance bottleneck, the cloud system needs to be monitored across the infrastructure, virtualization, NF, and, in some cases, the application layers. This requires collecting and reporting observed data for each layer and correlation of the data across the layers.

For 5G, the performance metrics can be collected and reported at the seconds level, enabled by the data streaming solution [18], [48]. For 6G, it is expected that performance metrics on both communication and computing will be used both for orchestration and for session control. Therefore a solution that allows real-time or near-real-time performance data collection and reporting is required and needs further research.

The monitored data needs to be able to reflect the compute status and performance, as well as the communication status and performance that affect the end users' experience.

With the service mesh of multiple microservices forming the application, it is important for the cloud M&O system to allow consumers to monitor the end-to-end performance rather than only the performance for each individual microservice.

Furthermore, to support intelligent, autonomous, and efficient orchestration and control, analytics for the monitored data is also indispensable for data correlation, analysis, statistics, and prediction.

2.6.4.4 Healing

Healing is a very fundamental capability for the cloud system to ensure reliability and fault tolerance. Traditionally the healing of the workload was basically done with consideration of the compute resource and capabilities. With the integration of communication and computing in the 6G cloud, the healing must take the capabilities and performance of both aspects into account. When the healing is done across compute nodes, the selected target compute resource for healing must meet the requirements for both communication and computing. The service continuity during the healing must be assured to the greatest extent.

2.6.4.5 Scaling

Scaling is the capability to ensure that the workload is allocated with the appropriate amount of compute resources. The scaling is to assure adequate resources for the workload and in the meantime to avoid excessive resource allocation to improve the resource utilization efficiency. The scaling could be allocating more resources for the workload via scale out (creating more instance(s) for the component of the workload) or scale up (adding more resources to the existing instance(s)), reducing the resources via scale in (deleting some instance(s)), or scale down (decreasing the resources for the existing instance(s)). The scale out and scale in have been supported but scale up and scale down are implausible for Virtual Machine-based solutions, while all of these kinds of scaling are supported for container-based solutions.

The scaling may occur within a compute infrastructure or across the compute infrastructures.

The scaling may be triggered automatically (i.e., auto-scaling) by the cloud M&O system or by a consumer on demand.

For scaling in the 6G cloud, the resource requirements and utilization efficiency must be met, and the communication performance must be achieved.

2.6.4.6 Traffic Routing Management

One key aspect of communication and computing convergence is routing the user's data via the best communication path to the most appropriate compute node. The traffic routing is conducted by control plane and/or user plane NFs according to the policy (e.g., performance or energy-efficiency priority) enabled by the cloud M&O system.

2.6.4.7 Management of Cloud Services

The services provided by cloud system can be in different forms, such as Infrastructure as a Service (IaaS,) Platform as a Service (PaaS), and Software as a Service (SaaS).

For IaaS, the virtualized or containerized resources are provided as a service to the consumer. By this service, the consumer may request a specific amount of cloud infrastructure and then deploy and manage the workload by its own solutions.

For PaaS, the cloud infrastructure, together with M&O capabilities for deploying, scaling, healing, monitoring of the workloads, and service mesh and traffic management for the applications, are all provided as a service.

For SaaS, some specific software (e.g., an NF, application) are provided as a service. The provider needs to allocate appropriate computing resources, manage and orchestrate the software in the cloud, and assure the performance of the software.

The cloud M&O system must be able to enable and manage these different types of cloud services.

2.6.4.8 APIs for Management Services

For different types of cloud services (i.e., IaaS, PaaS, and SaaS), the corresponding management capabilities described above need to be provided to consumers via management services.

To realize the management capabilities, the cloud M&O system may need to monitor a lot of data and make various configurations from different aspects (resource, networking, storage, functionality, etc.). These tasks can be very complex for consumers.

In order for the consumer to use the management capabilities more efficiently and more easily, the APIs for the management services exposing these management capabilities need to be easy to use. The consumer only needs to provide abstracted and high-level requirements or instructions (e.g., intent). The system will automatically translate the high-level requirements or instructions to the detailed configurations and then execute them in the cloud.

The cloud M&O system also must enable consumers to monitor and evaluate the performance of the management services and adjust the high-level requirements or instructions when necessary.

2.6.4.9 System Architecture Considerations

In 5G, both the 5GC and management plane are defined in the SBAs. However, the architecture for 5GC and management plane are sort of isolated and independent.

In 6G, it is expected that the cloud M&O related functions are still defined in SBAs. The network and management systems in 6G need to coordinate to enable the orchestration and traffic routing. For instance, the performance data need to be consumed by both networks and management systems.

2.6.5 Challenges and Research Directions

The convergence of communication and communication in the 6G wide-area cloud brings new requirements for managing and orchestrating the cloud-native and computing-native system. The M&O system must enable, provide, and manage the different forms of cloud services (IaaS, PaaS, and SaaS). The joint considerations of communication and computing capabilities and performance are required throughout the whole lifecycle of workload and resource management. A new challenge is how to accurately measure the capacity and performance of heterogeneous and continuously evolved hardware resources for processing traffic and workloads. The traffic route also needs to be managed with the appropriate policy to ensure the QoS, especially to support multi-stream routing for distributed computing. More importantly, cognitive solutions are needed to ensure the optimal end-to-end performance between the UE and application(s), including both communication and computing processes.

2.6.6 Conclusions

Integration of computing service into the 6G system creates an unprecedented opportunity to offer unmatchable end-to-end performance for computing-intensive applications. M&O are the key components of the cloud-native system. The convergence of communication and computing in the wide-area and distributed cloud poses new challenges to this area. Extending the M&O capabilities with joint consideration of communication and computing is essential to enable efficient and ubiquitous computing in the 6G system.

2.7 6G System Digital Twin

2.7.1 Overview

The 6G infrastructure is expected to support a level of trustworthiness not seen in previous generations. To enable a diverse set of applications and use cases, 6G technologies need to provide flexibility and transparency. A 6G system digital twin digitally represents a 6G system or subsystem, including physical and logical parts. There is a long history of using digital representations in industry and academia to model and simulate systems for feasibility and performance evaluation. A digital representation that is regularly synchronized with the real system is often referred to as a digital twin.

Since 3G, the network M&O have been using the NRM to manage mobile communication systems, including deployment, configuration, performance monitoring, fault

supervision, tracing, MDT, data analytics, Self-Organizing Networks (SONs), etc. The M&O of the network are done by controlling and monitoring the NRM data representing the network. This exemplifies an early stage of the digital twin concept.

There is ongoing work in several standards organization on the Network Digital Twin topic [53][54][55] 3GPP recently initiated a study on management aspects of Network Digital Twin for 5G systems. In 6G, the digital twin is expected to support some evolved and additional use cases. Therefore, the impacts of 6G system digital twins on NRMs need to be investigated.

The next sections discuss how the envisioned infrastructure for system digital twin support would largely change how the future 6G systems would be managed in terms of operations, observability, optimization, and enabling and realizing innovations.

2.7.3 System Observability

Observability is a prerequisite for efficiently monitoring and operating a complex system. A 6G system digital twin could potentially provide a holistic framework for observability. 6G systems need efficient technology to represent relevant data in different forms for both human and machine consumption. System automation and programmability, including new techniques (e.g., in the AI/ML area) will rely on data being available for ingestion. The amount of observable data will continually increase, and new tools and architectures for data collection and management will be needed.

2.7.4 System Optimization

A digital twin that provides an adequate level of accuracy can be used to optimize functions, such as radio transmissions and beamforming. A simulation framework build into the digital twin framework can benefit from the availability of data captured and thereby improve and accelerate operations such as dimensioning and feature tuning and training of AI/ML algorithms.

6G infrastructure is expected to offer a much higher level of reliability compared to 5G. A digital twin network system can be used to prepare for and test failure scenarios so that mitigating actions can be taken.

2.7.5 Innovation Platform

A digital twin system can accelerate the time it takes to bring new ideas and requirements implemented in products and systems. Providers could dedicate twin systems for different phases of the development cycle. A 6G twin system would encompass emulators for radio propagation, mobile devices, base stations, and several NFs. New functions and features would be added and tested in the twin system before being rolled out in the real 6G system. Operators could choose to partner with academic institutions and make digital twins available for fundamental research and enable innovations to be assessed in a realistic environment, allowing quick adaptation of new ideas.

2.7.6 System Operation

Operation of cellular wide-area network systems involves near-real-time monitoring of availability and performance of system components, traffic patterns, and security threats, as well as taking the necessary actions to tune the system and mitigate negative effects. Through digitalization, operations are continuously becoming more streamlined and automated. A digital twin system could provide benefits in terms of system operation, evolving the use of testbeds and CI/CD to prepare for new feature rollouts and verification of system upgrades/rollbacks. Digital twin platforms that are continuously fed with data from the real systems could support monitoring of 6G systems while simultaneously providing tooling for a variety of operations, including mitigation of security threat scenarios, recovery from system outages, and AI/ML training.

2.7.7 6G Digital Twin Use Cases

The development of mobile radio technology is highly dependent on digital representations, including advanced modeling and simulations. The vision of the digital twin as a platform where many technologies can merge to provide a truly versatile 6G ecosystem is tantalizing. Such a platform could build on tech from gaming, industry, and academia to provide modeling of cities, including building streets and vehicles. It also could use models based on physics to simulate and optimize propagation of signals and enhance 6G network performance.

2.7.8 Challenges and Research Directions

Digital representations of physical and logical objects are contributing to the increasing amount of data handled by digital systems. More research and development of standards and tools are needed to efficiently handle data in a variety of representations, while adhering to requirements for privacy, performance, and scale. At the same time, digital systems are becoming more complex, such as due to requirements to be more flexible and to support an increasing number of features. More research and innovation are needed in the area of building and maintaining digital systems. Modeling may likely play an important part both in how systems are built and how they are configured and deployed, covering functional aspects, security, privacy, performance, and availability. System interoperability is another critical area where improvements are needed (e.g., in the development of standards and the availability and quality of open-source components). AI/ML is expected to play an increasing part in how digital systems are built. More research is needed to further develop this area in terms of functionality and efficiency. The use of AI/ML in combination with digital twins provides a general framework for innovation.

2.7.9 Conclusions

6G systems are envisioned to be highly automated and flexible. This vision is contingent on advancing technologies for handling complex systems and data-driven architectures involving AI/ML. Frameworks and tools enabling efficient representation of physical and logical system components are key to enable the innovation needed to build, manage, and continually advance 6G systems.

2.8 Energy Efficiency

2.8.1 Overview

6G networks are expected to be more energy efficient than 5G networks because of the new architecture that brings selective infrastructure functionalities to the devices. This means that user devices will be able to act as active network nodes, helping distribute traffic and reduce the overall network energy consumption. 6G devices will be wearable, multi-sensory, and context-aware, with intuitive human-machine interfaces that can recognize gestures and speech. These devices will have a significant impact on the energy footprint of cellular systems, so it is important to develop new power management strategies that are both fundamental and holistic.

5G networks use a lot of power in their RAN components. In the past, researchers have measured the network's energy performance in terms of watts per data unit, data successfully communicated over a unit area, per user, etc. However, this doesn't give a complete picture of the network's performance because 5G RANs use much more power than their predecessors. Currently, there is no standard energy consumption test procedure that offers individual RAN hardware component-level energy cost estimates. This makes it difficult for mobile network operators to estimate energy costs and compare different access technologies. The following are some of the key aspects and best practices to monitor and achieve energy efficiency in 6G systems.

2.8.2 Energy-Efficient System Hardware

Energy-efficient or green 6G system hardware is the most important factor in reducing the carbon footprint. The hardware includes the cloud, computing, RAN, and UE. For example, the 6G system supports dense network use cases and requires more RAN base stations. The UEs that use Massive MIMO, THz, and HCI technologies, will be a major factor in energy cost. Therefore, the access and UE hardware and the access technologies should be energy efficient.

2.8.3 Dynamic Network Operation and Management

The system aims to support extreme and unpredictable user demands such as high data rate, low-latency, dense connectivity, etc. Therefore, management of dynamic hardware duty cycling and optimizing the use of intelligent communication techniques, such as adaptive modulation and coding, power-saving modes, and dynamic frequency selection, are important factors to realize overall energy efficiency.

2.8.4 Recognition of Energy Source

The energy source that will power 6G systems could be supplied by traditional fossil fuels and/or renewable energy. The cost of the energy may be different or the same, but the environmental impact of these energy sources is very different. M&O of the 6G system should consciously balance traditional KPIs, energy efficiency with environmental sustainability metrics of dynamic trade-offs for workload deployment, and run-time optimization such as energy saving, load balancing and traffic steering, all according to the stakeholder's requirements or preferences.

2.8.5 Scalable Network Signaling

6G systems need efficient control signaling schemes to orchestrate the operations of a massive number of system devices and keep track of the system's overall energy cost. This M&O overhead can be significant, especially in dense networks. Therefore, it is important to reduce the network signaling overhead in 6G using distributed and efficient signaling protocols and by optimizing the network architecture.

2.8.6 Challenges and Research Directions

The following are some of the challenges specific to energy efficiency in 6G systems that need to be researched.

One of the biggest challenges for 6G systems is meeting the extreme user experience requirements. These requirements include extremely high data rates, use of high-order MIMO, very low latency, and low-efficiency sub-THz radio frequencies. It is very challenging to meet all these requirements without increasing energy costs. The possible research directions include, but are not limited to, UP protocol stack, intelligent modulation, coding, and physical layer technologies.

The next challenge to achieve 6G energy efficiency is end-to-end power management and operational synchronization. The end-to-end power management strategies can be implemented across the network, from end devices to the core networks to achieve energy efficiency involving:

- Reduction in the activity periods and exploiting device/network off time that ensures that devices and networks are active only when necessary and that they remain in low-power modes or shut down when not in use.
- Synchronization and alignment of network and device inactivity periods by coordinating their activity periods to minimize energy waste. This also includes opportunistic communication.
- Effective utilization and smart management of network components, computing, communication, storage, and other network resources within certain proximity, edge, or the cloud.

2.8.7 Conclusions

6G networks are expected to bring large improvements in the energy efficiency of the underlying hardware, radio interface, and signalling. But the network will also face more demanding requirements, so optimized and intelligent use of resources to achieve goals is necessary. M&O must ensure the efficient matching of resources (including its own resource needs) to demands while optimizing system sustainability.

2.9 Security and Privacy

2.9.1 Overview

The previous generation network systems were designed to provide various fundamental security mechanisms exclusively for aspects such as mutual authentication, communication security establishment (such as confidentiality, integrity, and replay protection), key management, and authorization aspects. The management domain in such cases helps to

configure the prerequisites for the functioning of the related security mechanisms. But the heterogeneity and varied deployment options of the network come with more virtualization and allow the functions to be distributed across different cloud infrastructure and locations for better flexibility and scalability of the network services. The heterogeneity of the communication infrastructure makes it difficult to protect the network resources with traditional perimeter-based security.

The evolving zero-trust cyber security paradigms require the defense to be moved from the static network-based perimeters and focus on more fine-grained security controls at the level of users, assets, and resources, respectively. Here, trustworthiness is the new security approach that plays a vital role in addition to other existing security mechanisms to enable a defense-in-depth strategy to protect the network resources from both external and internal threats. The overview of the factors that can enable trustworthiness in 6G systems, the complexity involved, data privacy, and the best practices are described in the sections below. In addition to that, the threat model for AI/ML and its risk assessment is one the most important challenges in 6G networks because AI/ML has become the key enabling technology for all network and security-related analytics (e.g., for UE, 5G RAN, core). Therefore, a dedicated effort is needed to identify AI/ML-related threats and risks and recommendations for the potential security mechanisms to mitigate and safeguard the 6G system.

2.9.2 Trustworthiness Enablers for End-to-End 6G Systems

Zero-trust security in principle assumes no implicit trust (e.g., over any asset, network resource or end-user device based solely on its physical location or ownership). Instead, it recommends always verifying (i.e., to continuously analyze and evaluate the risk associated to the assets, resources, and services to estimate the respective trustworthiness). Once an attacker breaches the perimeter security, further lateral movement of the attack will be unhindered. Thus the need to adopt a zero-trust security approach becomes more critical. Meanwhile, it can be complementary to the perimeter security to provide 6G with a more trustworthy and resilient security system.

In general, a zero-trust architecture is uses a set of zero-trust security principles as described by NIST [49], NSA [50], CISA, [51] and so on. More specifically, zero-trust security principles can be applied to five pillars: [3] Identity, Devices, Networks, Applications & Workloads and Data. According to NIST SP 800-207 [49], there are seven zero-trust basic principles or tenets:

- Consideration of data sources and computing services as resources.
- Securing all communications regardless of network location.
- Access to individual resource granted on a per-session basis.
- Access to resources based on dynamic policy, which also considers behavioral and environmental attributes.
- Monitoring and measuring of integrity and security posture of all owned and associated assets.
- All resource authentication and authorization by dynamic and strict enforcement for access control (i.e., by scanning and assessing threats, adapting, and continually reevaluating the trust in ongoing communication).
- Collection of information about the current state of assets, network infrastructure, and communication and using it to improve the security posture.

Each of these seven security principles can be analyzed and applied to all domains of the 6G system that encompasses different resource and assets related to the end user devices, RAN, core network, application functions and servers, data storage/repository, management functions, etc. Based on the type of zero-trust security principle implemented, the type of enabling technology and the related security mechanism may differ specific to each of the domain based on the need and capability of the associated asset/resource. For example, with authentication, it can be any Extensible Authentication Protocol (EAP) method like EAP-Transport Layer Security (TLS) or EAP-Authentication and Key Agreement (AKA), or a 3GPP native protocol like 5G AKA or any enhancement of it for 6G. Authorization can be based on OAuth access token/ local authorization/subscription-based verification and access control etc. For continuous security evaluation and monitoring, if any asset in 6G system experiences abnormal behavior attempts (e.g., with malformed messages that deviate from the expected message format, or with message floods, excess resource consumption, excess service loads) toward it, such data can be collected and analyzed with AI/ML to check if there exists any attack trace or if it's due to a minor error.

The management services offered by the OAM domain already assists in data collection related to resource utilization, network performance, etc. from various managed entities in the network for non-security reasons. But this data can also be used for security monitoring along with the other security specific data described above. Further monitoring and measuring integrity of the asset can be enabled with remote attestation to verify the evidence specific to the target environment (e.g., configuration information, profiles) to check if the state of the asset is intact or deviates from the expected state.

Overall, the management domain can play a significant role in assisting in collection of specific data from each of the managed entity, which can be used by a function or tool, such as Security Information and Event Management (SIEM), to analyze and identify the risks and attack traces if exists. Furthermore, the security evaluation and monitoring results can be utilized to configure security policies to apply fine-grained access control decisions in the network to prevent any compromised asset from continuing operations and to prevent the threat's lateral movement. Meanwhile, such results can be used by the management domain to automate the security actions (e.g., impacted function sandboxing, replacement, and so on, as applicable.) and can further use the insights to improve the orchestration decisions. 5G Advanced systems have already started exploring ways to adapt zero-trust security principles [52]. The 6G system

design can integrate zero-trust security principles right from the initial phase to realize inbuilt trustworthiness in the system and the services toward a zero-trust architecture.

The management domain plays a critical role in data privacy because it can be primarily involved in collection, management, and aggregation of data (e.g., related to abnormal behaviors as security logs/reports, metrics, alarms) from the managed entities in different 6G system domains (such as RAN, core). Therefore, the 6G system should govern data privacy because data collected from different network entities can be privacy sensitive (e.g., subscriber identifiers, names, service information, MAC address, cell identifier, network topology) and it would fall under different jurisdictions and regulatory requirements. If such data is shared with applications across geographical boundaries, country-specific privacy requirements can be violated. To govern data privacy, sufficient privacy-preservation methods (e.g., based on operator policies) can be applied in the managed entities during the data collection process.

2.9.3 Challenges and Research Directions

Some of the challenges specific to zero-trust adoption is to design security mechanisms that do not demand more resources and computation. For example, continual data collection over the end-to-end system to perform repeated security evaluation and monitoring can be as resource intensive as the remote attestation verifications. There should be sufficient conditions defined to collect data only when things deviate from the normal or expected behavior. Based on the needs, there can be conditions that trigger data collection and security evaluation for periodic or occasional monitoring. There can be conditions defined to invoke attestation verifications of the functions and entities (e.g., during registration and configuration updates) to limit the frequency of attestation verification process.

Similarly, collected data can include any message or information that is gathered over any interface in the end-to-end network or gathered from a network function/entity. If there is sensitive data that is being collected, then exposing such data to a security evaluation and monitoring function that resides in a different geographical location may lead to violations of regional data protection acts or regulations. In such cases, sufficient privacy preservations need to be applied before exposure for processing related to security evaluation and monitoring. Another prime challenge is to precisely model the AI/ML threat systems in 6G. When threat modeling AI/ML systems, it is important to consider the following:

- **Unique characteristics of AI/ML systems:** AI/ML systems are often complex and opaque, which can make them difficult to threat model. Additionally, AI/ML systems can be trained on data that is sensitive or confidential.
- **Potential threats:** AI/ML systems can be vulnerable to a variety of threats, including adversarial attacks, data poisoning, and model stealing.
- **Attack surface:** The attack surface of an AI/ML system can be large and complex, including the training data, the model itself, and the deployment environment.

The following are the key considerations to precisely model the AI/ML threat models:

- The very first step is to identify which AI/ML assets in 6G systems need to be protected. This may include the training data, the model itself, and the deployment environment.
- The next step is to find the threats that could adversely impact the AI/ML system. This may include adversarial attacks, data poisoning, and model stealing. In addition to that, it is important to recognize the points of attack in the AI/ML system where an attacker could exploit a vulnerability or misconfiguration.
- Finally, there is the need to assess the risk of each threat to the AI/ML system in 6G systems. This should consider the impact of a successful attack and the likelihood of occurrence and then highlight the recommendation controls to mitigate the AI/ML risks and to secure AI. This may include technical controls, such as encryption and authentication, as well as operational and management controls, such as security training and risk assessment.

6G M&O needs to be updated to support zero-trust security principles This means that M&O must continuously assess the trustworthiness of all devices, users, network entities, and applications. M&O systems may also need to be able to dynamically enforce security policies based on this assessment. Here are some specific changes that are needed in 6G M&O to improve security and privacy:

- **Support for zero-trust security principles:** M&O systems must be able to implement zero-trust security principles by continuously assessing the trustworthiness of all devices, users, network entities, and applications. This can be done by collecting data about device behavior, user activity, and network traffic etc.
- **Dynamic policy enforcement:** M&O systems must be able to dynamically enforce security policies based on the trustworthiness assessment. This means that M&O systems must be able to quickly and easily change access control rules and other security configurations.
- **Integration with AI/ML systems:** M&O systems can be integrated with AI/ML systems to improve security and privacy. For example, AI/ML systems can be used to detect anomalous behavior and malicious activity. AI/ML systems can also be used to generate personalized security policies for each device, user, and application.

There are multiple new 6G functions that will rely on M&O for improved security and privacy. These include:

- **Network slicing** allows operators to create multiple virtual networks on a single physical infrastructure. This can be used to isolate different types of traffic and improve security. M&O systems can be used to provision and manage NSs.
- **Edge computing** brings computing and storage resources closer to end users. This can improve performance and reduce latency. However, it can also

introduce new security risks. M&O systems can be used to secure edge computing resources and protect data.

- **AI and ML** are expected to play a major role in 6G networks. However, AI and ML systems can also be vulnerable to attack. M&O systems can be used to monitor and manage AI and ML systems to detect and prevent attacks.

Overall, M&O will play a critical role in improving security and privacy in 6G networks. By implementing the changes described above, M&O systems can help to protect 6G networks from attack and protect the privacy of user data.

2.9.4 Conclusions

End-to-end 6G systems and M&O with the right integration of security and privacy features specific to trustworthiness, secure intelligence (secure AI/ML), and privacy-protected data processing have the potential to emerge as the robust and secure communication system. Furthermore, despite the heterogeneous deployments and various third-party service integrations, these features can enable 6G systems to offer the most secure and diverse services to meet various individual needs and business needs. They also help M&O to manage the 6G system as a resilient network.

3 CONCLUSIONS AND RECOMMENDATIONS

This report presented nine key areas that are fundamental to the evolution of M&O. It identified key research challenges that need to be addressed within each of these areas. M&O is a key enabler of mobile networks, without which the networks would become inefficient, fragile, unreliable, and unsecure. M&O is the only tool to combat the increasing complexity of future networks. M&O is also envisioned to be a key component to enable and provide some new services and capabilities (such as distributed computing) for 6G systems.

Continued research and investigation are encouraged in the areas identified as robust and flexible M&O will be required regardless of the eventual functional capabilities of a 6G system.

4 ABBREVIATIONS AND ACRONYMS

3GPP	3rd Generation Partnership Project
5GC	5G Core
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
AutoML	Automated Machine Learning
CapEx	Capital Expenditure
CD	Continuous Deployment
CI	Continuous Integration
CNN	Convolutional Neural Network
CSI	Channel State Information
CT	Continuous Training
DaaS	Data as a Service
DevOps	Software development and Information Technology operations
EAP	Extensible Authentication Protocol
EAP-AKA	EAP Authentication and Key Agreement
EAP-TLS	Extensible Authentication Protocol Transport Layer Security
eMBB	Enhanced Mobile Broadband
FL	Federated Learning
GAN	Generative Adversarial Networks
IaaS	Infrastructure as a Service
IoT	Internet of Things
JCAS	Joint Communication and Sensing
KB	Knowledge Base
KPI	Key Performance Indicators
KVI	Key Value Indicators
LCM	Life Cycle Management
LLM	Large Language Models
M&O	Management and Orchestration
MDA	Management Data Analytics
MDT	Minimization of Drive Tests
MEC	Multi-Access Edge Computing
MIMO	Multiple-Input and Multiple-Output
mIoT	Massive IoT
ML	Machine Learning
MLOp	Machine Learning Operation
MnS	Management Service

4 ABBREVIATIONS AND ACRONYMS

NF	Network Function
NFV	Network Function Virtualization
NG-RAN	Next Generation Radio Access Network
NRM	Network Resource Model
NS	Network Slice
NWDAF	Network Data Analytics Function
OAM	Operations, Administration, and Maintenance
O-CU	O-RAN Central Unit
O-DU	O-RAN Distributed Unit
OpEx	Operational Expenditure
OT	Operational Technology
PaaS	Platform as a Service
PM	Performance Metric
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RL	Reinforcement Learning
SaaS	Software as a Service
SBA	Service Based Architecture
SBMA	Service-Based Management Architecture
SDO	Standards Development Organization
self-CHOP	self-Configuring, Healing, Optimizing, and Protecting
SIEM	Security Information and Event Management
SL	Supervised Learning
SLA	Service-Level Agreement
SON	Self-Organizing Network
TinyML	Tiny Machine Learning
TL	Transfer Learning
TMN	Telecommunications Management Network
UE	User Equipment
URLLC	Ultra Reliable and Low Latency Communications
USL	Unsupervised Learning
VAE	Variational Autoencoder
VoIP	Voice over Internet Protocol
VR	Virtual Reality
XR	Extended Reality

5 REFERENCES

1. Next G Alliance. (2022). 6G Technologies. https://www.nextgalliance.org/white_papers/6g-technologies/.

2. Meriem, T. B., Chaparadza, R., Radier, B., Soulhi, S. LozanoLópez, S., Prakash, A. (2016). GANA - Generic Autonomic Networking Architecture. ETSI. ETSI White Paper No. 16. https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp16_gana_Ed1_20161011.pdf.

3. Ziegler, V., Yrjola, S. (2020). 6G Indicators of Value and Performance. 2020 2nd 6G Wireless Summit (6G SUMMIT), pp. 1-5, https://doi.org/10.1109/6GSUMMIT49458.2020.9083885.

4. 3rd Generation Partnership Project (3GPP). Management and orchestration; Management Data Analytics. (3GPP TS 28.104). https://www.3gpp.org/dynareport/28104.htm.

5. 3GPP. Architecture enhancements for 5G System (5GS) to support network data analytics services. (3GPP TS 23.288). https://www.3gpp.org/dynareport/23288.htm.

6. 3GPP. NR; NR and NG-RAN Overall description; Stage-2. (3GPP TS 38.300). https://www.3gpp.org/dynareport/38300.htm.

7. 3GPP. NG-RAN; Architecture description. (3GPP TS 38.401). https://www.3gpp.org/dynareport/38401.htm.

8. 3GPP. Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface. (3GPP TS 38.843). https://www.3gpp.org/dynareport/38843.htm.

9. 3GPP. Management and orchestration; Artificial Intelligence / Machine Learning (AI/ML) management. (3GPP TS 28.105). https://www.3gpp.org/dynareport/28105.htm.

10. 3GPP. Draft CR AIML_MGMT - TS 28.105; Enhancements for AI-ML management. (DraftCR S5-235988).

11. 3GPP. Management and orchestration; 5G performance measurements. (3GPP TS 28.552). https://www.3gpp.org/dynareport/28552.htm.

12. 3GPP. Telecommunication management; Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN). (3GPP TS 32.425). https://www.3gpp.org/dynareport/32425.htm.

13. 3GPP. Management and orchestration; 5G end to end Key Performance Indicators (KPI). (3GPP TS 28.554). https://www.3gpp.org/dynareport/28554.htm .

14. 3GPP. Telecommunication management; Subscriber and equipment trace; Trace control and configuration management. (3GPP TS 32.422). https://www.3gpp.org/dynareport/32422.htm.

15. 3GPP. Telecommunication management; Subscriber and equipment trace; Trace data definition and management. (3GPP TS 32.423). https://www.3gpp.org/dynareport/32423.htm.

16. 3GPP. Telecommunication management; Quality of Experience (QoE) measurement collection; Control and configuration. (3GPP TS 28.405). https://www.3gpp.org/dynareport/28405.htm.

17. 3GPP. Telecommunication management; Quality of Experience (QoE) measurement collection; Information definition and transport. (3GPP TS 28.406). https://www.3gpp.org/dynareport/28406.htm.

18. 3GPP. Management and orchestration; Generic management services. (3GPP TS 28.532). https://www.3gpp.org/dynareport/28532.htm.

19. 3GPP. Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and stage 3. (3GPP TS 28.541). https://www.3gpp.org/dynareport/28541.htm.

20. *3GPP. Study on AI/ML management. (3GPP TR 28.908). https://www.3gpp.org/dynareport/28908.htm.*

21. *Shafique, M., Theocharides, T., Reddy, V. J., and Murmann, B. (2021). TinyML: Current Progress, Research Challenges, and Future Roadmap. 58th ACM/IEEE Design Automation Conference, pp. 1303-1306. https://doi.org/10.1109/DAC18074.2021.9586232.*

22. *O-RAN Working Group 2. (2023). O-RAN Non-RT RIC Architecture 4.0. (O-RAN.WG2.Non-RT-RIC-ARCH-R003-v04.00).*

23. *O-RAN Working Group 3. (2023). O-RAN Near-RT RIC Architecture 5.0 (O-RAN.WG3.RICARCH-R003-v05.00).*

24. *O-RAN Working Group 1. (2023). O-RAN Use Cases Detailed Specification 12.0. (O-RAN.WG1.Use-Cases-Detailed-Specification-R003-v12.00).*

25. *Ericsson. (2023). Trustworthy AI - What it means for telecom. Ericsson White Paper. https://www.ericsson.com/en/reports-and-papers/white-papers/trustworthy-ai.*

26. *National Institute of Standards and Technology. (2020). NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management. U.S. Department of Commerce. (NISTIR 8312). https://doi.org/10.6028/NIST.CSWP.10.*

27. *Gerald M. Karam et al. (2022). The Evolution of Networks and Management in a 6G World: An Inventor's View. IEEE Transactions on Network and Service Management, Vol. 19, No. 4. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9815023.*

28. *James P.G. Sterbenz et al. (2010). Resilience and Survivability in Communication Networks: Strategies, Principles, and Survey of Disciplines. Computer Networks, Volume 54, Issue 8, pp. 1245 – 1265. http://doi.org/10.1016/j.comnet.2010.03.005*

29. *3GPP. (2023). Management and orchestration; Management services for communication service assurance; Requirements. (3GPP TS 28.535). Version 17.8.0. https://www.3gpp.org/dynareport/28535.htm*

30. *3GPP. (2023). Management and orchestration; Management services for communication service assurance; Stage 2 and stage 3. (3GPP TS 28.536). Version 17.6.0. https://www.3gpp.org/dynareport/28536.htm*

31. *Camelo, M., et al. (2022). Requirements and Specifications for the Orchestration of Network Intelligence in 6G. IEEE CNCC 1st International Workshop on 6G: Visions, Use Cases and Technologies (6G). https://doi.org/10.5281/zenodo.5771600.*

32. *Péerez-Valero, J., et al. (2022). AI-driven Orchestration for 6G Networking: The Hexa-X vision. IEEE GLOBECOM 2nd Workshop on Architectural Evolution Toward 6G Networks (6GArch). https://doi.org/10.1109/GCWkshps56602.2022.10008726.*

33. *Crowcroft, J. (2023). A True History of the Internet. AI4IP. http://paravirtualization.blogspot.com/2023/08/plenty-can-and-has-been-said-about.html.*

34. *Ahokangas, P., et al. (2023). Envisioning a Future-Proof Global 6G from Business, Regulation, and Technology Perspectives. IEEE Communications Magazine. https://doi.org/10.1109/MCOM.001.2200310.*

35. *Qin, Z., Lu, J., Tong, W., Li, G.Y. (2022). Semantic Communications: Principles and Challenges. ArXiv. https://doi.org/10.48550/arXiv.2201.01389.*

36. *Wheeler, D., Natarajan, B. (2023). Engineering Semantic Communication: A Survey. IEEE Access 11, 13965-13995. https://doi.org/10.48550/arXiv.2208.06314.*

37. *Chaccour, C., Saad, W., Debbah, M., Han, Z., & Poor, H. V. (2022). Less Data, More Knowledge: Building Next Generation Semantic Communication Networks. ArXiv. https://doi.org/10.48550/arXiv.2211.14343.*

38. *Joda, R., Elsayed, M., Abou-zeid, R., Atawia, R., Bin Sediq, A., Boudreau, G., Erol-Kantarci, Melike, Hanzo, L. (2023). The Internet of Senses: Building on Semantic Communications and Edge Intelligence. IEEE Network. https://doi.org/10.48550/arXiv.2212.10748.*

39. *Yang, W., Du, H., Liew,, Z.O., Lim, W.Y.B., Xiong, Z., Niyato, D., Chi, X., Shen, X., Miao, C. (2023). Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2022.3223224.*

40. *Next G Alliance. (2023). 6G Applications and Use Cases. https://www.nextgalliance.org/white_papers/6g-applications-and-use-cases/.*

41. *3GPP. System Architecture for the 5G System. (3GPP TS 23.501). https://www.3gpp.org/dynareport/23501.htm.*

42. *O-RAN Working Group 1. (2023). O-RAN Architecture Description 10.0. (O-RAN.WG1.OAD-R003-v10.00).*

43. *European Telecommunications Standards Institute (ETSI). (2014). Network Functions Virtualization (NFV); Architectural Framework. (ETSI GS NFV 002 V1.2.1). https://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf.*

44. *ETSI Industry Specification Group (ISG) ON Multi-Access Edge Computing (MEC). (2023). ETSI MEC: An Introduction. https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/ETSI-MEC-Public-Overview_Generic.pdf.*

45. *3GPP. Architecture for enabling Edge Applications. (3GPP TS 23.558). https://www.3gpp.org/dynareport/23558.htm.*

46. *Next G Alliance. (2022). 6G Distributed Cloud and Communications Systems. https://www.nextgalliance.org/white_papers/6g-distributedcloud-andcommunicationssystems/.*

47. *Next G Alliance. (2023). 6G Technologies for Wide Area Cloud Evolution. https://www.nextgalliance.org/white_papers/6g-technologies-for-wide-area-cloud-evolution/.*

48. *3GPP. Management and orchestration; Performance assurance. (3GPP TS 28.550). https://www.3gpp.org/dynareport/28550.htm.*

49. *Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture. (NIST SP 800-207). https://doi.org/10.6028/nist.sp.800-207.*

50. *National Security Agency. (2021). Embracing a Zero Trust Security Model. https://media.defense.gov/2021/Feb/25/2002588479/-1/-1/0/CSI_EMBRACING_ZT_SECURITY_MODEL_UOO115131-21.PDF.*

51. *Cybersecurity and Infrastructure Security Agency Cybersecurity Division. (2023). Zero Trust Maturity Model. https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf.*

52. *3GPP. Technical Specification Group Services and System Aspects. (2022). Study on applicability of the Zero Trust Security principles in mobile networks. Release 18. (3GPP TR 33.894). https://www.3gpp.org/dynareport/33894.htm.*

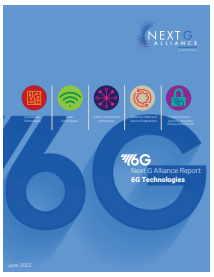
53. *Digital twin network - Requirements and architecture. (ITU-T Y.3090). https://www.itu.int/rec/T-REC-Y.3090-202202-I/en.*

54. *ETSI. Zero Touch and Service Management (ZSM); Network Digital Twin. (ETSI GR ZSM-015). https://docbox.etsi.org/isg/ZSM/Open/Drafts/015_Nwk_DTwin/ZSM-015_Nwk_DTwinv002.docx.*

55. *IETF. Network Management Research Group (nmrg). https://datatracker.ietf.org/rg/nmrg/about/.*

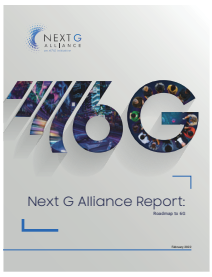
56. *3GPP. Study on data management phase 2. (3GPP TR 28.842). https://3gpp.org/dynareport/28842.htm.*

NEXT G ALLIANCE REPORTS



Next G Alliance Report: 6G Technologies

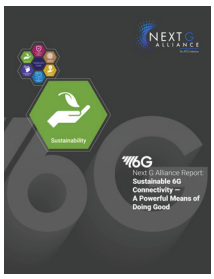
Next G Alliance Report: 6G Applications and Use Cases



Next G Alliance Report: Roadmap to 6G

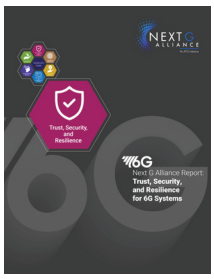
Green G: The Path Toward Sustainable 6G

Next G Alliance Report: AI-Native Wireless Networks



Next G Alliance Report: Sustainable 6G Connectivity — A Powerful Means of Doing Good

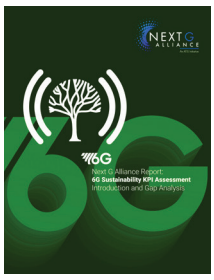
Next G Alliance Report: 6G Distributed Cloud and Communications System



Next G Alliance Report: Trust, Security, and Resilience for 6G Systems

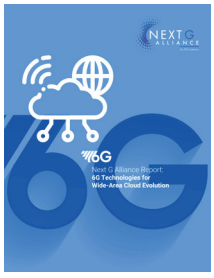
Next G Alliance Report: Digital World Experiences

Next G Alliance Report: Cost-Efficient Solutions

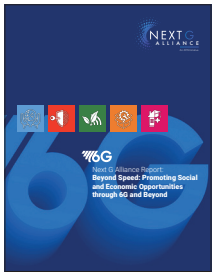


Next G Alliance Report: 6G Sustainability KPI Assessment Introduction and Gap Analysis

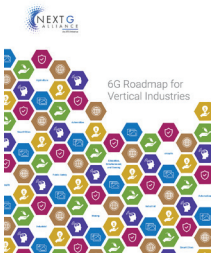
Next G Alliance Report: 6G Radio Technology Part I: Basic Radio Technologies



Next G Alliance Report: 6G Technologies for Wide-Area Cloud Evolution



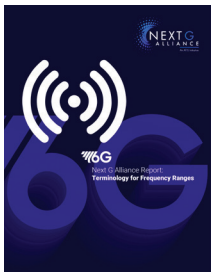
Next G Alliance Report: Beyond Speed: Promoting Social and Economic Opportunities through 6G and Beyond



Next G Alliance Report: 6G Roadmap for Vertical Industries

6G Market Development: A North American Perspective

Next G Alliance Report: 6G Spectrum Considerations



Next G Alliance Report: Terminology for Frequency Ranges

Next G Alliance Report: Distributed Sensing and Communications

Next G Alliance Report: Network-Enabled Robotic and Autonomous Systems

Next G Alliance Report: Multi-Sensory Extended Reality (XR) in 6G

Next G Alliance Report: Shaping Tomorrow: The Evolution of Personalized Digital Experiences Through 6G Technologies

COPYRIGHT AND DISCLAIMER

Published February 2024

Alliance for Telecommunications Industry Solutions
1200 G Street, NW, Suite 500
Washington, DC 20005

Building the foundation for North American leadership in 6G and beyond.

nextgalliance.org