Network control with AI/ML
– Standardization in ITU and related research–

Ved P. Kafle

National Institute of Information and Communications Technology
Tokyo, Japan

A part of this work includes an outcome of the project entitled “Research and development for **innovative AI network integrated infrastructure technologies (JPMI00316)**” supported by the Ministry of Internal Affairs and Communications, Japan.

Outline

- Overview
- Activities at ITU-T FG ML5G
- ITU-T Recommendations developed in SG13
- Related AI NW research work – an example
- Related activities in other SDOs
- Conclusion

NOTE -
FG ML5G: Focus Group on Machine Learning for future networks including 5G
ITU-T SG13: Study Group 13 is responsible for standardization of network architecture
AI/ML: Artificial Intelligence/Machine Learning
AI NW: AI-based network
SDOs: Standards development organizations

Overview

ITU activities on AI/ML based 5G/IMT2020 network control and management

Note: IMT-2020 is ITU's terminology for 5G network
Sup. = Supplement is like an ITU technical report
Rec. = Recommendation is ITU standard document.

FG ML5G activities overview

https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx

- Work scope

  Study of architecture, interfaces, use cases, protocols, algorithms, data formats, interoperability, performance, evaluation, security

  Three working groups (WGs):

  WG1: Use cases, services and requirements

  WG2: Date formats and ML technologies

  WG3: ML-aware network architecture

- Meetings and workshops
  - 1st : January 2018, Geneva
  - 2nd : April 2018, Xian, China
  - 3rd : August 2018, San Jose, USA
  - 4th : November 2018, Tokyo
  - …
  - 8th : March 2020, online
  - 9th (Final): June 2020, online

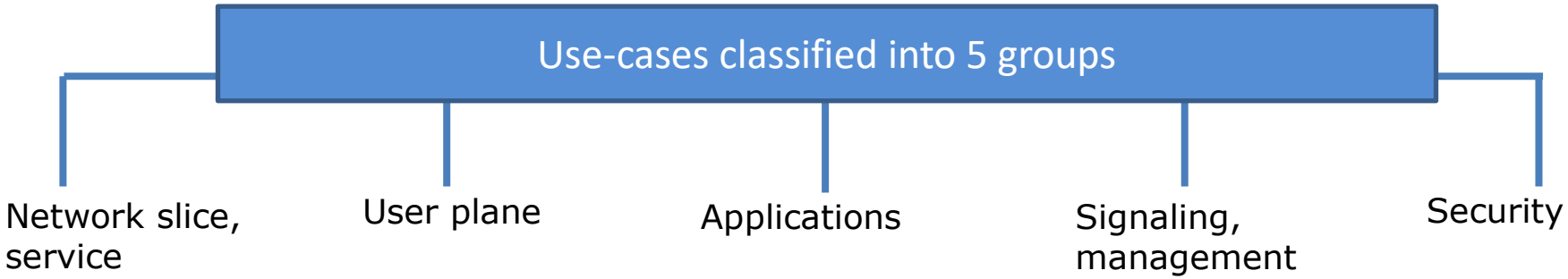
FG ML5G/ITU-T SG13 documents list

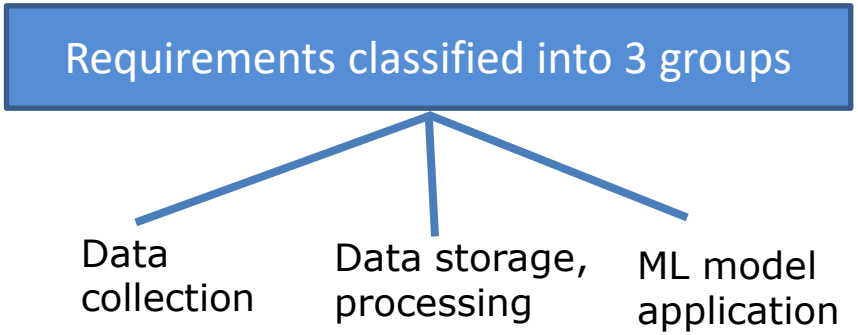
1. Y.sup55: "Machine learning in future networks including IMT-2020: use cases" (Oct 2019)
2. Y.3172: "Architectural framework for machine learning in future networks including IMT-2020" (Jan 2020)
3. Y.3173: "Framework for evaluating intelligence levels of future networks including IMT-2020" (Feb 2020)
4. Y.3174: "Framework for data handling to enable machine learning in future networks including IMT-2020" (Feb 2020)
5. Y.3175: "Functional architecture of machine learning based quality of service assurance for the IMT-2020 network" (Apr 2020)
6. Y.3176: "ML marketplace integration in future networks including IMT-2020" (Sep 2020)

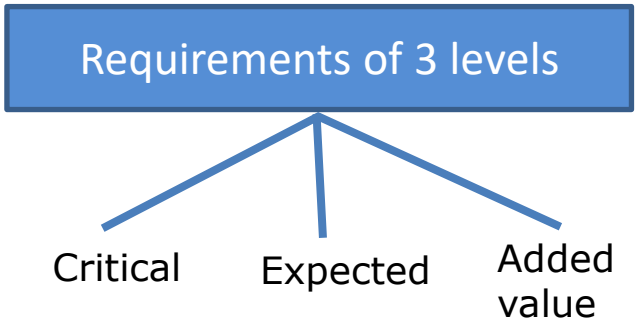
7. Y.ML-IMT2020-RAFR (Resource adaptation and failure recovery),
8. Y.ML-IMT2020-serv-prov (Network service provisioning)
9. …

ITU-T Y.sup55: Machine learning in future networks including IMT-2020: use cases

- More than 30 use-cases and their requirements compiled







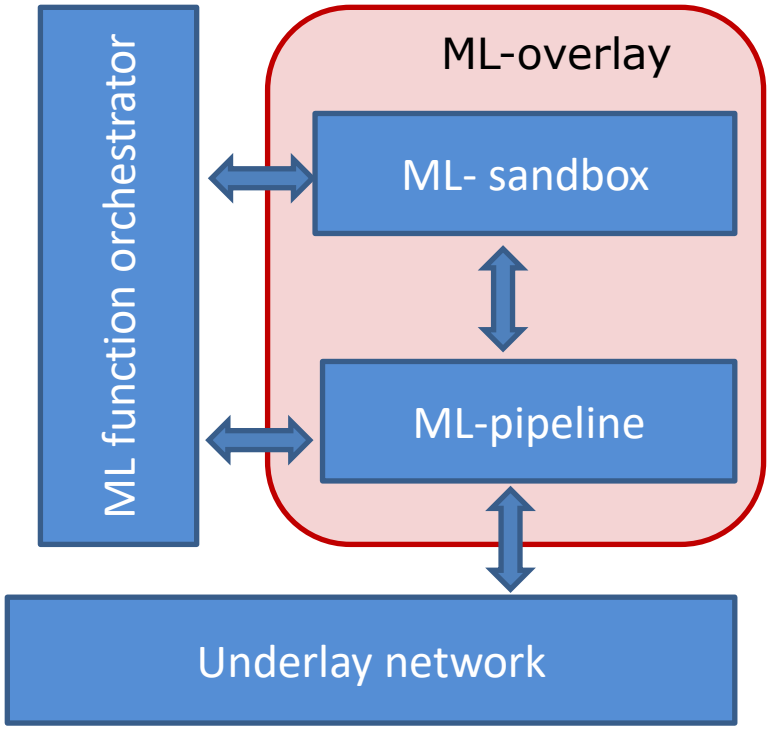
Use cases – some examples

Use-case titles	Description	Requirement examples
Radio resource management for network slicing	Providing performance guarantee with high reliability, while ensuring efficient utilization of radio resources	Support the continuous collection of data, analysis of network slice behaviour and resource utilization patterns
End-to-end network service design automation	Automatically translating service requirements of application services to network parameters/ requirements	Support data models to specify service requirements, integrate automated network configuration methods
End-to-end fault detection and recovery	Predictive detection and root cause analysis, and automated recovery decision making	Support collection of performance data on real-time basis, generation of training data using testing environments

ITU-T Rec. Y.3172: Architectural framework for machine learning in future networks including IMT-2020 (1/4)

Scope

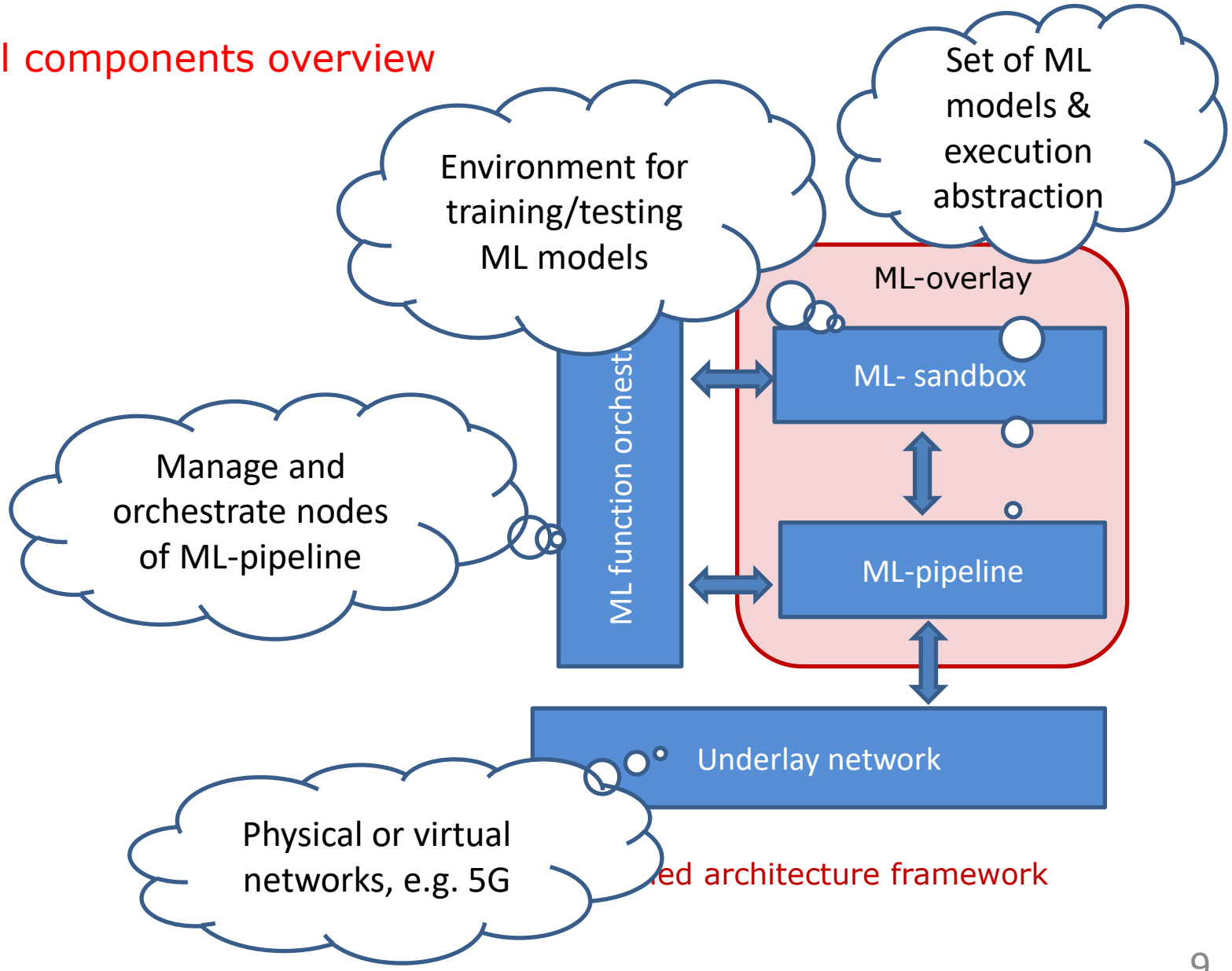
- Listing requirements
- Specifying architectural components needed to satisfy the requirements
- Specifying architectural framework with the integration of components
- Providing guidelines for applying architectural framework in networks



Simplified architecture framework

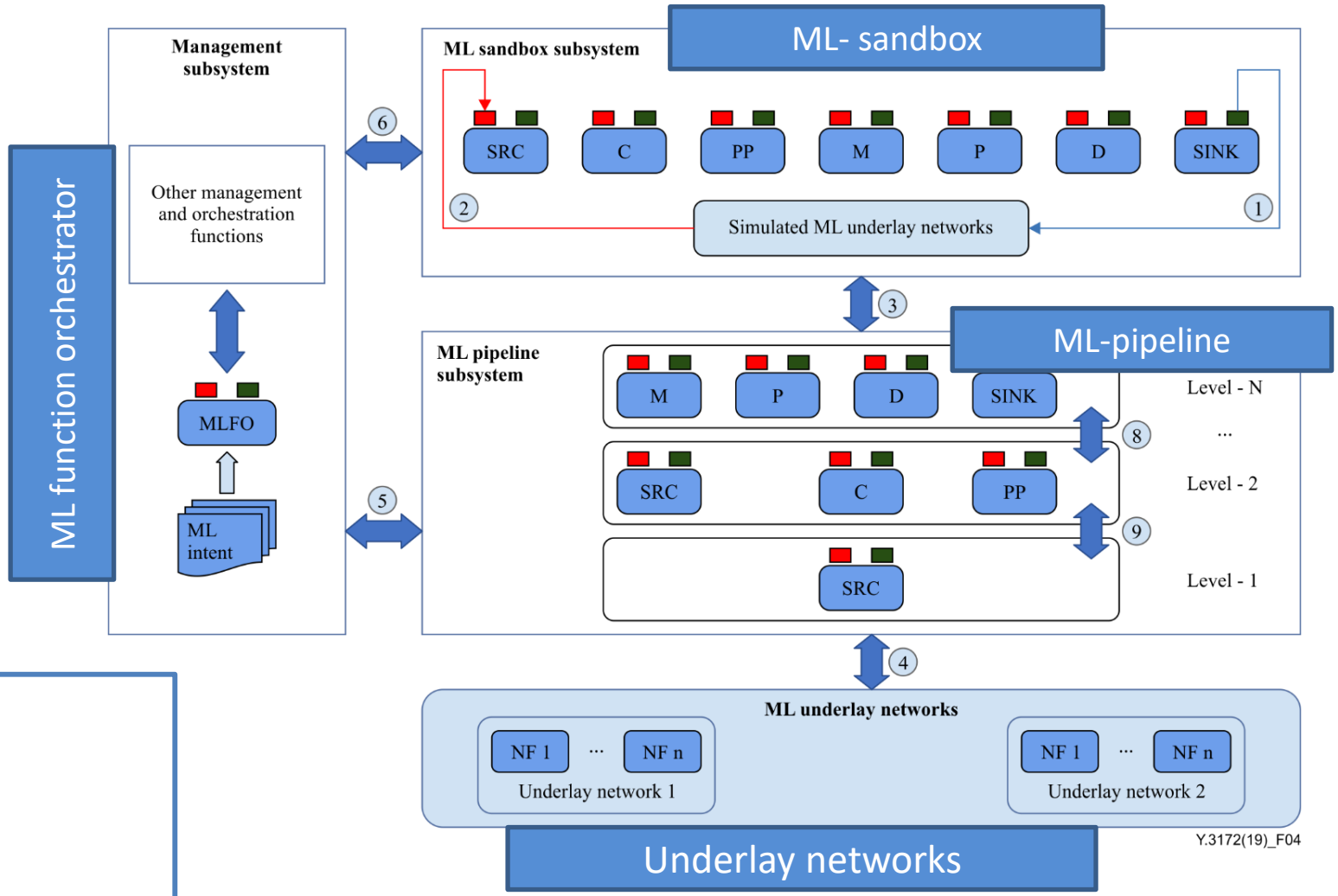
ITU-T Rec. Y.3172: Architectural framework for machine learning in future networks including IMT-2020 (2/4)

Functional components overview



ITU-T Y.3172: Architectural framework for machine learning in future networks including IMT-2020 (3/4)

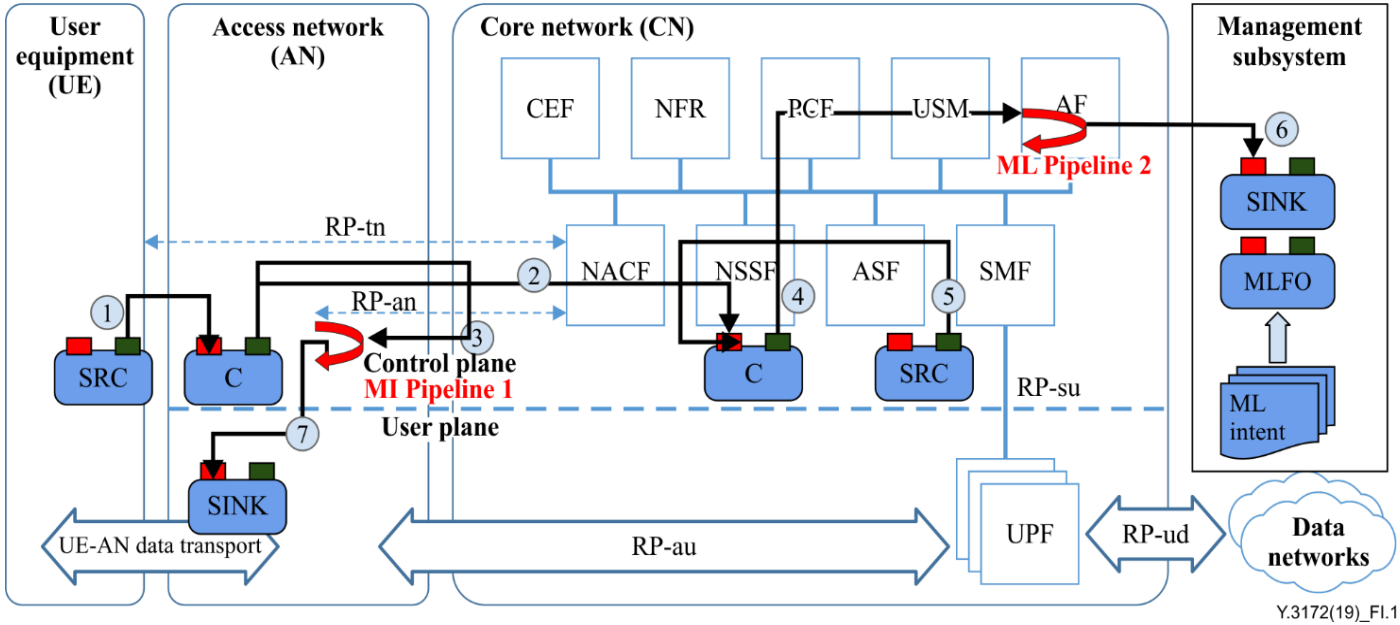
High-level architecture framework & components



SRC: Source
C: Collector
PP: Preprocessor
M: Model
P: Policy
D: Distributor
NF: Network function

ITU-T Y.3172: Architectural framework for machine learning in future networks including IMT-2020 (4/4)

Example of the high-level architecture realization in an IMT-2020 network



1. Collect location information from UEs
2. Collect channel measurement from AN
3. Analyze to make intelligent scheduling decisions and execute through 7.

4. Collect DL packet information from GW
5. Collect AN information
6. Analyze to make intelligent QoS configurations

ITU-T Y.3173: Framework for evaluating intelligence levels of future networks including IMT-2020 (1/4)

Scope

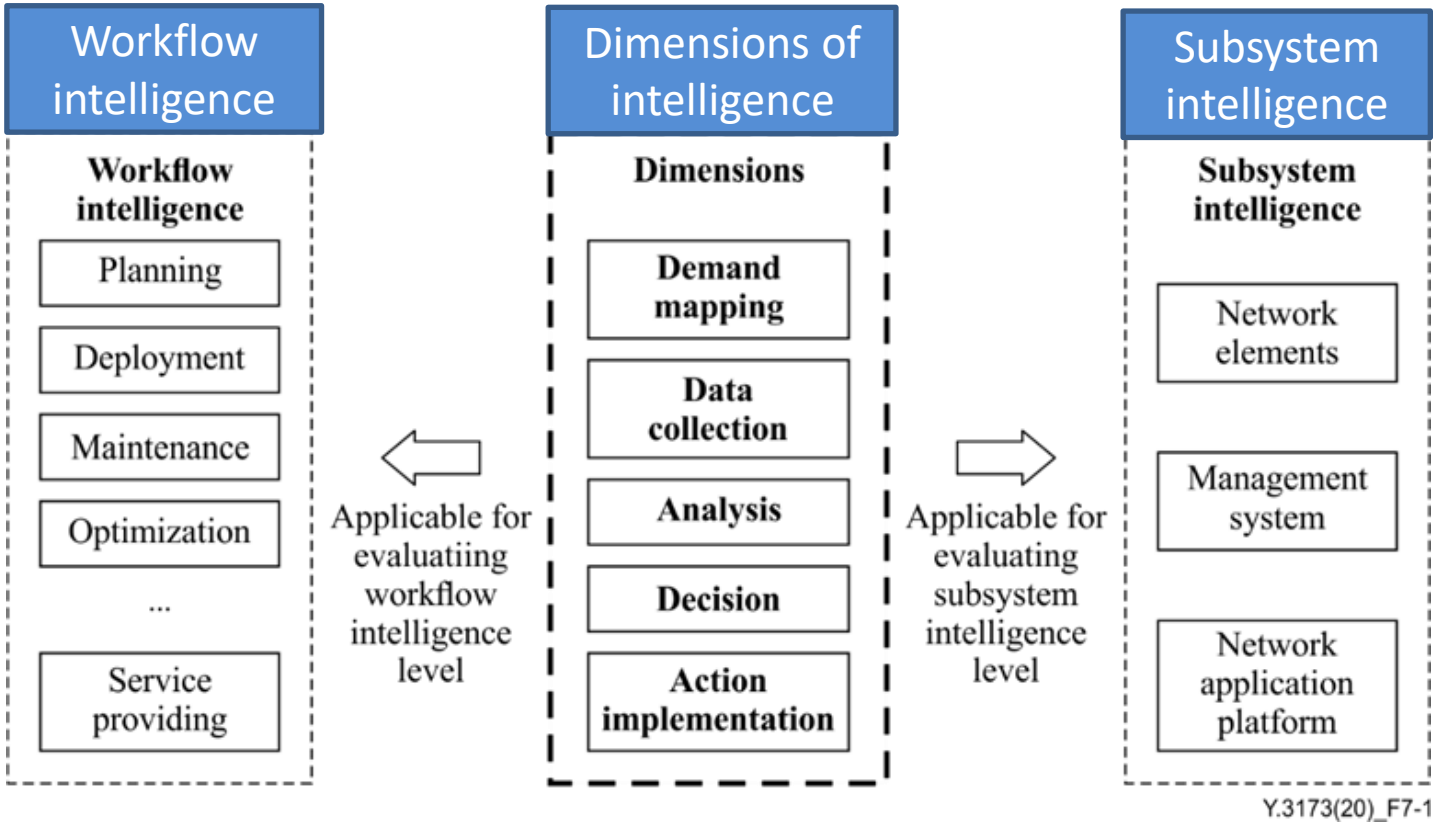
- Development trend for evaluating network intelligence levels
- Evaluation methods
- Architectural view

Importance of evaluation framework

- To provide an evaluation basis
- To help industry to reach a consensual and unify understanding, formulate relevant strategies and development plans
- To provide decision mechanisms to operators, equipment vendors and other network industry participants for planning of network technology features and products' roadmaps

ITU-T Y.3173: Framework for evaluating intelligence levels of future networks including IMT-2020 (2/4)

Dimensions for evaluating network intelligence levels



NOTE: SAE (Society of automotive engineers) documents on intelligence levels evaluation methods are taken as reference.

ITU-T Y.3173: Framework for evaluating intelligence levels of future networks including IMT-2020 (3/4)

Actors for classifying dimensions of network intelligence capability levels

Actors in network intelligence capability level	Roles
Human	Rules definition, decision and action implementation all carried out by human
Human and system	Rules definition by human, decision and actions implementation carried out by system automatically
System	Rules definition, decision and action implementation all carried out automatically by system

ITU-T Y.3173: Framework for evaluating intelligence levels of future networks including IMT-2020 (4/4)

Network intelligence level chart

Network intelligence levels		Dimensions of intelligence				
		Action implementation	**Data collection**	**Analysis**	**Decision**	**Demand mapping**
L0	Manual operation	Human	Human	Human	Human	Human
L1	Assisted operation	Human and System	Human & System	Human	Human	Human
L2	Preliminary intelligence	System	Human & System	Human & System	Human	Human
L3	Intermediate intelligence	System	System	Human & System	Human & System	Human
L4	Advanced intelligence	System	System	System	System	Human & System
L5	Full intelligence	System	System	System	System	System

ITU-T Y.3174: Framework for data handling to enable machine learning in future networks including IMT-2020 (1/2)

Scope:

- High-level requirements of data handling and data models
- Framework for data handling
- Guidelines and example usage

Challenges addressed:

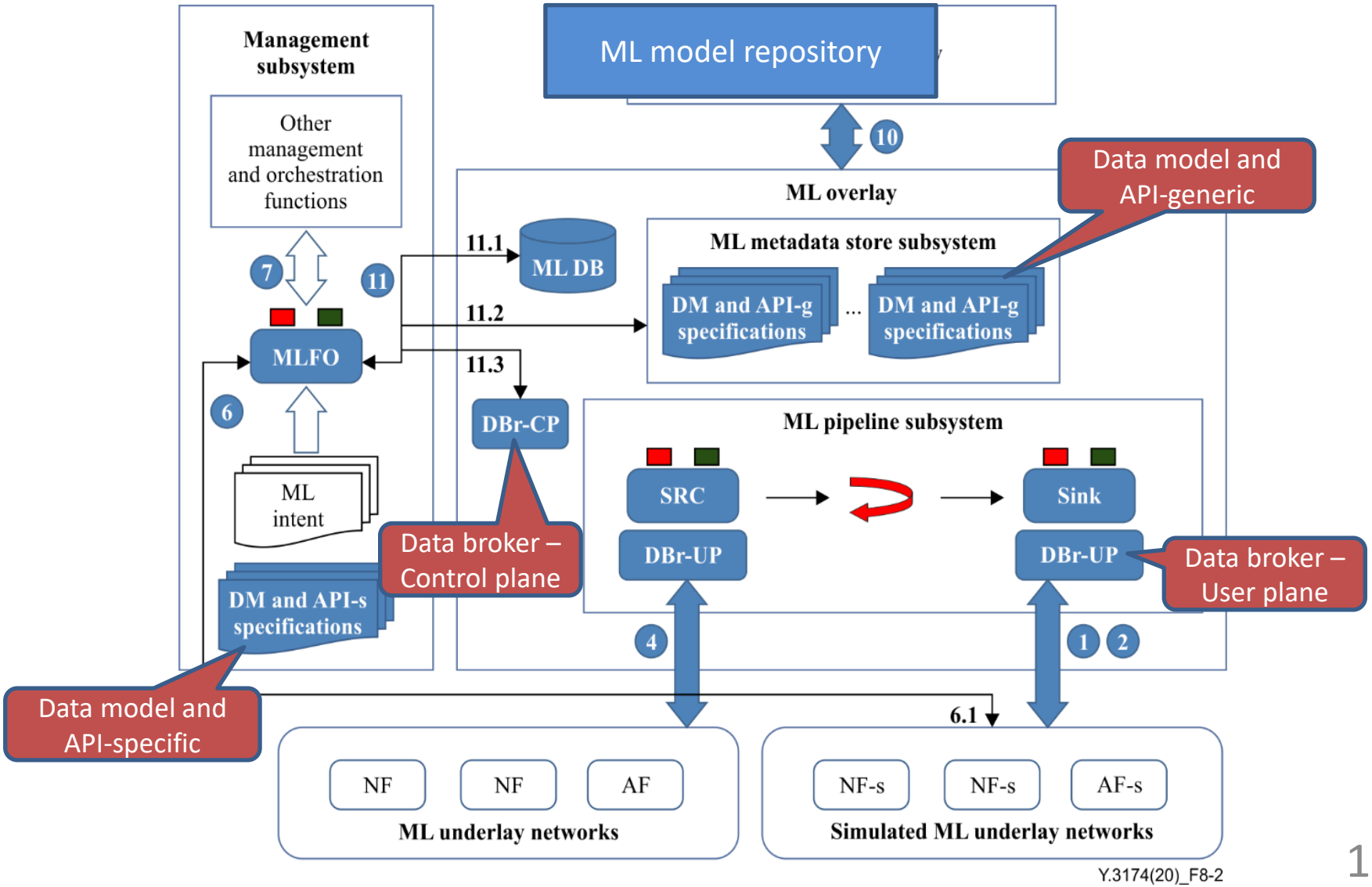
- Diversity in data produced by various components
  - Increased flexibility and agility leading to complicate configuration, dynamically evolving sources of data and applicable network configuration parameters and policy

Requirements listed:

- 56 requirements captured, analyzed and classified into 3 groups
  - ML input data collection
  - ML processing
  - ML data output

ITU-T Y.3174: Framework for data handling to enable machine learning in future networks including IMT-2020 (2/2)

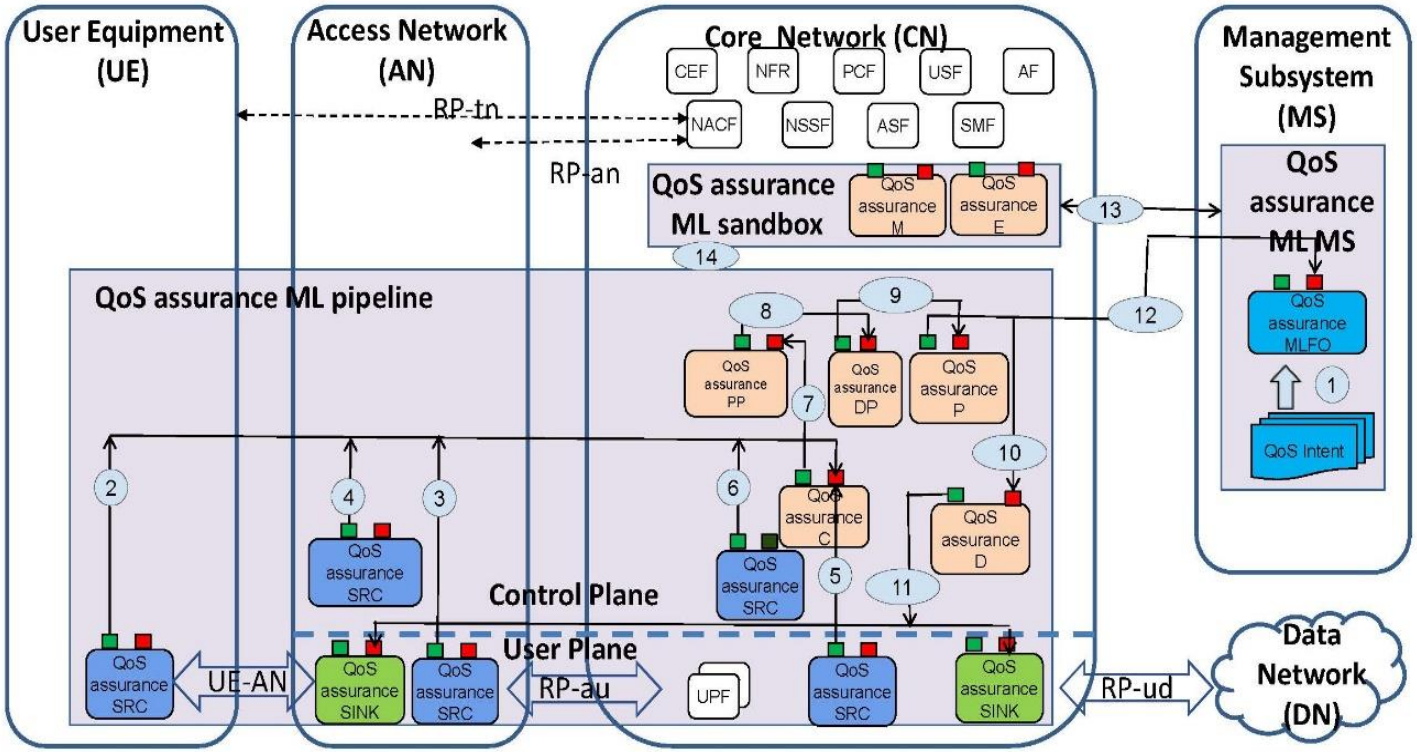
High-level architecture of data handling



ITU-T Y.3175: Functional architecture of ML based QoS assurance

Scope:

- Architecture framework
- Procedures



- QoS assurance Service for egress
- QoS assruance Service for ingress

SRC: QoS source of data
C: QoS data collector

PP: QoS data pre-processor
DP: QoS anomaly detection and prediction
SINK: QoS assruance target of ML output

P: QoS policy
D: QoS policy distributor
M: QoS assurance ML modelling
E: QoS assurance model evaluating

ITU-T Y.3176: ML marketplace integration in networks (1/2)

*** ML marketplace**: a repository of ML, interoperable AI models

Scope:

- Challenges and motivations
- High level requirements
- Architecture and interfaces

Challenges to address:

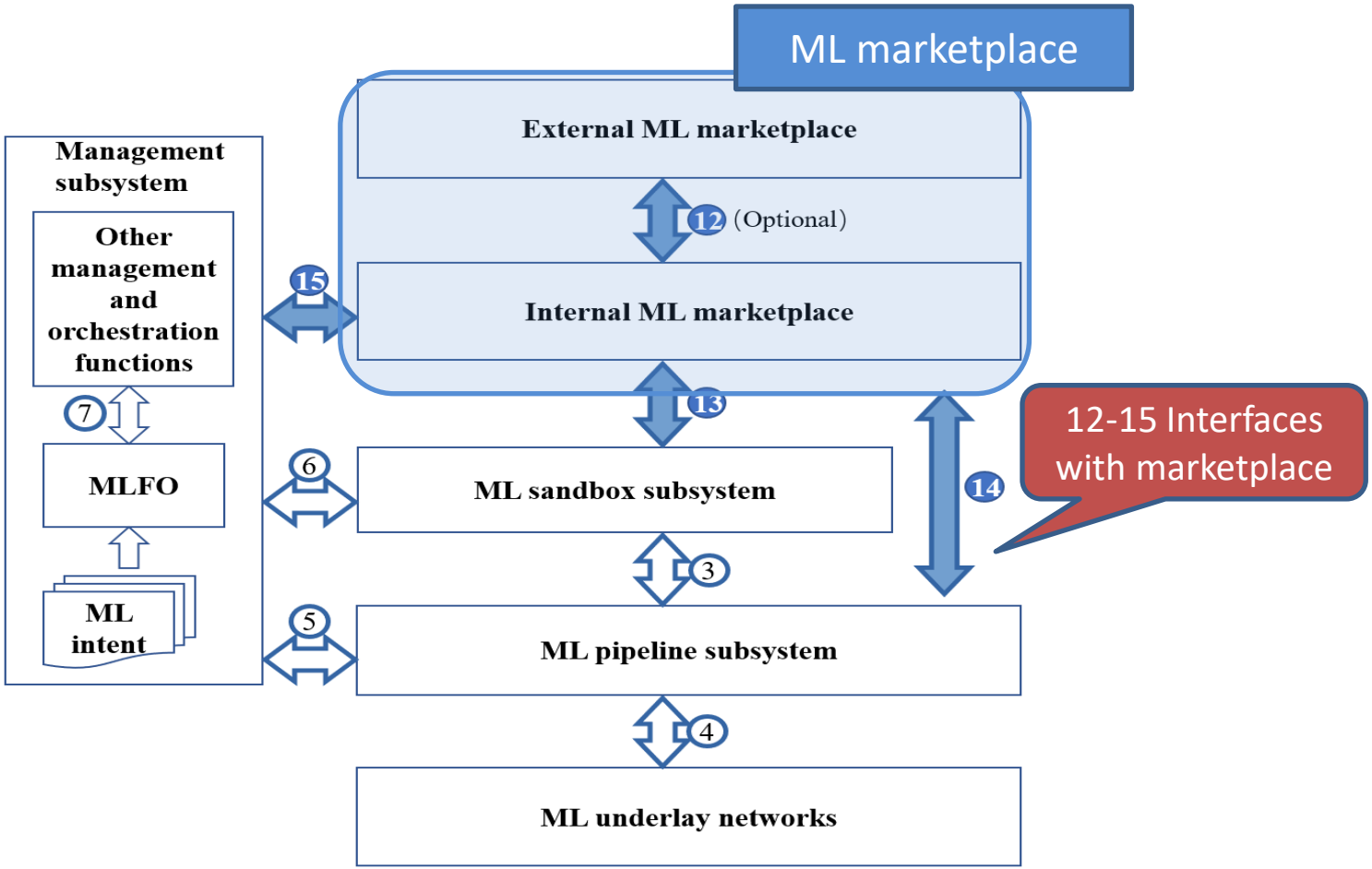
- Need of interoperable mechanisms for ML model identification, selection, chaining, testing and deployment from various ML marketplaces into the operators' networks.

Approach used:

- ML Intent and MLFO used to select ML models from marketplace
- Standard metadata used to interface between MLFO and ML marketplace
- Interfaces to push ML models from ML marketplace to ML-sandbox/ML-pipeline

ITU-T Y.7176: ML marketplace integration in networks (2/2)

Architecture for ML marketplace integration in network

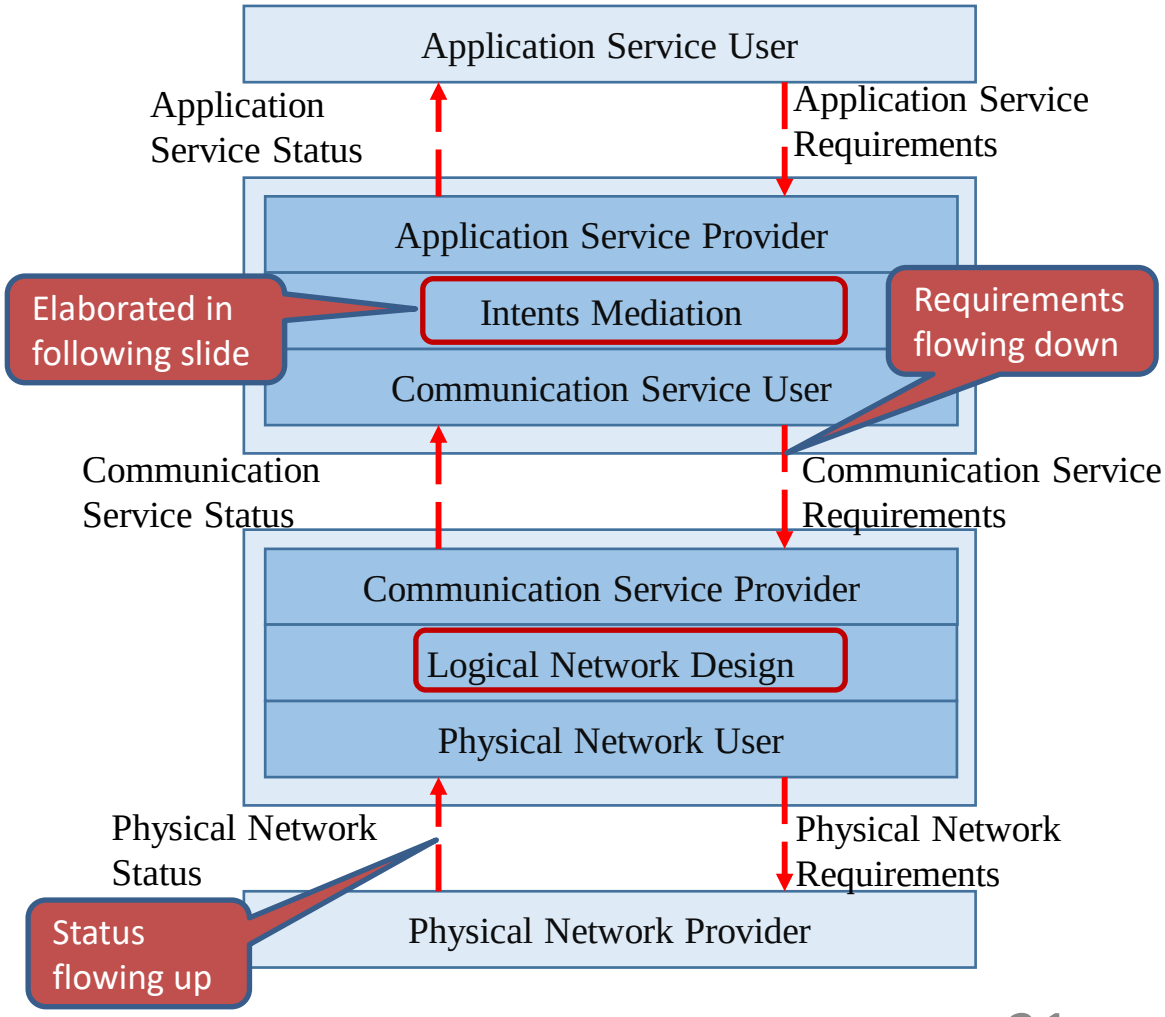


ITU-T Y.ML-IMT2020-serv-prov draft: Architecture framework of user-oriented network service provisioning (1/2)

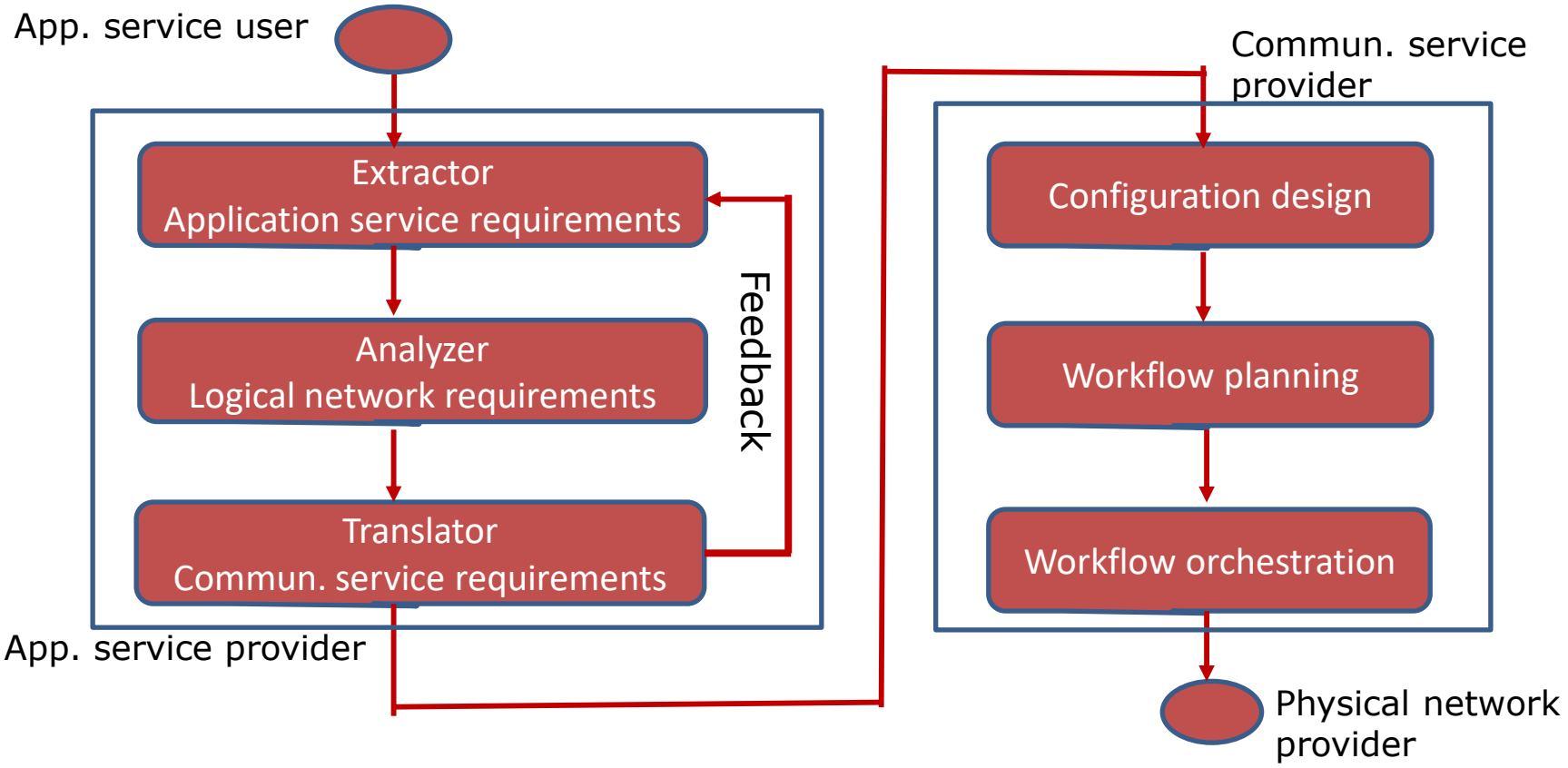
Scope:

- Architecture framework of user-oriented network service provisioning with AI-based automatic generation of
  - AI-based framework for different layers of provider/user interaction
  - Network requirements
  - Configuration and workflow

Framework for network service provider/user interaction



ITU-T Y.ML-IMT2020-serv-prov draft: Architecture framework of user-oriented network service provisioning (2/2)



Workflow for communication service requirements generation

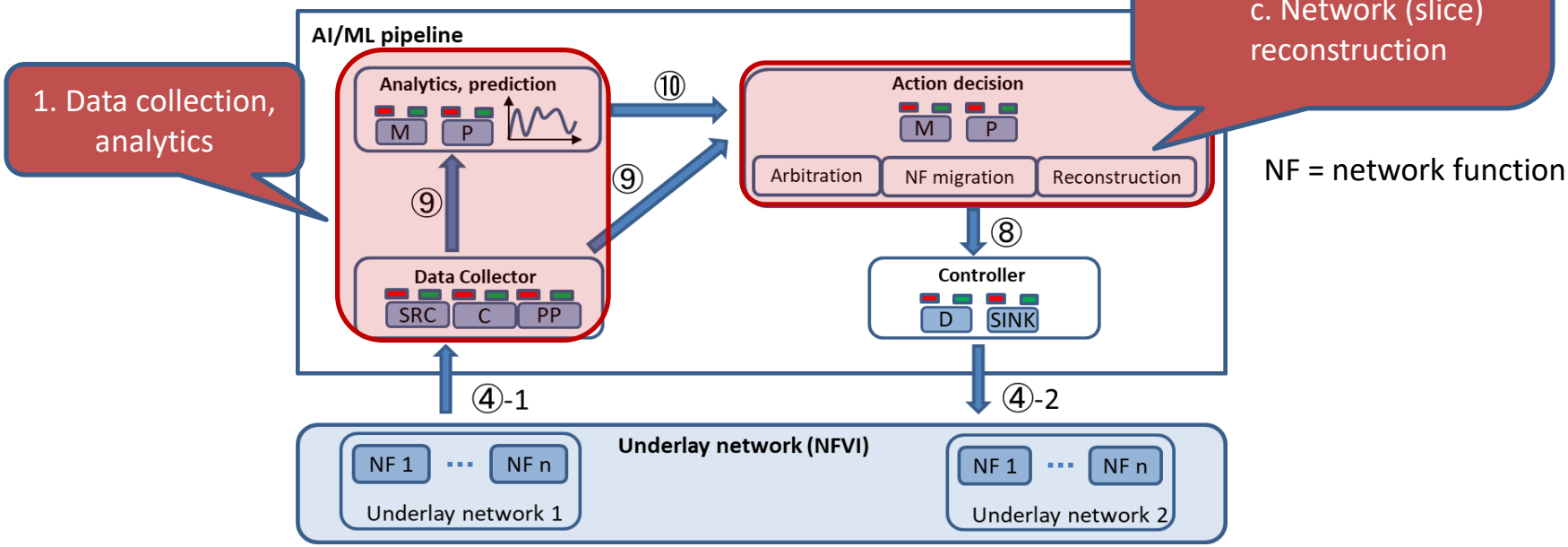
Workflow for logical network design and deployment

ITU-T Y.ML-IMT2020-RAFR draft: Architecture framework for resource and fault management (1/2)

Scope:

- Architecture framework of AI-based network automation for
  1. Resource management, and
  2. Fault management

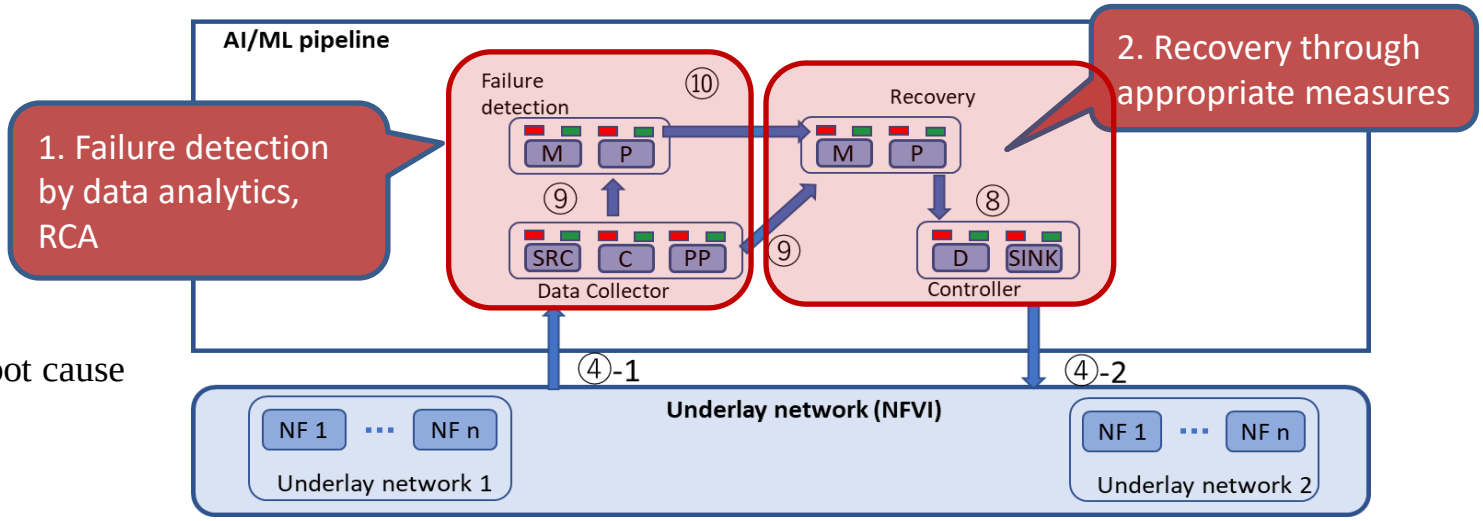
AI-based resource management (adaptation) framework



NF = network function

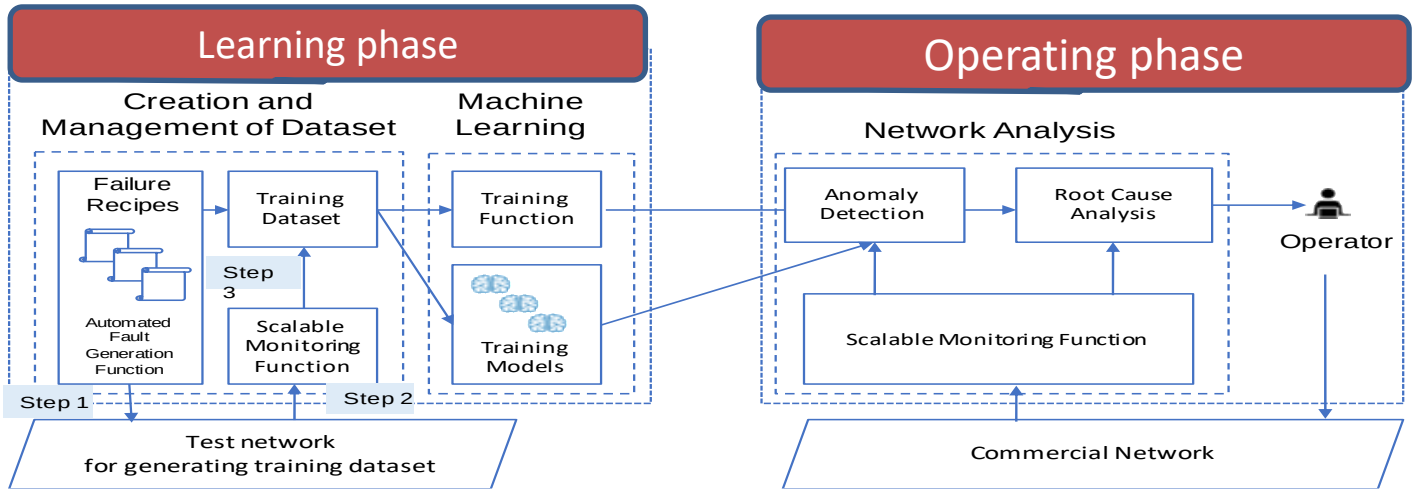
ITU-T Y.ML-IMT2020-RAFR draft: Architecture framework for resource and fault management (2/2)

AI-based fault management (failure detection and recovery) framework



RCA: Root cause analysis

ML model training, RCA, and action execution

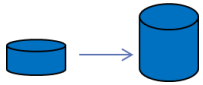


Y.ML-IMT2020-RAFR
Resource management related research overview

Resource adjustment approaches (an example of computational resources)

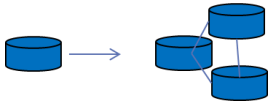
- Computational resources used in networks are **finite**
- Need to automatically control and manage the computational resources that operate the virtual network functions (VNFs).
  - Requires a control mechanism that **adjusts the amount of computational resources dynamically and quickly** for keeping up with utilization variations and fault conditions

● Vertical Scaling



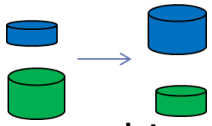
Dynamic increase/decrease in allocated CPU resource for VM
(Scale up & scale down)

● Horizontal Scaling



Dynamic adding/removing of allocated VMs for service
(Scale out & scale in)

● Internetwork Scaling



resource arbitration and function migration

- Inter-service computing-resource arbitration at node
- Inter-node VNF migration & SF chain reconstruction

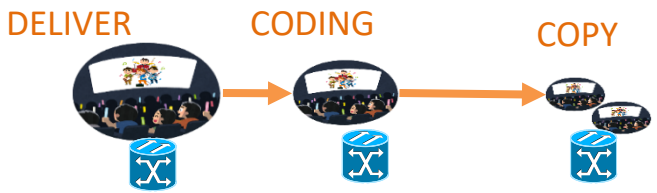
Cf) V. P. Kafle et al., *IEEE Commun. Stds. Magazine*, Dec 2018.

Motivation for ML usage in network service control

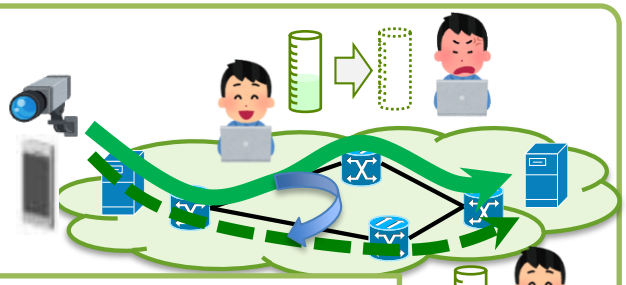
Speeding up processing by minimizing number of function-chain reconstructions for target services

Prerequisite knowledge

SFC use case
(e.g. video distribution service)



Resource arbitration & SF migration



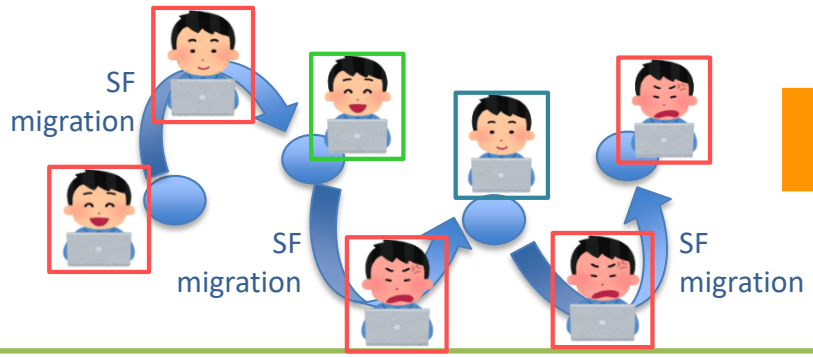
Example of SFC reconstruction

Conventional

For problems humans cannot solve w/ experience
- Sequential optimal solutions

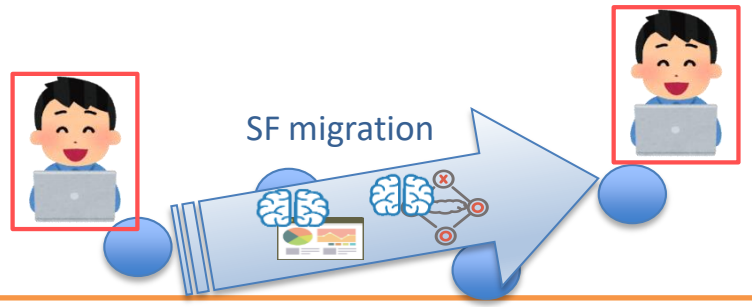
⇒ Larger number of reconstructions
- Optimal solution for the future

⇒ Requires hours to days



This work

ML (EDRNN) simultaneously achieves
- Minimizes number of reconstructions based on time series analysis
- Determines locations to migrate service functions in seconds

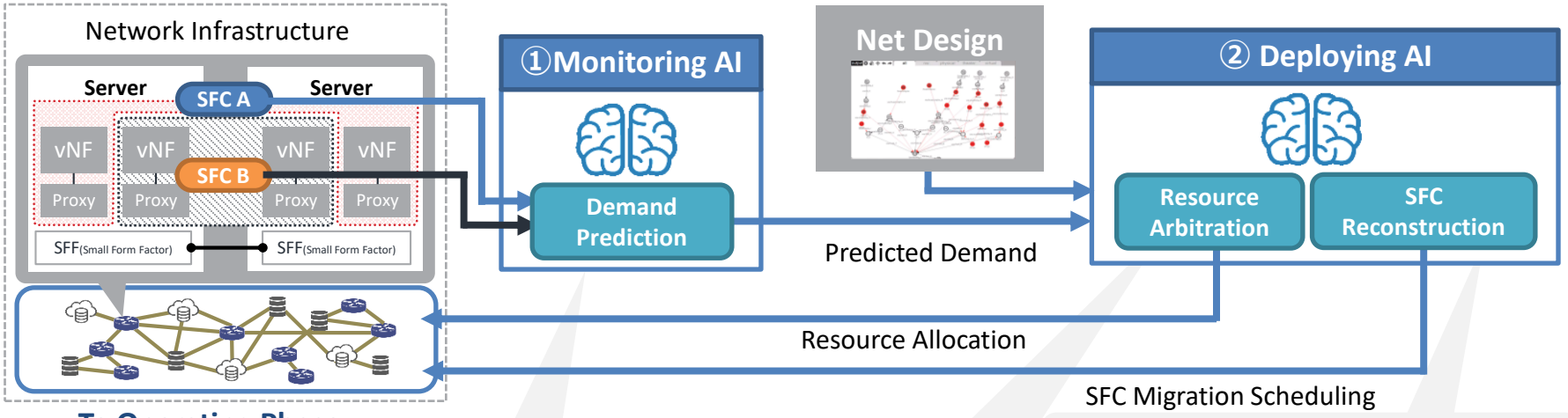


SF = Service Function

Dynamic adjustment of computing resources

Automatic AI-based control ensures resource arbitration for high-priority in seconds, VNF migration time minimization, and quality of service sustention

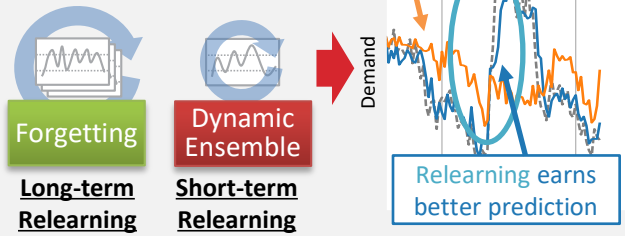
QoS × Effective utilization of resources × Agility in control — Simultaneous achievement！



To Operating Phase

Demand prediction

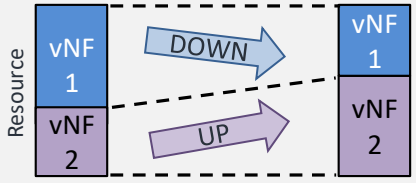
Relearning Framework



Cf) T. Hirayama *et al.*, IEEE NetSoft 2020.

Autonomic resource arbitration among VNFs

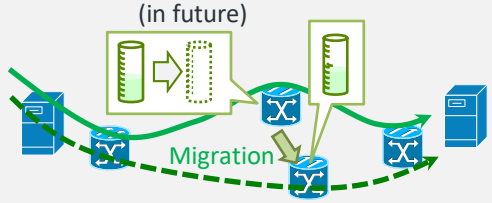
(Virtualized Network Functions) **within a server**



Cf) T. Miyazawa *et al.*, IEEE/IFIP NOMS 2018.

Autonomic migration scheduling

Decide which & where VNF should be migrated to avoid resource shortage

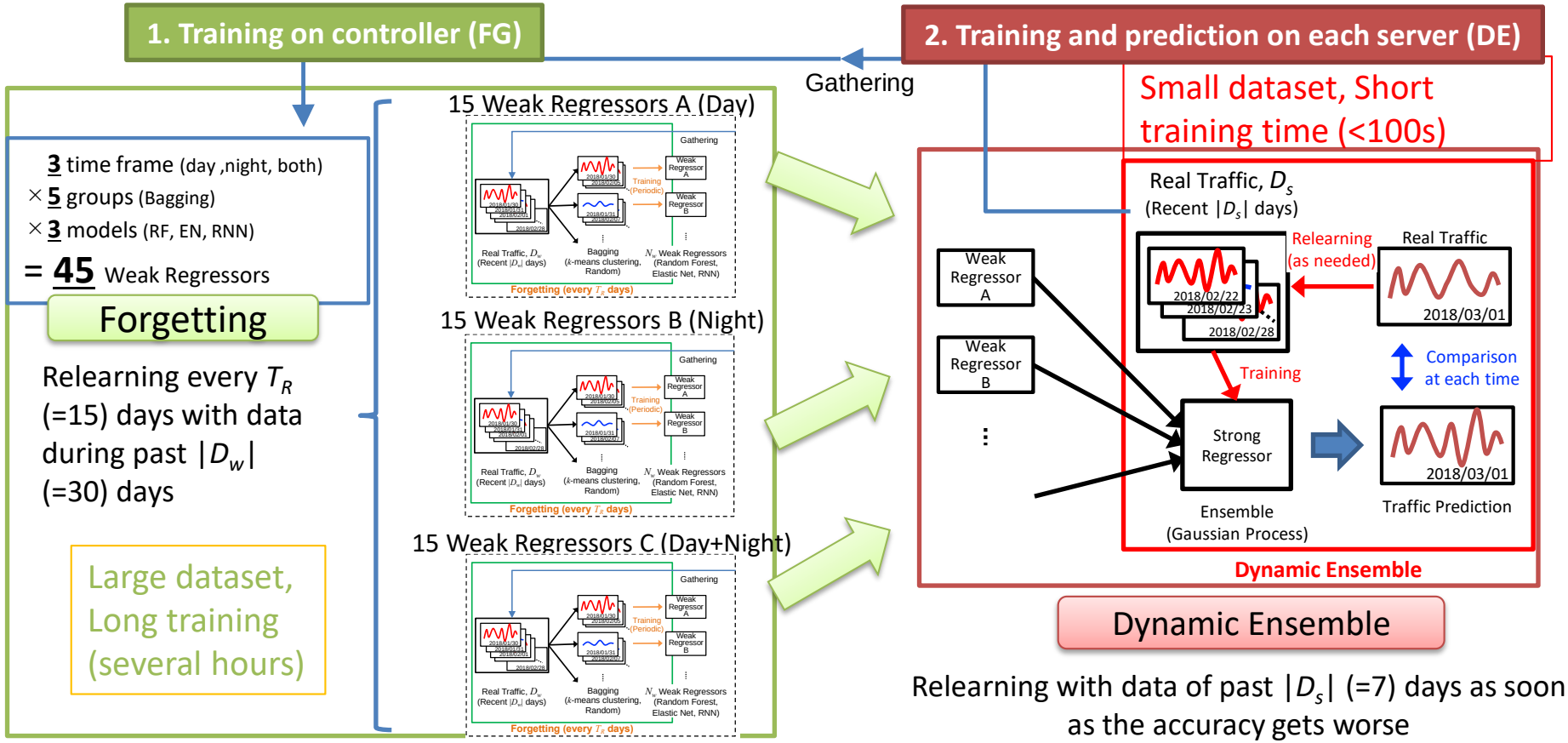


Encoder-Decoder Recurrent Neural Network

Cf) T. Hirayama *et al.*, IEEE NetSoft 2019.

Ensemble learning architecture and relearning mechanism

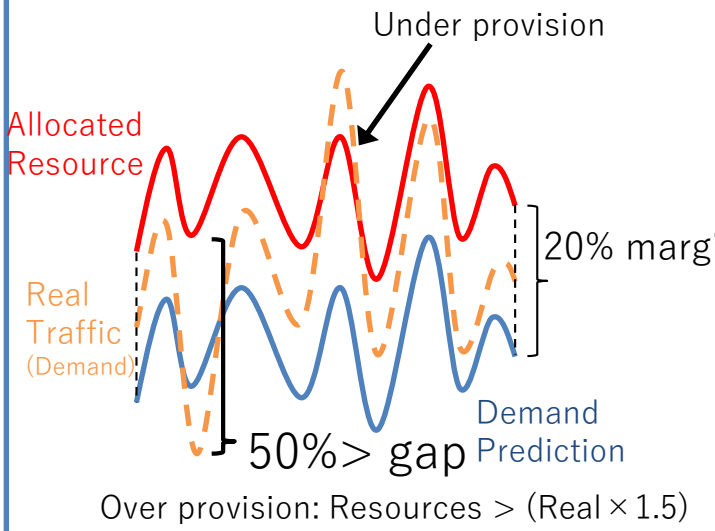
- Prediction framework based on ensemble learning with weak and strong regressors
  - **Weak regressors** trained with traffic dataset from the **whole of network**
    - **Forgetting (FG):** Periodical (week, month) retraining to follow **long-term** trend changes
  - **Strong regressors** trained with traffic dataset gathered on **each server**
    - **Dynamic Ensemble (DE):** Retrain as soon as accuracy declines to **short-term** trend change adoption
- Regressors are trained with the wider-variant data (3-time frames in an example)



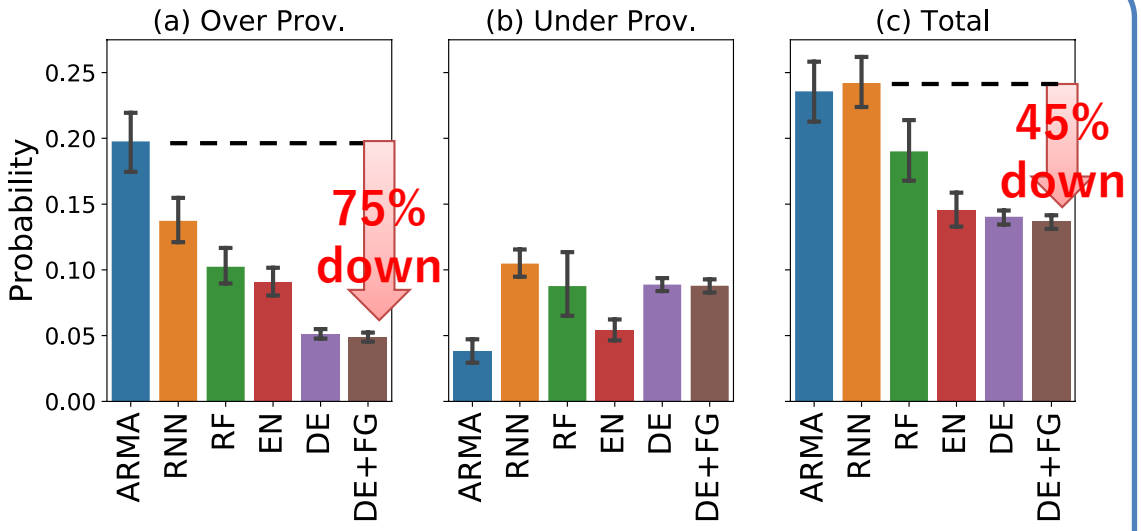
Cf) T. Hirayama *et al.*, IEEE NetSoft 2020.

Effect of weak-strong regression
Use case of CPU allocation to a VNF

- Regressors are trained with the wider-variant data (3-time frames)
- Allocated CPU resources of 120% of predicted value (i.e. margin of 20%)
- DE+FG framework reduced frequencies of over- and under-provisioning more than 45% in comparison to RNNs and ARMA



Allocation rule and definition of errors



Probability of over/under provisioning (and total)

RNN: Recurrent Neural Net
RF: Random Forest, EN: Elastic Net
DE: Dynamic Ensemble, FG: Forgetting

Cf) T. Hirayama *et al.*, IEEE NetSoft 2020.

ITU Liaisons on AI/ML networks

ETSI

Industry Specification Group (**ISG**)

- ZSM (Zero-touch network & service management)
- ENI (Experiential Networked Intelligence)

Linux Foundation

Acumos AI

ONAP
OPEN NETWORK AUTOMATION PLATFORM

O-RAN
ALLIANCE

ISO/IEC JTC 1/SC 42
Artificial intelligence

3GPP
A GLOBAL INITIATIVE

IRTF

NMRG

Conclusion

- ITU-T standardizing mainly requirements, frameworks, architectures of AI/ML supported network control and management.
- Architectural functional details and interface specifications are developed in other SDOs: ETSI, 3GPP, IETF, Linux Foundation, forums and open-source software communities.
- AI/ML supported network research and development work progressing rapidly.

Acknowledging the input of Takahiro Hirayama, Takaya Miyazawa, Masahiro Jibiki, and Hiroaki Harai in this presentation.