This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

6G-Core-in-the-Loop: Enabling Service and Network Orchestration in a Cloud-Native Ecosystem

Michail Dalgitsis*, Eftychia Datsika*, Cristina Santana Casillas†, and Angelos Antonopoulos*
*Nearby Computing, Barcelona, Spain
†Telefónica S.A., Madrid, Spain

***Abstract*—As 6G networks emerge, the boundaries between communication infrastructure and applications are rapidly dissolving, giving rise to an integrated, cloud-native ecosystem where services and network functions coexist and evolve together. Central to this transformation is the mobile core, which acts as the "brain" of the network, crucial for enabling intelligent and dynamic service delivery through the use of application programming interfaces (APIs), which facilitate seamless interaction between network functions and external applications. In this article, we explore two key dimensions of next-generation mobile core evolution: i) network-aware service orchestration, where real-time core network insights inform and optimize service lifecycle management (LCM); and ii) context-aware core network orchestration, where information from services and network conditions drive adaptive decisions in the cloud-native 6G core, such as network function scaling and slice reconfiguration. Our contribution is threefold: i) we assess the current state and standardization of service–network interplay, ii) we demonstrate proof-of-concept (PoC) implementations highlighting the practical value of "API-driven" integration, and iii) we outline the challenges and opportunities shaping the future of the mobile core in this converged network–service paradigm.**

***Index Terms*—6G Mobile Core, Service Orchestration, Network Exposure, CAMARA, Telco Edge-Cloud Continuum.**

I. INTRODUCTION

The transition from 5G to 6G highlights a significant shift in mobile core network architecture, moving beyond enhanced mobile broadband toward intelligent, ultra-low-latency, and highly adaptive systems. 5G introduced key advancements, such as 5G core (5GC) service-based architecture (SBA) and cloud-native software-defined infrastructures that enabled microservice deployment of core and radio access network (RAN) network functions (NFs) [1]. However, it still faces limitations, such as siloed and rigidly integrated NFs, lack of awareness of hardware or real-time network states, limited flexibility in dynamic scaling, and insufficient orchestration capabilities. To overcome these constraints, the envisioned 6G core (6GC) will consist of loosely coupled, independently managed NFs that dynamically interface to support elastic scalability, integrated orchestration, and real-time adaptability. Built on a cloud-native continuum spanning telco edge and cloud environments, 6GC aims to enable agile service provisioning, support innovative business models, enhance resource optimization, and reduce time-to-market for emerging applications [2].

The 6GC is also expected to adopt an AI-native framework that will manage, orchestrate, and operate all NFs autonomously. In contrast to current architectures, where AI exists as an external add-on, 6G will integrate AI as a core component of the system [3]. With direct access to all NF data, performance metrics and alerts, AI-native systems will drive decisions affecting function execution, resource allocation, prediction, security and service resilience. While the exact specifications are still under development, early research has begun exploring AI-native integration in the core network to enhance existing capabilities, e.g., customized network slicing that is based on near real-time traffic forecasting [4].

The 5GC/6GC cloud-native architecture, characterized by disaggregated, containerized, and stateless NFs, provides a fertile foundation for fine-grained monitoring and control. This architectural shift facilitates the exposure of granular operational data through application programming interfaces (APIs), enabling orchestration platforms to interact more intelligently with the network. By aligning orchestration logic with the inherent flexibility of the 5GC, network-aware service management can be more adaptive, resilient, and scalable. On the other hand, the service behavior, along with underlying network conditions, may affect the deployment and operation of 5GC NFs, which can be configured according to joint service-network context information [5]. This evolution naturally paves the way for more intelligent and autonomous coordination between the network and orchestration layers. Building on these principles, the 6GC is expected to further strengthen this interplay through a *6GC-in-the-loop* perspective, where the core both informs and is influenced by orchestration decisions.

Based on this perspective, this article explores two complementary paradigms: (i) **Network-Aware Service Orchestration (NASO)**, which leverages real-time core insights to optimize application deployment and operation, and (ii) **Context-Aware Core Network Orchestration (CCNO)**, which enables adaptive function scaling, slice reconfiguration, and edge placement driven by service context. The innovation lies in demonstrating how the 6GC can act both as an orchestration target and as an active source of intelligence that guides management decisions, achieving a unified, bidirectional coupling between the service and network domains.

Our contribution is threefold:

(i) We analyze the current state and standardization landscape of the service–network interplay.
(ii) We demonstrate proof-of-concept (PoC) implementations that highlight the practical feasibility and benefits of API-driven integration.

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

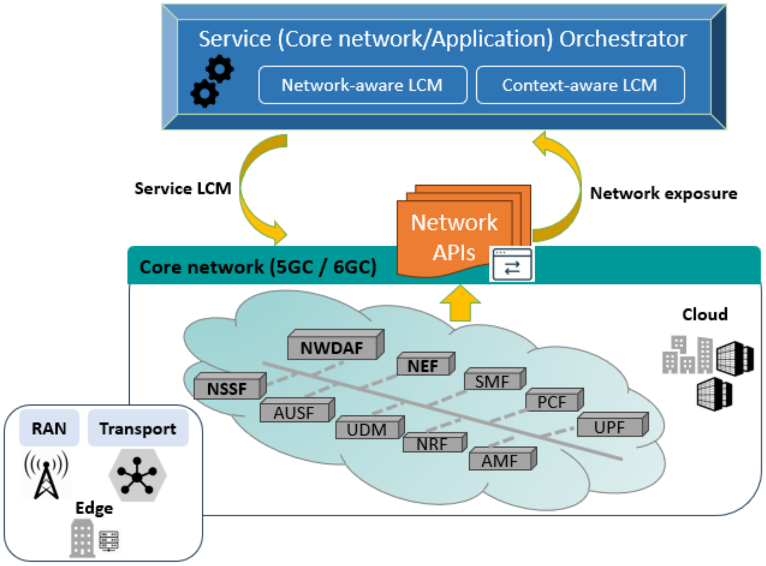

Fig. 1: Service (core network or application) orchestration in 6G network

(iii) We identify key challenges and research opportunities for realizing a tightly converged, intelligent network–service orchestration framework in future 6G systems.

The remainder of this article is organized as follows: Section II describes the 5GC and 6GC architectural principles. Sections III and IV introduce the NASO and CCNO paradigms, respectively, while Section V discusses standardization efforts that support them. Section VI showcases the network-service interplay with concrete PoC use cases. Section VII discusses emerging research challenges that will shape the 6GC, whereas Section VIII concludes the article.

II. 5G CORE AND ENVISIONED 6G CORE NETWORKS

The 5GC is central to network management, offering microservice-based NFs that interact via service-based interfaces using representational state transfer (REST) APIs and HTTP/2.0. Instead of dedicated point-to-point connections, NFs communicate over a logically shared infrastructure using either the Request/Response model or the Subscribe/Notify model to access services. This architecture enables flexible and scalable interactions across core components.

Despite its cloud-native design, 5GC still faces limitations. NFs are often deployed and managed independently, lacking a unified orchestration layer to translate high-level service intents into coordinated actions. This fragmentation hinders automation and end-to-end service management. For example, the Access and Mobility Management Function (AMF) handles user equipment (UE) registration and mobility, the Session Management Function (SMF) manages session setup and quality-of-service (QoS), and the User Plane Function (UPF) routes user data. The Network Exposure Function (NEF) enables external applications to access UE-related data and influence control plane behavior [6], while the Network Data Analytics Function (NWDAF) provides performance insights [7]. These analytics are consumed by the Policy Control Function (PCF), which makes real-time decisions on congestion control, QoS enforcement, and slice prioritization [8]. The Network Slice Selection Function (NSSF) assists in assigning the appropriate slice configuration during session setup. Although NFs communicate through APIs, orchestration remains disjointed, limiting dynamic capabilities such as real-time slicing and low-latency edge computing.

6G aims to overcome these gaps with a service-centric, API-driven core architecture (Fig. 1). It will expose real-time telemetry, QoS control, and slicing as programmable services accessible to third-party applications. The orchestration layer will act both as an *intelligent consumer of network insights*, leveraging real-time state (e.g., topology, load, latency, slice availability) for lifecycle management (LCM), and as a *context-aware decision engine* that optimizes NF behavior, slice assignment, and policy enforcement using session, user, and traffic data. This bidirectional interaction between the core and the orchestration layer reflects a *6GC-in-the-loop* perspective, in which the 6G Core is not only subject to orchestration but also continuously informs orchestration decisions through real-time feedback.

III. THE NETWORK-AWARE SERVICE ORCHESTRATION (NASO) PARADIGM

The dynamic nature of 5G and beyond requires orchestration tightly linked with real-time core network intelligence. NASO uses 5GC insights like load, session state, and mobility to optimize service lifecycle, scaling, and placement. This section examines how core NFs and telemetry improve service management.

Real-time core network telemetry can be directly mapped to key stages of the service lifecycle, i.e., instantiation, configuration, scaling, healing, migration, and termination, enabling

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

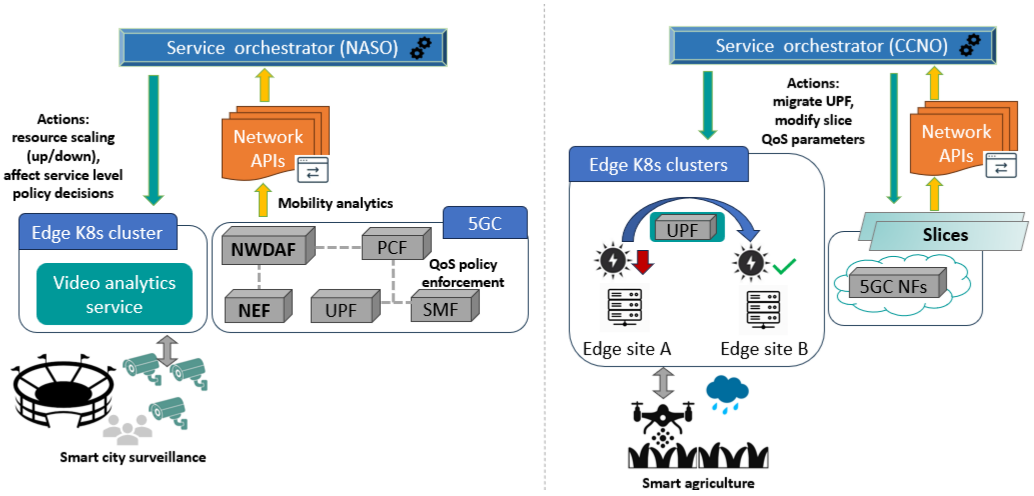

Fig. 2: 6GC-driven orchestration paradigms (left: NASO, right: CCNO)

closed-loop orchestration. For instance, network-aware autoscaling can be implemented by integrating NWDAF analytics on traffic load and session density with the orchestration logic in management and orchestration (MANO) platforms like ETSI NFV-compliant orchestrators or Kubernetes-based core NF managers. Service instantiation decisions can be influenced by current UPF load, ensuring placement in regions with available capacity and low latency to UEs. Healing mechanisms can incorporate real-time fault and degradation events received via service assurance interfaces or fault management functions, allowing service orchestrators to trigger targeted re-deployments or re-routing.

In practice, network configuration information (e.g., UE mobility patterns, 5GC load indicators or subscriber profiles), can be used to enforce resource policies, ensuring service-level objectives are met within their respective logical partitions. By correlating these data with application-level metrics, service orchestrators can implement multi-layer optimization strategies, e.g., proactively scaling services when a surge in UE mobility is detected in a given region. Integration with policy engines, such as the 5GC PCF, also allows for enforcement of orchestration policies based on subscriber profiles or service classes, supporting differentiated handling of services in line with operator-defined service level agreements (SLAs). These capabilities enable adaptive orchestration loops where services are dynamically optimized in real time based on both network state and business intent.

As an illustrative example of network-aware service orchestration, let us focus on a video analytics service deployed at the network edge for smart city surveillance (see left part of Fig. 2). The service is delay-critical, as it processes camera feeds and provides real-time event detection. The service runs on a Kubernetes-managed edge node and is orchestrated by a platform integrated with NWDAF and NEF from 5GC. NWDAF analyzes UE mobility and predicts increased device density near a stadium before an event. This prediction, made available through NWDAF's analytics interface or via NEF if exposure to external domains is required, is consumed by the service orchestrator, which triggers a proactive scaling action.

In particular, the service orchestrator is able to validate available compute and UPF load across edge locations, select an optimal node, and instantiate new pods. At the same time, it interacts with the PCF to influence service-level policy decisions, which are then enforced by SMF and UPF during data flow treatment, for instance, ensuring the surveillance service receives higher resource priority than non-critical workloads. After the event, updated analytics from NWDAF indicate normalized traffic, prompting the orchestrator to scale down excess resources. This example illustrates how dynamic network state and analytics imported by 5GC, combined with policy-driven orchestration, enable adaptive and efficient service LCM. Such integration supports fine-grained decisions that balance latency, resource usage, and SLA adherence across diverse and time-varying conditions.

IV. THE CONTEXT-AWARE CORE NETWORK ORCHESTRATION (CCNO) PARADIGM

As 6G evolves toward full cloud nativity, core orchestration must exploit contextual insights from service behavior and network conditions. Context-Aware Core Network Orchestration (CCNO) enables adaptive decisions by linking user intent, QoS needs, traffic, and mobility data with real-time core state, enabling dynamic actions such as NF scaling, slice adjustment, traffic steering, and edge offloading. CCNO continuously adapts network behavior based on multi-dimensional context collected from both the network and hosted services. "Context" covers i) network conditions (e.g., congestion, topology), ii) slice configuration and utilization (e.g., S-NSSAI data), iii) service requirements (latency, bandwidth), iv) user attributes (mobility, location), and v) environmental factors (energy, weather). Such continuous analysis triggers network reconfiguration, including vertical/horizontal NF scaling, on-the-fly slice adaptation, and NF migration to edge sites.

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

Within a cloud-native 5G/6G core, CCNO realizes real-time, closed-loop decision-making, where orchestration systems adapt the deployment and configuration of core NFs and slices based on continuously evolving multi-dimensional context (as explained above). Orchestrating 5GC NFs is guided not only by internal network telemetry but also by real-time service-layer information. This enables dynamic, context-aware deployments tailored to the performance demands of diverse services. The information provided by services can support dynamic, context-aware orchestration, where core NFs are deployed and adapted to meet the specific demands of various services. Each service has unique requirements for resource allocation and function placement, such as low latency, high bandwidth, or high availability. Service-specific data, such as traffic type, SLAs, or geographical location, facilitate orchestration decisions about where and how to place NFs. For instance, services requiring ultra-low latency, such as autonomous vehicles, will drive the placement of UPF instances closer to the edge. In contrast, latency-tolerant services like video streaming may benefit from placing UPFs in compute-rich regional data centers.

Leveraging distributed observability frameworks and analytics components (e.g., NWDAF extensions, telemetry collectors, ML inference engines), the orchestrator can dynamically scale microservice-based NFs, trigger slice-level adjustments (e.g., modifying QoS flow descriptors or re-allocating resources to S-NSSAIs), and instantiate or migrate NFs closer to users using edge clusters. Implemented through declarative policies (e.g., Kubernetes Operators, ETSI frameworks), CCNO enables SLA-aware, context-driven adaptation across heterogeneous service environments.

To further explain the CCNO paradigm, let us consider a smart agriculture scenario, where a network slice is provisioned to support autonomous drone operations for real-time crop monitoring over a wide rural area (see right part of Fig. 2). Each drone streams high-resolution video and telemetry to edge analytics services running in distributed micro-edge data centers. During midday, telemetry data from energy management systems (e.g., via NEF-integrated environmental sensors) indicates a rapid drop in solar energy availability at one of the edge nodes due to a developing weather front. Simultaneously, the NWDAF observes an increasing CPU load and rising service latency on the corresponding UPF and analytics functions.

In this scenario, the orchestrator is able to trigger a coordinated set of actions, possibly considering predefined energy-efficiency policies and real-time context: i) it migrates the overloaded UPF to a neighboring edge site with stable power and available resources, and ii) it modifies the slice profile by updating its QoS parameters via the PCF to temporarily relax latency requirements for non-critical drone feeds. These adaptations occur without service disruption, guided by real-time telemetry, infrastructure conditions, and service policies.

V. STANDARDIZATION LANDSCAPE REVIEW FOR CONVERGED SERVICE AND CORE NETWORK ORCHESTRATION

The advancement of 5G heavily depends on the availability of standardized interfaces and frameworks that enable real-time, adaptive, and policy-driven decision-making. Ongoing standardization efforts support both NASO, which adapts service deployment and LCM to network state, and CCNO, which orchestrates core NFs based on runtime context such as UE behavior, policy, and infrastructure conditions (Table I).

At the core of this evolution, 3GPP has defined key functions for exposing and analyzing network data. While any 5GC NF can expose APIs, a secure intermediary is required when orchestrators operate across domains. Within 3GPP, the NEF offers northbound APIs for network capabilities and event triggers (e.g., QoS changes, UE location, slice status), and the NWDAF provides analytics such as load forecasting, mobility patterns, and anomaly detection. The PCF complements these by enforcing dynamic policies based on user, subscription, or real-time network state, supporting closed-loop orchestration. These components are standardized to interface with service orchestration platforms, allowing orchestration decisions to be guided by actual network behavior rather than static configurations.

The Common API Framework (CAPIF) [9] standardizes how operators expose network capabilities via APIs to third-party services. It defines common functions for API publication, discovery, invoker onboarding, access control, routing, and event reporting, covering both 3GPP and non-3GPP interfaces. CAPIF, deployable by operators or third parties, enables orchestration platforms to uniformly manage and consume network APIs.

ETSI complements 3GPP by providing a flexible orchestration foundation. The NFV MANO framework enables policy-driven orchestration using telemetry and analytics, integrating fault and performance data via standardized interfaces (e.g., Os-Ma-Nfvo) and supporting cloud-native workloads through Kubernetes-based orchestration and observability pipelines (e.g., OpenTelemetry, Prometheus). The Zero-Touch Service Management (ZSM) framework introduces a closed-loop, intent-based architecture for telemetry ingestion, intent interpretation, and autonomous orchestration [10]. Meanwhile, the Multi-access Edge Computing (MEC) framework defines APIs such as RNIS and location API, offering authorized access to UE location and RAN status at the edge [11].

Furthermore, the TeleManagement (TM) Forum, a global telecom industry group, has introduced Open APIs to standardize interactions between service orchestrators and systems like network inventory, fault and performance management, and policy engines [12]. Its Open Digital Architecture (ODA) offers a modular, microservice-based framework aligned with cloud-native 5G/6G principles for vendor interoperability. Similarly, the open RAN (O-RAN) Alliance, though focused on RAN, supports real-time, analytics-driven orchestration via the RAN Intelligent Controller (RIC), which decouples control

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

from base stations. The near-real-time RIC hosts xApps for low-latency tasks like handovers, while the non-real-time RIC runs rApps for higher-latency functions such as policy control and ML training, regulating xApp behavior [13].

To facilitate the standardized exposure of network capabilities, GSMA Open Gateway defines standardized APIs exposing capabilities like QoS, service availability, location, and device reachability to third-party application developers [14]. The CAMARA project, developed with the GSMA Operator Platform Group, implements these APIs to support cross-operator interoperability via open-source, intent-based APIs [15]. Key APIs include the EdgeCloud set of APIs (for edge zone discovery, compute availability, and service continuity) that enables orchestration platforms to adapt service deployment to edge constraints, the Device Status API (for UE connectivity control), and Location Verification API (for UE geographic presence confirmation), enabling orchestration platforms to adapt dynamically to real-time conditions.

These standardization efforts enable practical implementation of NASO and CCNO. NASO uses real-time network exposure and analytics (e.g., NEF, NWDAF) and unified APIs (e.g., CAPIF, CAMARA) to optimize service placement, scaling, and lifecycle management. CCNO leverages telemetry, observability tools (e.g., Prometheus), and standardized interfaces for adaptive core network decisions. Both rely on standardized exposure, policy-driven automation, and cloud-native orchestration for flexible, context-aware, and interoperable 5G/6G networks.

VI. 6G-CORE-IN-THE-LOOP: SHOWCASING THE SERVICE–NETWORK INTERPLAY IN ORCHESTRATION

To highlight service-core interplay in cloud-native architectures, we first assess a service orchestrator using NASO principles that adapts deployment via real-time core insights. Then, as a CCNO PoC, we explore dynamic resource allocation across slices, reallocating bandwidth between critical and non-critical services based on demand. The evaluation emphasizes qualitative behavior, acknowledging that quantitative outcomes in real networks may differ with varying demand and load dynamics.

A. NASO Proof-of-Concept

For the NASO evaluation, we have developed an event-driven Python simulator modeling a cloud-native service deployment across a distributed edge-cloud topology comprising 5–50 regional nodes and a central data center. Each region hosts 10–50 service instances, with CPU capacities between 4 and 16 vCPUs and service demands ranging from 1–3 vCPUs. Traffic thresholds vary from 50–500 Mb/s, and user activity (10–400 users per region) is randomly generated over 1000 simulation events. Each event's traffic load determines the number of required service replicas, enabling realistic evaluation of dynamic scaling, migration and energy efficiency. NASO decisions are driven by real-time network load, orchestrating where and when services are deployed or migrated across regions. The framework includes components for service autoscaling and edge–cloud migration. Autoscaling adjusts the number of service instances and their resource allocation based on user activity and traffic metrics, and is evaluated in a scenario focused on energy-efficient scaling. The migration component ensures service continuity by relocating workloads to the central cloud when edge resources are limited, and is assessed in a second scenario simulating dynamic workload migration under edge constraints.

Focusing on the first scenario, NASO's network-aware autoscaling component is evaluated based on its ability to minimize overall energy consumption. The energy model takes into account both constant energy (i.e., baseline server power consumption) and dynamic energy (i.e., CPU activity tied to active workloads). We compare NASO to the network-unaware service orchestration (Network-unaware SO), i.e., a naïve approach that lacks any form of autoscaling and deploys all services uniformly, regardless of demand. Simulation results (Fig. 3) show that NASO reduces energy consumption by up to 50% in high-demand scenarios (e.g., with 40 or more services). By selectively deploying services where they are actually needed, NASO significantly outperforms the Network-unaware SO strategy, which wastes energy by maintaining idle service instances across the infrastructure. Interestingly, increasing the number of regions does not always lead to higher energy consumption in the NASO approach, as it deploys services only where there is active demand. As the number of regions increases, user demand becomes more sparsely distributed, and many regions may have few or no active UEs for specific services. Consequently, fewer service instances are deployed in total, and unused edge nodes remain idle, reducing overall energy usage despite the larger infrastructure footprint.

In the second scenario, we evaluate NASO's edge-cloud migration component (i.e., migration strategy), focusing on relocating containerized service instances between edge and central cloud. Migration is triggered by resource exhaustion or network constraints, ensuring service continuity while adapting to dynamic compute and network conditions. The baselines are alternative migration policies. The random service orchestration (Random-SO) migrates services arbitrarily, without considering the CPU load of edge nodes or the UE activity. Next, the high CPU-based service orchestration (High-CPU-SO) prioritizes migrating services with the highest CPU usage aiming to reduce the number of total migrations by offloading resource-heavy services. Based on the assumption that lower CPU usage correlates with lower radio traffic and, consequently, fewer UEs are expected to be affected by migration events, the low CPU-based service orchestration (Low-CPU-SO) migrates services that induce the lowest CPU consumption. The performance results shown in Fig. 4 demonstrate that NASO's migration strategy consistently maintains better service continuity and user experience. The percentage of UEs affected by service migrations reflects the proportion of UEs whose services experience disruption, such as increased latency or temporary disconnection, due to the relocation of services when edge node capacity is exceeded. This metric is calculated by observing, over time, how often UEs in each

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

Standardization initiative	Description and References	Functionality focus	AI/ML Support	Scope
3GPP NEF	Exposes core capabilities and events via APIs. TS 29.522 [6]	Exposure	Partial	Core
3GPP NWDAF	Core analytics (e.g., load, mobility, QoS). TS 23.288 [7]	Analytics	Yes	Core
3GPP PCF	Policy control for context-based decisions. TS 23.503 [8]	Policy Control	No	Core
ETSI ZSM	AI-native, intent-driven, closed-loop orchestration [10]	Orchestration	Yes	Cross-domain
ETSI MEC	LCM, telemetry, multi-cloud orchestration [11]	Orchestration	Partial	Core
TM Forum APIs & ODA	Open APIs and modular orchestration framework [12]	Exposure, Orchestration	Partial	Cross-domain
O-RAN Alliance	Real-time RICs, real-time RAN analytics sharing with core network [13]	Analytics, Orchestration	Yes	RAN
GSMA Open Gateway	Global APIs for QoS, slicing, location, etc. [14]	Exposure	No	Cross-domain
CAMARA	Open-source APIs for service-network integration [15]	Exposure, Orchestration	Partial	Cross-domain

TABLE I: Standardization initiatives (Partial: facilitates or supports AI/ML integration by providing data or APIs but does not perform analytics or model training itself. Cross-domain: enables orchestration across multiple domains (e.g., core network, RAN, edge, cloud, or domains of multiple operators))

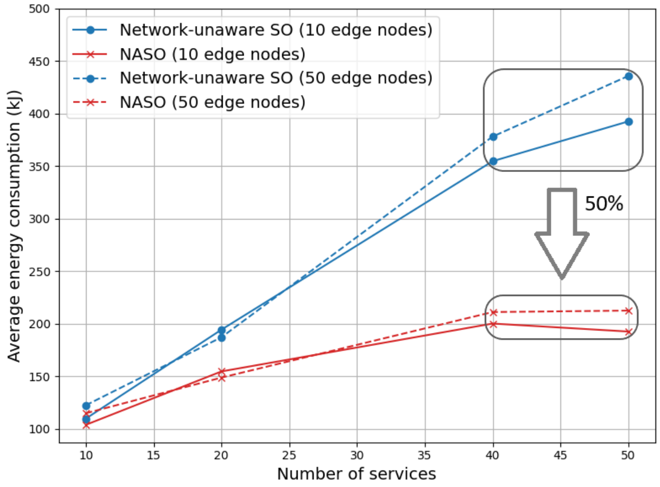

Fig. 3: Energy consumption vs. services

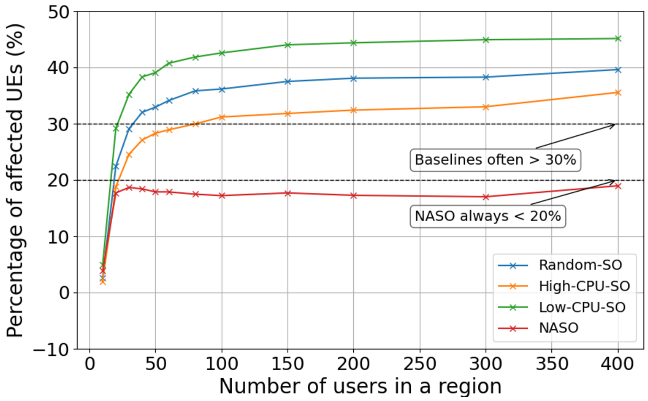

Fig. 4: UEs affected by service migration events

region are affected by these migrations. In each observation window, a UE is marked as affected if the service it is using is migrated. By averaging these counts over all UEs and time windows, the overall impact of service migrations on user experience is captured as a normalized percentage. It is observed that even under heavy traffic, NASO affects fewer than 20% of users, while baseline strategies lead to disruptions for more than 30% of users in most cases. By intelligently offloading workloads to the central cloud only when necessary, NASO enables responsive and efficient service orchestration across the distributed infrastructure.

B. CCNO Proof-of-Concept

To showcase the CCNO paradigm, we consider a dynamic core network slice reconfiguration scenario involving two slices with distinct priorities: one for critical services and another for non-critical services. The proposed approach is compared with a fixed slice allocation (FSA) baseline, where user requests have fixed bandwidth requirements and their distribution across slices changes over time. The orchestrator monitors service performance per time window and triggers reconfiguration if performance drops below a predefined threshold.

We have implemented a lightweight emulation environment with containerized components: a 5GC, a Prometheus-based telemetry module and a Q-learning agent, all deployed on Kubernetes via the service orchestrator. The agent receives real-time telemetry (CPU load, latency, packet drops), recommends slice allocations, and the orchestrator enforces them through API-driven reconfiguration for context-aware, real-time adaptation. It optimizes two slices: critical (priority 0.9) and non-critical (priority 0.1), using synthetic traffic from 50–500 Mb/s. Trained offline, the agent models total traffic dynamically distributed between slices, with states discretized into 10 intervals (100 states). Each action defines a vector of resource percentages allocated to each slice. The reward function prioritizes high acceptance for the critical slice.

The performance comparison between CCNO and FSA baseline, measured by total acceptance ratio of service requests

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

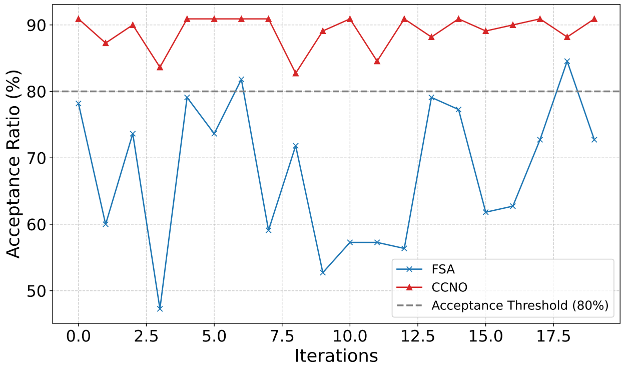

Fig. 5: Total acceptance ratio over a time period

over time, is presented in Fig. 5. The horizontal dashed line at 80% marks the acceptance ratio threshold used for triggering reconfiguration in CCNO. The results demonstrate that CCNO approach consistently maintains acceptance ratios above threshold, whereas the static FSA baseline frequently falls short under varying traffic loads. These observations highlight the limitations of static resource management in environments with fluctuating and heterogeneous slice demands, reinforcing the need for adaptive, context-aware reconfiguration of core network slices.

VII. CHALLENGES AND FUTURE DIRECTIONS FOR 6G MOBILE CORE NETWORKS

The evolution toward a 6G-ready mobile core requires not only new technologies but also addressing key challenges that affect orchestration, governance, and collaboration across stakeholders. While the proposed NASO and CCNO paradigms establish the foundations for intelligent, context-aware orchestration, their real-world adoption depends on overcoming architectural, operational, and business barriers. This section outlines the main challenges and highlights promising directions shaping the future 6G core.

A. Key Challenges

1) Data Exposure and Governance: Mobile network operators (MNOs) and service providers use orchestration platforms to deploy applications across distributed infrastructures, but sharing detailed network and infrastructure data raises governance and security concerns. Data exposure typically occurs in a controlled manner between MNOs and authorized service providers. Initiatives such as CAMARA and the GSMA Operator Platform illustrate how standardized APIs, combined with role-based access control and policy-based authorization, can enable secure and policy-compliant data sharing within the 6G ecosystem.

2) Economic and Trust Implications in Cross-Domain Orchestration: The realization of NASO and CCNO depends on the willingness of MNOs, cloud providers, and service providers to share operational data. Such collaboration introduces economic and governance challenges, as service providers may hesitate to reveal suboptimal performance, and infrastructure providers may fear exposure to SLA audits or contractual risks. Incentive and trust mechanisms, such as revenue-sharing models, independent auditing frameworks, and standardized API marketplaces (e.g., GSMA Open Gateway, CAMARA), can promote transparency and make awareness-driven orchestration viable and sustainable.

3) Resource Management and Service Stability: Dynamic orchestration may cause resource contention when new requests compete with ongoing workloads. NASO and CCNO address this through closed-loop monitoring, policy-based admission control, and slice-aware prioritization. Before scaling or deploying a service, the orchestrator evaluates real-time conditions and potential SLA impacts to maintain fairness and performance. Enhancing these mechanisms with AI-based prediction and adaptive policies will further strengthen future 6G orchestration resilience.

B. Future Directions and Enabling Technologies

Addressing these challenges opens several research directions and enabling technologies for the 6GC. Before outlining them, it is worth framing these directions under a unified orchestration vision. In this context, the two paradigms presented in this article (i.e., NASO and CCNO) can be interpreted as complementary pillars of a unified 6G orchestration framework. NASO represents a network-informed approach, where real-time telemetry from the 6GC guides service lifecycle and placement decisions, while CCNO embodies a service-informed approach, where application and contextual data influence core network configuration and scaling. Together they define a bidirectional orchestration loop in which the 6GC acts both as an orchestrated target and as an intelligent participant. This duality forms the foundation for a generalized 6G orchestration taxonomy along three dimensions: awareness (network, context, intent), scope (service, network, cross-domain) and intelligence (rule-based, closed-loop, AI-native), offering a conceptual roadmap for unified 6G orchestration.

1) Evolution of Network Functions: To support dynamic and intelligent 6G services, NFs must adopt richer, event-driven APIs and improved real-time analytics. Future work should enhance their programmability and interoperability for deeper integration with orchestration platforms and external applications.

2) AI-Driven Automation and Intent-Based Management: As networks grow more complex, manual management becomes impractical. AI-driven automation can analyze and optimize core performance in real time. Future work should develop intent-based models that act autonomously while ensuring transparency and trust.

3) Cross-Domain Orchestration and End-to-End Visibility: As networks become more distributed, orchestration must extend across domains, including the core, RAN, and transport networks. Future orchestration platforms should provide end-to-end visibility and coordination through unified data models, standardized APIs, and governance mechanisms that span administrative boundaries.

This article has been accepted for publication in IEEE Communications Standards Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCOMSTD.2026.3657234

4) Edge–Cloud Continuum and Vertical Integration: The 6G core will form a seamless continuum with edge–cloud infrastructures, enabling services closer to end users. Research should explore how orchestrators allocate resources across this continuum in response to real-time conditions. Key verticals such as smart manufacturing, autonomous vehicles, and digital health will benefit most from this agility.

VIII. CONCLUSION

This article elaborated on the 6G-in-the-loop vision that highlights the 6GC's importance as both an information provider for intelligent service orchestration and a cloud-native orchestrated network element. Building on this perspective, we introduced two complementary orchestration paradigms: Network-Aware Service Orchestration (NASO), which leverages real-time network insights to optimize service lifecycle management, and Context-Aware Core Network Orchestration (CCNO), which uses service and operational data for adaptive decisions such as function scaling and slice adaptation. Together, NASO and CCNO showcase in practice how 6GC involvement bridges network and service orchestration, enabling a converged, automated network–service architecture. Fully realizing 6G will require ongoing 3GPP alignment, deeper observability and more advanced AI-native automation.

ACKNOWLEDGMENT

This work has been partially supported by the project UNITY-6G (101192650) funded from EU's Horizon Europe SNS-JU R&I Programme.

AUTHOR BIOGRAPHIES

Michail Dalgitsis is Network Cloud Research Engineer with Nearby Computing, pursuing also a Ph.D. degree at Polytechnic University of Catalonia (UPC). His main research interests include network slicing and orchestration of cloud-native 5G/6G networks.

Dr. Eftychia Datsika is Senior Researcher at Nearby Computing. She has been involved in various European R&D ICT projects. Her main research interests include mobile network management, network slicing and edge computing. She received the UPC Best PhD Thesis Award in ICT, in 2020.

Cristina Santana Casillas is Head of Edge Computing at Telefónica's CTIO area, where she has been leading edge computing innovation initiatives since 2019. Her work focuses on integrating these research outcomes into commercial networks. She chairs the CAMARA Edge Subgroup and the GSMA Edge Platform Group, and actively contributes to strategic European initiatives such as IPCEI-CIS.

Dr. Angelos Antonopoulos is R&I Director at Nearby Computing. He received the Ph.D. from the Signal Theory and Communications department of UPC. He has authored over 150 papers on several topics, including network management, edge computing and 5G/6G networking. He has led several research projects and received awards for his work. He is an IEEE Senior Member.

REFERENCES

[1] R. Liebhart, M. Shafi, H. Tataria, G. Shivanandan, and D. Chandramouli, "Perspectives on 6G Architectures," *IEEE Wireless Communications*, vol. 32, no. 1, pp. 108–114, 2025.

[2] NGMN Alliance, "Network architecture evolution towards 6G," 2025, Accessed: July 24, 2025. Available at: `https://www.ngmn.org/wp-content/uploads/250218_Network_Architecture_Evolution_towards_6G_V1.0.pdf`.

[3] A. Dolas, R. Bodile, A. Kaushik, A. Kaur, R. Singh, and P. Chatzimisios, "Integrated AI and 6G Driven e-Health: Enabling Design, Challenges, and Future Prospects," in *IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2024, pp. 371–376.

[4] M.E. Kim, H. Bang, N. Ko, and M. Katz, "A Novel Radio Access Network Converged 6G Mobile Core Network Architecture," *IEEE Communications Magazine*, vol. 63, no. 5, pp. 70–77, 2025.

[5] R. Direito, D. Gomes, and R. L. Aguiar, "An Automated Testing and Orchestration Framework for Softwarized 5G and B5G Network Applications," in *2024 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2024, pp. 358–364.

[6] 3rd Generation Partnership Project (3GPP), "5G System; Network Exposure Function Northbound APIs; Stage 3 (Release 18)," *3GPP Technical Specification*, vol. TS 29.522, no. V18.9.0, 2025.

[7] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Services and System Aspects; Architecture enhancements for 5G System (5GS) to support network data analytics services; (Release 19)," *3GPP Technical Report*, vol. TS 23.288, 2024, Release 19.

[8] 3rd Generation Partnership Project (3GPP), "5G System; Policy and Charging Control Framework; Stage 2 (Release 18)," *3GPP Technical Specification*, vol. TS 23.503, no. V18.9.0, 2025.

[9] J. Fernandes *et al.*, "Pushing the Boundaries of Scalable 5G Core Networks: Cloud-Native NEF and CAPIF Interplay," in *2024 3rd International Conference on 6G Networking (6GNet)*, 2024, pp. 201–205.

[10] European Telecommunications Standards Institute (ETSI) Industry Specification Group (ISG), "Zero-touch network and Service Management (ZSM); Intent-driven autonomous networks; Generic aspects," *ETSI*, vol. GR ZSM 011, no. V2.1.1, Sept. 2024, Accessed: July 24, 2025. Available at: `https://www.etsi.org/deliver/etsi_gr/ZSM/001_099/011/02.01.01_60/gr_ZSM011v020101p.pdf`.

[11] European Telecommunications Standards Institute (ETSI), "MEC Application Developer Guidelines for Universal Access to Service APIs across the Industry," *ETSI White Paper*, vol. WP-68, June 2025, Accessed: July 24, 2025. Available at: `https://www.etsi.org/images/files/ETSIWhitePapers/ETSI-WP-68-MEC-app-dev-guidelinestracking.pdf`.

[12] TM Forum, "Open Digital Architecture (ODA): A blueprint for modular, cloud-based, open digital platforms," 2021, Accessed: July 24, 2025. Available at: `https://info.tmforum.org/rs/021-WLD-815/images/TMF\_ODA\_Guide\_2021.pdf`.

[13] A. Masaracchia *et al.*, "Toward 6G-enabled URLLCs: Digital Twin, Open Ran, and Semantic Communications," *IEEE Communications Standards Magazine*, vol. 9, no. 1, pp. 13–20, 2025.

[14] GSMA, "GSMA Open Gateway API Descriptions," *GSMA*, 2023, Accessed: July 24, 2025. Available at: https://www.gsma.com/solutions-and-impact/gsma-open-gateway/gsma-open-gateway-api-descriptions/.

[15] Linux Foundation, "CAMARA Project," *CAMARA Project*, 2023, Accessed: 2025-07-24. Available at: https://camaraproject.org/.