International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3182

(09/2022)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

Future networks

Machine learning based end-to-end multi-domain network slice management and orchestration

Recommendation ITU-T Y.3182

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Computing power networks	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	**Y.3000–Y.3499**
CLOUD COMPUTING	Y.3500–Y.3599
BIG DATA	Y.3600–Y.3799
QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800–Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	
General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.3182

Machine learning based end-to-end multi-domain network slice management and orchestration

Summary

Recommendation ITU-T Y.3182 describes an intelligent cost-effective network management and orchestration framework that can cope with the challenges of multi-domain network slicing, while minimizing human intervention towards full automation of slice lifecycle management and runtime operation.

It addresses the following subjects:

- Overview and interoperability requirements of machine learning based multi-domain end-to-end network slice management and orchestration;
- Functional requirements of machine learning based multi-domain end-to-end network slice management and orchestration;
- Framework of machine learning based multi-domain end-to-end network slice management and orchestration;
- Cognitive components for the framework.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.3182	2022-09-29	13	11.1002/1000/15059

Keywords

Artificial intelligence, automation, cognition, end-to-end, machine learning, multi-domain, network management, orchestration, network slice, QoE, QoS.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Y.3182

Machine learning based end-to-end multi-domain network slice management and orchestration

1 Scope

This Recommendation provides the framework and requirements of machine learning based end-to-end network slice management and orchestration in multi-domain environments. It addresses the following subjects:

- Overview and interoperability requirements of machine learning based multi-domain end-to-end network slice management and orchestration;
- Functional requirements of machine learning based multi-domain end-to-end network slice management and orchestration;
- Framework of machine learning based multi-domain end-to-end network slice management and orchestration;
- Cognitive components for the framework.

Use case examples are provided in Appendix I.

NOTE 1 – Multi-domain environments include those provided by the same or different network operators.

NOTE 2 – The framework described in this Recommendation is also applicable to single domain environments as appropriate.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T X.1601] Recommendation ITU-T X.1601 (2015), *Security framework for cloud computing.*

[ITU-T Y.2701] Recommendation ITU-T Y.2701 (2007), *Security requirements for NGN release 1*.

[ITU-T Y.3101] Recommendation ITU-T Y.3101 (2018), *Requirements of the IMT-2020 network.*

[ITU-T Y.3110] Recommendation ITU-T Y.3110 (2017), *IMT-2020 network management and orchestration requirements*.

[ITU-T Y.3111] Recommendation ITU-T Y.3111 (2017), *IMT-2020 network management and orchestration framework.*

[ITU-T Y.3112] Recommendation ITU-T Y.3112 (2018), *Framework for the support of network slicing in the IMT-2020 network.*

[ITU-T Y.3153] Recommendation ITU-T Y.3153 (2019), *Network slice orchestration and management for providing network services to 3rd party in the IMT-2020 network.*

[ITU-T Y.3156] Recommendation ITU-T Y.3156 (2020), *Framework of network slicing with AI-assisted analysis in IMT-2020 networks*.

[ITU-T Y.3170] Recommendation ITU-T Y.3170 (2018), *Requirements for machine learning-based quality of service assurance for the IMT-2020 network*.

[ITU-T Y.3172] Recommendation ITU-T Y.3172 (2019), *Architectural framework for machine learning in future networks including IMT-2020*.

[ITU-T Y.3178] Recommendation ITU-T Y.3178 (2021), *Functional framework of artificial intelligence-based network service provisioning in future networks including IMT-2020*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 closed loop [b-ITU-T Y.3115]: A type of control mechanism in which the outputs and behaviour of a system are monitored and analysed, and the behaviour of the system is adjusted so that improvements may be achieved towards definable goals.

NOTE 1 – Observe, orient, decide and act (OODA) [b-IEEE-2006], and monitor, analyse, plan and execute with knowledge (MAPE-K) [b-IEEE-2003] are examples of closed loop mechanisms.

NOTE 2 – Examples of definable goal types are optimization of network resources' utilization and automated service fulfilment and assurance. Goals may be defined using declarative mechanisms.

NOTE 3 – The system may consist of a set of managed entities, workflows and/or processes in a network.

3.1.2 control plane (CP) [b-ITU-T Y.2011]: The set of functions that controls the operation of entities in the stratum or layer under consideration and the functions required to support this control.

3.1.3 data plane (DP) [b-ITU-T Y.2011]: The set of functions used to transfer data in the stratum or layer under consideration.

NOTE – In the ITU-T International Mobile Telecommunications-2020 (IMT-2020) related standard Recommendations, "User plane" is used preferentially rather than "Data plane".

3.1.4 network slice (NS) [b-ITU-T Y.3100]: A logical network that provides specific network capabilities and network characteristics.

NOTE 1 – Network slices enable the creation of customized networks to provide flexible solutions for different market scenarios which have diverse requirements, with respect to functionalities, performance and resource allocation.

NOTE 2 – A network slice may have the ability to expose its capabilities.

NOTE 3 – The behaviour of a network slice is realized via network slice instance(s).

3.1.5 network slice instance (NSI) [b-ITU-T Y.3100]: An instance of network slice, which is created based on a network slice blueprint.

NOTE 1 – A network slice instance is composed of a set of managed run-time network functions, and physical / logical / virtual resources to run these network functions, forming a complete instantiated logical network to meet certain network characteristics required by the service instance(s).

NOTE 2 – A network slice instance may also be shared across multiple service instances provided by the network operator. A network slice instance may be composed of none, one or more sub-network slice instances which may be shared with another network slice instance.

3.1.6 role [b-ITU-T Y.3502]: A set of activities that serves a common purpose.

3.1.7 network slice service user (NSsu) [b-ITU-T Y.3103]: The NS service user uses the service(s) provided by the NS instance(s).

3.1.8 network slice service provider (NSsp) [b-ITU-T Y.3103]: The NS service provider is the user of the NS instance(s), and is responsible for providing services to its NS service users via the NS instance(s).

3.1.9 network slice provider (NSp) [b-ITU-T Y.3103]: The NS provider is the owner of the NS instance(s) and provides the NS instance(s).

3.1.10 network infrastructure provider (NIp) [b-ITU-T Y.3103]: The network infrastructure provider is the owner, the provider and the manager of the network infrastructure.

3.1.11 network slice management and orchestration provider (NSmop) [b-ITU-T Y.3103]: The NS management and orchestration provider is responsible for orchestrating NS(s) and managing the lifecycle of NS instance(s) based on NS blueprint(s) (a term defined in [b-ITU-T Y.3100] (see definition in clause 3.1.2)).

NOTE – The NS blueprints can be provided by third parties.

3.1.12 network sub-slice provider (NSSp) [b-ITU-T Y.3103]: The network sub-slice provider is the owner and provider of the network sub-slice instance(s).

3.1.13 network sub-slice management and orchestration provider (NSSmop) [b-ITU-T Y.3103]: The network sub-slice management and orchestration provider is responsible for orchestrating network sub-slice(s) and managing the lifecycle of the network sub-slice instance(s).

3.1.14 network infrastructure management provider (NImp) [b-ITU-T Y.3103]: The network infrastructure management provider integrates the infrastructures of multiple infrastructure providers to offer the combined resources to the NS management and orchestration provider.

3.1.15 management [b-ITU-T Y.3100]: In the context of International Mobile Telecommunications-2020 (IMT-2020), the processes aiming at fulfilment, assurance, billing of services, network functions, and resources in both physical and virtual infrastructure including compute, storage and network resources.

3.1.16 orchestration [b-ITU-T Y.3100]: In the context of IMT-2020, the processes aiming at the automated arrangement, coordination, instantiation and use of network functions and resources for both physical and virtual infrastructures by optimization criteria.

3.1.17 quality of experience (QoE) [b-ITU-T P.10]: The degree of delight or annoyance of the user of an application or service.

3.1.18 quality of service (QoS) [b-ITU-T Q.1741.9]: The collective effect of service performances that determine the degree of satisfaction of a user of a service. It is characterized by the combined aspects of performance factors applicable to all services, such as:

– service operability performance;
– service accessibility performance;
– service retainability performance;
– service integrity performance;
– other factors specific to service.

3.1.19 user plane (UP) [b-ITU-T Y.1714]: Refers to the set of traffic forwarding components through which traffic flows.

NOTE – "User plane" is referred to as "transport plane" in other ITU-T Recommendations.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ABAC	Attribute-Based Access Control
AI	Artificial Intelligence
BER	Bit Error Ratio
CDSSO	Cross-domain Single Sign-On
CP	Control Plane
CPS	Control Plane Services
CPU	Central Processing Unit
CSP	Customer Service Provider
DP	Data Plane
eMBB	enhanced Mobile Broadband
E2E	End-to-End
FCAPS	Fault, Configuration, Accounting, Performance, Security
FIDO	Fast Identity Online
IED	Intelligent Electronic Device
IMT-2020	International Mobile Telecommunications-2020
MAPE-K	Monitor, Analyse, Plan and Execute with Knowledge
ML	Machine Learning
MOS	Mean Opinion Score
NBI	Northbound Interface
NFVI	Network Functions Virtualisation Infrastructure
NIp	Network Infrastructure provider
NMR-O	Network Domain and Resource Orchestrator
NS	Network Slice
NSI	Network Slice Instance
NSmop	Network Slice management and orchestration provider
NSp	Network Slice provider
NSS	Network Sub-Slice
NSSmop	Network Sub-Slice management and orchestration provider
NSSp	Network Sub-Slice provider
NSsp	Network Slice service provider
NSsu	Network Slice service user
OODA	Observe, Orient, Decide and Act
OPEX	Operational Expenditure
OSA	One Stop shop Access
OSS	Operations Support System

PIN	Personal Identification Number
PNF	Physical Network Function
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
SIM	Subscriber Identity Module
SLA	Service Level Agreement
UAF	Universal Authentication Framework
UE	User Equipment
UP	User Plane
URI	Uniform Resource Identifier
URLLC	Ultra-Reliable Low-Latency Communications
vFW	virtual Firewall
vIDS	virtual Intrusion Detection System
vIPS	virtual Intrusion Prevention System
VNF	Virtual Network Function

5 Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option, and the feature can be optionally enabled by the network operator / service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with this Recommendation.

6 Introduction

Network operators and service providers are investigating network slicing for delivering services in the International Mobile Telecommunications-2020 (IMT-2020) and beyond networks to a wide range of vertical industries. The heterogeneity of these vertical businesses (e.g., eHealth, automotive, smart city, industry 4.0, energy and smart grid, etc.) poses a very different collection of requirements for their deployment, from infrastructure resources, network performance to service levels, and at the various phases of the lifecycle of the vertical services.

ITU has defined three typical classes of IMT-2020 network slices in terms of service requirements, including enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC) and massive machine type communications (mMTC). The provisioning of end-to-end (E2E) network slices with proper quality of service (QoS) / service-level agreement (SLA) and quality of experience (QoE) guarantee is one of the key enablers to meet the diverging quality requirements for the various vertical businesses.

Vertical users' satisfaction is of paramount importance for current network management and control infrastructures and is identified as a fundamental pillar for future generation telecom architectures. Indeed, verticals' QoE becomes a relevant metric that should be constantly gauged to understand the current performance of deployed network slices and services, allowing anticipating undesired situations that may result in poor quality scenarios.

However, the subjective nature of QoE, which is a user dependent metric, makes it difficult in a poor-quality situation to understand what are the lower-level infrastructure (virtual or physical) parameters that cause the situation. Thus, in cases of uncertainty of the root cause, a bottom-up approach becomes more efficient. Specifically, such an approach exploits network slice QoS parameters collected from given monitorable and quantifiable infrastructure, which network and service operators have access to, to derive the QoE of vertical users, e.g., using artificial intelligence (AI) / machine learning (ML) models.

If a poor QoE value is derived, network operators can effortlessly understand which are the QoS parameters that resulted in such a value, since these parameters are the ones employed as inputs for the QoE derivation, which are obtained at the network end. Hence, it informs the network operators of what slice QoS parameters should be re-configured to keep optimal quality levels during the whole lifecycle of the service / slice.

Thus, a framework that allows for the derivation of QoE values from monitored QoS parameters should be in place. Since the specific functionalities that analyse the QoS parameters may be use case dependent, the efforts should be focused on the overall framework together with a reference operational workflow, thereby setting up a common ground that could be adapted to a range of use cases.

Meanwhile, QoE-aware network slicing poses several challenges in the network slice management that need to be addressed for efficient end-to-end (E2E) services delivery, including estimating QoE values from monitored QoS metrics and reconfiguration operations (actuations) to support and maintain the desired quality levels. Measuring QoE directly is costly and complex due to human involvement in the process. Machine learning (ML) can be helpful when deriving or predicting QoE values from QoS metrics. By forecasting the QoE degradation based on QoS metrics, which is cost-efficient, the network can alert the orchestrator in advance, which allows it to take remedial actions and correct the problem before it occurs.

Therefore, cognitive network slice management that leverages AI / ML techniques are entailed to maintain the network reactively or proactively in the required state to assure QoE for the vertical users. Consequently, a framework that can satisfy these diverse requirements whilst delivering intelligence-enabled added value to the verticals will be highly beneficial to encourage verticals to embrace the network slicing technologies and IMT-2020, and beyond systems for optimising and evolving their business services of guaranteed and sustained high quality.

In addition to QoE management, ML application for network management automation becomes especially relevant for dynamic network slicing adaptation e.g., to enable on-demand re-configurations to be orchestrated quickly and efficiently even when sufficient human resources, experience and special skills are not available. In this context, ML can help introduce cognitive capabilities to provide the means to analyse the changing environments and adjust the network slice function behaviour accordingly.

An area where ML can also help network slicing management automation is to extract the insights from the previous network management actions and improve the management process over time, including QoS-QoE mapping, fault corrections and performance degradation predictions of provisioned network slices by leveraging on advanced data analytics. Examples of typical vertical use cases that can benefit from cognitive network slicing can be found in clauses I.1 (Smart grid) and I.2 (eHealth).

Furthermore, to achieve a truly E2E delivery of network slices, network operators and service providers shall consider the control and orchestration of heterogeneous resources (and services) deployed in different network domains to fulfil vertical service requirements in terms of performances, SLAs and geographical constraints. Coordinating the subscription, provisioning and operation of such complex network slice based vertical services entails E2E multi-domain service orchestration.

In addition, this E2E multi-domain network slice management and orchestration can benefit from ML to improve and enhance the runtime operation of the delivered IMT-2020 vertical services and network slices. Single-domain performance management evolves based on technology abstraction to enable domain specific technologies, infrastructures and deployments to be easily federated across multiple domains. The pool of available resources is aimed to be exploited by operations in the context of higher level, inter-domain business patterns. As technology related capabilities are abstracted to produce the domain offerings that are made available to multi-domain entities, a similar abstraction approach can be followed to exploit these offerings in the context of the design and provision of vertical-oriented service characteristics. At the level of E2E network slice management, slicing matches the vertical requirements with efficient management and control over the available multiple domains' offerings.

This Recommendation describes an intelligent cost-effective network management and orchestration framework that can cope with the challenges of multi-domain network slicing, while minimizing human intervention towards full automation of slice lifecycle management and runtime operation. A multi-domain cognitive network slice management and orchestration framework is thus presented. The overall architectural framework is described, and the cognitive components in the framework are highlighted focusing on the management of deriving QoE from QoS metrics for network slice instances.

7 High-level requirements

The high-level cognitive multi-domain network slice management and orchestration requirements are as follows:

7.1 General requirements

REQ-G1. The framework is required to enable closed-loop automated and autonomous management of network slices to allow a full automation pipeline without human intervention.

REQ-G2. The framework is required to manage the different levels (service, network slice and resource levels) (in line with the network slice instance definition in [b-ITU-T Y.3100]) of network slicing and network slice based services effectively.

REQ-G3. The framework is required to manage network slices across multiple network service domains, which may belong to different administrative domains, to achieve true E2E network slicing.

REQ-G4. The framework is required to have a built-in cognitive network management, with ML integrated as part of the workflows and operations to support intelligent operations.

REQ-G5. The framework is required to be QoE / QoS aware to meet the diverse QoE / QoS requirements and assure the runtime performance of a variety and wide range of vertical businesses.

7.2 Functional requirements

REQ-F1. Per-service performance monitoring and data analytics functionalities are required to assure that the services' requirements are continuously monitored and satisfied for the vertical services with service issues detected automatically.

REQ-F2. QoE evaluation is required to be conducted automatically and objectively especially for services where feedback from the verticals cannot be acquired or processed.

REQ-F3. A given set of monitored QoS metrics at runtime is recommended to be mapped to real-time or even predicted QoE for both reactive and proactive QoE management approaches.

REQ-F4. Slice-aware policy management is required to allow policy management at the slice level so that actions such as reconfiguration of a specific network slice instance can be applied to optimise its performance.

REQ-F5. Policy management is recommended to be able to generalise rules and comprehend their intent to deal with unknown yet similar conditions.

REQ-F6. Intent-based technology-agnostic control is required to apply the corrective or preventive actuations to a network slice instance regardless of the network slicing technologies employed over a specific network segment or domain.

REQ-F7. Multi-domain coordination is required to deliver tailored E2E services provisioned as a combination of a series of chained single domain network slices to the verticals.

REQ-F8. Either a hierarchical or a peer-to-peer architecture is required to enable multi-domain coordination.

REQ-F9. A common capacity exposure of monitoring and control capabilities is required to allow the evolution of multi-domain ML and automation modules to function properly for the harmonised and uniform exposure of each domain's monitoring and actuation offerings, allowing an optimised selection of domains to compose an E2E network slice.

REQ-F10. Coordination of multiple layers of orchestration logic is required to manage the lifecycle of specific managed entities at different logical layers, which are respectively vertical services, E2E network slices and the single domain network slices composed of technology and domain specific resources, with dedicated orchestration functions.

REQ-F11. Multi-level (service, network slice and resource) orchestration is required to coordinate different levels of decisions coming from layer-specific ML techniques applied over heterogeneous sets of data that are collected from different layers.

REQ-F12. The cognitive framework is required to be applicable to multi-domain FCAPS (Fault, configuration, accounting, performance, security) management regarding the network slice instances.

REQ-F13. Performance evaluation feedback is recommended to improve the ML techniques e.g., to refine their decision making or actuation selection, or trigger the orchestration to rollback workflows whenever the actuations did not provide the desired effect.

REQ-F14. Coordination of the collection of performance metrics and measurements is required to feed the data analysis for ML-based network intelligence.

REQ-F15. The monitoring system is required to be aligned with the orchestration logic to allow informed and fine-grained orchestration operations.

REQ-F16. A data platform is required to store metrics and other monitored metrics to support data-driven ML-based network intelligence.

REQ-F17. The data platform is required to support the import of trained ML models and other external data for an enhanced data-driven ML-based approach.

8 Framework

The overall logical framework proposed for ML-enabled closed-loop cognitive multi-domain network slice control, management and orchestration meets the listed general requirements (REQ-G1-5) in clause 7. It is a closed-loop (REQ-G1), multi-level (REQ-G2), multi-domain (REQ-G3), cognitive (REQ-G4) and QoE / QoS aware (REQ-G5) system.

The framework and its components are compliant with those defined in the relevant existing ITU Recommendations for network slicing, such as [b-ITU-T Y.3103], [ITU-T Y.3112] and [ITU-T Y.3153], for IMT-2020 network management and orchestration [ITU-T Y.3110] [ITU-T Y.3111], for the architectural framework for ML in IMT-2020 [ITU-T Y.3172], and ML-based QoS assurance [ITU-T Y.3170]. Specifically, the framework addresses the functional requirements for network management and orchestration defined in [ITU-T Y.3110], focusing on network slice management and orchestration in terms of lifecycle management, instance management and FCAPS management. Moreover, it is aligned with the conceptual plane-based IMT-2020 network framework from the network slicing perspective defined in [ITU-T Y.3111]. It is noted that the multi-domain network slicing and the capacity exposure to third parties are well aligned to the architecture and functions described in [ITU-T Y.3153], whilst the proposed capacity exposure in this Recommendation emphasizes the exposure of monitoring and control capabilities for the multi-domain autonomous loop. Furthermore, the framework is in line with the functional model of ML-based QoS assurance for the IMT-2020 network defined in [ITU-T Y.3170] and the high-level framework of AI-assisted analysis for network slicing defined in [ITU-T Y.3156], in terms of the overall logic and components, whilst it complements these frameworks by adding the multi-domain and cross-level perspectives, among others. The framework in this Recommendation is also compatible with the high-level architectural components defined in [ITU-T Y.3172] and [ITU-T Y.3178] in terms of the closed-loop ML pipeline.

The framework leverages autonomic principles to proactively enable remedial actuations to maintain the network in the required state to assure QoE, as perceived by the vertical subscribers. Cognition for slice management combines multiple data sources, multi-level and multi-domain to interpret and predict outcomes. A cognitive telecommunication network / service management platform can be represented, in terms of its logical architecture, with monitoring, information, cognition and orchestration sub-planes under the management and orchestration plane. It is noted that the data plane (DP) and the control plane (CP) are not highlighted for brevity, unless otherwise presented. Thereby, taking into consideration the business roles from a network slicing perspective, represented in [b-ITU-T Y.3103],

Figure 1 shows the players that need to implement the identified sub-planes aiming to achieve a cognitive multi-domain E2E network slice management and orchestration.

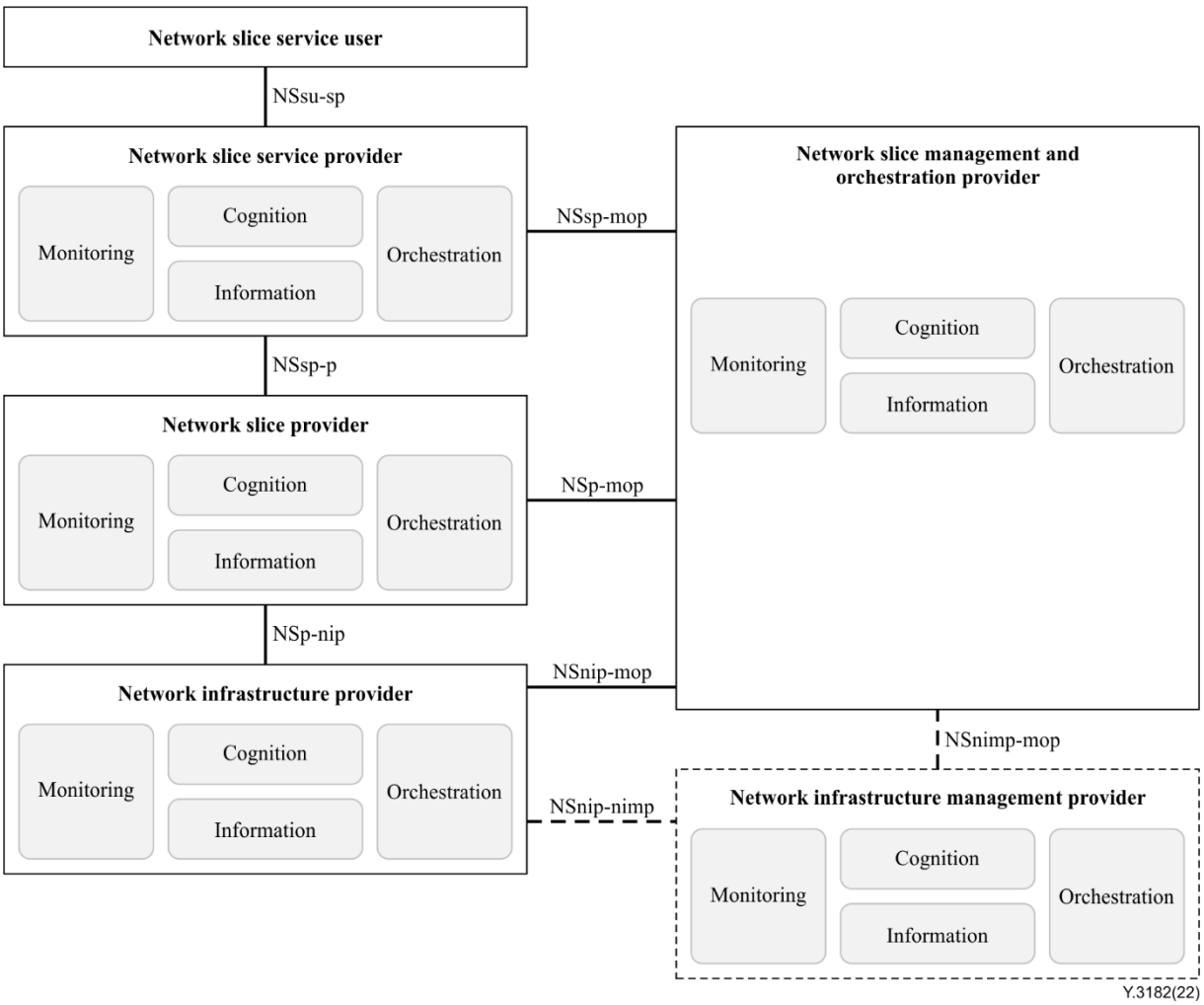


Figure 1 – Cognitive multi-domain E2E network slice management and orchestration framework with business roles

The model described above can be further developed to support hierarchical network slicing (via network slices and network sub-slices) according to Figure 2. This hierarchical architecture meets the multi-domain coordination requirements (REQ-F8 and REQ-F9) in the framework. In particular, the network slice service provider (NSsp) is able coordinate multiple network slice providers (NSps), which in turn can coordinate multiple network sub-slice providers (NSSps) to create E2E network slices across multiple domains and network segments.

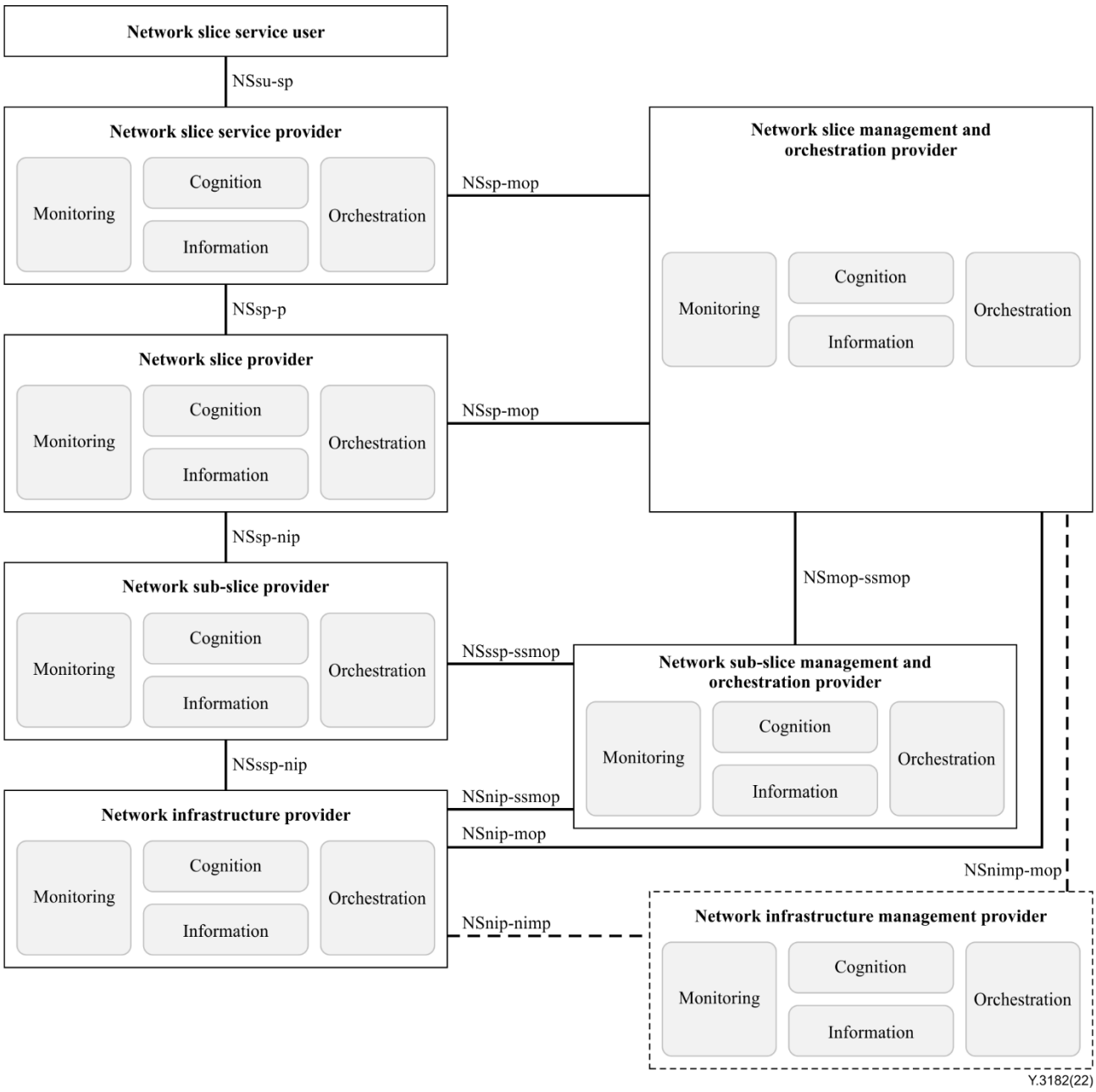


Figure 2 – Hierarchical cognitive multi-domain E2E network slice management and orchestration framework with business roles

Within each block of this framework, the monitoring, information, cognition and orchestration sub-planes compose a management and orchestration plane and together aim to achieve and go beyond the classical FCAPS management functions (REQ-F17) and the silos operations support systems (OSS). These sub-planes in the management and orchestration plane employ a closed-control loop approach (REQ-G1) as represented in Figure 3, together with the control plane and the data plane.

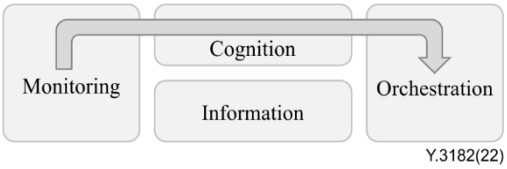


Figure 3 – Closed-control loop

Several players will implement this approach adapted to their specific realities, e.g., the monitoring sub-plane in the network infrastructure provider monitors their network resources in terms of alarms and performance whereas the network slice provider monitors the requirements that belong to a given network slice such as latency, bandwidth, etc. What is relevant for the closed-control loop to function properly is that several sub-planes can interact to ensure the high availability and security of the network infrastructure and respective network slices. This will also contribute in minimising the human effort in maintenance and troubleshooting tasks, thereby significantly reducing the operational expenditure (OPEX).

The architectural framework with the involved components is presented in Figure 4.

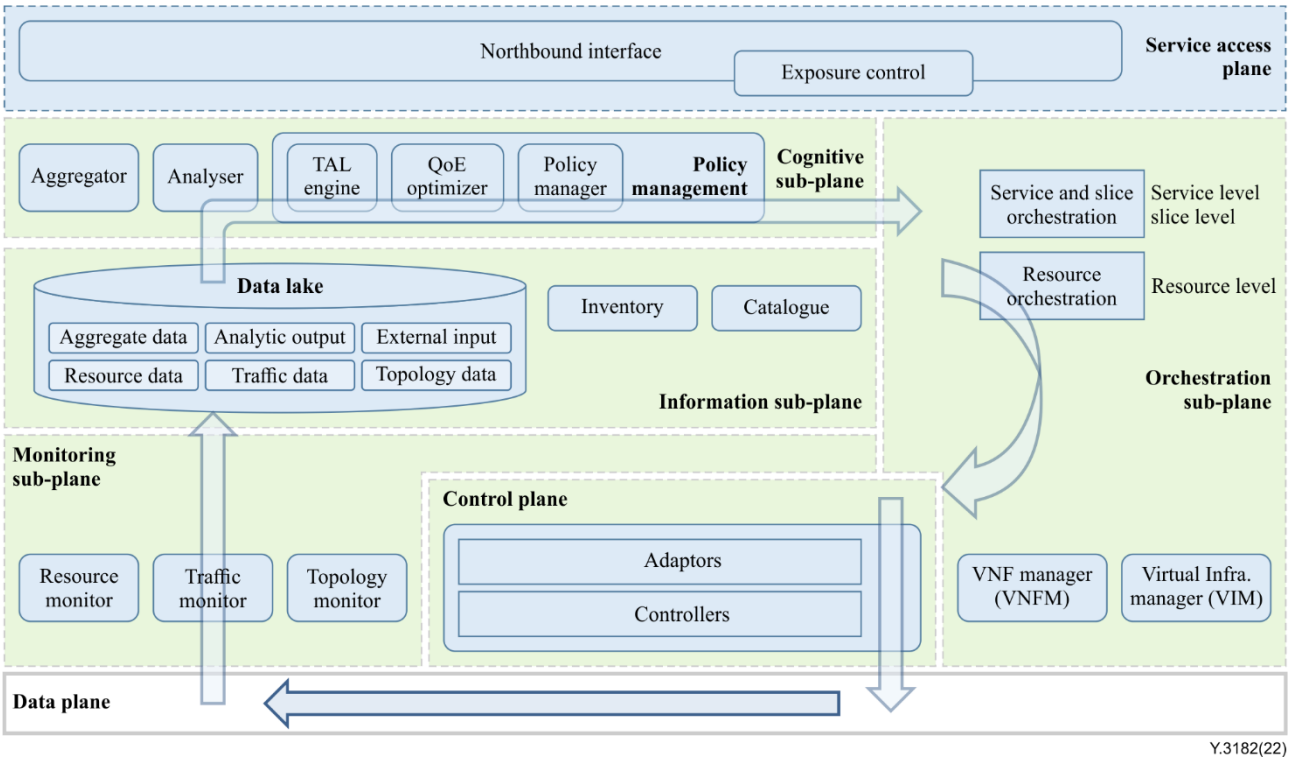


Figure 4 – Cognitive network slice management and orchestration architectural framework with closed-control loop

The **monitoring sub-plane** collects metrics and counters in terms of resources, traffic, topologies, etc. by the corresponding monitoring components (resource monitor, traffic monitor, topology monitor, etc.) from virtual and physical network elements at the network infrastructure provider level, and collects metrics and counters at the NSp and NSsp levels to achieve per-service and multi-level monitoring (REQ-F1). The same concept applies to the network slice management and orchestration provider (NSmop) and to the network infrastructure management provider (NImp) where the monitor sub-plane functions as a repository of events related to management and orchestration tasks to achieve the correlation between monitoring and orchestration (REQ-F15). The monitoring sub-plane integrates and combines heterogeneous sources of metrics and counters information (including possible performance evaluation feedback) in a common and coordinated way (e.g., by employing the TICK stack [b-Arbezzano]), where applicable providing preliminary aggregation of data, at the network infrastructure and the network slice levels exposes the collected and pre-processed data towards the cognitive sub-plane for further aggregation and analysis purposes (REQ-F13, F14). The monitoring sub-plane ingests all its raw data and the aggregated metrics into the information sub-plane, which functions as a cross-level "monitoring" database (REQ-F16, see below). This sub-plane thus meets the following requirements: REQ-F1, REQ-F13, REQ-F14, REQ-F15.

The **information sub-plane** provides a cross-level data platform (e.g., by using a data lake) to store metrics and counters, at the network infrastructure and the network slice and service levels, acting as a monitoring repository and stores all the catalogues and inventories available (REQ-F16). While the catalogues maintain the information related to the platform capabilities and offerings in terms of network infrastructure network elements (virtual and physical network functions, i.e., VNFs and PNFs), descriptors, templates and network slice service descriptors and templates, these inventories keep track of all provisioned instances of network infrastructure network elements and network slices. It comprises various types of data including resource data, traffic data, topology data directly from the monitoring sub-plane, and processed data such as aggregate data, analytic output and external input (REQ-F17). This sub-plane thus meets the following requirements: REQ-F16, REQ-F17.

The **orchestration sub-plane** provides a set of coordination functions required to onboard, provision and maintain network infrastructure network elements (virtual network functions (VNFs) and physical network functions (PNFs) and network slices. It provides functionalities to make the whole management and orchestration plane to work in a coherent way for both single- and multi-domain settings. The E2E multi-domain network slice orchestration is achieved through the hierarchical organisation as shown in the framework (Figure 2), where the orchestration sub-plane in each NSp's domain is responsible for orchestrating the corresponding slices whilst the orchestration sub-plane in the network slice service provider orchestrates the E2E multi-domain network slices (REQ-F7). It is noted that such a hierarchical orchestration system is highlighted although a peer-to-peer orchestration approach is an alternative, where e.g., the originating network slice provider also serves as the NSsp, depending on the business model (REQ-F8). The orchestration sub-plane also interacts with the information sub-plane which provides a heterogeneous set of catalogues and inventories. The orchestration sub-plane is organised into a layered structure consisting of orchestrators for the various levels (services, slices and resources), together with a virtual infrastructure manager and a VNF manager to achieve a coordinated multi-level orchestration and management (REQ-F10, REQ-F11). This sub-plane thus meets the following requirements: REQ-F7, REQ-F8, REQ-F10, REQ-F11.

NOTE 1 – More details of virtual infrastructure manager and the VNF manager can be found in [b-ETSI GS NFV 006].

NOTE 2 – More details of a possible implementation of such an orchestration sub-plane can be found in [b-Cabaça].

The **cognitive sub-plane** uses AI / ML techniques to ensure the operational optimisation for services, network slices and the underlying network infrastructure resources. The intelligence it provides is distributed among its inner modules. The integration of AI / ML techniques into the E2E network slice management and orchestration is essential to achieve an autonomous closed-control loop network. To enable the cognition pipeline in this sub-plane, several cognitive components are in place, including an aggregator, analyser, and policy framework (QoE optimizer, policy manager and a tactical autonomous language (TAL) engine for automation). The policy framework manages policies at the slice level (REQ-F4, F5). These components interact with each other to achieve the QoS to QoE mapping (REQ-F2, F3) for cognitive network slicing. Furthermore, this sub-plane collaborates with the other sub-planes and planes in the framework to achieve cognitive FCAPS management (REQ-F12). This sub-plane thus meets the following requirements: REQ-F2, REQ-F3, REQ-F4, REQ-F5, REQ-F12.

NOTE 3 – The primary objective of the TAL is to facilitate a formal definition of the autonomic behaviours in the framework based on an expandable syntax that aligns with the interactions of the planes and sub-planes. More details of a possible implementation of the TAL can be found in [b-SELFNET].

NOTE 4 – More details of the cognitive sub-plane especially the QoS to QoE mapping are provided in clause 9.

The **control plane** follows an intent-based approach to enforce the actuation intents deployed by the orchestration sub-plane over the networks (the data plane). Intent from the orchestration sub-plane is interpreted through the level of the adaptors into concrete actuations where the controllers can apply

to the specific underlying network segment (REQ-F6). This plane thus meets the following requirements: REQ-F6.

In addition, the **service access plane** provides the northbound interface (NBI) referred to as one stop shop access (OSA) for the users of this framework and controls the exposure of monitoring and control capabilities from a single domain to support multi-domain management and orchestration, through an exposure control component (REQ-F9). This plane thus meets the following requirements: REQ-F9.

9 ML-based cognitive management in the framework

9.1 ML approach and pipeline

The cognitive management in the framework described in clause 8 embraces a closed-loop approach. As an example, the MAPE-K approach [b-IEEE-2003] is adopted for automated and autonomic management, although an alternative closed-loop approach such as the observe, orient, decide and act (OODA) [b-IEEE-2006] is also applicable. MAPE-K is a loop of monitor-analyse-plan-execute governed by a knowledge base that encapsulates policies, rules, algorithms, etc. The monitor step separates the acquisition of monitoring data from the processing of that data and transforms it into network slice QoE metrics. The analysis step uses the acquired knowledge to assess the network slice QoE and the possible impact on the corrective actions. This is done by inferring both learned cognitive models and by applying more traditional automated management methods. The plan and execute steps (termed actuation) are governed by the policy framework.

The cognitive sub-plane employs a data-driven network operations methodology, also known as AIOPS (Artificial intelligence for IT operations) [b-Lerner]. Network analysis applications react to the collected operations data (both raw and processed) and generate new metrics and signals (e.g., QoE metrics and QoE-aware insights) that in turn trigger network operation actions. With this methodology, most components interact only with the data store, acting as consumers and producers. This approach minimises the direct interfaces, provides flexibility and easier integration of cognitive tools. It also allows existing techniques to be used with little change as the outputs of the cognitive tasks can be treated as advanced sensor metrics.

This data-driven approach is a continuation of the monitoring sub-plane to support QoE sensing, and data-operation applications may be deployed for each network slice to filter relevant data, apply security, add context, aggregate network slice metrics, etc. This addresses several of the design challenges related to the monitoring framework (e.g., the approach is scalable and allows attributing the cost of monitoring to each network slice). Moreover, flexible QoE sensors may be employed, from simple aggregation and transformation tasks to the inference of elaborate ML models.

ML algorithms learn and perform better with greater amounts of data. Commonly, network slices do not generate enough operation data to support their own learning processes. Thus, there is a need to combine multiple sources. In the proposed framework, multiple data sources are logically merged to provide all the required information for QoE-aware network slice management. Control and data sensor outputs are collected and persisted to support traditional monitoring through parsing, transformation and aggregation. Furthermore, this data is also employed for ML model training and for extracting QoS metrics. Additionally, feedback from end users, if available, can be combined to allow the data processing application to assume the role of QoE sensors, learning and estimating the end user perspective.

Moreover, the data source may be also fed from external data sources (both raw and processed) that are not created within the controlled system. Leveraging external data sources enables the training of ML models by means of historical data of the infrastructure or other deployed services. This allows for QoE network slice management under practical limitations, where some information must be curated to hide sensitive data or to anonymize it. For example, a network slice provider (NSp) may not be willing to provide some of its raw network metrics but may share processed alerts. Another

example would be data from multiple network slices may be merged and provided as an external source, allowing insights from one network slice to be applied to another.

Ingesting data from external sources is thus a crucial part of the knowledge acquisition process. However, data from other network slices or data from the underlying network slice infrastructure may be subject to confidentiality or privacy limitations. Data ingestion must enforce governance rules dictated by the data owner. In addition, external data may contain corrupt or partial data and thus it must be parsed, validated and cleaned before it enters the data lake. Finally, the data ingestion must receive the data on the sender's "terms", namely, it must handle the data volume and maximal data rates.

The cognitive sub-plane supports an ML pipeline. As a starting point and external to the pipeline, there is a data discovery and gathering phase, this is where the input for ML occurs. Logically, this step represents a data source from the pipeline point of view. Internally, it is divided into the following six different functional steps, covering all the phases from data collection to the ML models lifecycle:

1. **Ingest data**: this step enables the pipeline to read data and its responsibility is divided into two functions:
    a. **Readers**: data input can be multiple files containing observations or streaming data. Each reader abstracts the medium source of the observations and their nuances.
    b. **Normalisation**: data normalisation is the process of combining, merging and cleaning, according to the knowledge gathered from the data analysis, and it includes removing duplicate observations, and removing invalid and/or mis-formed data.
2. **Data analysis**: the initial analysis serves the purpose of gaining data insights and further problem contextualisation. This step runs statistical queries (i.e., counting, averaging, grouping), to check if the dataset is balanced, incomplete or how to focus its modelling.
3. **Transform data**: data transformation depends on data analysis and problem objectives. This step transforms the data into ML-ready. This is where features are extracted and their normalisation (e.g., ordinal, one-hot encoding) happens.
4. **Create model**: ML algorithms, which can cover the classification, prediction or clustering of ML areas are applied in this step. This is where models are effectively trained, optimised (i.e., hyperparameter tuning) and their testing strategies are put in place, including cross-validation, feature importance analysis, dimensionality reduction and so on.
5. **Deploy model**: during the training / testing phase, if a model shows significant fitness metrics values, it can then be deployed into production and start being used to predict, classify, or cluster data in the real-time problem domain.
6. **Monitor and maintain model**: deployed models can lose their effectiveness over time, especially when the data domain is too volatile and dynamic, which means that certain models may be unfit for usage since they no longer properly represent the real world. When models show fitness metrics that are below the configured acceptable values, they are archived, and a re-training task is scheduled to update them. When such a situation occurs, the process reverts to step 4 "**Create model**".

9.2 Workflows to derive QoE from QoS

The cognitive sub-plane provides a framework for the QoE-aware management of network slices on top of a shared IMT-2020 network infrastructure. QoE-awareness allows identifying degradation of the network slice service user (NSsu's) QoE by directly monitoring user feedback or by inferring QoE from QoS derived from measured slice infrastructure metrics. Examples of user feedback are quality metrics transmitted from the user equipment (UE), feedback about user satisfaction from the NSsu, etc. Examples of QoS parameters are those derived from infrastructure network traffic such as latency, throughput and packet loss, and resources such as central processing unit (CPU) and memory.

The following sections present two approaches to leverage ML to estimate the QoE of a network slice and then trigger a remedial action to re-configure the network slice.

9.2.1 Estimating QoE from network QoS

In this strategy, the relationship between the target application's QoE and the QoS is learned during the training phase, whilst at run-time only metrics from the provider's infrastructure are collected to derive the QoS. This run-time QoS is employed to infer the QoE from the training models to trigger remedial actuations when required. The premise is that although the provider can only observe partial network information, the inferred QoE is exposed in these observations.

With this goal in mind, an example scenario for the training of the ML model is described as follows. Two network slices that share infrastructure resources are demanded. One of the slices supports the target application e.g., real-time video streaming, whilst the other slice is running background network traffic to generate network congestion. Since the slices share infrastructure resources, the background traffic will interfere with the performance of the target application, which will in turn affect the target application's QoE. Training data can be created from simulations of various levels of traffic congestion running in parallel to the target application. While running these simulations, QoS metrics are collected from the provider's infrastructure monitoring framework. At the same time, QoE metrics are collected from the user application, such as mean opinion score (MOS) based on feedback from the actual end users or UEs running the target applications. An ML model correlating the QoS features with the target QoE metrics is generated which estimates the QoE from QoS.

At run-time, this QoE estimation model combined with current QoS measurements serves as triggers for remedial actions by the cognitive sub-plane. In the cases of unfavourable QoE, policies defined for the network slice would specify which remedial actions need to be triggered to address the unfavourable QoE. These remedial actions would be communicated to the orchestration sub-plane to execute the desired (re-)configurations. Examples of remedial actions are adjusting the network slice bandwidth, scaling overloaded VNFs, migrating VNFs, or handing the slice to another network slice provider.

The main difference between the run-time and training phase is that the QoE input from the NSsu, i.e., UE quality metrics is consumed in the training phase but not at run-time; instead, at run-time the QoE is derived from the QoS. In addition, no remedial actions are triggered during the training phase since they are not relevant to the generation of the QoE prediction model.

NOTE – UE quality metrics may include subjective QoE metrics collected from the NSsu and/or objective metrics measured at the UE. Appendix I provides examples of metrics for the described use cases.

Figure 5 and Table 1 illustrate the workflow of the estimating QoE from the network QoS during the ML model training phase.

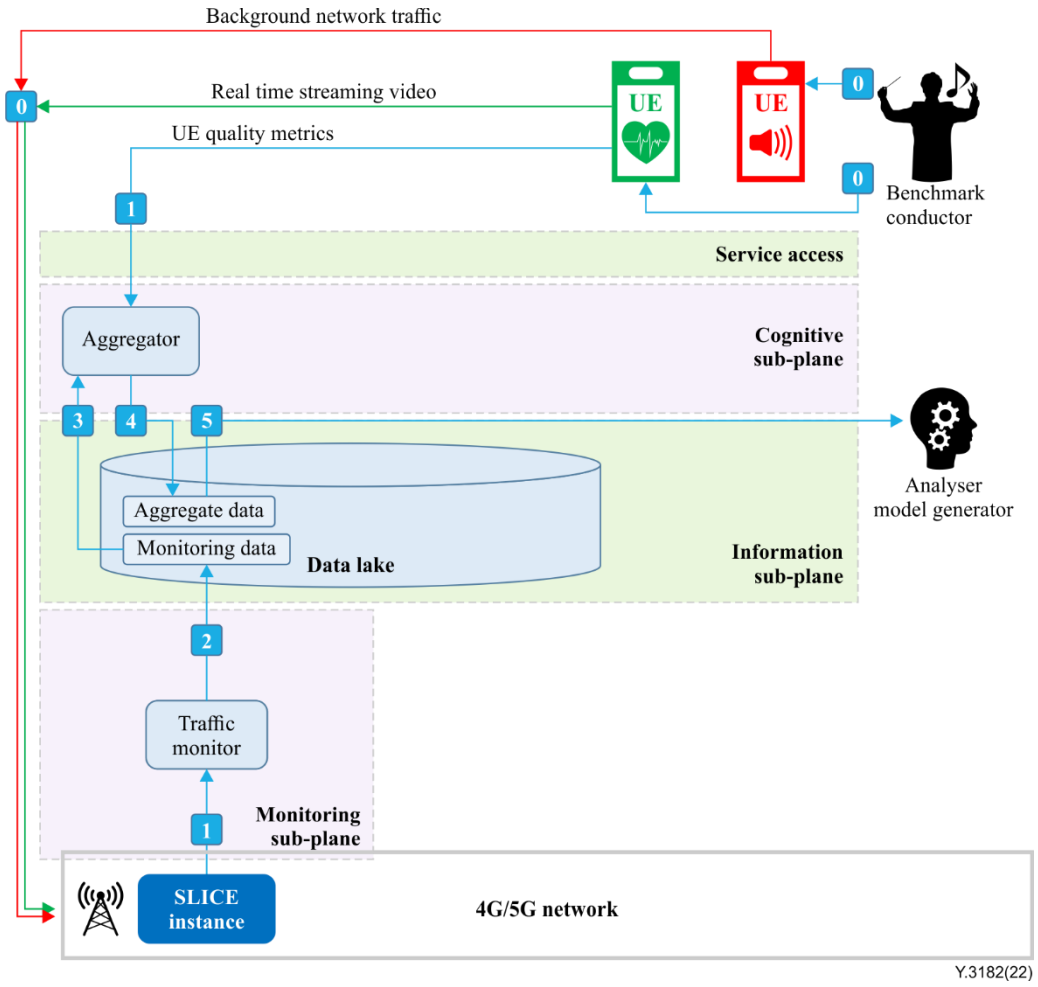


Figure 5 – Estimating QoE from the network QoS (training)

Table 1 – Estimating QoE from the network QoS (training)

Step	Description
0	The training phase requires the collection of many samples of a network slice for a real-time video streaming application with various levels of background network traffic (including no background traffic). A **benchmark conductor** is used to orchestrate the running and labelling of these benchmark samples. It runs numerous benchmarks and each benchmark consists of a real-time video stream and some background network traffic. Each benchmark is assigned a benchmark ID, which is used to map UE quality metrics from the UE's streaming video device, and network flow metrics from the network slice's VNF. NOTE – The **benchmark conductor** is for training phase purposes only, and is not an architectural component in the framework.
1	The **traffic monitor** (together with the **resource monitor**) continuously collects network flow metrics related to a VNF interface serving the stream as well as network flow metrics related to the rest of the network infrastructure. In parallel, the UE streams its quality metrics to the **aggregator**.
2	The **traffic monitor** stores the collected metrics in the data lake.
3	The **aggregator** consumes the traffic metrics.
4	The **aggregator** transforms the traffic metrics and the UE quality metrics into QoS ML-ready features and inserts the transformed data into the data lake.

Table 1 – Estimating QoE from the network QoS (training)

Step	Description
5	Once all the benchmarks are run, an **analyser model generator** will: • Read the accumulated QoS metrics and UE quality metrics from the **data lake**, • Derive target QoE estimations from the UE quality metrics aggregation, • Derive QoS features aggregation, which is highly correlated with the QoE estimations, • Generate ML model with target QoE estimations from QoS features, and • Persist the ML model. The ML model is then used at run-time to derive QoE estimations from the QoS metrics. NOTE – The **analyser model generator** is for training phase purposes only, and is not an architectural component in the framework.

Figure 6 and Table 2 illustrate the workflow of the estimating QoE from the network infrastructure QoS during run-time.

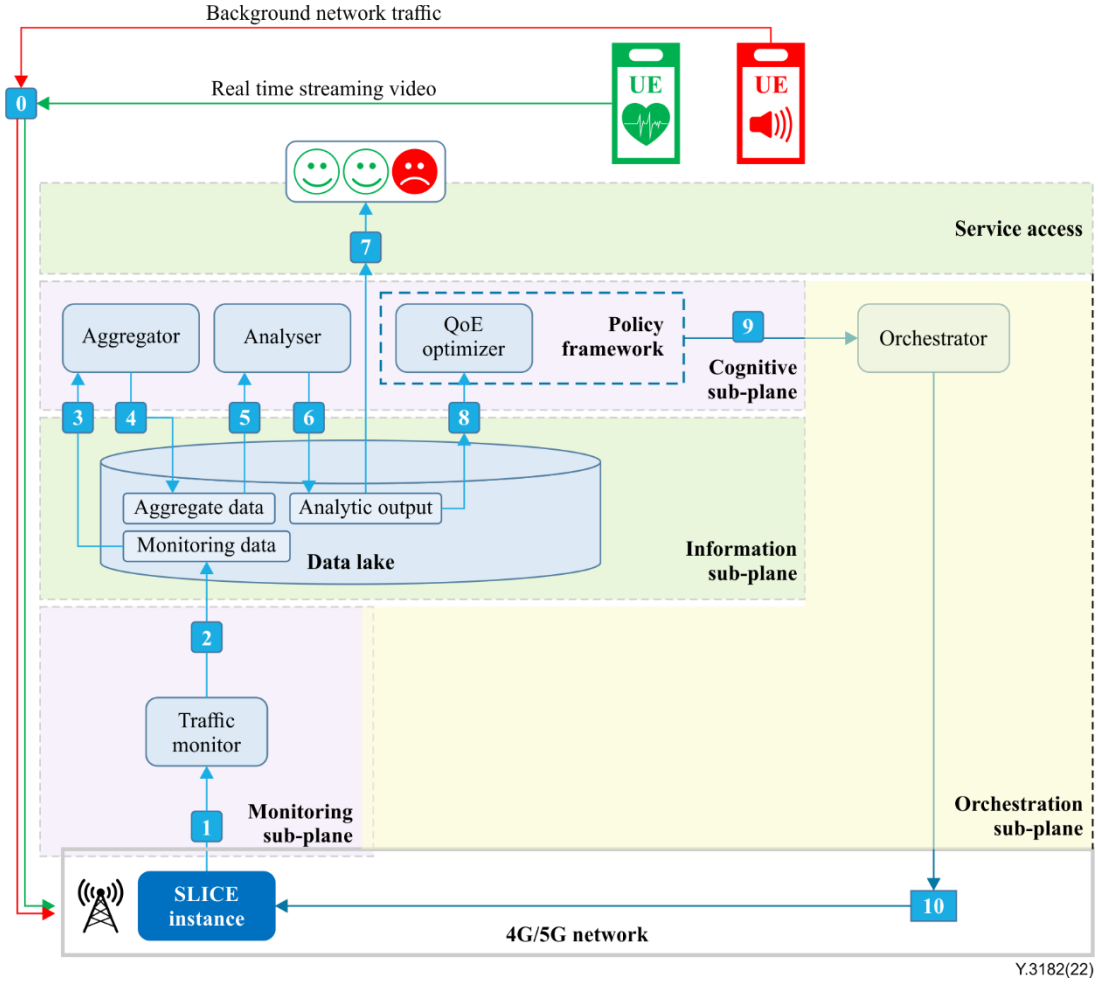


Figure 6 – Estimating QoE from the network QoS (run-time)

Table 2 – Estimating QoE from the network infrastructure QoS (run-time)

Step	Description
0	The UE streams video into the network slice assigned to the target application. In the background, network traffic is generated by the other UE devices to another slice.
1	The **traffic monitor** (together with the **resource monitor**) continuously collects the network flow metrics related to a VNF interface serving the video stream, as well as the network flow metrics related to the rest of the network infrastructure.
2	The **traffic monito**r stores the collected metrics in the **data lake**.
3	The **aggregator** consumes the traffic metrics.
4	The **aggregator** transforms the traffic metrics into QoS ML-ready features and inserts the transformed data into the **data lake**.
5	The **analyser** consumes the QoS data from the **data lake**.
6	The **analyser** (i.e., the ML model) analyses the QoS data, derives the QoE estimation of the target network slice, and produces a QoE estimation and inserts it into the **data lake**.
7	The QoE estimation is displayed over time.
8	The **QoE optimizer** consumes the QoE estimations.
9	When the **QoE optimizer** observes a series of unfavourable estimations, it then decides based on the policies specified for the slice when it was created, the most suitable remedial actions to overcome the undesired states. The decision is delivered to the **orchestrator**.
10	The **orchestrator** orchestrates the actual network re-configuration.

9.2.2 Estimating QoE from the UE quality metrics

In this strategy, the target application's QoE is inferred from the UE quality metrics. The QoE is inferred using the QoE estimation model trained using the UE quality metrics during the training phase. At run-time, the same UE quality metrics are collected and streamed into this QoE estimation model to trigger remedial actuations when required. Specifically, the UE transmits its quality metrics to the cognitive sub-plane's aggregator which feeds the metrics into the analyser's QoE estimation model. The QoE estimation model then triggers the QoE optimizer and policy framework to decide on the proper remedial actuations.

Figure 7 – Estimating QoE from the UE quality metrics (run-time) and Table 3 illustrates the workflow of estimating QoE from the UE quality metrics during run-time.

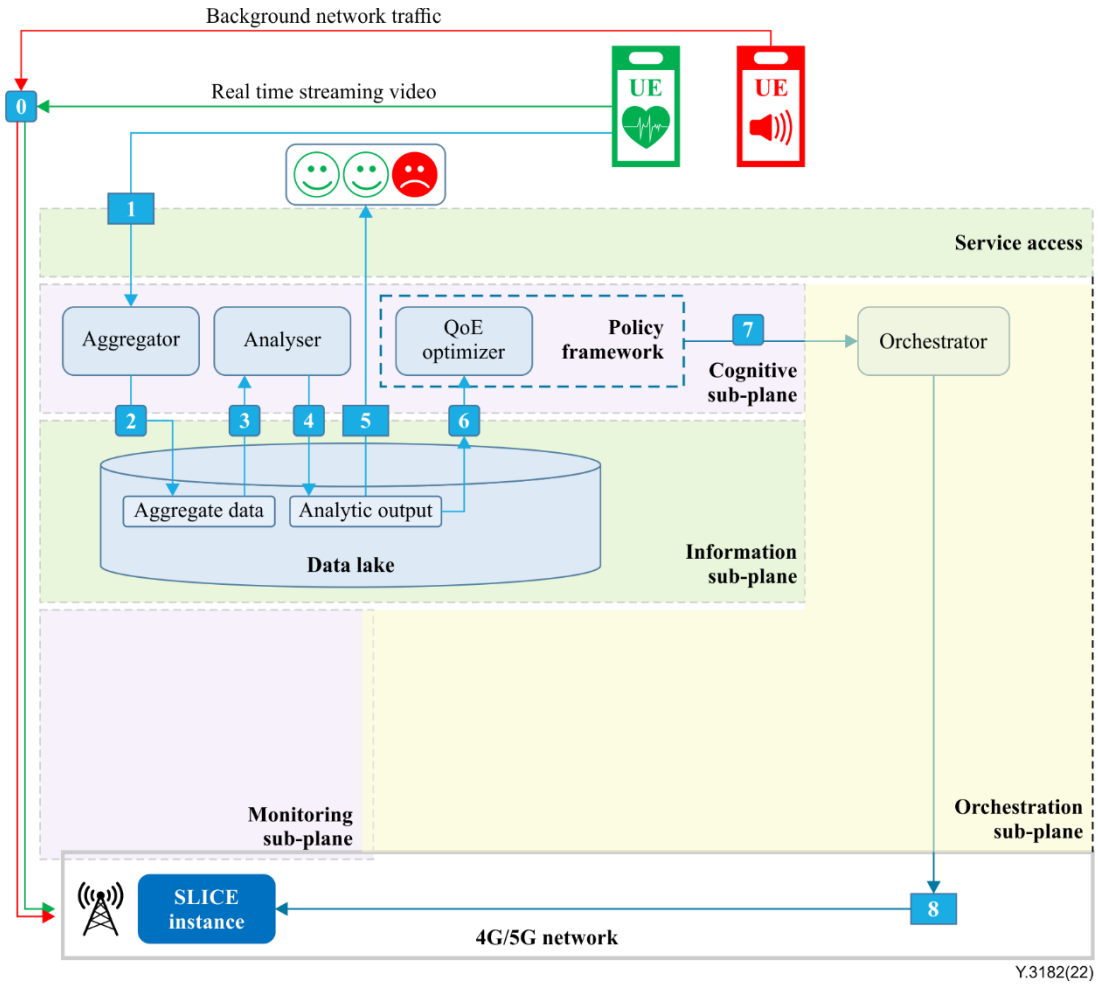


Figure 7 – Estimating QoE from the UE quality metrics (run-time)

Table 3 – Estimating QoE from the measured UE quality metrics (run-time)

Step	Description
0	The UE streams video into the network slice assigned to the target application. In the background, noise is generated by the other UE devices to another slice that is generating background network traffic to create network congestion.
1	The UE streams its quality metrics into the **aggregator**.
2	The **aggregator** aggregates the UE quality metrics and transforms the data into ML-ready features. Then, it inserts the transformed data into the **data lake**.
3	The **analyser** consumes the **aggregator**'s output.
4	The **analyser** (i.e., the ML model) analyses the **aggregator**'s output and inserts its QoE estimation into the **data lake**.
5	The QoE estimation is displayed over time.
6	The **QoE optimizer** consumes the QoE estimations.
7	When the **QoE optimize**r observes a series of unfavourable estimations, it then decides, based on the policies specified for the slice when it was created, the most suitable remedial actions to overcome the undesired states. The decision is delivered to the **orchestrator**.
8	The **orchestrator** orchestrates the actual network re-configuration.

NOTE – The cognitive multi-domain network slice management and orchestration framework can also be leveraged for the FCAPS management in relation to the multi-domain network slicing. For example, it considers security as a dedicated and specific constraint that verticals can express when they request the

provisioning of their services. More details of a possible implementation of such FCAPS management can be found in Appendix II.

10 Security considerations

This Recommendation presents the architectural framework for E2E multi-domain network slice management and orchestration enabled by machine learning, which is expected to be applied in future networks including IMT-2020. Therefore, general network security requirements and mechanisms in future networks should be applied [ITU-T Y.2701] [ITU-T Y.3101].

Furthermore, security aspects for consideration within the cloud computing environment, including data management are addressed by security challenges for the customer service providers (CSPs), which are described in [ITU-T X.1601]. [ITU-T X.1601] analyses security threats and challenges and describes security capabilities that could mitigate these threats and meet the security challenges.

Moreover, multi-domain security management should be in place. In a multi-domain context, security is required for multi-domain network slice orchestration and cognitive sub-plane security. The first aspect is detailed in Appendix II, while the second aspect covers securing the communication among the cognition sub-plane components. In particular, constraints among the cognitive sub-plane components are defined and translated into specific network security requirements. These requirements need to be fulfilled by security network functions to be deployed as part of the cognitive slices providers in support of encryption as a service (vEaaS), intrusion detection/virtual intrusion detection system (vIDS), intrusion prevention/virtual intrusion prevention system (vIPS), firewalling/virtual firewall (vFW) and so on.

In addition, it is required to prevent unauthorized access to, and data leaking from, the ML pipelines, whether they have a malicious intention with the implementation of mechanisms regarding authentication and authorization, external attack protection and so on.

Appendix I

Example use cases for multi-domain E2E slice management and orchestration

(This appendix does not form an integral part of this Recommendation.)

The following use cases are presented to help understand the applicability of the proposed framework in this Recommendation for vertical businesses, including the business roles, operation overview diagram, precondition and postcondition of the multi-domain E2E network slice management and orchestration operations, and derived requirements in the QoS / QoE.

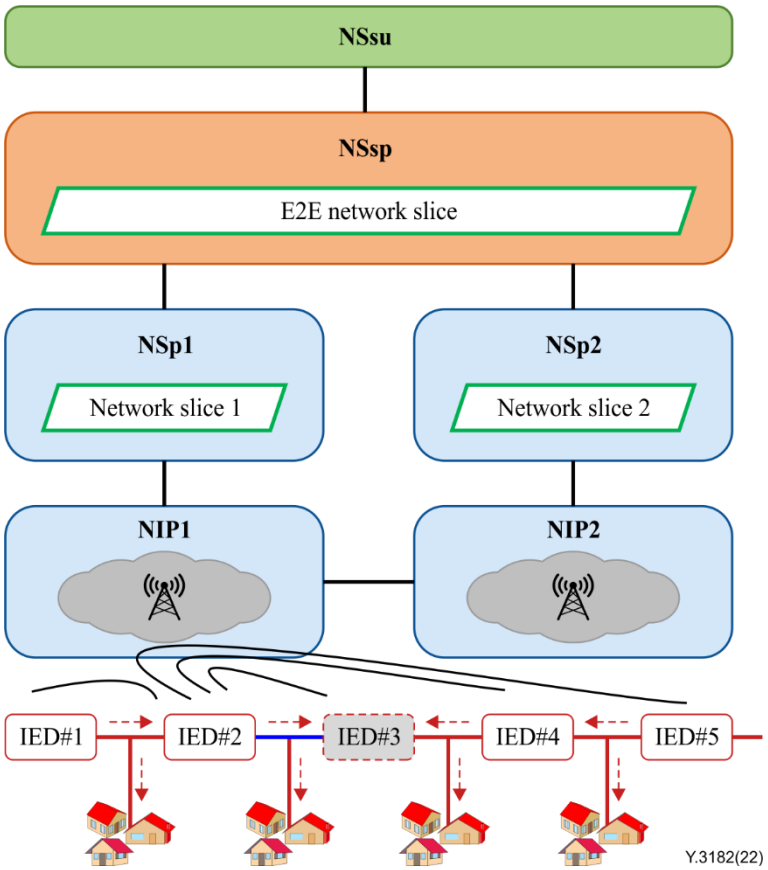
I.1 Smart grid vertical service use case

The smart grid use case is implemented in line with the IMT-2020 URLLC network slice to demonstrate a fully decentralized high-speed self-healing solution for electric power grids. These self-healing solutions rely on distributed automation and power system protection and aim at increasing energy supply QoS by reducing the number of customers affected by power outages, as well as the frequency and duration of these outages. The use case highlights the potential for IMT-2020 slicing to leverage critical systems supported by the IMT-2020 network infrastructures. Table I.1 summarises this use case.

Table I.1 – Smart grid vertical service use case

Title	Smart grid vertical service use case
Description	• Network slice service user (NSsu, in this case the smart grid operator) requests a URLLC service from a network slice service provider (NSsp). • NSsp is responsible for managing the service lifecycle, including its exposition to the NSsu and creation (composition of the E2E network slice (NS) across one or multiple administrative domains). • Network slice provider (NSp) responsible for managing the NSs lifecycle, including their creation and exposition to the NSsp, as well as their provision, monitoring and optimization. • Network infrastructure provider (NIp) is the owner, the provider and the manager of the network infrastructure. The NIp is responsible for managing the network resources lifecycle (VNFs / PNFs). One NIp represents one administrative domain. The smart grid use case aims to benefit from the IMT-2020 URLLC network to implement and demonstrate an advanced self-healing solution for electric power grids. The smart grid use case comprises three scenarios (1) Protection coordination, (2) Automatic reconfiguration and (3) Differential protection.
Roles	Player 1 – NSsu (Vertical); Player 2 – NSsp (E2E NS provider); Player 3 – NSp+NIp (administrative domain providing network slices to the NSsp). Figure I.1 shows the perspective of NSsu.

Table I.1 – Smart grid vertical service use case

Figure	 **Figure I.1 – Smart grid use case from the NSsu perspective**
Pre-conditions (optional)	The vertical power system must be operating normally (all target sections must be energized) and all field IEDs (intelligent electronic devices) protection devices communicating with each other using IEC 61850 R-GOOSE [b-IEC 61850] and with the control centre / substation.
Post-conditions (optional)	Once the use case has been executed, the entire system should maintain normal operation in terms of communication. The metrics defined for the pre-conditions should be used for the post-conditions as well. The use case scenario will be successful if the R-GOOSE events are received by the IEDs within the defined time, i.e., in time for the protection functions to coordinate.
Derived requirements	The system should be monitored for a pre-defined period while in normal operation before the use case scenario events take place. QoS must be evaluated for peer-to-peer communications between IEDs and for communications between the IEDs and the control centre / substation. The following metrics should be used for measuring QoS: • End-to-end latency; • Packet loss / Bit error rate (BER); • Out-of-order packets.

I.2 eHealth vertical service use case

The eHealth use case, using a connected ambulance, aims to provide support to medical emergency first responders by developing a platform that can rapidly provision dedicated IMT-2020 eMBB network slices to advance the emergency ambulance services through the design of better-connected,

integrated and coordinated healthcare. The connected ambulance will act as a connection hub for the emergency medical equipment and wearables, enabling storing and real-time streaming of video data to the awaiting emergency department team at the destination hospital. By providing prioritized life-critical video-streaming from inside a high-speed moving ambulance, the use case achieves "reliable and dependable QoS and QoE with 'zero perceived' downtime". It will use eMBB, requiring both extremely high data rates and low-latency communication in some areas and reliable broadband access over large coverage areas. Table I.2 summarises this use case.

Table I.2 – eHealth vertical service use case

Title	IMT-2020 network slicing for mission-critical services
Description	Vertical (eHealth ambulance service) – requests an eMBB service from an NSsp NSsp – responsible for managing the service lifecycle, including its exposition to the NSsu and creation (composition of the E2E network slice across one or multiple administrative domains), aggregating multi-domain FCAPS and hosting the optimisation over multiple administrative domains. NSp – responsible for managing the NSs and involved network resources lifecycle, including their creation and exposition to the NSsp, as well as their provision, monitoring and optimization. One NSp represents one administrative domain. eHealth use case aims at leveraging from the public safety service that takes priority over all other network traffic (e.g., industries 4.0, smart city, ad hoc access). It is crucial to guarantee the SLA for the service, e.g., availability, delay, bandwidth, coverage, security, etc. the proposed approach is meeting these requirements with the approaches below: • NBI OSA towards the vertical, with exposure control functionalities for service monitoring, reconfiguring and auto scaling. • Cross-domain, cross-plane orchestration to provide dynamic slicing and dynamic reconfiguration based on priority level. • Cognitive, agile QoE management of network slices for service assurance of vertical business. • End-to-end network slice FCAPS management to manage fault, configuration, accounting, performance, and security of all network slices across multiple planes and network operator domains.
Roles	NSsu, NSsp, NSp, NIp

Table I.2 – eHealth vertical service use case

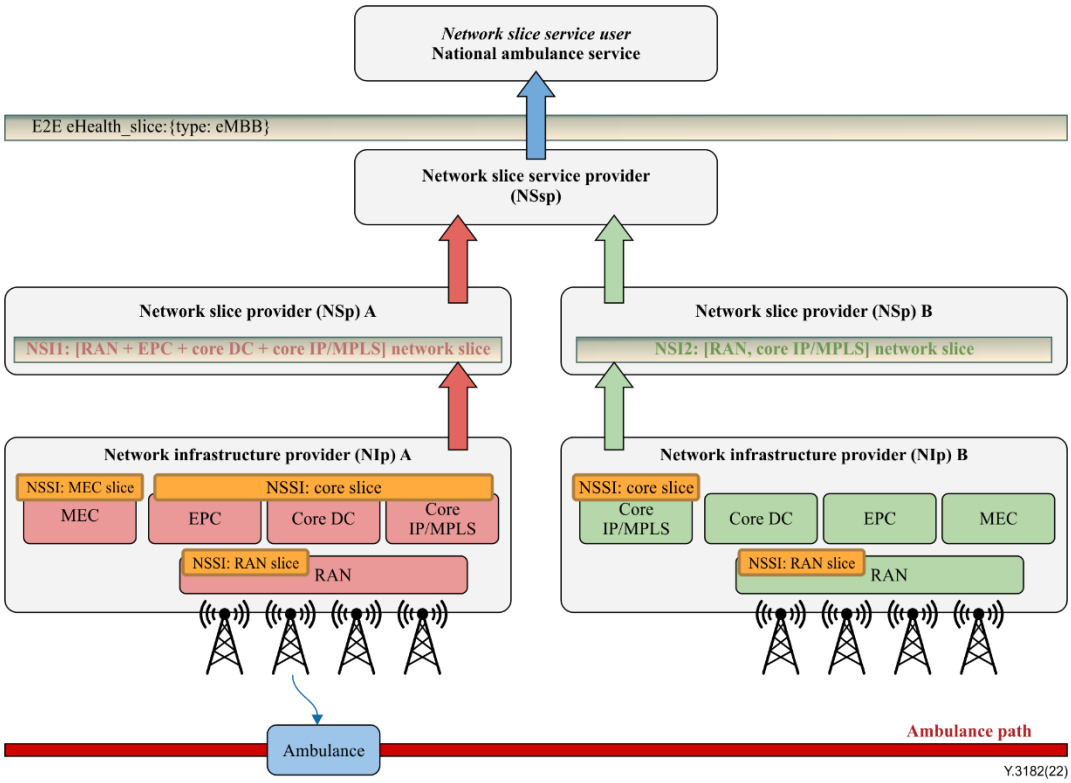
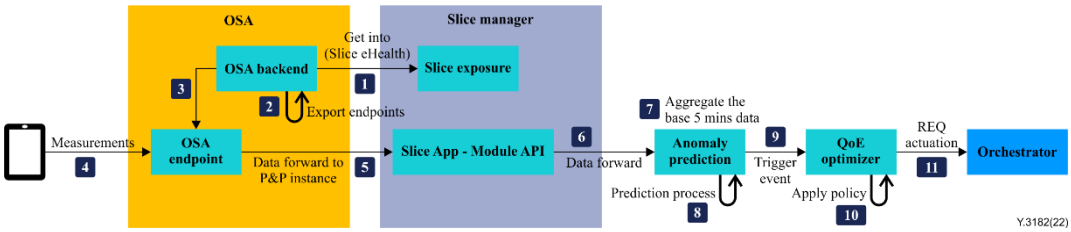
Figure (optional)	 **Figure I.2 – eHealth business model from the NSsu perspective**  **Figure I.3 – Continuance measurements** Figure I.2 shows the overall E2E deployment of the eHealth use-case. To maintain the QoE of the eHealth use-case, continuance measurements are feed from UEs to the anomaly prediction model as depicted in Figure I.3. In the NSsp domain, the OSA backend retrieves the eHealth slice information from the slice exposure in the slice manager. Then, the OSA backend creates an OSA endpoint that allows the UE to submit the collected measurements, which will be forwarded through the slice manager to the anomaly prediction model. The anomaly prediction model produces an event-based trigger to the QoE optimizer that includes a QoE value that characterises the link quality in the next five minutes. Based on the QoE value and the slice policy, the QoE optimizer may request an actuation from the orchestrator. The orchestrator uses the proper interface to request the actuation from the responsible NSp.
Pre-conditions (optional)	Subscriber identity module (SIM) cards need to be installed in ambulances and used by UEs that deliver the services.
Post-conditions (optional)	Once the use case has been executed, the entire system should maintain normal operation in terms of communication. The metrics defined for the pre-conditions should be used for the post-conditions as well.

Table I.2 – eHealth vertical service use case

Derived requirements	The system should be monitored for a pre-defined time period while in normal operation, before the use case scenario events take place. QoS must be evaluated for E2E communications between ambulances and eHealth services (ML and streaming). The following metrics should be used for measuring QoS and then later mapped to the QoE metrics: • End-to-end latency; • Packet loss / Bit error rate (BER); • Physical layer conditions.

Appendix II

Example of multi-domain FCAPS management

(This appendix does not form an integral part of this Recommendation.)

II.1 Multi-domain security management

This appendix presents a use case of applying the proposed framework in the FCAPS management operations for network operators, focusing on multi-domain security management including 1) security events monitoring, detection and actuation, 2) identity and access management, 3) cross-domain trust model, 4) independent domain authentication, and 5) cross-domain single sign-on.

When the security functions have been deployed, either by orchestration when initiating a network slice or by the network operators, the functions will go to the operation mode where security events will be monitored, relevant data will be collected and fed to the functions for security events detection, and later to the cognition sub-plane for analytics, decision-making and planning. At this point, the remediation can be triggered by the security functions themselves, e.g., running a script to add a new firewall rule to block an intruder based on the ML-based intrusion detection, or the functions will trigger an alert or another event either to the orchestrator for policies-based actuation or to the admin / operator for further investigation. For a service analysis for inter-domain, the security functions / FCAPS manager will be in the NSsp domain and information for those models will be injected from NSp to NSsp. Figure II.1 summarises the procedure. In that case the actions are exposed as a set of configurations (e.g., firewall rules, NF config).

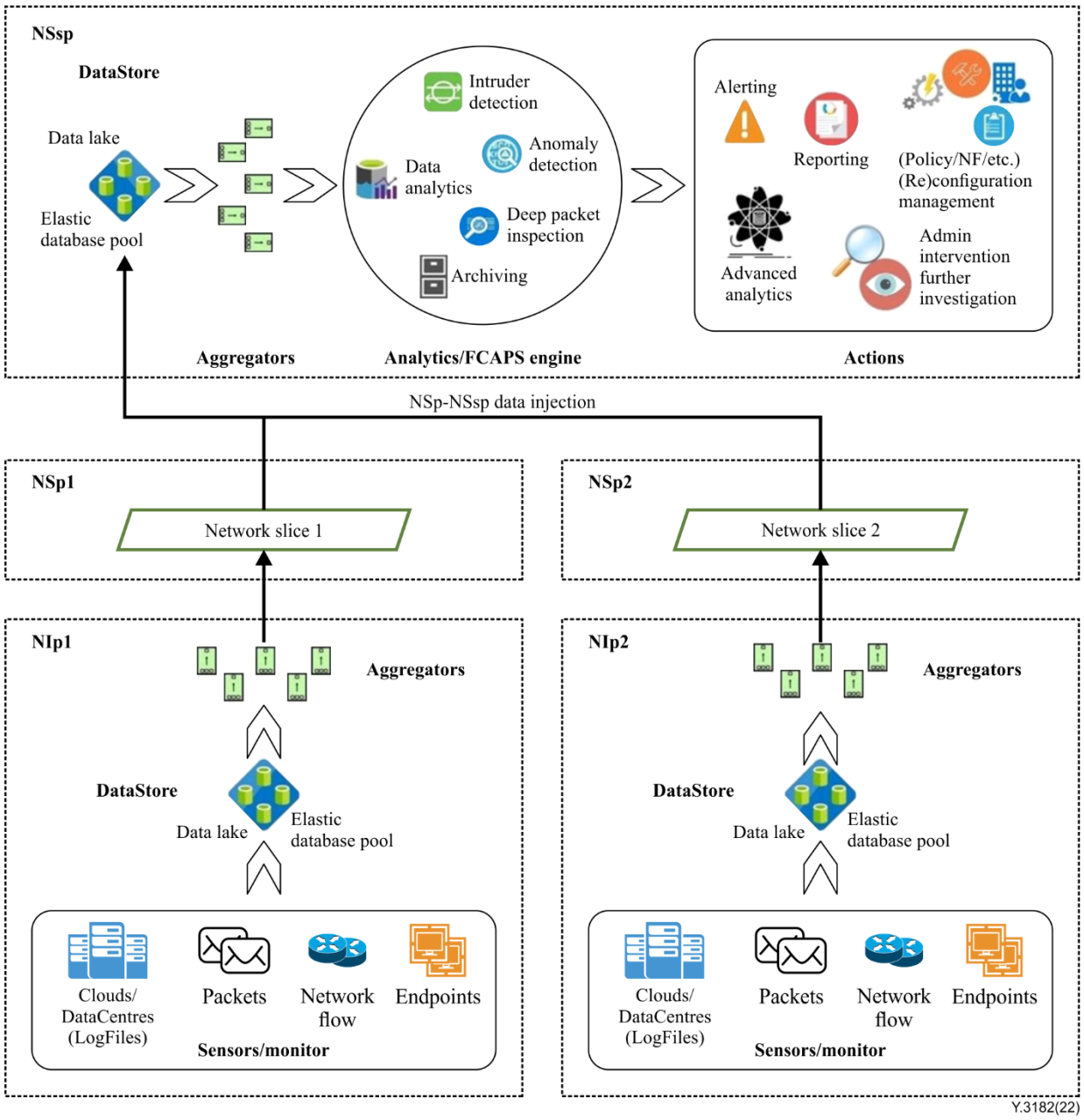


Figure II.1 – Security events monitoring, detection and actuation

II.2 Identity and access management

Inter-domain security management will ensure the required integrity and authorisation of the inter-domain access, which includes users' access to the management (NSsp admins, NSp admins, etc.) to design and onboard a network slice, construct policies for a network slice, security policies for the system, components and network slices, or to configure a feature, etc., and users access to the service offerings or applications (NSsu users, NSsp users to the NSp domain, etc.) in order to browse and subscribe for a service and to interact with their services at runtime, etc.

For this, the first phase will be for the system to identify the users or entities based on authentication technologies, e.g., performing identification authentication of users or entities by evaluating required login credentials (e.g., passwords, personal identification numbers (PINs), biometric scans, security tokens, etc.), or by multi-factor authentication which requires two or more authentication factors, is often an important part of the layered defence to protect access control systems.

After this identity authentication, the system will rely on an access control system that implements a process for defining security policy and regulating access to resources such that only authorized actors

are granted access according to that policy. Having an access control system is fundamental to mitigating the risk of unauthorized access from malicious external users and insider threats, as well as the risk of loss or exposure of critical assets, etc. In general, access control can be rule-based / role-based / attribute-based access where the access permission for the users or entities will depend on the conditions (rules), the role of these users or entities, the attributes of the actors, or it can be a combination of these types.

There are existing standardisation and related documents to refer to for a strong authentication ecosystem and access management, and below are some essentials:

- [b-ITU-T X.1277] with universal authentication framework (UAF) describes the components, protocols and interfaces that make up the fast identity online (FIDO) UAF strong authentication ecosystem. The goal is to provide a unified and extensible authentication mechanism that supplants passwords while avoiding the shortcomings of current alternative authentication approaches. Following the agile approach, it allows the relying parties to choose the best current authentication mechanism for the end user / interaction, while also preserving the option to leverage emerging device security capabilities in the future without requiring additional integration effort.
- NIST has developed an example of an advanced access control system, attribute-based access control (ABAC) [b-NIST 1800-3B], which can manage the access to networked resources more securely and efficiently, and with greater granularity than traditional role-based access management. It enables the appropriate permissions and limitations for the same information system for each user based on the individual attributes and allows for permissions to multiple systems to be managed by a single platform without a heavy administrative burden.
- [b-ITU-T X.812] defines the basic concepts for access control, demonstrates the manner in which the basic concepts of access control can be specialized to support some commonly recognized access control services and mechanisms, defines these services and corresponding access control mechanisms, identifies functional requirements for protocols to support these access control services and mechanisms, identifies management requirements to support these access control services and mechanisms and addresses the interaction of access control services and mechanisms with other security services and mechanisms.

II.3 Cross-domain trust model

This Recommendation introduces a cross-domain slicing architecture that creates E2E network slices operating across multiple network service provider domains. This means, a cross-domain trust model should be addressed so that the E2E network slices working across security domain boundaries can form trust. The trust model should clearly define all the actors, roles and rules that involve in cross-domain slicing architecture to ensure that the network slices can operate seamlessly from end-to-end. This trust model should also support the adaptation for different SLA agreements between business partners, at least to address the business use cases it supports.

To carefully design a cross-domain trust model, it needs to address the basic questions, including:

- Who are the entities / actors involved?
- What are the resources that need to be protected?
- Who provides the identity authentication service?
- Who provides the authorisation service?
- What is the trust relationship among business partners?
- What is the business model?

Table II.1 shows a list of resources.

Table II.1 – List of resources and their description

Resources	Description
End-to-end network slice instances	End-to-end network slice instances are provided by the NSsps to the verticals, depending on the vertical needs specified in the SLA agreement between the verticals and the NSsps.
Network slice / sub-slice instances	Network slice (NS) / Network sub-slice (NSS) instances are provided by the NSps to the NSsps based on the NS / NSS offerings, and the subscription from the NSsp for the NSps offerings to fulfil the E2E network slice requirement that the verticals require from the NSsps.
Control plane services (CPS)	CP services that are identified by RESTful uniform resource identifiers (URIs), allowing clients to access the CP services such as QoS control, NF config, etc.
Other services	Other services include monitoring services, data-plane services, management / orchestration services, etc.
Vertical data	Vertical data that is stored or generated during the runtime of the vertical's E2E network slices. For example, in the eHealth use-case, the vertical data can be the video in the ambulances, or patient information, etc.
Network data	The data related to the network performance (e.g., signal strength, bandwidth, packet loss, jitter, etc.) or data related to the traffic flows, etc.
Infrastructure	Network infrastructure includes the access network (e.g., radio access network (RAN), edge, routers, switches, etc) and core network (core data centre, routers, etc). Each NSp has its own network infrastructure and it is out of the scope of this Recommendation to secure the infrastructure.

II.4 Independent domain authentication

The simplest solution is that the authentication service is handled independently in each domain (Figure II.2), NSsu user will register with the NSsp domain to have a user account, then use this user account to authenticate with the NSsp via the NSsp OSA to access the NSsp services (e.g., service subscription). For NSsp and NSp, the NSsp will register with the NSp to have an account in each NSp domain (e.g., "nssp_nsp_1" for NSp1, "nssp_nsp_2" for NSP2). Using the corresponding account, the NSsp can have access to the services that the NSp provides, for example, with "nssp_nsp_1", the NSsp can view the list of the network slice instances that it has subscribed with the NSsp 1 previously, and can perform NSp functions through the one stop shop access (OSA) on this NSp1. The NSp1 will have to maintain a repository for all the user accounts it has granted and a repository for maintaining the access control for the users. Independently, the NSp2 (and other NSps) has its own identity management system. This model requires the NSsp to sign-on separately in each domain which is not practical.

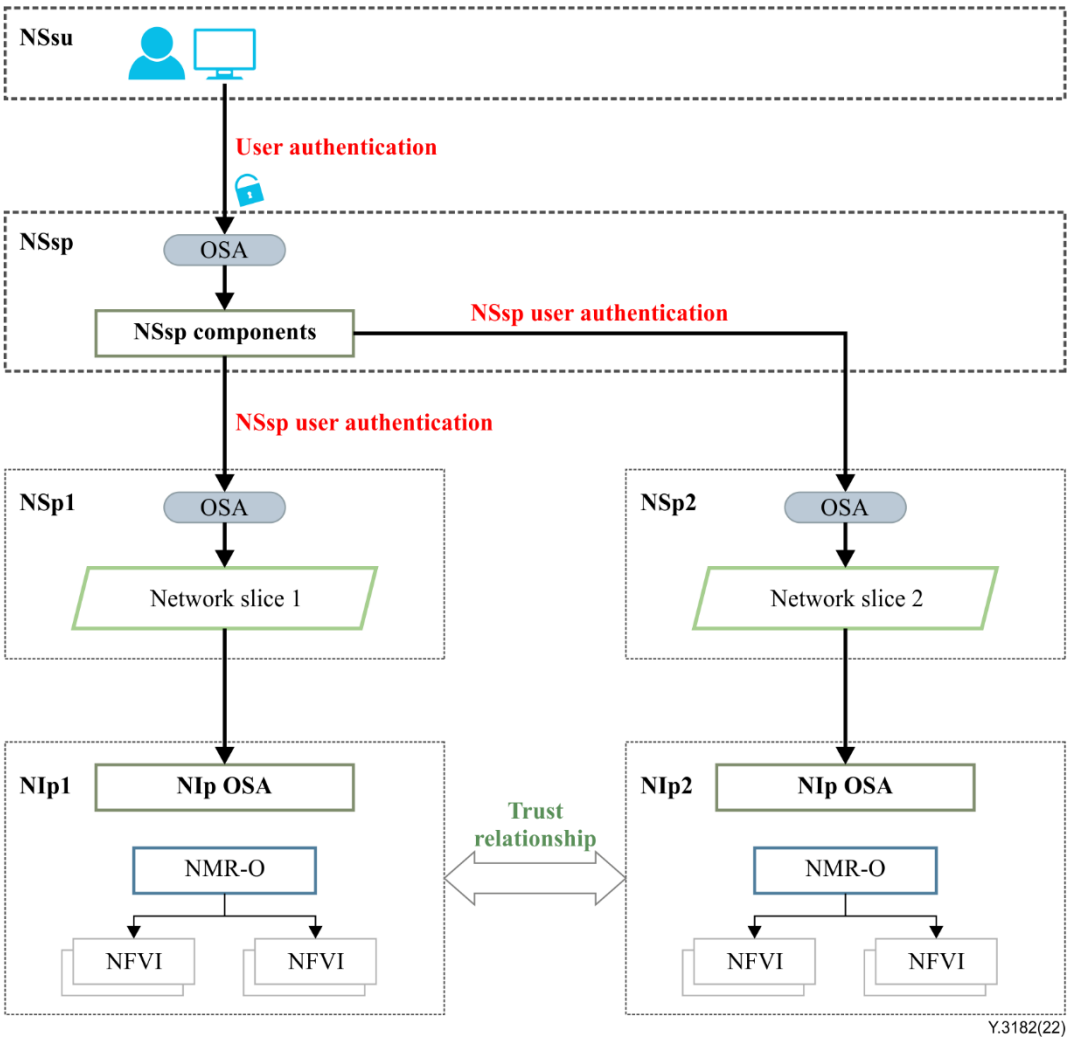


Figure II.2 – Independent domain authentication

Alternatively, a cross-domain single sign-on (CDSSO) model allows the movement of users between multiple domains with a single sign-on. With CDSSO, once logged-in to a domain, a user can make a request to the protected resources that are located in the second domain without being forced to perform another login. In general, the CDSSO mechanism will transfer the encrypted user identity token from the first domain to the second domain, and the second domain now has the user's identity that has already been authenticated in the first domain.

Bibliography

[b-ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience*.

[b-ITU-T Q.1741.9] Recommendation ITU-T Q.1741.9 (2015), *IMT-2000 references to Release 11 of GSM evolved UMTS core network.*

[b-ITU-T X.812] Recommendation ITU-T X.812 (1995) | ISO/IEC 10181-3:1996, *Information technology – Open Systems Interconnection – Security frameworks for open systems: Access control framework*.

[b-ITU-T X.1277] Recommendation ITU-T X.1277 (2018), *Universal authentication framework.*

[b-ITU-T Y.1714] Recommendation ITU-T Y.1714 (2009), *MPLS management and OAM framework.*

[b-ITU-T Y.2011] Recommendation ITU-T Y.2011 (2004), *General principles and general reference model for Next Generation Networks*.

[b-ITU-T Y.3100] Recommendation ITU-T Y.3100 (2017), *Terms and definitions for IMT-2020 network.*

[b-ITU-T Y.3103] Recommendation ITU-T Y.3103 (2018), *Business role-based models in IMT-2020.*

[b-ITU-T Y.3115] Recommendation ITU-T Y.3115 (2022), *AI enabled cross-domain network architectural requirements and framework for future networks including IMT-2020.*

[b-ITU-T Y.3502] Recommendation ITU-T Y.3502 (2014) | ISO/IEC 17789: 2014, *Information technology – Cloud computing – Reference architecture*.

[b-Arbezzano] Arbezzano, G. (2018), *How to Use the Open Source TICK Stack to Spin Up a Modern Monitoring System for Your Application and Infrastructure*. <https://www.influxdata.com/blog/how-to-use-the-open-source-tick-stack-to-spin-up-a-modern-monitoring-system-for-your-application-and-infrastructure/>

[b-Cabaça] Cabaça, J., and Neves P. (2020), SliceNet Deliverable D7.1, *Cross Plane Slice and Service Orchestrator*. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cf9b54e3&appId=PPGMS>

[b-ETSI GS NFV 006] ETSI GS NFV 006 V2.1.1 (2021), *Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Architectural Framework Specification.* <https://www.etsi.org/deliver/etsi_gs/nfv/001_099/006/02.01.01_60/gs_nfv006v020101p.pdf>

[b-IEC 61850] IEC 61850:2022, *Communication networks and systems for power utility automation – ALL PARTS.* <https://webstore.iec.ch/publication/6028>

[b-IEEE-2003] Kephart, J. O., and Chess, D. M. (2003), *The vision of autonomic computing*, IEEE Xplore, Computer, Vol. 36, No. 1, January, pp. 41-50. <https://ieeexplore.ieee.org/document/1160055>

[b-IEEE-2006] Thomas, R. W., Friend, D. H., Dasilva, L. A., and Mackenzie, A. B. (2006), *Cognitive networks: adaptation and learning to achieve end-to-end performance objectives*, IEEE Communications Magazine, Vol. 44, No. 12, pp. 51-57. <https://ieeexplore.ieee.org/document/4050101/authors#authors>

[b-Lerner] Lerner, A. (2017), *AIOps Platforms*. <https://blogs.gartner.com/andrew-lerner/2017/08/09/aiops-platforms/>

[b-NIST 1800-3B] NIST SPECIAL PUBLICATION 1800-3B (2017), *Attribute Based Access Control*. <https://www.nccoe.nist.gov/sites/default/files/legacy-files/abac-nist-sp1800-3b-draft-v2.pdf>

[b-SELFNET] Koutsopoulos, K. (2016), SELFNET Deliverable D5.1, *Report and Software Libraries to deal with the Tactical Autonomic Language*. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ab24a400&appId=PPGMS>

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2022