I n t e r n a t i o n a l   T e l e c o m m u n i c a t i o n   U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series Y

Supplement 55

(10/2019)

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

ITU-T Y.3170-series – Machine learning in future networks including IMT-2020: Use cases

ITU-T Y-series Recommendations – Supplement 55

ITU-T

International Telecommunication Union

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
Network control architectures and protocols	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.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.

Supplement 55 to ITU-T Y-series Recommendations

ITU-T Y.3170-series – Machine learning in future networks including IMT-2020: Use cases

Summary

This Supplement describes use cases of machine learning in future networks including IMT-2020. For each use case description, along with the benefits of the use case, the most relevant possible requirements related to the use case are provided. Classification of the use cases into categories is also provided.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y Suppl. 55	2019-10-25	13	11.1002/1000/14100

Keywords

Future networks, IMT-2020, machine learning, requirements, use cases.

* 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

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication 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 publication development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this publication. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Supplement 55 to ITU-T Y-series Recommendations

ITU-T Y.3170-series – Machine learning in future networks including IMT-2020: Use cases

1 Scope

This Supplement analyses use cases for machine learning in future networks including IMT-2020, and presents them in a unified format. It provides use case descriptions and indicates the basic set of possible requirements for each use case. The use cases are divided into categories.

2 References

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

[ITU-T Y.3104] Recommendation ITU T Y.3104 (2018), *Architecture 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.3172] Recommendation ITU-T Y.3172 (2019), *Architectural framework for machine learning in future networks including IMT-2020*.

3 Definitions

3.1 Terms defined elsewhere

This Supplement uses the following terms defined elsewhere:

3.1.1 machine learning (ML) [ITU-T Y.3172]: Processes that enable computational systems to understand data and gain knowledge from it without necessarily being explicitly programmed.

NOTE 1 – Definition adapted from [b-ETSI GR ENI 004].

NOTE 2 – Supervised machine learning and unsupervised machine learning are two examples of machine learning types.

3.1.2 machine learning pipeline [ITU-T Y.3172]: A set of logical nodes, each with specific functionalities that can be combined to form a machine learning application in a telecommunication network.

NOTE – The nodes of a machine learning pipeline are entities that are managed in a standard manner and can be hosted in a variety of network functions [ITU-T Y.3100].

3.2 Terms defined in this Supplement

This Supplement defines the following terms:

3.2.1 base-ML model: Machine learning model which is not trained with data sets.

3.2.2 machine learning output: Policies or configurations to be applied in the network, based on the output from the machine learning model.

NOTE – The target of machine learning output may be functions in the network. Such application of machine learning output may be based on use case requirements and controlled by network operator policies.

3.2.3 **training pipeline**: Machine learning pipeline which is used for training machine learning models.

4 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

1DCNN	One-Dimensional Convolutional Neural Network
2DCNN	Two-Dimensional Convolutional Neural Network
AI	Artificial Intelligence
AN	Access Network
API	Application Programming Interface
AR	Augmented Reality
ARIMA	Autoregressive Integrated Moving Average
ASN	Abstract Syntax Notation
BH	Backhaul
BLER	Block Error Rate
BS	Base Station
BTS	Base Transceiver Station
CAP	Common Alerting Protocol
CAPEX	Capital Expense
CAT-M	Category M
CDM	Caching Decision Module
CDR	Call Detail Record
CE	Customer Experience
CEP	Complex Event Processing
CLSTM	Convolution Long Short-Term Memory
CN	Core Network
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CQI	Channel Quality Indicator
CSP	Communication Service Provider
CU	Central Unit
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DN	Data Network
DNN	Deep Neural Network
DPI	Deep Packet Inspection
DRB	Dedicated Radio Bearer

DU	Distributed Unit
E2E	End-to-End
eMBB	enhanced Multimedia Broadband
EPC	Evolved Packet Core
E-RAB	E-UTRAN Radio Access Bearer
ETAs	Estimated Times of Arrival
FH	Front-Haul
FLR	Frame Loss Rate
gNB	(Next) Generation Node B
GNSS	Global Navigation Satellite System
GPM	Global Predictor Module
GPS	Global Positioning System
GPU	Graphical Processor Unit
HSDPA	High Speed Downlink Packet Access
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technology
IMEI	International Mobile Equipment Identifier
IoT	Internet of Things
ISP	Internet Service Provider
KPI	Key Performance Indicator
LPM	Local Predictor Module
LSTM	Long Short-Term Memory
LTE	Long Term Evolution
MAC	Medium Access Control
MANO	Management and Orchestration
MCS	Modulation and Coding Scheme
MDAF	Management Data Analytics Function
MEC	Multi-access Edge Computing
MH	Mid-Haul
MIMO	Multiple Input Multiple Output
mIoT	massive IoT
ML	Machine Learning
MML	Man-Machine Language
mMTC	massive MTC
MPP	Mobility Pattern Prediction
MTC	Machine Type Communication
NFV	Network Functions Virtualization

NFVI	NFV Infrastructure
NGC	Next Generation Core
NMS	Network Management System
NR	New Radio
OAM	Operation and Maintenance
OLLA	Out Loop Link Adaptation
ONAP	Open Network Automation Platform
OSS	Operational Support System
PDCP	Packet Data Convergence Protocol
PER	Packet Error Rate
P-GW	Packet Gateway
PMEC	Personalized Mobile Edge Caching
PMI	Precoding Matrix Indicator
PRB	Physical Resource Block
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RA	Resource Allocation
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RCA	Root Cause Analysis
RI	Rank Indicator
RNN	Recurrent Neural Networks
RRC	Radio Resource Control
RRM-NS	Radio Resource Management for Network Slicing
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RTWP	Received Total Wideband Power
RU	Radio Unit
SBA	Server Based Architecture
SDN	Software-Defined Networking
SDNC	Software-Defined Networking Controller
SD-WAN	Software-Defined Networking in a Wide Area Network
SFC	Service Function Chaining
SIM	Subscriber Information Module

SINR	Signal to Interference plus Noise Ratio
SLA	Service Level Agreement
SOHO	Soft Handover
SON	Self-Organizing Network
SRM	Slicing RA Master
TAC	Tracking Area Code
TAI	Tracking Area Identifier
TCO	Total Cost of Ownership
TCP	Transmission Control Protocol
TN	Transport Network
TPM	Trusted Platform Module
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra-Reliable and Low Latency Communications
VIM	Virtualized Infrastructure Manager
VLAN	Virtual Local Area Network
VNF	Virtualized Network Function
VNFM	Virtual Network Function Manager
VoIP	Voice over Internet Protocol
VR	Virtual Reality
XSD	XML Schema Definition

5 Conventions

In this Supplement, possible requirements which are derived from a given use case, are classified as follows:

The keywords "it is critical" indicate a possible requirement which would be necessary to be fulfilled (e.g., by an implementation) and enabled to provide the benefits of the use case.

The keywords "it is expected" indicate a possible requirement which would be important but not absolutely necessary to be fulfilled (e.g., by an implementation). Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

The keywords "it is of added value" indicate a possible requirement which would be optional to be fulfilled (e.g., by an implementation), without implying any sense of importance regarding its fulfilment. Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

6 Use cases and their requirements

This clause describes the use cases. The use cases are classified into five categories as below. For each use case, the requirements are further classified into those for data collection, data storage and processing, and application of ML output.

6.1 Network slice and other network service related use cases

This category of use cases is related to the creation or management of network slices (e.g., resource management for network slices). Similarly, the use cases related to the creation or management of network services have also been classified into this category.

6.1.1 Cognitive heterogeneous networks and ML-based SON

6.1.1.1 Use case description

As described in [ITU-T Y.3111], an IMT-2020 network provides network services to support diverse requirements, by using network functions instantiated as appropriate. Current networks are flexible and rely a lot on automation and virtualization, for instance self-organizing networks (SONs). Requirements on future networks will not only be high data rates and low latency, but the networks themselves should be smart and intelligent to keep all aspects of a telecommunication company continuously and optimally connected: users, services and machines. This requires the intervention of artificial intelligence and machine learning concepts and algorithms.

The existing common SON solution defined by the 3GPP [b-3GPP-32.500] covers three main aspects: self-configuration, self-optimization and self-healing. Cognitive heterogeneous networks are built on artificial intelligence technologies and allow the networks to be more aware about network problems, user behaviours, environmental aspects, etc.

Self-configuration includes plug-and-play configuration of newly deployed radio access nodes, where the access nodes configure their identity, transmission frequency and power, leading to faster cell planning and rollout. Functions for self-optimization include optimization of coverage, capacity, handover and interference. Self-healing includes features like automatic detection and removal of failures and automatic adjustment of parameters.

However, the first generation of SON solutions is facing limitations in the performance they could achieve since they do not utilize machine learning algorithms. An ML-based SON solution monitors network alarms and key performance indicators (KPIs), and takes proper action to clear alarms, enhance network KPIs or give network design recommendations without human intervention.

6.1.1.2 Use case requirements

6.1.1.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support an SON framework for global network optimization including the following:

- different radio access network technologies;
- multi-vendor environment;
- all network aspects including radio access network (RAN), core network (CN) and transport network (TN) related functions.

Expected requirements

It is expected that ML-enabled networks support the integration of probing and monitoring systems connected to the network elements in order to get accurate results for transport KPIs (e.g., frame loss rate (FLR), delay and jitter).

It is expected that ML-enabled networks support an SON framework which is connected to a centralized performance monitoring system to collect and consolidate data from all network management systems (NMSs) and probes.

It is expected that ML-enabled networks support an SON framework which takes all aspects of management and operation of the network into account and provides a solution for the global network optimization problem.

It is expected that ML-enabled networks support the collection of the following RAN data:

- geographical location of UE;
- measurement report for radio condition per UE and/or access node;
- resources allocated per UE and/or access node;
- connection signalling events per UE and/or access node;
- traffic type classified per UE and/or access node;
- service type classified per UE and/or access node, e.g., ultra-reliable and low latency communications (URLLC), massive machine type communications (mMTC), enhanced mobile broadband (eMBB) [ITU-T Y.3111];
- cell and/or access node resources utilization;
- cell and/or access node KPIs, e.g., KPIs related to accessibility, drops, throughput, power.

It is expected that ML-enabled networks support the collection of the following CN data:

- core network KPIs, e.g., KPIs related to success rates, failure rates and causes;
- core traffic figures, e.g., figures about traffic aggregated per service, traffic insights (about used applications), sessions per service, signalling per radio access technology (RAT), signalling of charging;
- network utilization KPIs, i.e., utilization KPIs of relevant network entities such as access nodes, links;
- call detail records (CDRs) and log files that can include customers' IP addresses, services and location.

It is expected that ML-enabled networks support the collection of the following TN data:

- transport utilization;
- transport failures;
- transport failure rate;
- transport bandwidth.

6.1.1.2.2 Use case requirements related to data storage and processing

<u>Critical requirements</u>

It is critical that ML-enabled networks support a centralized orchestration node, whose main function will be for actions which need an end-to-end and higher level view.

It is critical that ML-enabled networks support distributed orchestration nodes whose main function is to support fast and latency sensitive functions.

<u>Expected requirements</u>

It is expected that ML-enabled networks support network slices which span across multiple domains.

It is expected that ML-enabled networks support all types of network resources including software defined and hardware resources.

It is expected that ML-enabled networks support ML training in the centralized orchestration node and sending the trained model frequently to the distributed orchestration nodes.

<u>Added value requirements</u>

It is of added value that ML-enabled networks support API-based interfaces for the exchange of data and control messages related to the use case.

NOTE 1 – API-based interfaces may be used between centralized and distributed orchestration nodes.

NOTE 2 – API-based interfaces may be RESTful.

6.1.1.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support an SON framework enhanced with machine learning and intelligent algorithms.

It is critical that ML-enabled networks support both centralized and distributed orchestration nodes that have a privilege to modify network parameters and settings automatically, and then generate logs and reports for the network operator.

It is critical that ML-enabled networks allow the allocation of resources from different access network nodes and provide dynamic adjustments of resource allocation parameters to achieve gains in coverage, capacity and quality of service.

It is critical that ML-enabled networks support the handling of node resources such as dynamic adjustment of transmit power level and dynamic turn on/off of the embedded node resources, in order to optimize connection and quality of service (QoS) performance.

Expected requirements

It is expected that ML-enabled networks support the following SON functionalities in the network:

- For fully automated functions, SON has a privilege to modify network parameters and settings directly.
- For computer-aided functions, SON generates scripts and recommendations for their execution.

6.1.2 Radio resource management for network slicing (RRM-NS)

6.1.2.1 Use case description

In today's networks, the radio resource allocation is based on a per flow or per radio bearer QoS profile. However, such an approach is not efficient to provide the service quality guarantee and resource isolation needed by network slicing, thus, leading to over-reservation and hence under-utilization of resources. One of the major challenges with future networks and deployment of network slicing is providing performance guarantee in terms of minimum dedicated bandwidth with high reliability, while ensuring efficient utilization of the scarce radio resources.

6.1.2.2 Use case requirements

6.1.2.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the continuous collection of data to update the prediction models in order to improve in real time the accuracy and effectiveness of the prediction models.

Expected requirements

It is expected that ML-enabled networks support the usage of historical data and analysis of network slice behaviour and radio resource utilization patterns for high accuracy prediction of the network slice resource requirements.

It is expected that ML-enabled networks support the collection of measurement data including:

- Network slice (per cell):
  - KPIs, including per network slice uplink /downlink physical resource block (PRB) [b-3GPP 36.321] usage, and others as indicated in network slice level objectives (e.g., latency).
- UE:
  - location context, e.g., global navigation satellite system (GNSS), cell [b-3GPP 36.305];
  - mobility context, e.g., cell information history;
  - per network slice per QoS/dedicated data radio bearer (DRB) throughput;
  - reference signal received power (RSRP) [b-3GPP 36.214], reference signal received quality (RSRQ) [b-3GPP 36.214], received signal strength indicator (RSSI) [b-3GPP 36.214], signal-to-interference-plus-noise ratio (SINR) [b-3GPP 36.214];
  - beam state information, channel quality indicator (CQI) [b-3GPP 36.213] [b-3GPP 38.214], modulation and coding scheme (MCS) [b-3GPP 36.213] [b-3GPP 38.214].
- Cell:
  - position estimates, e.g., GNSS;
  - KPIs.
- Core network:
  - UE communication pattern;
  - UE mobility trajectory.

It is expected that ML-enabled networks support the collection of the following context information data:

- Network state data:
  - identification information, e.g., for cell identification, beam identification, cell tracking area code (TAC) [b-3GPP 23.003];
  - cell status query data;
  - list of network slices supported in the cell;
  - real-time network performance data;
  - network slice specific radio network configuration (e.g., cell level, network slice level).
- User state data, e.g., UE identifier, various logs, KPIs.
- Network slice level objectives, network slice profiles (e.g., target user name, latency, throughput, availability, reliability).

Added value requirements

It is of added value that ML-enabled networks support simulated data generated to further improve the training of ML models.

NOTE – Simulated data may be used for instance to learn rare events, and before general deployment of network slices.

It is of added value that ML-enabled networks support the collection of the following data:

- network slice level performance measurement parameters and alarm data;
- cell status query, user mobility and radio channel conditions data;
- user service request, user QoS and user data monitoring.

6.1.2.2.2 Use case requirements related to data storage and processing

Expected requirements

It is expected that ML-enabled networks support the analysis of network slicing requirements and radio network conditions, and fulfilment of the requirements of each network slice while achieving optimal radio resource utilization.

6.1.2.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support prediction, in real time, of the network slice radio resource requirements and optimal radio resource allocation per network slice to optimize the mobile network operation.

NOTE – Services enabled by RRM-NS include real-time predictive network slice performance monitoring and resource planning, real-time predictive network slice traffic increase/decrease and network slice provisioning, and real-time deployment of new network slices.

It is critical that ML-enabled networks support algorithms which predict per network slice and per cell radio resource requirements and required allocation trajectory.

It is critical that ML-enabled networks execute the network slice resource allocation plan, generate man-machine language (MML) commands for communication with the access network (AN) nodes, other elements of the core network (CN) and network managers, e.g., operational support system (OSS).

Expected requirements

It is expected that ML-enabled networks support NMSs which are involved in:

- data collection, including network slice level and cell level data collection;
- model training;
- non-real-time network slice related policies and configuration.

It is expected that ML-enabled networks support ANs and/or near real-time RAN control nodes, which support the following functions:

- data collection;
- near real-time network slice radio resource scheduler.

It is expected that ML-enabled networks support AN with an ML-based dynamic radio resource scheduler function.

Added value requirements

It is of added value that ML-enabled networks support ML algorithms to predict per network slice performance and to generate network slice planning and optimization strategies.

6.1.3 End-to-end network operation automation – Service design

6.1.3.1 Use case description

Vertical industry services imply a wide range of service requirements [ITU-T Y.3111]. Network configurations to satisfy these service requirements have many variations. For example, a high-throughput broadband network is necessary to satisfy the service requirements of remote diagnostic service and low-latency network configuration is needed to satisfy the needs of robotic surgery.
Therefore, it will be a challenging task to make agile and appropriate decisions regarding the configuration of critical services. The redesigning of networks, if done manually, is time consuming. ML may provide solutions for agile service design and automated network design by automatically translating service requirements of use cases to network requirements. ML can also provide

automated, scalable, customized solutions for network (re)design. This can shorten the service delivery time.

Network slicing is one approach to satisfy the diverse service requirements of vertical industries. As mentioned in [ITU-T Y.3111], the request from a network slice customer to create a network slice includes the specific catalogue of service requirements on network slice, e.g., service type (eMBB, mIoT, URLLC, etc.), network slice priority, network slice sharing option. The template for network slice provisioning is provided by the IMT-2020 service provider as a form of service requirement catalogue through its service portal.

6.1.3.2 Use case requirements

6.1.3.2.1 Use case requirements related to data collection

None.

6.1.3.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks possess the capability of automatically translating service requirements of the use cases to network requirements for network deployment conforming to network specification documents (e.g., data model).

It is critical that ML-enabled networks support data models used to specify service requirements of use cases including characteristics of logical networks (i.e., characteristics of a logical representations of the networks).

NOTE – Characteristics of a logical network may include the connectivity between various nodes of the logical network.

6.1.3.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks possess the capability to efficiently integrate automated network configuration methods to fulfil the requirements of IMT-2020 systems.

It is critical that ML-enabled networks support APIs which are exposed towards the service users and which support service requirements from service users.

It is critical that ML-enabled networks support automatic composition of logical network requirements from service requirements.

It is critical that ML-enabled networks support APIs which are exposed towards the network providers and which notify network providers of logical network requirements.

It is critical that ML-enabled networks support functions which translate logical network requirements into physical network deployment specifications.

6.1.4 End-to-end network operation automation – Network resource adaptation

6.1.4.1 Use case description

Network resource adaptation is necessary for maintaining QoS requirements of various application services offered via IMT-2020 network slices. Here, resource adaptation refers to the process of dynamically increasing or decreasing resources (e.g., CPU, storage, memory of nodes or bandwidth of links) allocated to a network slice so that QoS requirements are always met despite the fluctuation in the workload or sudden changes in network capacity (e.g., due to interferences or link/node failures). In a network slice, network functions (NFs) can be arranged in service function chaining (SFC) [b-ITU-T Y.2242] to process network traffic for offering the intended communication service.

The QoS requirements of services offered in various network slices may differ, e.g., QoS requirements of eMBB and URLLC services offered in different network slices are different. Some services (e.g., autonomous driving) require very stringent QoS requirements, while other services (e.g., best effort social networking services) may have tolerable ranges of QoS. A network resource adaptation mechanism dynamically monitors the performance and resource utilization of all network slices and increases/decreases resources of a network slice (assigned to NFs) in the case of changes in network conditions to meet the QoS requirements.

ML-based scaling for network resource adaptation provides an effective approach to meet the diverse QoS requirements of various services offered in different network slices. This involves regular monitoring and analysis of performance data, time-varying workload, resource utilization and available resources. Based on these, intelligent decisions for resource arbitration can be made to appropriately allocate the available resources of a given node or link to the network slices so that each network slice can satisfy the required level of QoS. In case all resources have already been allocated to the network slices, the intelligence decision addresses if some resources can be taken from some network slices without hampering their QoS and allocated to the network slice which requires stringent QoS to be maintained. It also addresses the case where an NF requiring more resources can be migrated from its current node to a new node.

Thus, ML-based scaling techniques for resource adaptation can be employed for both purposes of enhancing QoS satisfaction and achieving agile operation of resource control functions. These ML techniques can be applied as follows:

1) For dynamic resource arbitration among services being hosted in each network node, ML techniques are applied to determine services that require resource adjustment, and to determine the required amount of resources to be added or removed to them.
2) For NF migration from one network node to another along an already-established service function path, ML techniques are applied to determine the candidate nodes for the migration of the NF from its current node.
3) For resource reconfiguration including change of network topologies, ML techniques are applied to determine the new network topologies and new order of NF placement for all service function chains in the network.

6.1.4.2 Use case requirements

6.1.4.2.1 Use case requirements related to data collection

<u>Expected requirements</u>

It is expected that ML-enabled networks support the collection of the following data for agile closed-loop operation:

- failure status information obtained from monitoring systems in the underlying network;
- resource feedback information from controllers or orchestration nodes in the underlying network;
- status information of resources related to network services, e.g., computational resources such as CPU, memory and storage for network functions, including identification and configuration information;
- status information of network infrastructure, e.g., network resources such as physical switch and link, including identification and configuration information.

6.1.4.2.2 Use case requirements related to data storage and processing

<u>Critical requirements</u>

It is critical that ML-enabled networks possess the capability to continuously meet diverse levels of QoS requirements for various types of services.

It is critical that ML-enabled networks possess the capability of controlling resources of the NFV infrastructure in an agile and closed-loop operation manner to adapt the network resources according to the dynamically changing environment.

6.1.4.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks produce the following ML outputs (predicted and/or inferred):

- future CPU utilization associated with each NF;
- optimal allocation of computational resources.

It is critical that ML-enabled networks support the following interfaces in order to effectively utilize ML output:

- interfaces between the controller and underlay network infrastructure to realize a computational resource adaptation mechanism (including resource arbitration and NF migration) continuously meeting diverse QoS requirements of various types of services;
- interfaces between the controller and underlay network infrastructure to realize agile closed-loop resource control operations adapting the infrastructure to the dynamically changing environment.

6.1.5 End-to-end network operation automation – Logical network design and deployment

6.1.5.1 Use case description

Services such as software-defined networking in a wide area network (SD-WAN) services are bringing many advantages to customers, who can deploy and customize their logical network in a flexible and easy manner. The main motivation to automate logical network design and deployment is to mitigate operator's tasks when customers want flexible customization in a short delivery time. This design and deployment should be performed in a declarative manner based on design intent, or network specification produced during the service design phase, to help operators reduce frequent template/script updates. It has to be noted that this automation could also be used for ML pipeline design and deployment.

During this logical network design and deployment phase, "logical network designer" takes design intent as an input and outputs design result, which is given to a "workflow generator" to generate executable workflows needed to (re)configure network infrastructure. Workflows are executed by a set of "executors", which would include a machine learning function orchestrator [ITU-T Y.3172], NFV orchestrator [b-ETSI-NFV-MANO], virtualized infrastructure manager [b-ETSI-NFV-MANO], provisioning scripts, service orchestrator component [b-ONAP] or other tools used to deploy logical networks.

The design of logical networks can be automated if network operators employ their own predefined templates or scripts, but this approach would significantly limit flexibility in customization. ML gives the possibility to automate the design of flexible logical networks. One possible use case of ML is an appropriate selection of templates or scripts, where operators predesign a number of templates/scripts to allow different kinds of customization. An ML-based algorithm would evaluate them according to customer requirements and select the one which maximizes end user satisfaction.

Another use case of ML is more flexible design automation. Every network design is essentially a combination of logical network components (e.g., links, nodes, virtualized network functions (VNFs)), their deployment location at physical infrastructure, and a set of configuration parameters. But as the number of such combinations can become quite huge, searching for the best combination could take quite a long time. ML, using either supervised learning or reinforcement learning, helps to realize this search in real time.

6.1.5.2 Use case requirements

6.1.5.2.1 Use case requirements related to data collection

None.

6.1.5.2.2 Use case requirements related to data storage and processing

None.

6.1.5.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks possess the capability to allow a "logical network designer" to accept network requirements in a declarative form, as Design Intent, which provides logical network descriptions with functional and non-functional requirements, and to output logical network design results, which are deployable on a physical network infrastructure.

It is critical that ML-enabled networks possess the capability to generate executable workflows to safely configure and reconfigure the underlay network infrastructure.

It is critical that ML-enabled networks possess the capability to execute generated workflows by sending commands or configurations to the underlay network infrastructure, or its control/management entity, to deploy the desired logical networks.

Expected requirements

It is expected that ML-enabled networks support the capability to design logical network inputs and outputs in a machine readable format.

NOTE 1 – An example of a machine readable format is OASIS TOSCA [b-OASIS-TOSCA].

It is expected that ML-enabled networks support mechanisms for design intent which provide the ability to describe logical networks including their functional/non-functional characteristics:

- in an abstracted way to easily express customers intent;
- in a declarative way;
- in machine readable format.

It is expected that ML-enabled networks support network design capabilities to produce logical network descriptions with concrete design information:

- in a concrete manner to be deployable on physical infrastructure without any ambiguities;
- in a declarative manner;
- in machine readable format.

NOTE 2 – Examples of concrete design information include components, mapping to physical nodes, configuration parameters.

It is expected that ML-enabled networks support deployment workflows which include a sequence of commands and configurations to deploy designed logical networks in a procedural way.

It is expected that ML-enabled networks support underlay network infrastructure which allows the deployment of designed logical networks using the following aspects:

- Underlay network infrastructure is virtualized so that logical components in designed logical networks can be deployed.
- Underlay network infrastructure supports APIs to accept commands or configurations generated by design and deployment entities.
- Underlay network infrastructure may include orchestrators to configure multiple components of the infrastructure.

- Underlay network infrastructure may support infrastructure controllers, e.g., OpenStack [b-OpenStack] or Kubernetes [b-Kubernetes], to manage computing resources which are needed to deploy logical network components.

6.1.6 End-to-end network operation automation – Fault detection and recovery

6.1.6.1 Use case description

Automated closed loop operation has become a top priority for network operators to manage and maintain stable IMT-2020 networks. To achieve closed loop automation on IMT-2020 networks, ML is expected to support fault management such as predictive detection and root cause analysis, and automated recovery such as decision making of the recovery procedure.

Firstly, the management system should be automated to the extent possible, to promptly detect failures in network functions virtualization (NFV) environments. However, a network service involving a collection of virtualized network functions (VNFs) could cause unexpected behaviour even though each component works properly. In such cases where there is a collection of VNF software components, the number of software bugs in the code itself also increases, and thus detection of failures with unexpected behaviours gets even more difficult. Network operators want to promptly detect such failures, which may cause increasingly unstable behaviours before the process escalates into critical failure. Hence, it is worth tackling the above problems with the power of ML in order to process huge volumes of various types of management data (e.g., alarm, performance data, operation logs, and network topology).

Root cause analysis is also important to identify the failure types and locations, and to properly convey that failure type and location mapping information to the automation function. However, this task is complex and takes considerable time as the NFV systems consist of a large number of hardware and software components. In order to address this problem, ML can be applied to identify the root cause of failures based on previous experience obtained from actual or test environments.

Continuous automated improvement of the fault recovery process is also difficult and time-consuming. Generally, the sustainability of an automated system depends on the implementation of the predefined workflows dealing with several types of events. Although NFV systems have the potential to simplify fault detection and recovery processes, in reality it is not likely to support all types of events, e.g., unexpected failures, and hence exceptional manual operations are still needed. Therefore, reducing the exceptional conditions by continuous improvement of the workflows is also an expected role of ML.

6.1.6.2 Use case requirements

6.1.6.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the following:

- collection of performance data on real-time basis;
- common topology information model and configuration model to easily understand network topology for root cause analysis (RCA);
- generation of training data using testing environments to obtain enough training data.

Expected requirements

It is expected that the following data is used for predictive detection:

- performance data related to network function (NFs) (e.g., CPU utilization, memory utilization, disk I/O etc.) described in a common data format;
- performance data related to NFV infrastructure (NFVI) (e.g., CPU utilization, memory utilization, disk I/O etc.) described in a common data format;

- traffic information (e.g., traffic volume, discard counter error counter etc.) described in a common data format.

It is expected that ML-enabled networks support the collection of the following data for root cause analysis:

- alarm information generated by NFs and NFVI;
- topology and configuration information described in a common data format;
- historical knowledge data described in a common data format.

It is expected that ML-enabled networks support the collection of the following data for continuous automated improvement of fault recovery process:

- RCA results collected in a common data format.

6.1.6.2.2 Use case requirements related to data storage and processing

<u>Critical requirements</u>

It is critical that ML-enabled networks possess the capability to detect faults of several types.

It is critical that ML-enabled networks possess the capability to promptly analyse the root cause of failures based on past experience obtained from actual or test environments.

It is critical that ML-enabled networks possess the capability to automatically select optimal recovery actions, and to continuously improve the automated workflows.

<u>Expected requirements</u>

It is expected that ML-enabled networks possess the capability to proactively detect impending failures.

6.1.6.2.3 Use case requirements related to application of ML output

<u>Expected requirements</u>

It is expected that ML-enabled networks support APIs for fault recovery functionalities, defining the recovery tasks.

6.1.7 Application-specific network slicing through in-network machine learning

6.1.7.1 Use case description

Network slicing has been considered as one of the most significant technologies for IMT-2020 networks, where multiple network slices that support different categories of services with different QoS requirements are supposed to be deployed in the same physical infrastructure. However, how to effectively identify and classify applications in real time is still an open issue especially in the RAN context. Contextual information regarding the data from the UE (e.g., application to which the data belongs and device which generated the data) is hidden from the network functions. Conventionally, there are several ways to achieve application identification and classification, e.g., packet header marking, and deep packet inspection (DPI) to detect signature per application from packet payloads. However, packet header marking fails to identify the scope of applications while DPI is becoming harder due to the fact that application-specific information conveyed in payload is often encrypted.

In order to handle the above issues, a mechanism for an application-specific network slicing, utilizing in-network ML is proposed. This mechanism aims to apply application-specific radio resource scheduling in RAN, QoS control and various network functions on a per application and per device basis in CN.

6.1.7.2 Use case requirements

6.1.7.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support a number of customized UEs used to generate training data.

NOTE – Customization may include the generation of packets tagged with the information about the application payload and specific application in the data network (DN) to handle these tagged packets.

6.1.7.2.2 Use case requirements related to data storage and processing

None.

6.1.7.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support base stations with application-specific resource allocation mechanisms based on the identified application information and management policies.

Expected requirements

It is expected that ML-enabled networks apply ML decisions on a per flow basis and extract the useful features in a train of packets contained in a given flow, without looking into the payload of packets of that flow.

NOTE – To protect users' privacy, the flow parameters mentioned in [b-ITU-T Y.1540], i.e., source host address (SRC), destination host address (DST), class of service, and session identification, can be utilized as flow features without looking into the payload of packets.

It is expected that ML-enabled networks support identification of the application at the UPF [ITU-T Y.3104], and classification of the uplink packets from the UE to the MEC system [b-ETSI MEC001] for application-specific processing.

It is expected that ML-enabled networks support tagging of downlink packets from the MEC system [b-ETSI MEC001] with the identified application name before transmitting them to the base stations.

NOTE 1 – Base stations can apply application-specific resource scheduling based on the received attached application name.

It is expected that ML-enabled networks support UPF which has a feature extraction and a classifier updating capabilities for executing, training AI models, and has traffic classification and tagging functions utilizing trained deep neural network (DNN) models.

It is expected that ML-enabled networks support MEC which applies application-specific data processing.

NOTE 2 – Concerning application-specific MEC optimization, packets from UEs are classified to different network slices (e.g., VLANs) and sent to the MEC system. In each network slice, application-specific optimization policies may be applied, e.g., HTTP caching service for web browsing, video transcoding service for video streaming, and bandwidth control for tethering traffic in separated network slices.

6.1.8 Smart traffic mirror – an ML-assisted network service

6.1.8.1 Use case description

A smart traffic mirror is intended to prevent traffic accidents by installing a video camera which is assisted with ML capabilities in order to detect and monitor vehicles or pedestrians in an area where accidents might occur, such as a blind intersection or a blind curve. This use case, utilizes the expected low-latency characteristic of IMT-2020 networks. Very short response times are important for such control applications (e.g., factory control, automated driving), where short communication turnaround times are essential.

In this use case, ML can extract the useful knowledge from the imputed information and execute useful action with this knowledge. Concerning knowledge acquisition, it is important to perform all processing quickly enough after the occurrence of an event, to prevent the accident and notify the potentially impacted vehicle (s) or pedestrians. Since it is important to complete this process in a short period of time, performing functions such as moving image processing at the edge becomes critical to support the use case. For this use case, executing processing operations in a centralized computing platform rather than at the edge implies to carefully consider the delays required for transferring the large amount of video information to that centralized computing platform.

6.1.8.2 Use case requirements

6.1.8.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support knowledge acquisition to extract useful knowledge from collected information.

Expected requirements

It is expected that ML-enabled networks support knowledge acquisition which is executed as a capability outside the network or inside a network.

It is expected that ML-enabled networks support improving (or retraining) knowledge based on only a part of the collected real-time data.

6.1.8.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support the capabilities that are provided in underlying networks to perform knowledge execution based on extracted knowledge.

It is critical that ML-enabled networks support short turnaround times to turn the input information into the action taken by the knowledge execution.

Added value requirements

It is of added value that ML-enabled networks support the deployment of knowledge acquisition and execution in a distributed manner, especially knowledge execution in edge computing platforms.

6.1.8.2.3 Use case requirements related to application of ML output

None.

6.1.9 ML-based end-to-end network slicing for 5G

6.1.9.1 Use case description

The current process for resource allocation (RA) is mainly based on QoS provisioning techniques. These QoS guarantees are usually given at the access layer level. With the targeted E2E network slicing concept, a customer is granted a guaranteed part of the network resources usable across all network levels, including the radio access, transport and core networks. This offers service consistency in terms of latency and delay for critical applications such as autonomous driving and remote surgeries.

ML is envisioned to play a key role in a number of IMT-2020 networks, e.g., clustering services to allocate network slices accordingly, service classification and possibly prioritization for minimum QoS guarantees, predictive user allocation/reallocation to network slices based on the users' activity history and the system dynamics. This use case describes a mechanism which monitors the traffic from the UE and determines the best RA based on ML.

6.1.9.2 Use case requirements

6.1.9.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support an end-to-end orchestration functionality which collects all network data, counters and KPIs.

Expected requirements

It is expected that ML-enabled networks support the integration of active probing monitoring systems connected to the network elements in order to get accurate results for transport KPIs.

NOTE – Examples of transport KPIs include frame or packet loss ratio, delay and jitter.

It is expected that ML-enabled networks support the collection of the following inputs for the network slicing resource allocation capability including a list of counters and KPIs, collected from the network nodes working in the network slice at the radio access, transport and core network levels.

It is expected that ML-enabled networks support the collection of following radio access network data:

- UE location information;
- measurement report for radio condition per UE/access node;
- resources allocated per UE/access node;
- connection signalling events per UE/access node;
- traffic type classified per UE/access node;
- service type classified per UE/access node, e.g., URLLC, mMTC and eMBB service types;
- cell/access node resources utilization;
- cell/access node KPIs (accessibility KPIs, drops, throughput, power).

It is expected that ML-enabled networks support the collection of the following data from CN:

- core network KPIs, e.g., KPIs related to success rate, failure rates and causes;
- core traffic figures, e.g., figures about traffic aggregated per service, traffic insights on used applications), sessions per service, signalling per RAT, signalling of charging;
- network utilization KPIs, e.g., KPIs of relevant entities (e.g., nodes, links);
- CDRs and log files that can include customers' IP addresses, services and location.

It is expected that ML-enabled networks support the collection of the following transport network data:

- transport utilization;
- transport failures;
- transport failure rate;
- transport bandwidth.

6.1.9.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support a network slice resource allocation functionality which addresses the following:

- deciding the suitable level of resources for each service type;
- using the model, enhancing its accuracy via testing it on diverse datasets and studying its performance.

NOTE 1 – Deciding on the suitable level of resources for each service type can be done using clustering techniques in the unsupervised machine-learning paradigm. The decision also involves training the model to choose the best resources.

Expected requirements

It is expected that ML-enabled networks support an end-to-end orchestrator which continuously optimizes entire network slices including all associated resources such as access, transport and core network resources.

It is expected that ML-enabled networks support decentralized execution of the resource allocation is with some functionality delegated to the network edge for latency considerations and saving core-computing resources.

NOTE 2 – With network slicing, the user may have a new service identifier (linked to their ID) per service, which identifies the network slice the user is connected to. This would be very useful in use cases such as quantifying the user load per network slice for traffic monitoring and offloading decisions or efficiently managing the network slice resources.

Added value requirements

It is of added value that ML-enabled networks support the network slicing resource allocation output which is a specific range of dynamically changeable resource parameters allocated to each node in network slices.

6.1.9.2.3 Use case requirements related to application of ML output

None.

6.1.10 ML-based utility maximization of sliced backhauls

6.1.10.1 Use case description

Future networks including IMT-2020 networks will serve applications with very distinct performance requirements: ultra-high definition video streaming and immersive applications (AR/VR) demanding high throughput; delay sensitive applications (e.g., autonomous vehicles); IoT services (e.g., smart metering) with best-effort policies.

Network slicing enables such diverse services to be logically separated while sharing the physical infrastructure. However, finding optimal allocation of airtime resources to individual network slices, so as to maximise their utilities with agile decision making is non-trivial.

Deep learning can tackle the complexity of solving arbitrary combinations of utility functions. In particular, by employing stacks of convolutional blocks, it is possible to learn the relationships between traffic demands and optimal flow rate allocations. Once trained, such algorithms could make close-to-optimal inferences within milliseconds.

6.1.10.2 Use case requirements

6.1.10.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the collection of detailed measurements of per service demand at network entry points so that machine learning driven resource allocation algorithms can maximise the utility of multi-service backhaul networks.

It is critical that ML-enabled networks support dedicated measurement infrastructure deployment at the AN level.

NOTE 1 – The deployment may require hardware upgrades or software updates in the network.

It is critical that ML-enabled networks support measurement granularity which is adjusted depending on the delay sensitivities of each application.

NOTE 2 – Measurement granularity adjustment may result in large volumes of metadata to be stored at the network entry point, e.g., gNB [b-3GPP 38.401] level. Based on this information, a neural network can be trained centrally and optimal allocations signalled back to each hop along relevant paths.

It is critical that ML-enabled networks support data collection at each network entry point (e.g., gNB) which comprises a node identifier, flow identifier, type of services, KPI needed, and information about the capacity available on the link to the next hop.

It is critical that ML-enabled networks support algorithms responsible for airtime allocation on a per-backhaul link basis also has precise knowledge of the routing topology.

NOTE 3 – Knowledge of the routing topology includes information about what flows share which links and what capacity constraints arise as a result of this.

Expected requirements

It is expected that ML-enabled networks support for each uplink and/or downlink flow in the backhaul, the collection of the following information:

- identifier of the network entry point (e.g., gNB);
- flow path within the backhaul (i.e., sequence of traversed nodes);
- flow type, e.g., video, AR/VR, best-effort data;
- target KPIs, i.e., KPIs about delay, throughput, packet loss rate.

It is expected that ML-enabled networks support the monitoring of flow demand at all times and their storage along with timestamp summaries, both for training and inferences.

It is expected that ML-enabled networks support a centralized node involved in decision making which also gathers information about:

- backhaul topology;
- link capacities;
- eventual scheduling conflicts.

6.1.10.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support neural network models employed for the purpose of utility maximization that are retrained on a per-deployment basis.

NOTE – Optimal allocation of resources depends on the particularities of a given topology.

It is critical that ML-enabled networks support a centralized node, e.g., GW responsible for allocating resources, i.e., how much airtime should be allotted to each flow on each wireless link along the path taken, which is provisioned with appropriate GPU hardware, so as to facilitate rapid training and inference.

Expected requirements

It is expected that ML-enabled networks support measurements per service that are obtained through signalling or estimated based on historical data.

It is expected that ML-enabled networks support extensive storage and parallel computing capabilities (GPUs) that are supported at the CN.

6.1.10.2.3 Use case requirements related to application of ML output

None.

6.1.11 Energy efficient trusted multi-tenancy in IMT-2020 cross-haul

6.1.11.1 Use case description

IMT-2020 networks promise to deliver per user capacity along with minimum latency while ensuring flexible network control and orchestration. More specifically, these networks aim to serve diverse traffic profiles including enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable latency communications (URLLC). These services require to revisit air interfaces with cost-efficient coordinated management and operation of the network. This includes network components at access, transport and core. To provide these services, describes a cross-haul architecture that enables common control and data planes to achieve integration of front-haul (FH) and back-haul (BH). The architecture offers benefits of homogeneous, flexible, and automated interconnection of the radio, access and core network with the help of open interfaces. This also contributes to several cost-efficient system optimizations. To realize these features, the cross-haul architecture relies on software-defined networking (SDN) and network functions virtualization (NFV), where the former enables dynamic control, management and configuration, while the latter decouples network functions from underlying physical infrastructure.

The described cross-haul architecture can support diverse functionalities including multi-tenancy, service prediction and resource optimization. Multi-tenancy has the potential to host multiple operators over a cross-haul infrastructure, where virtual resources are mapped to a substrate infrastructure. The cross-haul is capable of sharing BH and FH resources to host multi-tenancy. A multi-tenant network can host multiple operators resulting in maximum utilization of cross-haul resources while minimizing cost overheads, e.g., spare physical resources can be leased to virtual operators. This additional leasing of resources to virtual operators may cut down the overall capital expenses (CAPEX) and operational expenses (OPEX) for both the cross-haul owners and network operators while ensuring high availability services to end users.

It is essential that the cross-haul multi-tenancy provides service level agreements (SLAs) to offer the services to a tenant with the help of a virtual network, e.g., virtual routers and switches.

Based on SLAs, network owners enable flexible network slice provisioning to meet individual tenant requirements. Virtual network slicing offers optimal utilization of physical resources by scaling virtual resources up and down based on dynamic traffic loads in the network.

However, to be successful, implementations of multi-tenancy based on cross-haul have to use tenant/operator-aware switching devices and a unified data plane. These may be used with essential optimizations on certain network segments to improve the overall network efficiency.

It has been revealed in a report by the 5G PPP security workgroup that IMT-2020 requires novel security capabilities in addition to meeting heterogeneous business requirements. The implementation of future networks including IMT-2020 will use innovative network infrastructure and services, e.g., autonomous vehicles and smart factories to support multi-tenant and multi-stakeholder scenarios. It is critical that these scenarios use appropriate security and trust strategies as part of network infrastructure. This is one of the major challenges which should be taken into consideration as most of the current trust models from the state-of-the-art are applicable within a single administrative domain only. Further, adversarial ML-enabled security threats with high heterogeneity of data sources makes it critical for ML-enablers to adapt secure business models, e.g., isolation of multi-tenancy and slicing across the cyber-physical networks.

6.1.11.2 Use case requirements

6.1.11.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support multi-tenancy and classifies traffic in real time.

It is critical that ML-enabled networks support multi-tenancy along with situational awareness of the environment which implies gathering information from the monitored environment.

Expected requirements

It is expected that ML-enabled networks support data collection to acquire the operational environment data.

NOTE 1 – The set of activities may comprise of appropriate knowledge representation and reasoning strategy to fetch, manage and process operational and factual knowledge.

It is expected that ML-enabled networks collect the following information across the network for each network slice:

- throughput;
- congestion;
- transmission delays;
- availability;
- risk level;
- registered events.

NOTE 2 – Sensors and external data repositories may act as critical sources of factual knowledge.

6.1.11.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support model learning and adaptation to dynamic traffic loads to allocate virtual resources/network slices.

It is critical that ML-enabled networks support the representation of network information in such a way that it can be input to the ML models.

NOTE 1 – This representation may need data preprocessing, normalization, aggregation/fusion, verification, prioritization and reconstruction.

It is critical that ML-enabled networks support the calibration of lightweight ML algorithms at the mid-haul (MH) to filter security threats.

It is critical that ML-enabled networks are equipped with self-awareness, self-healing, self-optimization, and self-protection capabilities.

NOTE 2 – This is ensured with the help of lightweight ML models which can detect, predict and react to security threats and ensure trust to tenants hosted by an infrastructure owner.

It is critical that ML-enabled networks support counter measures to cloud-level threats at the tenant-level.

NOTE 3 – This will help to assure trust offered by the ML-enabled multi-tenancy at the tenant level. These threats can propagate to cloud and may compromise its overall security posture.

It is critical that ML-enabled networks support the management of the underlying infrastructure.

It is critical that ML-enabled networks support the assembling of virtual resources into a virtual topology.

It is critical that ML-enabled networks maintain and configure a virtual network over the virtual topology according to the requirements of a service provider.

It is critical that ML-enabled networks support the allocation of resources to train/retrain the machine learning models.

It is critical that ML-enabled networks support dedicated physical infrastructure that is allocated with computing resources and virtualization capabilities to deploy ML mechanisms.

It is critical that ML-enabled networks support specific SLAs for multi-tenancy that are technology independent.

Expected requirements

It is expected that ML-enabled networks support preprocessing that produces a specific event-related correlation.

NOTE 4 – Case-based/attribute-based reasoning may be considered to find dependencies between events and resource-vectors.

It is expected that ML-enabled networks support a reinforcement learning-based ML model in the FH to enable adaptation to heterogeneous data sources.

NOTE 5 – These lightweight ML algorithms will facilitate context-aware recalibration based on self-decisions operating on resource-constrained systems to enable feasible deployment at the FH.

It is expected that ML-enabled networks support the reinforcement learning-based ML model which is monitored and controlled by optimization functionalities to adapt its operation under unforeseen events/threats.

NOTE 6 – These functionalities should be able to dynamically acquire and validate the consequences of decision making with the help of operational knowledge.

It is expected that ML-enabled networks support efficient onboarded security and energy policies complemented with situational awareness to take appropriate resource allocation decisions for each tenant.

It is expected that ML-enabled networks support a flexible data model along with the intent for an appropriate data collection.

It is expected that ML-enabled networks support security mechanisms that comply with the trust model and energy-saving policies needed by different actors from IMT-2020 networks.

Added value requirements

It is of added value that ML-enabled networks support the augmentation of ML models with a closed loop system, where expected key performance indicators (KPIs) can be assessed and appropriate corrective tactics can be taken.

NOTE 7 – These tactics may include modification of statistical models, data preprocessing rules and inference rules, and recalibration of heuristics.

It is of added value that ML-enabled networks support flow-level security driven by authorization/privacy policies at the edge in real time.

6.1.11.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support ensuring that different tenants are guaranteed appropriate quality of service (QoS) or quality of experience (QoE) requirements according to corresponding SLAs.

It is critical that ML-enabled networks support the allocation of virtual resources to different tenants according to SLAs with appropriate configurations while ensuring appropriate isolation.

It is critical that ML-enabled networks support implementing adaptive machine learning mechanisms or ML models, e.g., reinforcement learning, deep learning or hybrid instantiated at the BH.

NOTE 1 – This will essentially aid to countermeasure decisions with considerations to future events, threats, its replicas, and possible propagation holes in a multi-stakeholder multi-tenant environment.

It is critical that ML-enabled networks support an ML model, e.g., supervised learning, reinforcement learning or hybrid to enforce contextual adaptation of trust models.

NOTE 2 – This will ensure timely detection of virtual network function (VNF) anomalies and will prevent the propagation of threats.

It is critical that ML-enabled networks support lightweight ML models, to entail energy-efficient security measures.

NOTE 3 – ML models such as convolutional neural networks (CNNs) can be used to enable trust by enforcing appropriate levels of security and trust, thus enabling a trusted execution environment (TEE) [b-TEE Management]. To achieve this, ML-enabled trust mechanisms can be deployed at mid-haul and back-haul. E.g., an ML-enabled virtual dynamic root of trust for measurement (DRTM) can be deployed at the mid-haul to execute trust-related functions. An ML-enabled trusted platform module (TPM) can be deployed at the backhaul to ensure trust management.

Expected requirements

It is expected that ML-enabled networks support appropriate ML mechanisms for efficient cross-haul resource orchestration, i.e., to provide resource orchestration at the FH, BH, and mid-haul at the tenant level.

It is expected that ML-enabled networks support ML models that provide anticipatory responses as actions according to future expected events, and evolution, and propagation of discovered threats.

It is expected that ML-enabled networks support ML mechanisms to facilitate concurrent and seamless allocation of virtual resources to the tenants dynamically without any service disruption to virtual network operators (VNOs).

NOTE 4 – Virtual domains should be isolated across tenants where each VNO is able to specify its addressing space, deploy a network operating system of its own choice, along with their own virtual topology.

It is expected that ML-enabled networks support ML mechanisms to audit data integrity of each network slice, e.g., for resilience and VNF anomaly detection.

Added value requirements

It is of added value that ML-enabled networks support a continuous and collaborative decision-making process to offer a service in a multi-stakeholder scenario.

It is of added value that the ML-enabled networks overcome the limitations of classic multi-tenant architecture alternatives by offering an energy-efficient and trusted execution environment.

6.1.12 Network slice SLA assurance based on ML

6.1.12.1 Use case description

In the era of future networks including IMT-2020 networks, networks can be offered as a service and there will be a new demand for industry sectors (e.g., factories, stadiums, public transportation places, airports, power plants) to be connected to the network. Different service characteristics need to be offered and assured accordingly. The offer is realized by different network slices with different properties e.g., number of served customers, network performance, network availability, service experience. Assurance of the network slices is needed to justify the business model of price differentiation on a per network slice baseline. So a significant part of the operator's network slice offer will be to ensure the network slice assurance, e.g., per network slice:

- number of served subscribers;
- service KPIs' measurements, which may be summarized to a QoE vector or QoE score;
- guarantee levels and measurement thresholds.

The network slice operator needs to constantly measure SLA relevant (service) KPIs from RAN, CN, transport or combinations thereof and take (immediate) action in case the agreed conditions for the network slice are not fulfilled. It is important that the solution for SLA assurance be made cost effective by increasing the level of automation in the reconfiguration of the network.

6.1.12.2 Use case requirements

6.1.12.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support SLA management procedures that utilize continuous collection of large amounts of traffic data or measurements provided by OAM tooling.

NOTE 1 – Collected data is needed for calculation of the service measurements (KPIs) which indicate the level of assurance for the network slice.

It is critical that ML-enabled networks have SLA measurement systems that are connected to the data collection framework.

NOTE 2 – Data may be collected from interconnection between the SBA nodes and used by a network data analytics function. An example of a network data analytics function is the 3GPP MDAF [b-3GPP 28.533].

It is critical that ML-enabled networks support SLA management functionalities that collect the measurements and provide an execution environment for the trained ML model.

6.1.12.2.2 Use case requirements related to data storage and processing

None.

6.1.12.2.3 Use case requirements related to application of ML output

None.

6.1.13 Service management for smart cities

6.1.13.1 Use case description

Smart cities will integrate heterogeneous services to the operator's array of verticals. Usage of standard mechanisms for the management of machine learning functions will allow operators and regulators to seamlessly integrate ML-based services into the main array of service offerings for smart cities.

6.1.13.2 Use case requirements

6.1.13.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support inputs from all traffic categories and existing services provide input to the ML services offered by the operator.

NOTE 1 – For example, mIoT, URLLC and eMBB types of traffic may be analysed using the ML capabilities hosted in the operator's network. This will allow the operator to take full advantage of IMT-2020 features and provide analytic services on top of it.

It is critical that ML-enabled networks support data from verticals being used across network slices to not only optimize per network slice behaviour, but also cross-slice pollination of smart behaviour in the network.

NOTE 2 – This may require interoperable transfer of knowledge from one operator to another.

NOTE 3 – For example, data from network slices may be used for resource allocation in the network slices but across verticals like in smart city use cases.

6.1.13.2.2 Use case requirements related to data storage and processing

None.

6.1.13.2.3 Use case requirements related to application of ML output

None.

6.1.14 Automated testing of services

6.1.14.1 Use case description

An open testbed is a must for testing advanced services in IMT-2020, including ML-enabled network services and mechanisms in the IMT-2020 network. This will enable interoperability of ML-enabled network services and allow transparent standards development and regulation.

6.1.14.2 Use case requirements

6.1.14.2.1 Use case requirements related to data collection

None.

6.1.14.2.2 Use case requirements related to data storage and processing

None.

6.1.14.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks expose interoperable and open testing interfaces and monitoring interfaces.

NOTE 1 – Service definition should allow automated testing of ML-enabled network services in a standalone environment without affecting the operator KPIs.

It is critical that ML-enabled networks support the management of ML-enabled network services and data compliant with network operator policies.

NOTE 2 – The compliance should be verified in a standalone environment before deploying ML-enabled network services in the general availability network.

6.2 User plane-related use cases

This category of use cases is related to the user plane of the network. The use cases which belong to this category may use the user plane in different manners, for example as a source of data or sink for configurations (e.g., traffic classification).

6.2.1 Traffic classification

6.2.1.1 Use case description

Future networks including IMT-2020 networks transport traffic for heterogeneous services and applications with different QoS and QoE requirements. There is a growing trend towards encrypted traffic in mobile networks. To provide efficient management of network resources, which are tailored to specific traffic types, it is necessary to have efficient and accurate traffic classification methods. This may involve obtaining useful profiling information from the traffic data of the UE.

ML-based traffic classification aims at classifying large amounts of network traffic in a real-time manner. This ML-based classification overcomes the limitations of classical solutions, such as port-based methods and deep-packet inspection [b-ITU-T Y.2774]. ML-based traffic classification is used to enable the treatment of different services or applications according to their QoS or QoE requirements. It also provides key profiling information to operators for e.g., personalized advertising.

6.2.1.2 Use case requirements

6.2.1.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the continuous collection of a large amount of traffic data and learning the patterns of the collected data to build traffic classification models.

It is critical that ML-enabled networks collect real-time traffic data and labels it according to the application and protocol type.

It is critical that ML-enabled networks utilize the following input data:

- transport-layer payload of the first few bytes after the PDU session establishment;
- protocol fields of the first few packets after the PDU session establishment.

It is critical that ML-enabled networks normalize all the collected data before being input into the model, irrespective of the ML model being used.

NOTE 1 – Input data formats may be different depending on the ML models being used. For example, one-dimensional convolutional neural network (1DCNN) and long short-term memory (LSTM) ML models require a one-dimensional sequence of the form 1*784, while two-dimensional convolutional neural network (2DCNN) model requires a two-dimensional matrix of the form 28*28.

It is critical that ML-enabled networks support analysis based on the traffic packets that are unencrypted.

NOTE 2 – In an autonomous system, the traffic packets may be decrypted using a decryption key obtained via key exchange mechanisms, depending on the network operator policies and agreement with the application provider.

It is critical that ML-enabled networks support the collection of training and collected training samples are counted by category and labelled.

It is critical that ML-enabled networks support the collection of user specific data which follows applicable regulations and network operator policies.

Expected requirements

It is expected that ML-enabled networks support as part of data collection for ML-based traffic classification, the payloads from layers above the transport layer are extracted in the form of bytes and used as input data used for analytics.

NOTE 3 – The extracted data may be stored in the form of matrices or sequences in a database.

It is expected that ML-enabled networks support information from different OSI layers in the user traffic data, in the input data used for analytics.

NOTE 4 – For example, the information included in the input for analytics may include transport layer payload and header fields.

It is expected that the ML-enabled networks use diversified traffic categories as input data for analytics.

It is expected that ML-enabled networks are utilized to balance the traffic categories as unbalanced categories distribution may affect the performance of the ML model.

NOTE 5 – Over-sampling or under-sampling may be used as balancing mechanisms.

6.2.1.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support real-time flexible classification methods.

NOTE 1 – In this context, ML algorithms can be used, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) or their hybrid combinations. To classify real-time traffic data, an ML algorithm, during its training phase, automatically extracts relevant features for traffic classification without expert intervention. Then, based on the trained ML model, the classification method can intelligently classify services or applications from real-time traffic data.

It is critical that ML-enabled networks support monitoring capabilities to view in real time, the status of data, collectors, packets, and flows, as well as to obtain high-level network analytics.

NOTE 2 – These monitoring capabilities allow network administrators (e.g., using a web client) to control the quality and accuracy of the ML-based traffic classification.

It is critical that ML-enabled networks support retraining the ML model after the update of datasets.

It is critical that ML-enabled networks support storage of the collected data in a database.

It is critical that ML-enabled networks store traffic data that is refined to align the input requirements of the dataset for the ML training in the ML pipeline.

NOTE 3 – Refined traffic data sets can be used for training ML models. It is critical that ML-enabled networks support mechanisms so that overall efficiency in network performance can be achieved in real time.

It is critical that ML-enabled networks follow applicable regulations and network operator policies for storage and processing of user specific data.

Expected requirements

It is expected that ML-enabled networks support storage of data used for analytics.

It is expected that ML-enabled networks support classification of traffic data into different categories on the basis of application types.

NOTE 5 – Examples of application types are chat, email, streaming, peer-to-peer, voice over IP and file transfer.

It is expected that ML-enabled networks support trade-off of computation vs. level of information by using the first few bytes of the payload as input data.

It is expected that ML-enabled networks support distributed machine learning [b-FED-LEARN], with different ML models using separate sub-training datasets to train the base ML models.

It is expected that ML-enabled networks support the measurement of ML models performance.

NOTE 6 – Accuracy, precision and recall are examples of performance measurements of the ML models.

It is expected that ML-enabled networks support the optimization of ML model parameters, based on the type of inference and data types, improvement of the ML model performance and reduction of training time.

NOTE 7 – Batch-size, epochs, learning rate and regularization parameters are examples of ML model parameters. These parameters may vary based on type of inference and data types.

It is expected that ML-enabled networks support the selection of the ML model for traffic classification considering the characteristics of the traffic data and the metadata associated with the ML model.

NOTE 8 – As an example of the selection, to discover the horizontal and vertical connections of input data, 1DCNN, 2DCNN and LSTM models are selected as the base-ML models.

It is expected that ML-enabled networks support the customization of specific ML model configurations according to use case specific requirements.

NOTE 9 – Examples of configuration parameters include the type of operating system being used, the size of memory, the type of CPU, the type of specific processors like GPU, the type of machine learning library being used.

It is expected that ML-enabled networks use machine learning techniques to intelligently classify large amounts of traffic carried over the network.

It is expected that ML-enabled networks support learning a pattern from the collected traffic data to build the traffic classification model.

Added value requirements

It is of added value that ML-enabled networks support the traffic classification in real time.

6.2.1.2.3 Use case requirements related to application of ML output

Added value requirements

It is of added value that ML-enabled networks support pluggable deployment methods for enabling ML-based traffic classification in the network.

It is of added value that ML-enabled networks support the planning and design of network services using historical data of classified traffic at different points of the network.

It is of added value that ML-enabled networks support network devices enhanced with ML-based traffic classification capabilities.

6.2.2 Long-term traffic forecasting

6.2.2.1 Use case description

The annual mobile traffic consumption will exceed half a zettabyte by 2021. To meet the stringent performance requirements of emerging applications such as augmented/virtual reality, assisted living robotics, and autonomous vehicles, mechanisms such as precision traffic engineering and demand-aware allocation of cellular network resources are essential.

It is important that accurate traffic forecasting capabilities are available to fulfil these tasks. Accurate traffic forecasting is however challenging to implement and relies on specialized equipment, e.g., probes. Deploying these is expensive and involves storing locally massive amounts of logs that later have to be transferred for non-trivial post processing and analysis. Timely and exact mobile traffic forecasting is further complicated by the complex spatiotemporal patterns of user demand that arise due to user mobility.

Machine learning-based traffic forecasting can overcome these challenges and significantly outperform traditional predictive modelling techniques, e.g., autoregressive integrated and moving average (ARIMA) and exponential smoothing), which ignore important spatial correlations associated with user movement, only work well in estimating trends, and cannot be used for network-wide forecasting. By exploiting the exceptional feature extraction abilities of deep learning, structures such as 3D-CNNs (which work across space and time) and convolutional LSTMs (that are good at processing sequential data, while also capturing spatial relationships) can be used for precision long-term network-wide mobile traffic forecasting, only relying on limited observations.

6.2.2.2 Use case requirements

6.2.2.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks collect traffic consumption data at base station level, for a sufficiently long duration.

NOTE 1 – This will ensure ML-based traffic forecasting algorithms can be trained with high accuracy.

It is critical that ML-enabled networks support dedicated measurements infrastructure at the AN level.

NOTE 2 – The deployment of this dedicated infrastructure may require hardware and/or software upgrades or updates.

6.2.2.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support geographical location tagging of the collected data, i.e., the traffic consumption summaries collected periodically are accompanied by information about the geographical coordinates of the base station from where these measurements originate.

It is critical that ML-enabled networks provision sufficient storage in the core network.

NOTE 1 – Storage may be provisioned at gateway level where inferences will be made.

Expected requirements

It is expected that ML-enabled networks provision parallel computing capabilities (e.g., GPUs) in the core network.

It is expected that ML-enabled networks forecast future data at base station level, (and not at per user and per service level) aligned with privacy guidelines.

It is expected that ML-enabled networks support the streaming of measurements collected at different edge locations to a central location where traffic forecast can be performed.

NOTE 2- Gateways are examples of central location and an AN is an example of edge location.

NOTE 3 – This means there is no significant computational overhead at the edge locations for this use case.

It is expected that ML-enabled networks support bulk data storage and neural network model training where parallel processing resources are available.

It is expected that ML-enabled networks collect timing information (start and end of observation period) and geographical coordinates of the collecting entity along with measurements.

It is expected that ML-enabled networks support additional storage capabilities at the AN level in order to store and transfer measurements to a central location e.g., CN.

NOTE 4 – Additional storage is useful when the overheads incurrent with measurements transfer to a central location are considered unmanageable.

Added value requirements

It is of added value that ML-enabled networks support the following information as input for the forecast:

- cell identification;
- number of antennas;
- average number of connected users over each observation interval.

6.2.2.2.3 Use case requirements related to application of ML output

Expected requirements

It is expected that ML-enabled networks use ML models that, once trained, will only work with episodic observations of the traffic demand.

NOTE 1 – The episodic observations are represented using metadata as summaries sampled, e.g., every 5-10 minutes.

It is expected that ML-enabled networks support the sending of traffic forecasts to the edge of the network, e.g., gNB, so that agile resource allocation mechanisms can be executed at the edge.

Added value requirements

It is of added value that ML-enabled networks use neural network architectures that generalize well to different network deployments.

NOTE 2 – An example of generalization is an ML model which is trained at city scale and deployed in rural settings. It is important that validation of ML output across different cities is performed. This will also indicate whether further data collection is needed for model retraining.

6.2.3 Emergency services based on ML

6.2.3.1 Use case description

Telecommunications can play a crucial role during emergency cases of natural hazards, earthquakes, cyclones, floods, tsunamis, volcanic eruptions and fires. Machine learning can have a critical role in disaster risk reduction and management, through the following proposed steps:

- Detection: detection can be triggered through different methods, for example by IoT sensors, cameras, people alerts or warning alerts from authority.
- Analysis: ML-based analysis is done on images captured by cameras in different spots or uploaded manually by people. Information coming from different sources is correlated in order to validate the emergency situation. Analysis is done for finding optimum rescue route and transport routes for rescued users to get to safety zones.
- Warning: warning messages and notifications are sent to people on specific location based on the severity and impacted area.

NOTE – This use case uses the following deployments of application servers for emergency services in the network:

- Centralized application server: application function hosting coordination of emergency services for users in the network, in a centralized location e.g., CN or central cloud. It handles all inputs from different locations in order to process and train ML models in a centralized manner.
- Edge application server: application function hosting the coordination of emergency services for users in the network, in an edge location e.g., AN or edge cloud.

6.2.3.2 Use case requirements

6.2.3.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support emergency services that integrate different detection facilities in order to unify ways of communicating between such capabilities and avoid interoperability problems.

NOTE 1 – Examples of detection facilities include IoT sensors, building management systems, cameras, people alerts or authority warning alerts.

Expected requirements

It is expected that ML-enabled networks support the common alerting protocol (CAP) [b-OASIS-CAP] which is used for exchanging all-hazard emergency alerts and public warnings over different kinds of networks.

NOTE 2 – CAP provides an open, non-proprietary digital message format for various types of alerts and notifications. It allows a consistent warning message to be disseminated simultaneously over many different warning systems, thus increasing warning effectiveness while simplifying the warning task. CAP also facilitates the detection of emerging patterns in local warnings of various kinds, such as might indicate an undetected hazard or hostile act. CAP also provides a template for effective warning messages based on best practices identified in academic research and real-world experience.

NOTE 3 – CAP 1.2 [b-ITU-T X.1303 bis], technically equivalent to the OASIS Common Alerting Protocol v.1.2, provides an XSD specification and an equivalent ASN.1 specification (that permits a compact binary encoding) and allows the use of ASN.1 as well as XSD tools for the generation and processing of CAP messages. This Recommendation enables existing systems, to more readily encode, transport and decode CAP messages.

It is expected that ML-enabled networks support collection of the following data:

- UE measurement data, e.g., UE location information;
- users' contribution data, e.g., text, image or video;
- other data, e.g., external network detection information.

Added value requirements

It is of added value that ML-enabled networks support the methods of communication as interface for the external network elements in the detection phase in the order of priority:

NOTE – The following technologies may be used as interface for the external network elements:

a) CAT-M channel with a higher QoS for emergency sensors/IoT devices;

b) normal CAT-M channel for e-health sensors (optionally for personal health alerts);

c) a higher QoS connection for warning applications from the UE;

d) other types of internet connection e.g., IMT-2020 – which is neither emergency, nor health alert nor a warning.

6.2.3.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks use machine learning techniques in the analysis of the following:

- image processing for emergency images from different streams (public cameras, users' contributions, etc.) and comparison to previously reported incidents and known emergencies (fires, floods, etc.);
- correlation of the information collected from IoT sensors, building warning systems alarms, and other geographically collocated sensors in order to manage incident severity;
- analysis for the collected movement routes of rescued users to get to safety zones, in order to recommend the best route to get to safety or recommend alternative routes.

Expected requirements

It is expected that ML-enabled networks support signalling that gives highest priority for emergency IoT devices to communicate directly to the centralized application server without the need of CN signalling.

It is expected that ML-enabled networks support the following in the radio network:

- new downlink radio channel or higher quality for existing channel to broadcast or multicast based on location;
- new uplink radio channel or higher quality for existing channel for uploading user input regarding the emergency;
- new uplink and downlink radio channel for IoT and external systems communication towards ML model or higher quality for existing channel.

Added value requirements

It is of added value that ML-enabled networks support an edge application server that works as a standalone server in case connection to the CN is down during emergency situations.

NOTE 1 – Change in signalling is needed in order for IoT devices to communicate directly to the edge application server without the need to include CN signalling in this scenario.

It is of added value that ML-enabled networks use a separate communication channel (than the one used towards the CN) to connect the IoT device directly to the peer side (e.g., edge application server).

NOTE 2 – This architecture is more robust in case of catastrophic incidents but requires major changes to the network and signalling flows.

NOTE 3 – Edge application servers will eliminate all unnecessary information and communicate only, for example, the training sequence to the centralized application server. An edge application server can work also on standalone mode in case the link to the centralized application server is down.

It is of added value that ML-enabled networks support the breakout of emergency data to ML models at the edge application server directly without including CN nodes.

NOTE 4 – CN communication can be limited to only the authentication and initial connection establishment phase.

NOTE 5 – The local routing of emergency data to an ML model is value-add only in case the emergency server is not hosted in cloud. In case it is hosted in cloud, emergency data needs to be routed to the cloud.

6.2.3.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support the broadcasting of warning messages on a separate channel with a higher QoS class identifier (QCI) to have a higher priority.

It is critical that ML-enabled networks support the warning messages that are location-based and include text, image, video and navigation information.

Expected requirements

It is expected that ML-enabled networks support warning channels that are available even in case the link between the edge application server and centralized application servers or between edge application servers are down.

It is expected that ML-enabled networks support a common application server that is connected to all external network elements for detection via APIs and also connected directly to UEs.

It is expected that ML-enabled networks support edge application servers that are connected to communication networks through standard APIs.

NOTE 1 – It is possible to use an optional radio channel specified for this purpose. Edge application servers are also connected close to the access network in order to act standalone in case connection to the centralized application server is down. The centralized application server is handling machine learning and enhanced prediction collected from other edge application servers.

Added value requirements

It is of added value that ML-enabled networks support location-based local news content based entirely on auto-generated machine learning news.

NOTE 2 – The generated content may be a mix of the captured images of the event (e.g., local highlights, road blocking incidents, accidents, etc.) and the content generated automatically based on machine learning analysis of the situation constructed on information collected from the data sources e.g., detection cameras, IoT, sensors.

6.3 Application-related use cases

This category of use cases is related to the applications running on the network, e.g., using application data for machine learning in the network.

6.3.1 AN-assisted transmission control protocol window optimization

6.3.1.1 Use case description

The radio condition of a mobile network fluctuates in the order of milliseconds and may occasionally result in packet losses even without network congestion. However, the application data rate is adjusted on a larger timescale in the order of seconds and tends to attribute packet loss to network overloading and congestion. As a result, there is a misalignment between the RAN and the application, which may lead to non-effective usage of available radio resources and a degraded user experience.

The proposed AN-assisted TCP [b-IETF RFC 793] window optimization is used to inform the applications about the radio air-interface channel status in real time. This allows the applications to adjust their transmission data rate. Specifically, the TCP window size can be optimized to better match the radio channel variations, based on information provided by the AN.

NOTE – Examples of information provided by the AN may be buffer size in the base station, load of the base station (BS), the link throughput and packet error rate (PER).

ML-based TCP window optimization may offer an effective solution to solve the misalignment between the AN and the applications by matching the TCP window and the wireless channel condition. This will significantly improve system throughput and buffer utilization.

6.3.1.2 Use case requirements

6.3.1.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support collection and use of real-time data for AN-assisted TCP window optimization.

NOTE – The real-time data includes but is not limited to the MAC layer data [b-3GPP 38.321], PDCP layer data [b-3GPP 38.323], MAC data rate, PDCP buffer size, BS load, number of active UEs, service type and service status and application data.

6.3.1.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support TCP window prediction that utilizes:

- service type and status data fed back to the access network from the core network;
- ML prediction model which runs per UE;
- caching and/or processing and/or database in access network for storage of intermediate and/or final output from the ML models;
- training real-time predictions model hosted in access network;
- real-time data collection;
- real-time predictions;
- calibration of ML models.

Expected requirements

It is expected that ML-enabled networks support a supervised learning based model that is trained offline for TCP window prediction.

6.3.1.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support output data that includes the TCP window decision model and TCP packet transmission rate.

6.3.2 Retention and storage intelligence function

6.3.2.1 Use case description

Several use cases in IMT-2020 require massive data collection and processing at various nodes in the network architecture. Most of this data is stored in or near real-time processing engines and therefore these storages are going to have very limited capacity and need to clear the memory quickly.

Conventional methods and approaches for data clean up or retention cannot be used in IMT-2020, since most of the use cases are dependent upon historical data and the futuristic decisions need to be made based on some relationship among datasets. Hence, many fixed rules cannot be applied, and intelligence needs to be derived from the data itself, which is based on properties associated with the data. Dynamic rules regarding retention need to be formulated based on historical learnings, storage and the context.

Data at various network nodes, stored particularly at edge, require dynamic retention and storage solutions, so that the overall network and system architecture can efficiently manage storage, comply with all regulations and provide all relevant insights to run IMT-2020 use cases with all data dependencies handled in a transparent fashion.

6.3.2.2 Use case requirements

6.3.2.2.1 Use case requirements related to data collection

Expected requirements

It is expected that ML-enabled networks use the following data from the CN nodes for analysis:

- fault and performance data;
- service quality/DPI and sensor data;
- network parameters and service KPIs;
- service usage data;
- reference data from external sources, e.g., business specific, legal specific and geographic specific data.

6.3.2.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support data retention policies at a granular level in compliance with all regulations and privacy laws.

It is critical that ML-enabled networks support policy rules for data sharing with external agencies and other communication service providers are defined.

It is critical that ML-enabled networks support on-demand queries around data deletion or archiving and to have the capability to find an alternative mechanism in real time to avoid undesired data archiving or prevent loss of current data.

It is critical that ML-enabled networks support time sensitivity analysis of storage to make sure that it is proactively performing data clean-up to support any storage requirement scenario.

NOTE 1 – For example, random weather change or storm may trigger significant MTC data.

It is critical that ML-enabled networks are able to perform location sensitivity analysis of storage including hierarchical storage analysis from edge to core.

It is critical that ML-enabled networks support data predictability.

NOTE 2 – If data is highly predictable then that can be easily deleted and/or archived.

It is critical that ML-enabled networks support analysis of data relations so that proactive decisions on data archiving/deletion can be performed.

NOTE 3 – For example, in snowfall in a particular location, all installed weather sensors will start reporting a sudden fall in temperature. Most of these sensors will have informative data and that can be ignored, whereas a sensor reporting data from an operating ambulance in that area may be important. Though the data coming from medical devices need to always be preserved, it is important that ML-enabled networks have the ability to identify which data can be deleted from storage in case of application of priority in service data preservation.

It is critical that ML-enabled networks support standard mechanisms for data retention and storage management so that interoperability can be established and learning can be transferred from one CSP to another.

NOTE 4 – The transfer of learning can be achieved using transfer or model parameters such as weights. Data will go through various analytical lifecycles depending upon scenario and data type.

Expected requirements

It is expected that ML-enabled networks classify data based on characteristics such as service impact, user impact, regulatory impact and future influence basis.

6.3.2.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support rules and data processing at all levels so that best and optimal decision around data use, storage and state transition can be taken.

NOTE 1 – The levels include application/service, network and radio levels.

It is critical that ML-enabled networks take decisions at all levels with some control at the higher level in the hierarchy.

NOTE 2 – For example, an access node which has rule delegation from the core network, therefore different or similar rules regarding data processing and analytics will be applied at all levels of the network architecture.

It is critical that ML-enabled networks support national and/or regional regulations, e.g., those on security and privacy, along with the use case requirements.

Expected requirements

It is expected that ML-enabled networks use the intelligence derived from the data to dynamically tune the policies and/or processing rules.

NOTE 3 – The above-mentioned tuning may be controlled by a network node at a higher level in the hierarchy.

It is expected that ML-enabled networks support data processing based on static policies and/or dynamically assigned policy/business rules based on the dynamic network conditions.

6.3.3 Data-driven architecture for ML at the edge

6.3.3.1 Use case description

The complexity associated with future networks including IMT-2020 networks requires new paradigms for network management and orchestration. The usage of machine learning (ML) and artificial intelligence (AI) techniques to perform autonomous operations in cellular networks has been widely studied in recent years. This trend is coupled with the application of big-data analytics that leverage the huge amount of monitoring data generated in mobile networks to provide more insights on the behaviour of networks at scale.

It is important for mobile networks to support data-driven approaches to self-organize the network, achieve cost savings, but also to offer new services to the end users. For example, the prediction of the number of users in a base station can be used to optimize the performance of the network in a number of different ways: it can enable predictive load-balancing, bearer pre-configuration, scaling of RAN resources, sleeping periods for base stations, and so on. Moreover, by exploiting the knowledge on the mobility pattern of users, it is possible to understand hidden structures in the network and, for example, cluster together base stations which are visited by users with similar mobility patterns, to reduce the control plane latency.

In terms of new services to the end users, network operators can exploit the prediction to offer novel services to the end users. For example, consider a vehicle that has to travel from point A to point B in an area covered by cellular service. While on the journey, the passengers may want to participate in a conference call, or, if not driving, surf the web or stream multimedia content. Therefore, given the choice of multiple routes with similar estimated times of arrival (ETAs), the passengers may prefer to choose an itinerary with a slightly higher ETA but with a better network performance, because, for example, the itinerary crosses an area with a better coverage, or with fewer users. This becomes particularly relevant in view of the envisioned transition to an autonomous driving future, in which active driving might not be required and working or getting entertained in the car will become a common trend. In order to address this need, cellular network operators can exploit ML/AI architectures and the prediction of the number of active users in the cells to offer anticipatory services to the end users and inform them on which is the best route for their journey.

6.3.3.2 Use case requirements

6.3.3.2.1 Use case requirements related to data collection

Expected requirements

It is expected that ML-enabled networks perform the operations of collecting and processing real-time network generated data, in order to maximize the integration of ML pipelines and networking elements.

It is expected that ML-enabled networks support ML architecture that collects the data generated by the network.

NOTE – The data will be used to perform analytics and extract relevant metrics.

6.3.3.2.2 Use case requirements related to data storage and processing

Expected requirements

It is expected that the ML-enabled networks support input of the collected data to intelligent algorithms for network control and new services offerings to the users.

Added value requirements

It is of added value that ML-enabled networks are able to process data without losing awareness of the spatial information that is introduced by user mobility.

NOTE – The knowledge of spatial information improves the performance of ML algorithms for the use cases.

It is of added value that ML-enabled networks support RAN controllers, deployed at the edge, that are associated with a cluster of base stations, and host ML functionalities that use spatial information correlation.

6.3.3.2.3 Use case requirements related to application of ML output

Expected requirements

It is expected that ML-enabled networks support network controller, placed in the operator's cloud, that orchestrates the operations of the RAN controllers based on the ML output.

6.4 Signalling or management related use cases

This category of use cases uses either information from signalling messages as input, or uses signalling messages to control the behaviour of network functions or interface with management plane, e.g., it uses an interface with NMS to configure NFs.

6.4.1 ML-based mobility pattern prediction

6.4.1.1 Use case description

Many next-generation applications and mobile network optimization schemes require knowledge of mobility patterns of mobile users. Mobility pattern prediction (MPP) based on ML is a method to predict mobile users' trajectory and their service pattern usage, from available network data and user data, as well as context information. ML-based MPP schemes allow for an optimized mobile network with adaptive real-time network configuration including (but not limited to): proactive resource allocation, improved handovers, predictive caching, and advanced energy saving schemes. Furthermore, MPP will enable several new applications, such as adaptive public transportation solutions, adaptive streetlights, smart home heating systems, location-based advertisements etc. These new mobile network applications require an automatic real-time prediction of the users' mobility and service pattern. In this use case, the focus is on using ML methods to predict future trajectories of individual users (which could be used to predict the future condition of the mobile network).

Compared to current model-based prediction methods, a data-driven MPP approach using ML will increase intelligence of the system and significantly improve prediction results in terms of performance and accuracy. Consequently, the improved prediction results allow for the implementation and provisioning of better (autonomous, real-time etc.) applications and services.

Such services include personalized mobile edge caching (PMEC), which adopts a flexible and distributed computing and storage capabilities to provide ultimate QoE and personalized service access for next generation user-centric networks. PMEC is able to solve the problems of the existing edge caching such as limited sink location, limited application scenarios, limited service range and limited operational capability. A solution could be based on a two-tier deep learning based caching architecture for MEC. At a lower layer, the local predictor module (LPM) collects and analyses hourly user requests from one base station, while at the upper layer, the global predictor module (GPM) for network-wide analysis and prediction of each day is based on the preliminary process of each LPM using a deep learning technique. In order to determine which content should be placed in the cache appliances, the caching decision module (CDM) is designed to integrate the results from the LPM and GPM.

Another application based on MPP is handover optimization, where two conflicting objectives are pursued: minimizing the unnecessary handovers and minimizing the likelihood of dropped calls. Compared to LTE networks, the handover management for 5G networks is much more complicated and with much more stringent requirements, namely almost "zero" latency handover and consistent user experience. The ML model can be trained based on the history information such as reference signal received power/quality (RSRP/RSRQ) in handover, network load, throughput, interruption time, as well as the awareness and prediction of the UE position, moving direction, speed, etc.

It should be emphasized that ML can also be used to enhance or facilitate existing (model-based) approaches to the MPP problem. In fact, purely data-driven approaches might not be suitable for robust real-time MPP applications, in which case it might be necessary to use hybrid ML methods that include an additional context information in form of e.g., models, constraints etc.

The MPP will provide the possibility for predicting the condition of the future network by analysing users' behaviour patterns. In order to improve the accuracy and effectiveness of the prediction model in real time, the MPP collects training data continuously to update the prediction models accordingly.

The MPP consists of three main functionalities: data collection, strategy and execution. The data collection functionality is responsible for real-time collection of network performance data, cell status query, sleep/wake-up instruction, context data and other current network interaction functions. The core algorithm achieves the user trajectory prediction in the strategy functionality. The execution functionality is to generate man-machine language (MML) and communicate with the base transceiver stations (BTSs), other elements of the core network and managers such as VNF managers (VNFMs) and SDN controllers (SDNCs).

6.4.1.2 Use case requirements

6.4.1.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks collect position estimates (e.g., GNSS data, cell identifier, tracking area identifier) for mobility pattern prediction.

It is critical that ML-enabled networks collect real-time network performance data to support network performance optimization.

It is critical that ML-enabled networks support the collection of the following data:

- measurement data, such as position estimates;

  NOTE 1 – Examples of measurement data include GNSS data, RSRP, timing advance, beam state information.

- context information data, which includes network state data and user state data.

  NOTE 2 – Examples of context information data include UE identifier, various logs, KPIs and environmental information (e.g., maps). For the network state data, various identification information at the AN (e.g., cell identifier, beam identifier), configuration information (e.g., number of antennas), cell status query data and real-time network performance data should be included.

<u>Expected requirements</u>

It is expected that ML-enabled networks support collection of network measurements, e.g., RSRP to support network performance optimization (e.g., handover optimization) based on mobility pattern prediction.

It is expected that ML-enabled networks are able to utilize different data sets collected from different network nodes.

<u>Added value requirements</u>

It is of added value that ML-enabled networks support the collection of environmental information (e.g., maps) for mobility pattern prediction.

It is of added value that ML-enabled networks support enhanced training of ML models, for instance to learn rare events, by generating simulated data.

6.4.1.2.2 Use case requirements related to data storage and processing

<u>Critical requirements</u>

It is critical that ML-enabled networks support a core network with a database to store collected data and possibly store predictions which may be applicable long term.

NOTE 1 – The stored predictions may be provided to other consumers in the future.

It is critical that ML-enabled networks support long-term prediction in the core network which aims to predict large-scale activity and UE mobility at a coarse level.

It is critical that ML-enabled networks support databases to store collected data at the AN.

It is critical that ML-enabled networks support short-term predictions which aim to predict user group activity and UE mobility at a finer level at the AN.

It is critical that ML-enabled networks support databases to cache simulation data or sensor data at application functions.

It is critical that ML-enabled networks support real-time prediction which aims to predict user activity at application functions.

<u>Expected requirements</u>

It is expected that ML-enabled networks support offline databases and ML model training during the network planning and design phase.

It is expected that ML-enabled networks support online data collection and online predictions during the network deployment phase.

It is expected that ML-enabled networks support online model calibration and network management during the network operation and management phase.

It is expected that ML-enabled networks are able to make predictions at different time granularity, such as real-time predictions (user activity), short-term predictions (user group activity) and long-term predictions (large-scale activity).

It is expected that ML-enabled networks support different MPP models at different domains.

NOTE 2 – Examples of MPP models include UE-level ML model related to application and service levels, network-element-level ML model (e.g., CU, DU) in access network and network-level ML model (e.g., CN, NMS) in the core network.

Added value requirements

It is of added value that ML-enabled networks use new data sets to calibrate the MPP model and apply the MPP model to network management.

6.4.1.2.3 Use case requirements related to application of ML output

Added value requirements

It is of added value that ML-enabled networks enhance different network levels [ITU-T Y.3172] with ML-based MPP capabilities.

NOTE 1 – MEC deployments at various levels of the network may be examples. MEC deployment at the core network may include the implementation of global prediction with the LSTM model and data from the MEC at the access network. The MEC at the access network may include collecting and analysing the hourly user requests from one base station and executing caching decision and caching the video data.

It is expected that ML-enabled networks support the following ML capabilities as part of ML output:

- Classification: An ML capability to separate input data into predefined groups.

  NOTE 2 – The objective of this ML output capability is closely related to supervised learning.
- Clustering: An ML capability to separate input data into some non-predefined groups.

  NOTE 3 – The objective of this ML output capability is closely related to unsupervised learning.
- Prediction: An ML capability to forecast the occurrence of predefined phenomena.

  NOTE 4 – The objective of this ML output capability is focusing on future phenomena.
- Inference: An ML capability to reach a certain conclusion with input data.

6.4.2 Load balance and cell splitting/merging

6.4.2.1 Use case description

The mobile network traffic is rapidly increasing and more micro-base stations are needed to meet the requirement of high capacity. The coordination between macro and micro-base stations is one of the most important ways to deal with high traffic load. In particular, tidal effects of UEs require adequate approaches to deal with a large number of people gathering for instance in a stadium, on a square for a special event, a shopping street etc. There are two main techniques to cope with this problem: load balance among cells and cell splitting and/or merging.

The currently existing load balance algorithms are only able to switch a few UEs to another cell in a short period of time. The performance of the load balance mechanisms yields a poor performance, since network optimization engineers need to modify manually the network parameters to balance UEs among different cells.

In order to meet the requirements due to the rapid traffic increase, cell splitting is another solution without deploying a new physical network device. A cell can be split into two or more cells to increase the network capacity. On the other hand, cell merging leads to reduced operating expenses for situations with low traffic demands.

Based on the analysis and learning of a large amount of data, an algorithm will be trained to describe when to execute cell splitting or merging to improve the quality of service and energy efficiency.

6.4.2.2 Use case requirements

6.4.2.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support load balancing schemes which consider automatically the current number of UEs and the traffic speed of the UEs, and schedules UEs among different cells for a guaranteed QoE.

Expected requirements

It is expected that ML-enabled networks support the collection of the following measurement data for load balancing and cell split/merge schemes:

- number of UEs in the current cell and the neighbouring cells;
- RB utilization of current cell and neighbouring cells;
- PDCP [b-3GPP 38.323] package number of current cell and neighbouring cell;
- number of RRC [b-3GPP 38.331] connection requests in current cell and neighbouring cells;
- active radio bearer number of current cell and neighbouring cells.

6.4.2.2.2 Use case requirements related to data storage and processing

Expected requirements

It is expected that ML-enabled networks support schemes for load balancing which use an ML model to predict the load.

It is expected that ML-enabled networks support, based on the prediction result on the load, ML models which indicate a feasible map between UEs and cells with the specific measurement data reported from the cells.

Added value requirements

It is of added value that ML-enabled networks support the deployment of ML functionalities in NMS so that load balancing among cells or cell splitting/merging can be achieved.

6.4.2.2.3 Use case requirements related to application of ML output

Expected requirements

It is expected that ML-enabled networks apply the output generated from the ML model to the NMS.

It is expected that ML-enabled networks apply the output from the ML model to the NMS to balance the UEs connected to the base stations.

It is expected that ML-enabled networks support different ML models for cell splitting concerning load prediction and cell status analysis.

It is expected that ML-enabled networks enable, based on the prediction result, the NMS to decide whether to perform cell splitting/merging and select a set of configuration parameters for the splitting/merging process.

6.4.3 ML-based QoE optimization

6.4.3.1 Use case description

Traditional network planning and optimization methods rely on the optimization of KPIs to improve service quality. However, multimedia services and emerging forms of content are increasingly complex and diverse, and the network performance related to QoS cannot necessarily meet the needs of network users. Moreover, the requirements of different users, services and content are widely divergent in terms of QoS.

QoE is used to measure the user's evaluation of the performance of communication services and reflects the true level of user satisfaction and subjective feelings with different network services. QoE depends on QoS, and is one of the main optimization objectives of mobile networks providing complex services.

ML-based QoE may optimize the network configuration to maximize the QoE score for all network users while meeting the network capacity and service requirements.

6.4.3.2 Use case requirements

6.4.3.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support privacy and security mechanisms to limit the use and accessibility for authorized usage regarding the collection of data that relate to specific users.

It is critical that ML-enabled networks input different types of data from various parts of the network to the ML functionalities, including network status measurement data and QoS/QoE feedback data.

Expected requirements

It is expected that ML-enabled networks use KPIs such as bandwidth, bit rate and cache size as input to the ML model.

It is expected that ML-enabled networks use real-time network states and perceived service quality.

NOTE – Since some of the QoS/QoE feedback information may not be available, the ML model needs to handle incomplete input data.

It is expected that ML-enabled networks carry out accurate network measurements and recording of raw network status samples, as well as data processing to remove protocol dependent features.

It is expected that ML-enabled networks are capable of handling network status data changes at a high rate and extracting data at a high rate.

It is expected that ML-enabled networks support the collection of the following measurement data:

- bandwidth;
- cache size;
- bit rate;
- user experience:
  - re-buffering time;
  - staying time (duration of the user session);
  - blockage frequency (data blockage).

It is expected that ML-enabled networks support the following QoS/QoE feedback data:

- re-buffering time;
- staying time;
- blockage frequency (data blockage);
- latency;
- jitter.

6.4.3.2.2 Use case requirements related to data storage and processing

Added value requirements

It is of added value that ML-enabled networks store data alongside the obtained KPIs in a network status database.

6.4.3.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support mechanisms to decide the granularity of configuration according to the abstraction level of information or data model in the level of network in which the ML output is applied.

It is critical that ML-enabled networks support the deployment of ML functionalities in different parts of the network.

NOTE – For example, ML models may be deployed at core network functions.

It is critical that ML-enabled networks enable interface between machine learning functionalities and network functions at multiple levels of the network.

NOTE – Examples of levels in the network are edge network and core network.

It is critical that ML-enabled networks enable stable and proactive network optimization.

<u>Expected requirements</u>

It is expected that ML-enabled networks support a global constrained optimization for network configuration considering the QoE model and the behaviour model of the services.

It is expected that ML-enabled networks provide QoE scores under various service requirements as output of ML models.

<u>Added value requirements</u>

It is of added value that ML-enabled networks support optimization functionalities by the use of reinforcement learning mechanisms where prediction and feedback based learning models are employed.

6.4.4 ML-based network management for Industry 4.0

6.4.4.1 Use case description

Industrial IoT is one of the main drivers for the transition into the next industrial revolution, known as Industry 4.0. In industrial IoT processes are automated using connected intelligent devices which share data and provide interfaces enabling remote management. Meeting the stringent requirements of industrial IoT is becoming more and more challenging in the face of the added complexity of smart manufacturing and in the presence of diverse communication technologies. Additionally, the amount of data being produced is increasing exponentially year on year and will become increasingly difficult to support and manage with traditional approaches.

Use of machine learning enables dynamic and continuous management of the industrial IoT network's operational behaviour based on environmental observations and manufacturing patterns, resulting in optimized operation. Machine learning is expected to be an integral part of the solution for network and data management in smart manufacturing.

Industrial IoT networks are unique in their high standards and requirements when it comes to reliability, availability, security and determinism requirements they need to satisfy.

6.4.4.2 Use case requirements

6.4.4.2.1 Use case requirements related to data collection

<u>Critical requirements</u>

It is critical that ML-enabled networks support data formats and protocols which enable interoperability among IoT devices from different manufacturers.

It is critical that ML-enabled networks support data security and protection.

NOTE 1 – For example, industrial plants may have specific data which needs to be protected.

Expected requirements

It is expected that ML-enabled networks support network planning capabilities that incorporate the overhead imposed by data collection.

It is expected that ML-enabled networks act upon real-time events from industrial plants with minimal turnaround time.

NOTE 2 – This imposes stringent timing requirements on data collection, ML processing and delivery of the output to the network. Depending on the scenario, the latency requirements for communication, processing and responses can vary from multiple seconds to a few milliseconds.

It is expected that ML-enabled networks support the collection of the following data:

- measurement data
  - signal strength;
  - packet loss rate;
  - spreading factor/multiplier;
  - modulation, coding, spatial streams/ antenna usage.
- context information data
  - network state data:
    - topology routes, mesh instances;
    - traffic flows;
  - environmental information.
- planning data
  - production requirements;
  - resource availability;
  - production timelines.
- collection of information at various levels of the network (operational data);
- collection and transport of planning data to the data processing logic.

6.4.4.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks use ML algorithms that meet the accuracy, latency and functional requirements of the specific use cases.

Expected requirements

It is expected that ML-enabled networks support a resilient framework in processing events from/to industrial plants so that any failures in the framework does not impact the management of the IoT devices.

It is expected that ML-enabled networks support a data processing logic which takes planning and operational data as input and creates forecasting models for the generation of an optimal network configuration.

It is expected that ML-enabled networks support the processing of information at various levels of the network (operational data).

6.4.4.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support IoT devices with interfaces that enable setting/monitoring/measuring of device parameters and network related functions.

NOTE 1 – This includes physical and link layer interfaces, network formation/routing interfaces, power and device state interfaces.

It is critical that ML-enabled networks support IoT devices with protocols and interfaces to control various network management functions.

NOTE 2 – Examples of network management functions include schedule delivery, network resource allocation and gateway association information.

Expected requirements

It is expected that ML-enabled networks support dynamic network configuration in real time or near real time based on the output of the data processing logic.

6.4.5 ML-based correlations between transport KPIs and radio KPIs

6.4.5.1 Use case description

In future networks including IMT-2020 networks, the correlation between transport KPIs and radio KPIs using ML techniques over the collected historical data is needed. The ML output formulates the relation between the selected KPIs in both domains in order to be able to optimize their thresholds.

The transport network performance is measured against some KPIs, e.g., frame loss ratio (FLR), delay and jitter), with different thresholds for each KPI according to the radio services running. For example, the FLR threshold for 2G voice traffic is different to the FLR threshold for 4G background traffic.

Accordingly, it is necessary to intelligently analyse the complex relations between the KPIs in both domains in order to have a descriptive mapping that defines the exact correlation between them. This is used for a better understanding of the impact on radio services due to changes that happen in transmission KPIs.

NOTE – The output of the analysis may include accurate thresholds and criteria to be used in the planning and design phase, and optimization proposals for the operation and management phase.

6.4.5.2 Use case requirements

6.4.5.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the correlation of the transport KPIs and radio KPIs to continuously collect the performance measurements at different layers of transport and radio domains.

It is critical that ML-enabled networks support NMSs connected to network elements from different vendors in the transport and radio domains to continuously collect the performance data.

Expected requirements

It is expected that ML-enabled networks support the collection of raw data with variable granularity.

NOTE 1 – The granularity could be 5 minutes or less (i.e., the timer generates a value every 5 minutes or less to catch near real-time results).

It is expected that ML-enabled networks support the integration of probing and monitoring systems connected to the network elements in order to get accurate results for transport KPIs.

It is expected that ML-enabled networks support performance monitoring systems, collecting and consolidating the data from all NMSs and active probes.

It is expected that ML-enabled networks support the collection of the following transport network topology data:

- number of traversed microwaves in hops till fibre termination;

- travelled fibre distance;
- sectional representation of cascading sites.

It is expected that ML-enabled networks support the collection of the following transport measurements data:

- end-to-end latency;
- end-to-end jitter;
- end-to-end packet drops;
- sectional capacity;
- sectional utilization;
- sectional buffer size.

It is expected that ML-enabled networks support the collection of the following radio measurement data:

- time advance and propagation delay;
- RRC attempts;
- U RSSI and RTWP;
- SOHO indicator;
- utilization power;
- utilization CE and RB;
- latency, packet loss and Iub related counters;
- CQI;
- number of users;
- throughput;
- data volume uplink and downlink;
- voice traffic;
- HSDPA drop rate.

NOTE 2 – The above measurement data summary represents data readily available and stored for network performance monitoring, usually stored on an hourly basis for a multiple weeks' time period.

<u>Added value requirements</u>

It is of added value that ML-enabled networks support higher frequency data readings (e.g., 5 min, 1 min) and storage for longer time periods in order to give more accurate observations and results.

It is of added value that ML-enabled networks support the collection of real customer experience measurements from application-specific readings (e.g., web browsing, video streaming, VoIP, online gaming, VR/AR).

NOTE 3 – This data may be retrieved from crowd sourcing applications, or region-specific drive tests.

6.4.5.2.2 Use case requirements related to data storage and processing

<u>Critical requirements</u>

It is critical that ML-enabled networks support analytics functionalities which handle consolidated data and derive correlations between the various input data.

<u>Expected requirements</u>

It is expected that ML-enabled networks support a big data-capable analytics platform to manage the big volume of data collected over long periods with low granularity.

6.4.5.2.3 Use case requirements related to application of ML output

None.

6.4.6 ML-based end-to-end network management

6.4.6.1 Use case description

Network slicing in future networks including IMT-2020 enables operators to build virtual end-to-end networks tailored to various application requirements. Network slices are usually implemented over various network domains such as radio access, transport and core networks. Hence, end-to-end network management becomes more critical and tight interworking is necessary to guarantee user throughput and low latency.

However, most providers have separate management planes for their network domains, such as IP (access/core), transport (access/core/), IT (cloud computing platforms) and network functions (DU/RUs/EPC/NGC [b-3GPP 23.501]). Since the management planes still rely on human operators, this increases the total cost of ownership (TCO). With separate management environments, different operators should communicate with each other to identify the root cause of the failure and fix it together. However, most operators are only interested in their own domains and stick to their operation procedures, while there is a need to perform combined root cause analysis. Hence, providers should consider a unified end-to-end network management system covering all domains. This system should be based on matured ML technologies instead of the current human-oriented operation environment.

The new management architecture needs to have big data capabilities to collect performance data, events and alarms occurred from all involved domains. Then, the root cause analysis (RCA) on top of the platform detects automatic failures, processed by ML without human intervention. The automatic RCA services can be realized with ML such as complex event processing. With the large amount of performance data, failure prediction services are also feasible.

6.4.6.2 Use case requirements

6.4.6.2.1 Use case requirements related to data collection

<u>Expected requirements</u>

It is expected that ML-enabled networks use the following information for big data capabilities:

- information about alarm and event data;
- information about inventory and topology data;
- normalized format for the data;
- storage duration and deletion cycle;
- storage type (distributed or centralized).

It is expected that ML-enabled networks use the following information for end-to-end network management:

- service request information, e.g., about bandwidth, service type;
- information about network resource, e.g., available, required;
- path characteristics, e.g., packet-based or circuit-based, physical or virtual;
- topology information, e.g., intra-domain, inter-domain;
- connection information between domains.

6.4.6.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks support the use of a common information model to manipulate data of the objects which are managed in different network management and operation systems.

It is critical that ML-enabled networks support the use of a common data model to achieve the compatibility of data obtained across related network management and operation systems.

Added value requirements

It is of added value that ML-enabled networks are capable of deriving the timing threshold needed by services or applications based on the following:

- prediction time consumed by ML capabilities;
- required time to carry ML information between ML capabilities and service components;
- time to take action using ML information.

6.4.6.2.3 Use case requirements related to application of ML output

None.

6.4.7 ML-aided channel modelling and channel prediction

6.4.7.1 Use case description

With future networks including IMT-2020, it is envisioned to have various deployment scenarios and frequency bands, as well as new physical layer techniques, such as millimetre wave and massive MIMO. It is well acknowledged that wireless propagation measurement and channel modelling are fundamental to the success of any wireless system design and evaluation. Meanwhile, accurate and timely channel prediction dictates how well the physical layer techniques can achieve the optimal transmission capacity. Innovative methods in the big data and ML domain are expected to help in the challenging research of wireless channel modelling and channel prediction for future networks including IMT-2020. Channel modelling requires huge amounts of channel data in various propagation environments and frequency bands; meanwhile, channel prediction has very stringent latency and accuracy requirement.

This use case is divided into two parts: one is building a channel model that is purely based on neural networks through supervised learning which can capture the underlying relations between the physical environment (e.g., 3D environmental data, scatters/reflectors, propagation mechanism) and the resulting channel impulse response (from measurement campaign or real network channel estimation data); the other is to use historical and user-specific channel data to predict the future channel response in real time, which will potentially assist physical layer processes such as channel coding and modulation and reducing the need for pilot/preamble transmission.

6.4.7.2 Use case requirements

6.4.7.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support the collection of a large amount of scenario-specific channel measurements.

NOTE 1 – The collected data could be the channel impulse response from a measurement campaign or real channel estimations from the live RAN. "Scenario-specific" means that the measurements are performed for various frequency bands across 0-100GHz, multiple antenna elements at both the RAN and UE, different propagation environments (e.g., urban macro, urban micro, rural, high-speed train and indoor).

It is critical that ML-enabled networks record the actual propagation environment during measurement is recorded.

It is critical that ML-enabled networks record the system configuration for the measurement set-up.

NOTE 2 – System configuration may include for instance, the antenna pattern and configuration, MIMO mode, operating frequency, transmit power, bandwidth, sampling rate, thermal noise level and waveform used.

It is critical that ML-enabled networks support the collection of UE and BS location information, e.g., GNSS or channel information inferred distance and angular information.

It is critical that ML-enabled networks support the collection of system configuration information for the measurement equipment or the BS and UE configuration, e.g., antennas, MIMO mode, frequency band, transmit power, bandwidth.

It is critical that ML-enabled networks support the implementation of ML-based channel modelling which meets the necessary latency and accuracy requirements to estimate, transfer, receive and store channel information across the network and data centre.

Added value requirements

It is of added value that ML-enabled networks support the prediction of UE location and mobility.

It is of added value that ML-enabled networks support standard data formats for environmental data, which enable information sharing and promote joint research between different organizations.

NOTE 3 – Examples of environmental data include a 3D map of the propagation environment, transmitter and receiver location, identified reflectors and scatters, material EM properties and surface roughness.

6.4.7.2.2 Use case requirements related to data storage and processing

Critical requirements

It is critical that ML-enabled networks use ML models with clear input and output definitions.

NOTE – Based on the input and output definitions, chaining of ML models can be performed to make a unified channel modelling process.

It is critical that ML-enabled networks support the storage of UE-specific channel impulse responses estimated by the BS or reported by the UE along with time stamps.

It is critical that ML-enabled networks support protocols enabling the trained models to be subscribed, distributed and maintained.

It is critical that ML-enabled networks are able to handle time-series data and data with different dimensions, and meet the latency and accuracy requirements.

It is critical that ML-enabled networks support close-loop monitoring of the ML model performance, and are able to fall back to existing channel estimation mechanisms [b-3GPP-36.211].

Expected requirements

It is expected that ML-enabled networks support federated learning and distributed learning to distribute the computational load across processing nodes.

Added value requirements

It is of added value that ML-enabled networks store channel data and categorize them according to certain attributes, such as environment, frequency and purpose.

It is of added value that ML-enabled networks use ML mechanisms to accomplish different tasks in the channel modelling process, such as clustering and feature extraction and classification, as ML sub-models.

It is of added value that ML-enabled networks support dedicated protocols for data handling.

It is of added value that ML-enabled networks support ML models with a standard interface to the physical layer processes for collecting the data for channel modelling and applying the output of the ML processing.

6.4.7.2.3 Use case requirements related to application of ML output

None.

6.4.8 ML-based link adaptation optimization

6.4.8.1 Use case description

In cellular communication systems, the quality of the signal received by a UE depends on the channel quality from the serving cell, the level of interference from other cells, and the noise level. To maximize the throughput while maintaining target reliability on wireless communication channels, link adaptation is introduced to adjust the transmission parameters such as modulation and coding scheme (MCS) [b-3GPP 36.213] [b-3GPP 38.214] which determines the transmit block size of each stream, as well as MIMO transmission rank (number of spatial data streams for each user) and precoding to match the variation of channel conditions.

The state of the art link adaptation mechanism relies on the CQI [b-3GPP 36.213] [b-3GPP 38.214] feedback from the UE. Using predefined look-up tables for various link quality metrics, the proper MCS would be found. To further enhance the performance, out loop link adaptation (OLLA) is also introduced to modify the MCS based on the link feedback. However, the CQI feedback is usually out of date and cannot capture the inter-subcarrier and multi-user/multi-layer interference in the actual downlink transmission, which leads to the mismatch between the CQI feedback and the actual CQI for the downlink data; in addition, it is hard to measure accurately the link quality. Link adaptation in 5G with large numbers of antennas and channels is further challenging due to the high channel state information (CSI) dimension, which makes it even harder to find the proper mapping tables between link qualities and link adaptation parameters. OLLA is perfect for full buffer services but it is hard to converge for services with small burst and fast variation channel conditions. All the above-mentioned issues lead to performance degradation.

An ML-based link adaptation scheme is desired to enhance the performance by using the historical channel condition data and the corresponding KPIs to find the optimized MCS and rank.

6.4.8.2 Use case requirements

6.4.8.2.1 Use case requirements related to data collection

<u>Expected requirements</u>

It is expected that ML-enabled networks support the collection of the following layer 1 (L1) UE level feedback data:

- CQI;
- rank indicator (RI);
- pre-coding matrix indicator (PMI);
- hybrid automatic repeat request (HARQ) acknowledgements/negative acknowledgements;
- downlink RSRP measurement.

It is expected that ML-enabled networks support the collection of the following L1 BS measurement data:

- MIMO pre-coder;
- uplink RSRP measurement;
- CQI/PMI/RI feedback delay;
- estimated Doppler frequency offset.

It is expected that ML-enabled networks support the collection of the following layer 2 (L2) BS measurement data:

- PRB usage ratio;

- scheduled MCS;
- scheduled rank;
- number of users;
- interference from neighbouring cell.

It is expected that ML-enabled networks support the collection of the above-mentioned data regarding L1 UE level feedback, L1 and L2 BS measurement for model training.

6.4.8.2.2 Use case requirements related to data storage and processing

Expected requirements

It is expected that ML-enabled networks support the learning of relationship between empirical observations of the CSI related values and their associated ACK/NACK flows including their relationship with BLER.

It is expected that ML-enabled networks support the deployment and update of the trained model into the BS scheduler.

It is expected that ML-enabled networks support a BS scheduler which supports the ML model output and executes the MCS and RI selection policy based on the ML output.

6.4.8.2.3 Use case requirements related to application of ML output

None.

6.5 Security related use cases

This category of use cases is related to the security aspects of the network.

6.5.1 Combating use of counterfeit ICT devices – ML-assisted network service

6.5.1.1 Use case description

This service is expected to identify cloned IMEIs where the ML capability is utilized in the radio access part of the network and the core network. From the radio access part of the network, several device characteristics like the number of antennas and MIMO are available and in the core network these attributes as per the IMEI/model are available. ML capability can be used to identify the cloned IMEI using this information and can be deployed in the edge network nodes to prevent such mobiles from accessing the network.

6.5.1.2 Use case requirements

6.5.1.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support cloned IMEI detection capability which continuously collects and classifies the traffic data in terms of capabilities associated with the IMEI.

NOTE – Classified data can be matched against the capabilities associated with the IMEI available in the core network.

6.5.1.2.2 Use case requirements related to data storage and processing

None.

6.5.1.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks support the continuous updating of information gained from the classification in the edge nodes for faster action against cloned IMEI devices.

6.5.2 ML-based identification of illegal exchanges using SIM boxes

6.5.2.1 Use case description

Use of SIM boxes as illegal exchanges is a serious issue which causes heavy losses to the government and to the telecommunication service providers and also causes problems concerning law enforcement. To identify and stop such communications effectively, it is necessary to have an efficient and accurate method to identify such traffic.

ML-based identification of illegal exchanges using SIM boxes collects traffic data and classifies the real-time traffic as from an illegal SIM exchange or not and conveys the information to an appropriate node for further action. This ensures that such traffic does not pass through the network.

6.5.2.2 Use case requirements

6.5.2.2.1 Use case requirements related to data collection

Critical requirements

It is critical that ML-enabled networks support continuous collection of real-time data and learning of collected data patterns to build a traffic model for traffic from illegal exchanges, including patterns for identifying VoIP traffic from illegal exchanges.

6.5.2.2.2 Use case requirements related to data storage and processing

None.

6.5.2.2.3 Use case requirements related to application of ML output

Critical requirements

It is critical that ML-enabled networks, on the basis of classification according to the traffic model for traffic from illegal exchanges, support the sending of appropriate information to operator configured authorities.

Bibliography

[b-ITU-T X.1303 bis] Recommendation ITU-T X.1303 (2014), *Common alerting protocol (CAP 1.2)*.

[b-ITU-T Y.1540] Recommendation ITU-T Y.1540 (2016), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[b-ITU-T Y.2242] Recommendation ITU-T Y.2242 (2018), *Service function chaining in mobile networks*.

[b-ITU-T Y.2774] Recommendation ITU-T Y.2774 (2019), *Functional requirements of deep packet inspection for future networks*.

[b-FG ML5G use cases] ITU-T FG ML5G 17289-TD1-R1 (2019), *Machine learning in IMT-2020 and future networks: use cases and basic requirements*.

[b-3GPP 23.003] 3GPP TS 23.003 (2019), *Technical Specification Group Core Network and Terminals; Numbering, addressing and identification*.

[b-3GPP 23.501] 3GPP TS 23.501 (2019), *Technical Specification Group Services and System Aspects; System Architecture for the 5G System (5GS)*.

[b-3GPP 28.533] 3GPP TS 28.533 (2019), *Management and Orchestration; Architecture Framework.*

[b-3GPP-32.500] 3GPP TS 32.500 (2018), *Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements*.

[b-3GPP-36.211] 3GPP TS 36.211 (2019), *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation*.

[b-3GPP 36.213] 3GPP TS 36.213 (2019), *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures*.

[b-3GPP 36.214] 3GPP TS 36.214 (2019), *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements*.

[b-3GPP 36.305] 3GPP TS 36.305 (2019), *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Stage 2 functional specification of User Equipment (UE) positioning in E-UTRAN*.

[b-3GPP 36.321] 3GPP TS 36.321 (2019), *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*.

[b-3GPP 38.214] 3GPP TS 38.214 (2019), *Technical Specification Group Radio Access Network; NR; Physical layer procedures for data*.

[b-3GPP 38.323] 3GPP 38.323 (2019), *Technical Specification Group Radio Access Network; NR; Packet Data Convergence Protocol (PDCP) specification.*

[b-3GPP 38.331] 3GPP 38.331 (2019), *Technical Specification Group Radio Access Network; NR; Radio Resource Control (RRC) protocol specification*.

[b-3GPP 38.401] 3GPP TS 38.401 (2019), *Technical Specification Group Radio Access Network; NG-RAN; Architecture description*.

[b-5G-Crosshaul] X. Costa-Perez et al. (2017), *5G-Crosshaul: An SDN/NFV Integrated Fronthaul/Backhaul Transport Network Architecture*, IEEE Wireless Communication, pp. 38-45.

[b-5GPP Sec] 5 G PPP Security WG (2017), *5G PPP Phase 1 Security Landscape, White Paper*.

[b-ARIMA-Traffic] Hyun-Woo Kim, Jun-Hui Lee, Yong-Hoon Choi, Young-Uk Chung, and Hyukjoon Lee (2011), *Dynamic bandwidth provisioning using ARIMA-based traffic forecasting for Mobile WiMAX,*. Computer Communications 34, 1, pp. 99–106.

[b-Action-recog] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu (2013), *3D convolutional neural networks for human action recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1, pp. 221–231.

[b-Conv-LSTM] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo (2015), *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*. In Proc. NIPS.

[b-DELMU] R. Li, C. Zhang, P. Cao, P. Patras, J. S. Thompson (2018), *DELMU: A Deep Learning Approach to Maximising the Utility of Virtualised Millimetre-Wave Backhaul*, in Proceedings International Conference on Machine Learning for Networking (MLN), Paris, France.

[b-ETSI GR ENI 004] ETSI GR ENI 004 V1.1.1 (2018), *Experiential Networked Intelligence (ENI); Terminology for Main Concepts in ENI*.

[b-ETSI-NFV-MANO] ETSI GS NFV-MAN 001 V1.1.1 (2014), *Network Functions Virtualisation (NFV); Management and Orchestration*.

[b-ETSI NFV-Sec] ETSI NFV-SEC 013 v3.1.1 (2017), *Network Functions Virtualisation (NFV) Release 3; Security; Security Management and Monitoring specification*.

[b-FED-LEARN] Koneˇcn´y, et al. (2017), *Federated learning: strategies for improving communication efficiency arXiv:1610.05492v2 [cs.LG]*.

[b-IETF RFC 793] IETF RFC 793 (1981), *Transmission Control Protocol*.

[b-Kubernetes] Kubernetes, https://kubernetes.io/

[b-ML Edge] Michele Polese, Rittwik Jana, Velin Kounev, Ke Zhang, Supratim Deb, Michele Zorzi (2019), *Machine Learning at the Edge: A Data-Driven Architecture with Applications to 5G Cellular Networks*, submitted to IEEE Transactions on Mobile Computing, available on https://arxiv.org/abs/1808.07647

[b-OASIS-CAP] OASIS (2010), *Common Alerting Protocol Version 1.2*.

[b-OASIS-TOSCA] OASIS (2013), *Topology and Orchestration Specification for Cloud Applications Version 1.0*.

[b-ONAP] Open Network Automation Platform, https://www.onap.org/

[b-OpenStack] OpenStack, https://www.openstack.org/

[b-TEE Management] GlobalPlatform, Inc. (2016), *TEE Management Framework v1.0, Public Release GPD_SPE_120.*

[b-TEE Sys Arch]	GlobalPlatform, Inc. (2018), *TEE System Architecture v1.2, Public Release GPD_SPE_009*.
[b-Testbed]	I. Oliver et al. (2018), *A testbed for trusted telecommunications systems in a safety critical environment,* International Conference on Computer Safety, Reliability, and Security, pp. 87-98.
[b-Traffic-forecast]	C. Zhang, P. Patras (2018), *Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks*, in Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Los Angeles, USA.
[b-Traffic-predict]	D. Tikunov and T. Nishimura (2007), *Traffic prediction for mobile network using Holt-Winters exponential smoothing,* In Proc. SoftCOM. Split, Croatia.
[b-Trust]	GlobalPlatform, Inc. (2018), *Root of Trust Definitions and Requirements v1.1,* Public Release GPD_REQ_025.
[b-Trusted Cloud]	F.A.M. Ibrahim and E.E.Hemayed (2019), *Trusted cloud computing architectures for infrastructure as a service: Survey and systematic literature review,* Computers & Security, vol. 82, pp. 196-22.

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, 2019