NOKIA

6G orchestration and automation: A system software view

White paper

This white paper is a sequel to an earlier Nokia white paper also covering 6G orchestration [5]. Here, we take a look at orchestration and automation from the perspective of system software. We shall argue that a properly designed system software platform caters for not only the evolution from 5G to 6G, but also for novel value networks beyond the state of the art.

Contents

Introduction

Orchestration and automation (O&A) are the machinery for optimally configuring a service provider's assets in view of its chosen business models. At the moment, business models are diversifying, especially at the edge of the network. The emergence of neutral hosts and the offloading of enterprise and end user workloads to edge clouds are but two examples of the ongoing developments. Consequently, O&A needs to cater to a variety of business models and value chains. Further, 6G needs to support interactions between stakeholders participating in service provisioning, which is more dynamic than the current paradigm based on relatively static service level agreements (SLAs) and contracts. Given the expected proliferation of business models and use cases, interoperability will be an important factor for the 6G era.

6G multi-access networks will often be built on an existing 5G network, allowing for an easy introduction of 6G capabilities for service providers as well as the ability to monetize resources right from the launch, an important factor in the success of 6G. 3GPP standards provide a global backbone for interoperability in 6G, supported by ancillary standards, best current practices (BCPs) and de facto standards.

Automation is a key requirement for 6G, which, because it is expected to natively incorporate artificial intelligence and machine learning (AI/ML), the entire architecture has been designed with AI/ML support in mind. This also includes data architectures supporting training and inference. The impact of AI/ML nativity on O&A is twofold: on the one hand, O&A itself will have advanced AI/ML capabilities; on the other, it will need to manage AI/ML features for the rest of the 6G system.

Support for end-to-end (E2E) services is important and is facilitated by an automation architecture built on AI/ML and intent-based management principles, as will be discussed in later sections. Approaching automation on an architectural level in 6G represents an advance as compared to bespoke AI/ML and analytics use cases employed in previous mobile network generations.

As we shall see in the next section, a suitably designed system software platform supports not only a top-down O&A paradigm but, also, truly decentralized services. This means that 6G service provisioning can be a mix of traditional top-down orchestration, decentralized orchestration and hybrid orchestration paradigms. Addressing future networking requirements holistically in system software is key to creating future-proof O&A systems [1].

Unified Networking Experience Technology (UNEXT) vision for a system software platform

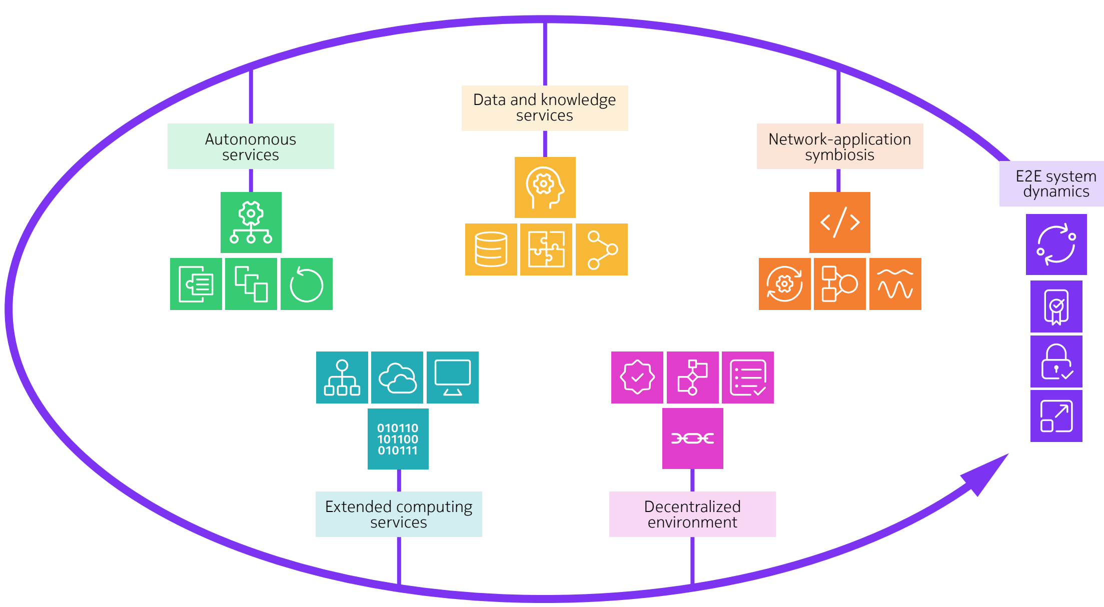
The UNEXT vision of Bell Labs [1] describes a self-managing network. UNEXT has been designed to support a broad palette of use cases, of which 6G O&A use cases discussed in this white paper are a subset. The first goal of UNEXT has been to provide a basis for automation and security of current and future use cases to support top-down, decentralized and hybrid orchestration models — the last two facilitating participation in novel value chains. The second, key goal for UNEXT is to increase the level of automation, including multi-stakeholder scenarios.

The UNEXT system approaches the requirements for future networking holistically, instead of solving disparate problems separately. Its architecture envisions six areas:

1. Autonomous services
2. Data and knowledge services
3. Network-application symbiosis
4. Extended computing services
5. Decentralized environments
6. E2E system dynamics

The areas are illustrated in Figure 1. The E2E system dynamics tie together the rest of the areas. We shall highlight some aspects of the UNEXT areas relating to O&A below. A more comprehensive account can be found in the UNEXT white paper [5] previously referenced.

Figure 1. UNEXT vision architecture of Bell Labs



The six architectural areas in UNEXT perform the following functions:

- Autonomous services develop automation and autonomy for services
- Data and knowledge services develop services related to data and AI/ML
- Network-application symbiosis develop enhancements in application performance and network efficiency
- Extended computing services develop heterogeneous orchestration methods for cloud continuum
- Decentralized environments develop secure and trusted orchestration across heterogeneous devices, components and actors
- E2E system dynamics ensure system level properties such as the reliability, security and scalability of UNEXT.

The UNEXT vision has automation and autonomy at the center of the concept supporting features such as enhanced network and application interaction and decentralization. It also features a data architecture to support automation across stakeholder boundaries, including data collection, distribution and information extraction.

We shall take a look at some O&A-related aspects of UNEXT and further elaborate UNEXT concepts below.

Service composition

Service composition is a key building block for UNEXT and is developed by autonomous services. It allows for automated composition of a service in run time, in either multi-vendor or multi-stakeholder scenarios. This represents a significant improvement over typical service management tooling involving templates. Service APIs are declarative and consumed through intents. Semantic modelling of service interfaces allows for automation of composition as well as consumption of services via intent interfaces regardless of the abstraction level of the services or the business intents.

Other UNEXT areas relevant to O&A include:

- Decentralized services are important for decentralized and hybrid service provision scenarios
- Extended computing services support extreme edge capabilities and integrated network and compute
- Knowledge and data services include data collection and fusion and semantic extraction
- Network-application symbiosis enhances the interaction between applications and the network, increasing the effectiveness of the network and improving the quality of application (QoA) characteristics
- E2E system dynamics ensure system-level properties.

6G architecture aspects

Below, we shall highlight a few aspects of the emerging 6G network architecture developed after the first release of this white paper. More architectural aspects can be found in the previously referenced O&A white paper [5].

A 6G CSP multi-access network typically builds on an existing 5G network. The 5G network may employ one or both variants of a purpose-built radio access network (RAN) and cloud RAN. Consequently, both purpose-built and cloud RAN variants will be supported at the launch of 6G.

Cloud RAN features far edge clouds, which provide a platform for extending the cloud continuum towards the extreme edge in 6G. Cloud-native methods are the key to effective management of far edge and extreme edge capabilities. The concept of extreme edge allows for integration of resources at an enterprise or at home to be used as a part of an end-to-end 6G service. Specialized non-public networks (NPNs) can be built for industry segments leveraging the functional architecture of 6G RAN. Integrated network and compute (INC) introduces use cases such as application compute offloading where application software components need to be deployed on-demand in optimal locations from both a network and compute perspective. This requires orchestration between the 6G network and computing platforms.

A generalization of the cloud architecture in UNEXT is based on the concept of extended computing services (ECS). Service assurance and support for dynamic value networks are supported in UNEXT by the concept of decentralized environments (DE). The concept of autonomous services (AS) supports composite services distributed across multiple platforms.

There are non-functional requirements for the 6G architecture that affect the design of O&A. These include:

- Sustainability
- Service assurance for highly automated service creation
- Support for dynamic value networks — important for capturing future business opportunities, which SLA-based arrangements are too slow a mechanism to support.

Orchestration

Orchestration is the machinery for monetizing a service provider's resources. This involves not only external service exposure, but also the resource-facing services supporting it. Orchestration needs to link increasingly elaborate value networks to increasingly complex resources so that changes on the business side are seamlessly translated into resources. Orchestration consists of service orchestration and resource orchestration. Service orchestration is responsible for E2E service performance.

In evolving to 6G, orchestration will cater to a widening array of use cases and value network instances for which an increased level of automation is required. Increased automation will also be needed for cross-stakeholder operations. As discussed, business agility will benefit from the ability to go beyond SLA-based agreements between stakeholders. Similarly, automation is crucial for E2E services support.

Some use cases involve existing stakeholders interacting in a novel manner. The INC aspect of 6G requires cross-orchestration between the 6G network and computing platforms, even in real time. Depending on the ownership of the network and computing resources, multiple stakeholders may be involved.

The Bell Labs UNEXT vision supports orchestration across diverse value chains through the concept of composable autonomous services. Stakeholders involved in end-to-end provisioning can leverage the capabilities of UNEXT to automate their processes and enable new business models based on decentralization.

Service layer

Automation is critical for 6G-ready orchestration. It allows automated alignment of resources to support intent-based service instantiation as well as service provisioning constellations. Given the central role of automation, it is essential that automated service fulfilment is accompanied by automated configuration of corresponding service assurance.

Service exposure is a capability crucial to monetization of a communication service provider's (CSP's) resources. The most obvious example is anything-as-a-service (XaaS) interfaces.

Providing access to developers is an increasingly important channel for monetization and has picked up pace with organizations such as CAMARA [2]. Nokia provides a network-as-code platform and developer portal for an integrated developer experience across CSP capabilities [4]. Developer exposure is expected to stay relevant during the evolution to 6G, adding capabilities such as sensing in the course of this evolution.

Arrangements between service providers and ecosystem stakeholders have been fairly static and based on SLAs. In view of the expected value network diversification, it is expected that more dynamic arrangements are called for by the dawn of the 6G era. The ability to cooperate in run-time allows for effective utilization and monetization of service provider resources via a value network. This is expected to be important for orchestration of the extreme edge involving multiple stakeholders, e.g. multi-tenant industrial scenarios. The UNEXT vision provides mechanisms facilitating dynamic cross-stakeholder service provisioning, including decentralized trust and AI/ML, in addition to decentralized service composition discussed earlier.

Resource layer

As a part of the intent-based management automation architecture, controllers are central to the automation of resource orchestration. For each resource in the architecture, goal state is derived from intents that cascade down the hierarchy to the controller that is tracking the goal state for the resource in question. The UNEXT architecture is fully compatible with intent-based management controller architecture.

Cloud-native methods are relevant to core, cloud RAN and extreme edge-related functionalities in the 6G cloud continuum; they provide the basis for the controller architecture. The management of physical network elements should be based on the same principles as their cloudified counterparts.

In view of expected higher performance requirements in 6G, orchestration of related functionalities such as the user plane (UPF) may take place closer to the network edge than in current use cases.

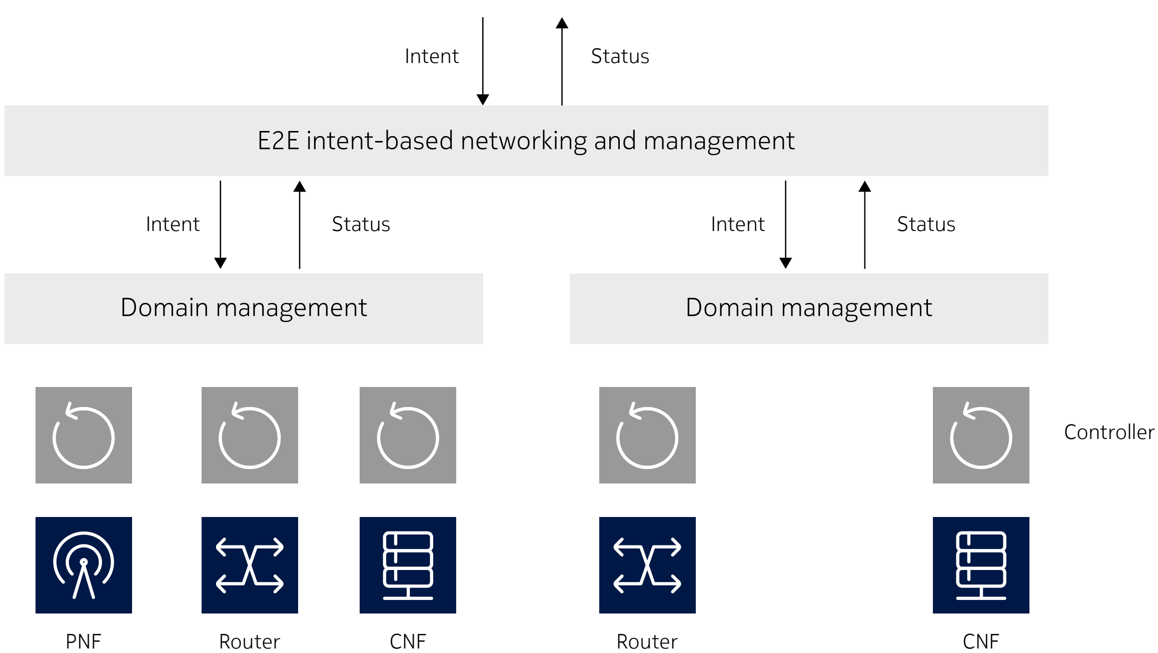
The extreme edge extends the 6G cloud continuum beyond the scope of either a public-land mobile network (PLMN) or non-public network (NPN); this also holds for application workloads. Extreme edge orchestration may require performance optimizations that differ from the principles of a PLMN's or NPN's internal orchestration based on a combination of intent-based management hierarchy and cloud-native principles. Optimal deployment of application workloads as envisioned in INC calls for sharing metrics and status information between network and compute resources and even making deployment decisions in real time.

New exogenous triggers such as sustainability require the availability of metrics to facilitate corresponding orchestration actions. As well, the 6G RAN architecture itself may impose performance requirements for orchestration beyond those in effect for 5G evolution.

Automation

A hierarchical controller architecture is envisioned as the basis for implementing intent-based management in 6G automation, which is described in the first release of the Nokia 6G O&A white paper [5]. The controller architecture supports services end-to-end by leveraging the service layer, through an intent-based management hierarchy, to control the underlying cloud and hardware resources. The controller architecture is illustrated in Figure 2. It comprises a hierarchy of layers, each mapping higher-level intents into a goal state, which is tracked by closed-loop automation (CLA). The leaf nodes of the hierarchy manage physical or virtual resources. The UNEXT vision architecture leverages the intent-based management pattern, as well as technologies referred to below.

Figure 2. Intent-based controller architecture



A few of the key aspects of automation on which we will elaborate are:

- Network digital twins (NDTs) play an important role in the controller architecture
- Semantic methods allow for automation of system integration
- Generative AI can be used in a “return channel” for intent-based management of the business interface.

Network digital twins

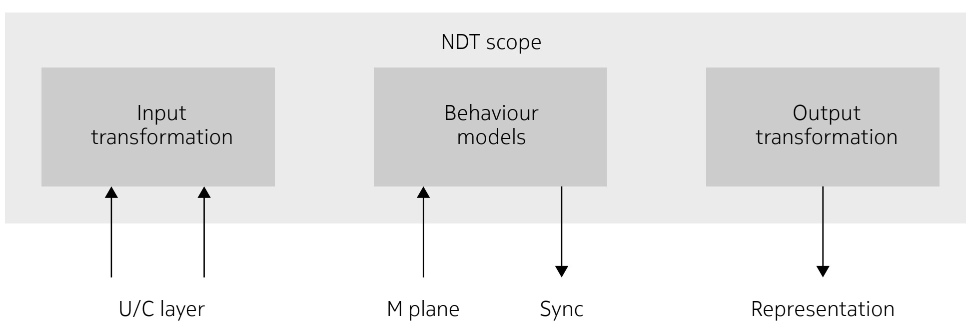
For the controller architecture, NDTs provide the ability to map intents into goal states that are tracked by the controllers, with support from AI/ML. A network digital twin for 6G is a logical capability incorporated into the controller architecture at different layers. It is a particular use of a more general concept relevant to the controller architecture. Some concepts relevant to NDTs are illustrated in Figure 3.

The availability of consistent information about the state of the system relevant to the abstraction level of the NDT instance is key. This information is used to model, optimize and to plan changes made to the system.

An NDT instance needs to be able to model the effect of changes made to the system and facilitate autonomous operation. This requires incorporation of virtual entities mimicking aspects of the system being managed.

An NDT needs the capability to learn and to adapt. Thus, it is important that it can compose new models out of existing building blocks and incorporate them into individual NDT instances as well as the system-level distributed NDT. An NDT may also interact with application-level digital twins, allowing for validation of application interactions prior to committing to a configuration.

Figure 3. An illustration of the logical architecture of an NDT and its interactions with the rest of the system



Digital twins are a part of the UNEXT vision, allowing for contextualization of triggers for network operations (e.g., intents).

Semantic methods for system integration

System integration, especially involving multi-vendor components, involves facilitating communication between them. This requires matching data models and service interfaces, which may contain vendor specific extensions in addition to standardized features. Multi-vendor-capable operations support systems (OSS) like Nokia's MantaRay Network Management (NM) [3] have adaptations for vendor systems to facilitate the above.

The future is expected to require agile composition of components and shorter integration time. The use of semantic descriptions of module interfaces and data models allows for automatic composition of capabilities using a set of modules. Semantic composition lends itself naturally to intent-based management principles.

Semantic methods are a key enabler for service management in UNEXT, facilitating discovery and composition of service components in a multi-vendor, multi-stakeholder scenario.

Generative AI in intent-based management

Large language model (LLM) generative AI has received a lot of attention recently. LLMs can often generate a plausible text on a given general topic. They also have a downside: apparently plausible text may actually contain factual errors.

The business-intent interface of the controller architecture allows for expressing the desired state of the system using natural language relevant to the domain of application. The "return channel" of business intents — called "insight" — is important for monitoring the status of the system and understanding whether issuing the intent has produced the desired results. The goal of using business intents is to avoid needing to use involved telecommunications concepts. The insight interface should have the same characteristics. The task of expressing the status of one domain in the language of another seems like a good match for LLMs, which is why they are promising candidates for this interface.

Summary

In the preceding sections, we have summarized the present understanding of some key aspects of 6G orchestration and automation in light of the current status of research. We have discussed drivers, identified some key aspects of 6G architecture relevant to O&A and linked it to key technological topics in the areas of orchestration and automation.

6G proposes to introduce new capabilities that will make business ecosystems more dynamic and provide more advanced technologies. It is also important to retain good things from previous generations: 3GPP will be a global backbone for interoperability, and the capabilities of 5G SA provide a powerful basis for monetization of 6G right at launch.

System software plays a key role in facilitating future requirements for O&A and is best approached holistically. The Nokia Bell Labs UNEXT vision of the network has been designed for automation from the ground up. Automation is a key architectural component, incorporating key technologies like intent-based management, network digital twins and generative AI. It supports automation of use cases in top-down orchestration scenarios, and also opens the possibility of decentralized and hybrid value chains.

Abbreviations

AI/ML	Artificial Intelligence / Machine Learning
BCP	Best Current Practice
C layer	Control layer
CSP	Communication Service Provider
E2E	End-to-end
INC	Integrated Network and Compute
JCAS	Joint Communication and Sensing
LLM	Large Language Models
M plane	Management plane
MEC	Multi-Access Edge Cloud
NDT	Network Digital Twin
NEF	Network Exposure Function
NPN	Non-Public Network
O&A	Orchestration and Automation
PLMN	Public Land Mobile Network
RAN	Radio Access Network
SA	Stand-Alone
SLA	Service Level Agreement
U layer	User layer
UNEXT	Unified Networking Experience
UPF	User Plane Function
XaaS	Anything as a Service

References

1. Azimeh Sefidcon, C. V., "UNEXT - A unified networking experience," Nokia, 2023. Available: https://www.bell-labs.com/institute/white-papers/unext-a-unified-networking-experience
2. CAMARA, "APIs enabling seamless access to Telco network capabilities," The Telco Global API Alliance, Linux Foundation, 2023. Available: https://camaraproject.org
3. Nokia, "MantaRay NM: Intelligent network management for 5G," Nokia web site, 2023. Available: https://www.nokia.com/networks/mobile-networks/mantaray-nm
4. Nokia, "Network as Code," Nokia web site, 2023. Available: https://www.nokia.com/networks/network-as-code
5. Vilho Räisänen, "6G orchestration - a scalable application driven network optimizing approach," Nokia Bell-Labs web site, Feb 2023. Available: https://www.bell-labs.com/institute/white-papers/6g-orchestration-a-scalable-application-driven-network-optimizing-approach

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID213955 (April)