Financiado por la Unión Europea
NextGenerationEU

MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

Plan de Recuperación,
Transformación
y Resiliencia

uc3m

6G-CLARION-OE Entregable E12

Use cases definition

PROGRAMA DE UNIVERSALIZACIÓN DE INFRAESTRUCTURAS DIGITALES PARA LA COHESIÓN UNICO I+D 5G 2021

UNICO I+D

Fecha: 31/7/2022

Versión: 1.0

Propiedades del documento

Id del documento	E8			
Título	Use cases definition			
Responsable	UC3M			
Editor	Albert Banchs			
Equipo editorial	**Partner**	**Name**	**Surname**	**Sections**
	UC3M	Albert	Banchs	All
	UC3M	Francisco	Valera	All
Nivel de diseminación	Público			
Estado del documento	Final			
Versión	1.0			

Historial

Revisión	Fecha	Por	Descripción
1.0	31/07/22	Editor	Final version

Revisor

Equipo revisor	**Partner**	**Name**	**Surname**	**Sections**
	UC3M	Marco	Gramaglia	All

Descargo de responsabilidad

This document has been produced in the context of the 6G-CLARION Project. The research leading to these results has received funding from the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO 5G I+D programme. The information contained in this document is provided "as is" without any express or implied warranties, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. The document writer shall not be liable for any damages, whether direct or indirect, arising out of or in connection with the use of this information. The user of this document assumes all risks and liabilities associated with its use and shall indemnify and hold harmless the document writer from any and all claims, losses, damages, or expenses, including attorney's fees, arising from the use of this information.

Table of Contents

Lista de acrónimos

AI - Artificial Intelligence

API - Application Programming Interface

CDN - Content Delivery Network

CPU - Central Processing Unit

LTE - Long-Term Evolution

ML - Machine Learning

NAT - Network Address Translation

OS - Operating System

QoS - Quality of Service

RAM - Random Access Memory

VPN - Virtual Private Network

WAN - Wide Area Network

Resumen ejecutivo

La integración de la operación de red y la provisión de servicios de red con la llegada de 5G ha resaltado las limitaciones de las arquitecturas de red actuales para proporcionar medios para dicha integración. Los diseños actuales de redes móviles solo permiten una optimización continua dentro de dominios específicos, lo que resulta en una automatización de estilo "silos" que no cumple con los requisitos para la automatización de bucle cerrado. En este trabajo, se analiza la viabilidad de un marco de exposición de capacidades a nivel de red para la automatización de bucle cerrado mediante el análisis de los dominios productores y las diferentes capacidades que deben exponer cada uno de ellos.

Abstract

The integration of network operation and network service provisioning with the advent of 5G has highlighted the limitations of current network architectures in providing means for such integration. Existing mobile network designs only allow for continuous optimization within specific domains, resulting in a "silo-style" automation that falls short when aiming for closed-loop automation. In this work, we analyze the feasibility of a network-wide capability exposure framework for closed-loop automation by analyzing the producing domains and the different capabilities that shall be exposed by each of them.

1. Introduction

The continuous development towards more flexible networks[1] makes softwarization a key enabling technology that has also impacted standardization during the last few years; starting from the introduction of software-defined networking (SDN) and followed by the network function virtualization (NFV) concepts[2]. Recently, many network-related standards adopted service based architecture (SBA) principles that reference service-oriented architecture (SOA) paradigms. Here, a network function (NF) can flexibly communicate with other NFs to consume the provided services, thus overcoming the limitations of a reference point-based interaction, where an interface is only defined between two NFs.

Moreover, this has opened up new opportunities[3], e.g., the recently developed network slicing6 paradigm. On the one hand, the flexibility given by a programmatic approach to network management and control has allowed the possibility of creating network instances (i.e., the network slices), tailored to various applications and service provider needs on the same infrastructure. However, on the other hand, this introduced additional complexity in the management side, due to the dependencies between slices that require different optimal operating points but have to share a common underlying network infrastructure.

The emergence of new paradigms represents a significant development step from the monolithic structure of legacy mobile networks, which were designed to provide mobile broadband services over a single physical network instance. In order to manage a potentially large number of network instances, Big Data and Artificial Intelligence (AI) techniques are being considered as potential enablers for autonomous network management. Features such as auto-scaling, self-optimization, and intent-based networking are well-suited to a data-driven approach to the network environment, in which elements can be configured according to service-related policies in a jointly optimized way.

1.1. Data Driven Network Management

However, achieving autonomous network management is a challenging task that involves overcoming a number of technical complexities. One of these is data heterogeneity, as the data collected from the network includes a wide range of metrics, from network key performance indicators to general-purpose resource utilization. Another challenge is temporal timescale heterogeneity, as decisions may need to be made on different temporal scales depending on the type of element being assisted.

In addition to scalability challenges, there is a fundamental technical barrier relating to the interfaces needed across different network domains. The operation of the network poses new challenges that require a revision of how different parts inside and outside the network interact with each other. This involves a much higher granularity that goes beyond the pure ownership of the infrastructure and involves different network domains.

1.2. The need for a new exposure capability

Traditionally, network management, orchestration, and control procedures were developed separately in

[1] Sciancalepore V, Mannweiler C, Yousaf FZ, et al. A future-proof architecture for management and orchestration of multi-domain NextGen networks. IEEE Access. 2019;7:79216-79232. doi:10.1109/ACCESS.2019.2923364

[2] Bega D, Gramaglia M, Bernardos CCJ, Banchs A, Costa-Perez X. Toward the network of the future: from enabling technologies to 5G concepts. Trans Emerg Telecommun Technol. 2017;28(8):e3205. doi:10.1002/ett.3205

[3] Marsch P, Bulakci Ö, Queseth O, Boldi M. E2E architecture; 2018:79-114

various standardization tracks, resulting in domain-specific and often proprietary designs of executing functions such as OSS functions, element managers, orchestrators, and NFs. As a result, optimization only occurred in a per-domain manner, with limited interaction between domains. In legacy networks, this approach was adequate due to their limited configurations. However, it falls short in the 5G environment, where network slicing requires a more modular design of NFs and interfaces for data acquisition, processing, and AI integration. The Self-Organizing Network (SON) was the first step towards flexible network management without human intervention, but it is insufficient for 5G network slicing. To close the loop in an automated manner, new interfaces and functionalities are needed, including flexible data exchange across domains and reliable and scalable configurations for NF parameters. Current network architectures are defined in a silo-based way, lacking open exchange among domains. This work aims to identify possible data exchanges among different domains, find an architecture to overcome this issue, and exemplify it compellingly.

2. Network Domains

2.1. Function Domain

The functions and interactions within the RAN and between the RAN and CN have been defined by 3GPP Releases 15,16,17 and 18. With 5G New Radio, the Service Data Adaptation Protocol (SDAP) layer is included, enabling QoS flows to be mapped to radio bearers and providing greater QoS enforcement in the RAN. The F1 interface is introduced to enable the split between CU and DU and promote flexible centralization of RAN functions and network function virtualization. Network slicing support is also provided in the RAN via Network Slice Selection Assistance Information (NSSAI). The O-RAN Alliance has proposed a more flexible layout based on the RAN Intelligent Controller (RIC), allowing custom control plane functions to be implemented and operated.

The 5G CN (5GC) follows new design paradigms with Control Plane - User Plane (CP/UP) separation, modular NFs, and SBA since 3GPP Release 15. The 5GC comprises NFs and Network Entities (NEs) that include CP functions such as AMF, SMF, PCF, AF, UDR, and UP functions such as UPF. The 5GC also includes new NF/NEs supporting service-based communication between 5GC CP NFs/NEs, such as the Network Repository Function (NRF) and Network Exposure Function (NEF), enabling service registration and service discovery in the same domain. In some cases, 5GC CP NFs/NEs may need to communicate with other network domains. For such cross-domain communications, special service communication restrictions and information translation are necessary.

The NWDAF is a new NF introduced in 5GC for network analytics, providing network analytics on the load level of an NF. In 5GC SBA, data analytics services of NWDAF can be consumed by any NF. 3GPP Release 16 extends the usage of NWDAF also to use cases beyond load level, e.g., network performance analytics, slice load level related network data analytics, observed service experience related network data analytics, UE related analytics, quality of service (QoS) sustainability analytics, etc.

2.2. The management domain

Beginning with 3GPP Release 15, the Operations, Administration, and Maintenance (OAM) domain, also referred to as the management plane, has implemented the Service-Based Management Architecture (SBMA). Under this framework, a management service provides management capabilities, with essential services including generic provisioning, fault supervision, and performance assurance management services, typically produced by the Network Function (NF) or a lower management layer, such as the Network Function Management Function or Network Slice Subnet Management Function. Management services are accessed by management service consumers through a standardized service interface, composed of management service components that describe the management operation, the managed entity, and the managed data associated with that entity, such as performance information. These management services can generally be exposed to any authorized consumer, but the Exposure Governance Management Function (EGMF), introduced by 3GPP SA5, enables access restrictions based on policies.

Additional management service abstraction may be necessary when a trust relationship between the management service producer and consumer is lacking, such as when the consumer resides outside of the operator's administrative domain. The "Network Slice as a Service" (NSaaS) business model requires specific sets of management services to be exposed to and consumed by an operator's customer. Additionally, 3GPP21 lists lifecycle management services for a communication service instance, including activation, modification, Management Data Analytics (MDA)-assisted service-level specification (SLS) assurance, and termination. The MDA service (MDAS) possesses the ability to process and analyze raw network data, such as performance measurements, trace reports, Quality of Experience (QoE) reports, alarms, configuration data, and other network analytical data, to detect specific events and predict network performance. The provided analytics reports may include recommended actions. MDAS instances can be tailored for a specific use case and exposed to external consumers through EGMF, enabling subscription to customized analytics reports.

2.3. Orchestration domain

The introduction of SDN, NFV, and containerization technologies has significantly boosted the orchestration and advanced configuration of traditional network domains (NFs and management plane) that were already present in legacy networks up to 4G/LTE. Resource virtualization has enabled on-demand provisioning of network services, following the successful paradigm of Infrastructure as a Service (IaaS), allowing seamless creation, re-configuration, and termination of services. The success of network programmability has led to the adoption of API-based access to network configuration, facilitating flexible re-configuration of NFs in contrast to traditional network control methods. However, the orchestration and lifecycle management procedures of general-purpose cloud resources have not been tackled by 3GPP. Instead, vendorspecific, standard-compliant solutions such as NokiaCloudBandSuite or open-source initiatives provided by standardization bodies (such as ETSI NFV MANO provided by OSM) or other fora (e.g., ONAP) are used for orchestration of network resources. Radio resources are less widespread in terms of orchestration solutions, although industry initiatives like O-RAN have defined radio orchestration procedures. Besides orchestration, these resources need to be configured based on network conditions such as load and traffic patterns. For example, inter- and intra-datacenter networks can be configured through an SDN controller, and the transport network can use an SD-WAN approach. Such controller-based approaches followed by SDOs can also be extended to the network domain, as proposed by O-RAN for the access network.

2.4. Service domain

The emergence of the novel network softwarization paradigm introduced by 5G and beyond 5G networks has allowed for a diverse and heterogeneous landscape of tenants. This includes various service providers, such as industrial verticals, that are now an essential part of network operations. This is a significant departure from the legacy 4G networks, which were characterized by a full Over The Top (OTT) service delivery model.

The new network provisioning model provides a more integrated view of the system from the tenant and service provider perspective. This is made possible by leveraging novel configuration primitives to act on the underlying network slices. From the standardization point of view, this concept has been gaining more attention, as evidenced by 3GPP's definition of two management models: Network Operator Internals and NSaaS.

Industrial initiatives, such as 5G-ACIA, have been instrumental in defining the requirements for a 5G exposure reference point towards enterprise tenants. The goal is to promote better integration between third-party service providers and network operators, especially for Industrial Internet of Things (IIoT) use cases. The proposed exposure framework involves the selective exposure of specific functionalities from a 5G non-public network or from the UE towards IIoT applications in the IT enterprise domain.

While 5G-ACIA does not mandate any specific solutions for the exposure reference point, usability, simplicity, modularity, and extensibility are listed as key requirements. The specific capabilities to be exposed depend on the envisioned use case. Device provisioning and onboarding, device connectivity management, device connectivity monitoring, device group management, device location information, network monitoring, and network maintenance are some of the capabilities that can be exposed.

The quest for a tighter interaction between service providers and network operators is also being pursued by industrial consortia, such as NetApps. These initiatives promote the use of open APIs between traditionally separated domains. Open interactions between applications and underlying access technology are essential, especially for providers that make use of edge computing technologies to offload use cases that require very low latency, such as vehicular ones, to provide autonomous driving services.

However, current efforts, such as the integration of MEC with relevant architectures like O-RAN, are targeting specific architectural elements and lack exposure functionality towards other domains, such as the Core. As such, there is a need for more flexible interfacing between service providers and network operators to cater to different business models, including hybrid private-public deployments.

3. Capabilities

In previous discussions, we acknowledged that while existing state-of-the-art network architectures have data-driven functionalities, their limited scope often hinders automated network operation involving cross-domain activities, such as reactive orchestration and service-driven network re-configurations essential to autonomous network operation. To enable cross-domain interaction, we introduce four types of capabilities facilitated by inter-domain message buses. Note that authorization levels for potential consumers from the same or different domains are associated with registered capabilities.

Capability type 1 pertains to monitoring and data collection, involving raw monitoring data provision between different network elements, including network functions that provide key performance indicators (KPIs) related to cell and user-centric performance, end-to-end service performance, and infrastructure monitoring data such as CPU and RAM utilization. It also covers customized measurement jobs, trace collection configurations, and real-time performance measurements on monitored elements.

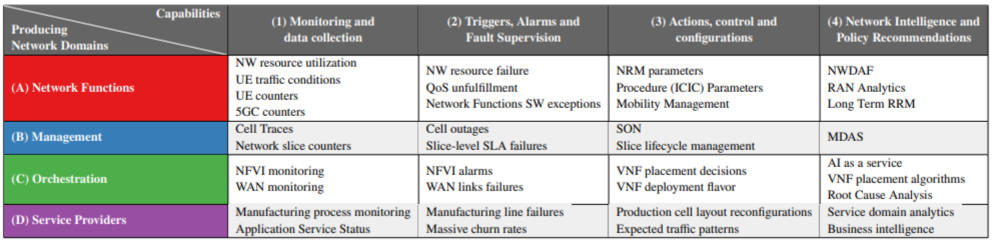
Capability type 2 focuses on triggers, alarms, and fault supervision, offering a more refined way to access data collected from network elements. Network elements often react according to well-defined state machines upon triggering events, making it important to provide event notifications such as alarm information, alarm state changes, alarm correlation information, and notifications on associated troubleshooting actions.

Capability type 3 involves exposing configuration and control capabilities of network elements of different domains to other elements. It covers capabilities to create, modify, delete objects and their parameters and configuration attributes, and the definition of an object varies depending on the domain exposing this capability.

Finally, Capability type 4 pertains to network intelligence and policy recommendations. Future network operation requires intelligence, meaning tasks that currently require human intervention for optimal network performance will be automated with some kind of AI in the loop. Network elements will thus expose capabilities to perform complex analytics on inputs coming from other elements (i.e., those exposed by capability types 1-3). Intelligent network elements will perform root cause and impact analysis and derive novel information required for enhanced operations, such as failure prediction to prevent faulty network states. Policy recommendations exposed through this capability are consumed by other elements in each domain or across domains.

3.1. A wider exposure functionality

Mobile network systems can implement a closed-loop control system by opening flexible exposure of capabilities among domains. This can be achieved by following a producer-consumer or publish-subscribe approach. Artificial intelligence, particularly machine learning (and especially deep learning) solutions, are considered as one of the most important tools to achieve this vision. To enable network automation through the usage of such solutions, a large amount of data is needed for correct training and configuration of these models. Thus, domain interactions for network data exchanges are summarized in Table 1 and discussed next, by producing domain.



Capabilities Producing Network Domains	(1) Monitoring and data collection	(2) Triggers, Alarms and Fault Supervision	(3) Actions, control and configurations	(4) Network Intelligence and Policy Recommendations
(A) Network Functions	NW resource utilization UE traffic conditions UE counters 5GC counters	NW resource failure QoS unfulfillment Network Functions SW exceptions	NRM parameters Procedure (ICIC) Parameters Mobility Management	NWDAF RAN Analytics Long Term RRM
(B) Management	Cell Traces Network slice counters	Cell outages Slice-level SLA failures	SON Slice lifecycle management	MDAS
(C) Orchestration	NFVI monitoring WAN monitoring	NFVI alarms WAN links failures	VNF placement decisions VNF deployment flavor	AI as a service VNF placement algorithms Root Cause Analysis
(D) Service Providers	Manufacturing process monitoring Application Service Status	Manufacturing line failures Massive churn rates	Production cell layout reconfigurations Expected traffic patterns	Service domain analytics Business intelligence

Table 1 Producing domains and their capabilities

In the context of mobile network management and orchestration, raw counters and aggregated performance metrics are collected from network functions (NFs) in both RAN and CN domains. This is currently supported by many functions, and continuous monitoring data coming from network probes belongs to this category. Similarly, standardization efforts such as the one carried out by GSMA have defined the Generic Slice Template (GST) that provides attributes, including key performance indicators (KPIs), a network slice should fulfill. When a GST is filled with values, i.e., with the customer requirements, the Network Slice Type (NEST) is constructed. Control plane NFs and exposure of their capabilities have received great attention in mobile network standardization. 5GC provides a mechanism for dynamically accessing CP services for NFs through the Network Exposure Function (NEF) to the applications. Analytics services are also provided, mostly through the NWDAF in the 5GC, which provides information about the load of an NF that may be used by other NFs to adjust settings, such as load balancing purposes.

On the other hand, management functions can expose monitoring data at any of the granularity levels that is used for handling the lifecycle management of objects. This includes counters regarding UE-related events, cell loads, or network slice loads for management and orchestration purposes. Traditionally, the interaction between service providers and network management happens through the OSS and BSS, which are usually involving customer care services and require “human-alike” timings. The management system also produces analytics, and the interaction between the service provider and the network management can be direct, which can allow timely network management upon changes, without indirect policy configurations.

4. Conclusion

In conclusion, the implementation of closed-loop control systems in mobile network systems through the usage of Artificial Intelligence (AI) and Machine Learning (ML) solutions is a rapidly evolving field, with significant attention from both industry and academia. To enable the automation of network functions through AI and ML, a large amount of data is required, which can be gathered from various domains within the mobile network. These interactions are summarized in Table 1 and discussed in detail in the text. Overall, the ability to implement closed-loop control systems and network automation through AI and ML solutions is a promising direction for mobile network systems, with the potential to significantly improve network performance and efficiency, reduce downtime, and enhance the user experience. However, it also poses significant technical and operational challenges, which require further research and development efforts to be addressed.