AI/ML Standardization Status in 3GPP R18

Presenter: *Tricci So* **(OPPO)**

tricci.so@oppo.com

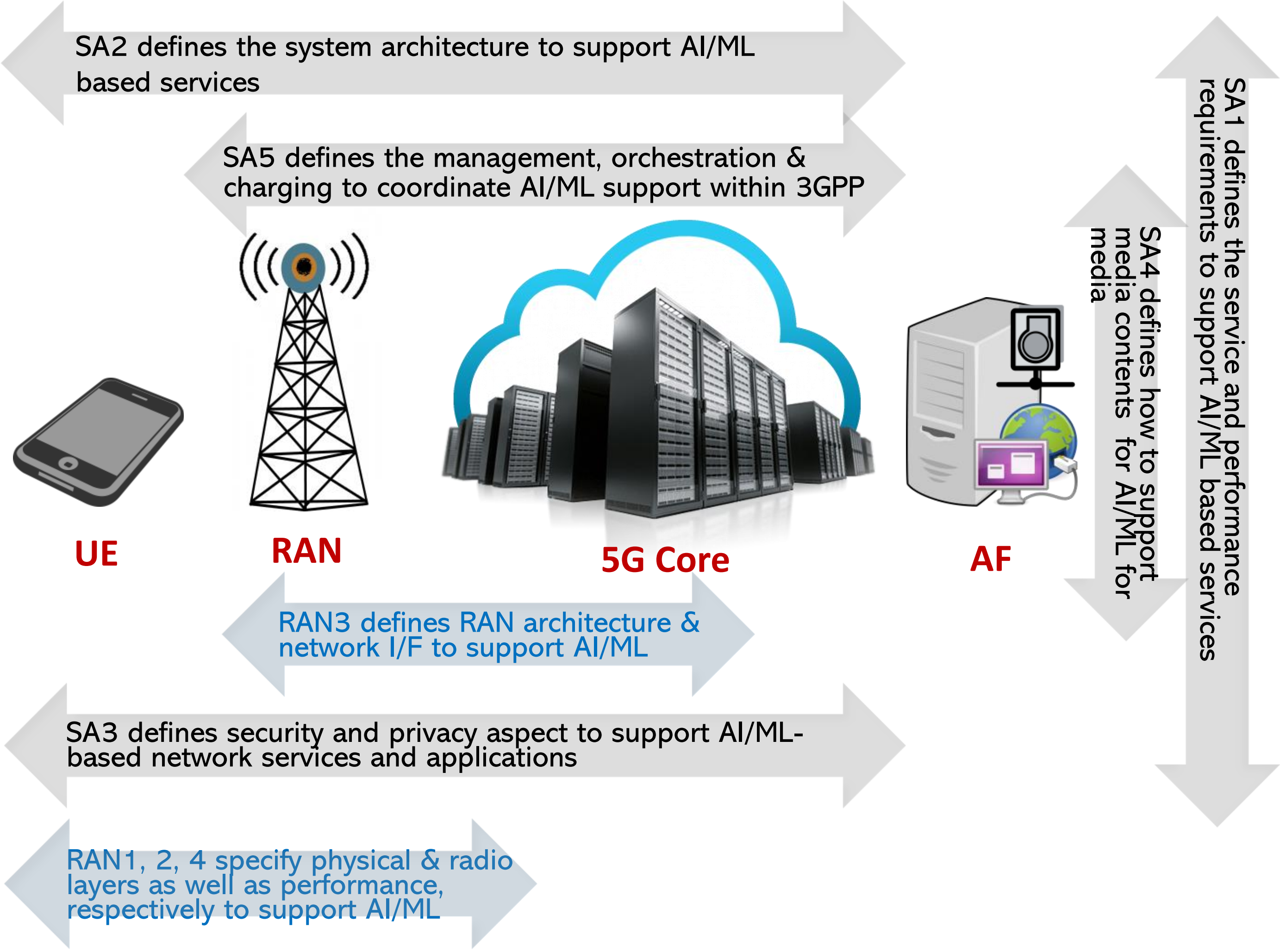
OPPO standardization research department

AGENDA

General	3GPP Working Groups for AI/ML Support Development
SA1	SA1 Services & Performance Definitions & Requirements
SA2	SA2 Architecture Enhancement for Application AI/ML Operation (AIMLsys)
SA2	SA2 Architecture Enhancement for Federated Learning Application AI/ML Operation (AIMLsys) – HFL
SA2	SA2 System Architecture – NWDAF Network Operation AI/ML (eNA)
SA2	SA2 System Architecture – NWDAF Support for Horizontal Federated Learning (eNA)
SA3	SA3 Security & Privacy Support for Network Analytics (eNA)
SA3	SA3 Security & Privacy support for network AI/ML operation （eNA_Sec）
SA4	SA4 AI/ML for Multi-media
SA4	SA4 Supporting UE Data Collection, Reporting & Exposure
SA5	SA5 AI/ML Management in 5G System
SA5	SA5 AI/ML Management Operation
RAN3	RAN3 AI/ML-enabled NG-RAN
RAN1&2	RAN1&2 AI/ML for Air Interface (pave the way to 6G)
RAN1&2	RAN1&2 AI/ML for Air Interface (LCM Framework)
SIDs/WIDs	3GPP Rel-18 AI/ML Related Study/Work Items
Backup Slides	

About 3GPP Working Groups

- ✓ **SA WG-1 (SA1):** Responsible for identifying service and performance requirements for 3GPP systems, in Rel-18, SA1 focused on defining the AI/ML model transfer in 5G.
- ✓ **SA WG-2 (SA2):** Responsible for developing system architecture, in Rel-18, SA2 worked on 5G system support for intelligent transport for the AI/ML-based services.
- ✓ **SA WG-3 (SA3):** Responsible for security and privacy aspects. For AI/ML, SA3 examined and determined the system security and privacy impacts towards 5G Core when supporting AI/ML-based network services and applications.
- ✓ **SA WG-4 (SA4):** Responsible for defining media codec for the system and delivery aspects of the media contents, inRel-18, SA4 defined the AI/ML for media.
- ✓ **SA WG-5 (SA5):** Responsible for management, orchestration, and charging for 3GPP systems, in Rel-18, SA5 defined AI/ML management to coordinate AI/ML functions across 5G system.
- ✓ **RAN WG-3 (RAN3):** Responsible for the overall RAN architecture and the specification of protocols for the related network interfaces, in Rel-17 and 18, RAN3 defined the initial support for AI/ML for next-generation RAN (NG-RAN).
- ✓ **RAN WG-1, 2, and 4 (RAN1, RAN2, and RNA4):** Responsible for physical layer, radio layer and performance of the radio Interfaces for UE, Evolved UTRAN, NG-RAN, and beyond, respectively, in Rel-18, these WGs define AI/ML for new radio (NR) air interface which is led by RAN1.



SA2 defines the system architecture to support AI/ML based services

SA5 defines the management, orchestration & charging to coordinate AI/ML support within 3GPP

SA1 defines the service and performance requirements to support AI/ML based services

SA4 defines how to support media contents for AI/ML for media

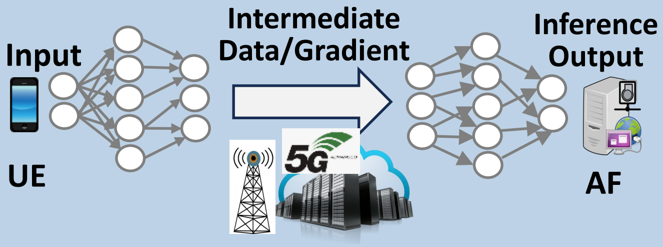
RAN3 defines RAN architecture & network I/F to support AI/ML

SA3 defines security and privacy aspect to support AI/ML-based network services and applications

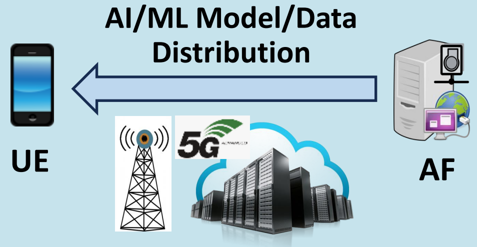
RAN1, 2, 4 specify physical & radio layers as well as performance, respectively to support AI/ML

SA1 Services & Performance Definitions & Requirements

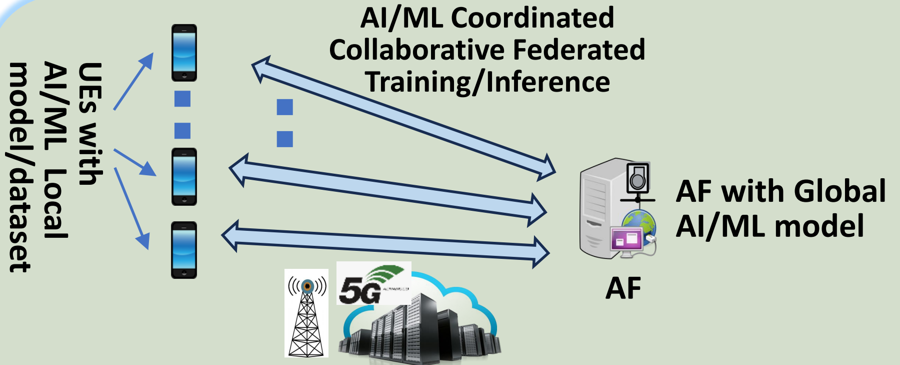
□ Defining 3 AI/ML Model Transfer use cases:



(1) AI/ML Splitting Operation between endpoints



(2) AI/ML model/data distribution over 5G system


UEs with AI/ML Local model/dataset
AI/ML Coordinated Collaborative Federated Training/Inference
AF with Global AI/ML model
AF
5G

(3) AI/ML distributed collaborative Federated Learning over 5G system

□ Defining AI/ML Service Requirements:

✓ Identify the AIML related key requirements to Uu interface, including

- *Candidate member selection for Federated Learning (FL)*
- *Aggregated QoS management for Federated Learning*
- *In-time exposure of Network status, Event alerting (e.g. QoS prediction) to the authorized AIML application*
- *Network resource monitoring for an authorized AIML application*

NOTE: The applicability of the requirements is subject to operator policy, user consent, and regulatory requirements

□ Defining AI/ML Performance Requirements:

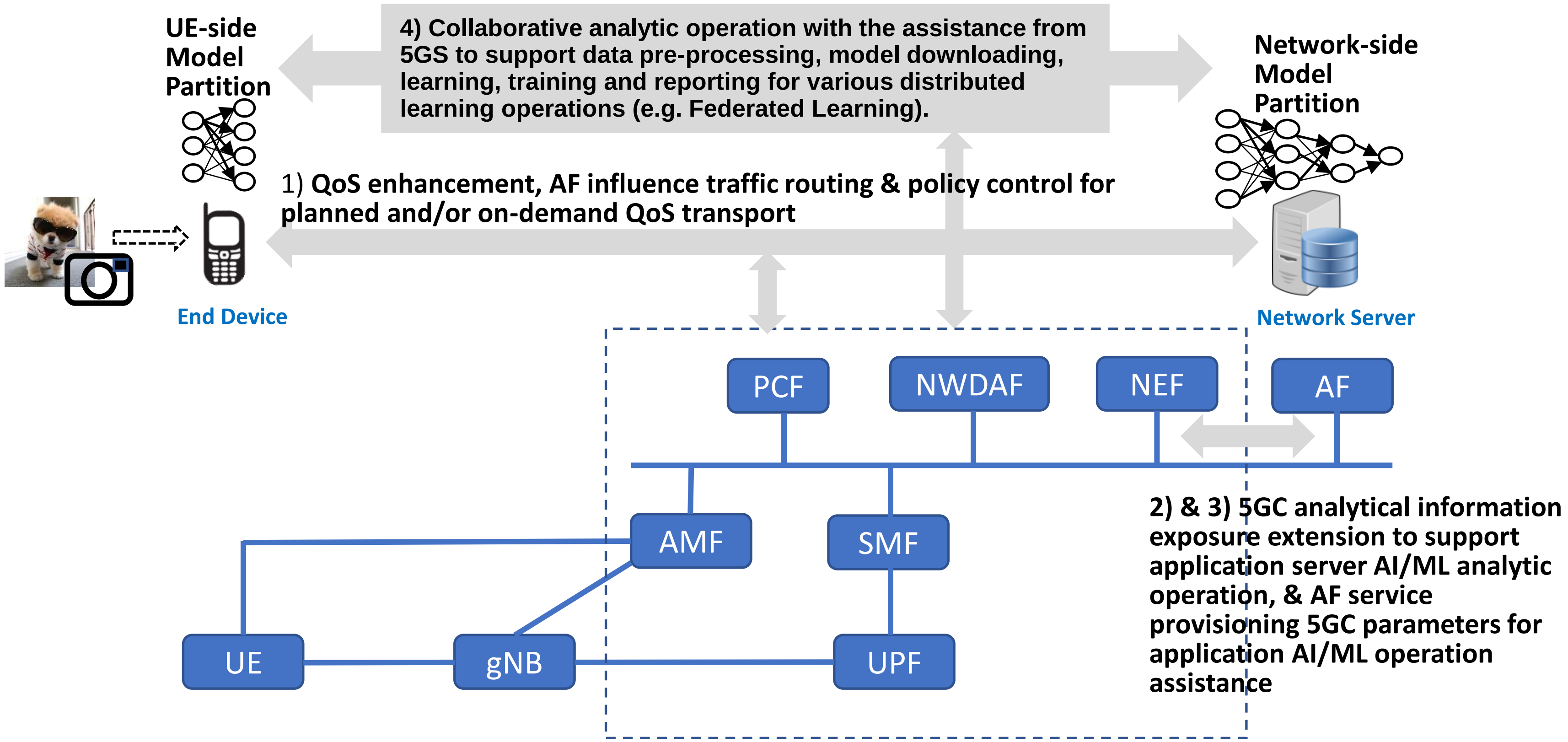
✓ Specify KPIs for AI/ML model transfer in 5G system, including end-to-end latency, experienced data rate, reliability, and communication service availability, among others.

<u>NOTE:</u> 3GPP SA1 Requirements for AI/ML are specified in TS 22.261.

SA2 Architecture Enhancement for Application AI/ML Operation (AIMLsys)

In Rel-18, 5G Core is extended to assist Application AI/ML operation. AF remains to control the logic of the application layer AI/ML operation while 5GC:

1) **Enabling application influence on traffic routing and policy control to provide planned or on demand QoS transport.**
   - **Policy framework is extended to leverage the data analytics of the target AoI capacity and performance for the corresponding UE(s) to determine the viable schedule for the application AI/ML data transport**
2) **Extending the network exposure function (NEF) in 5GC to support monitoring and configuration capability for detection and/or reporting of monitoring events to authorized external party**
   - **New monitoring network resource events include the measurement of data rate or prediction of the network resource utilization for the support of application layer AI/ML operation.**
   - **Extending 5GC information exposure to authorized third party to indicate the UE or network conditions and performance predictions on, e.g., UE location, load, and QoS.**
3) **Enhancing provisioning capability to allow the external party to provision information to 5GC to facilitate the support of application layer AI/ML operation in 5G system.**
   - **One example of the external parameter provisioning information is expected UE behaviors such as expected UE mobility and communication characteristics.**
4) **Enabling 5G system assistance to assist application layer federated learning operation (see next slide for more info).**



4) Collaborative analytic operation with the assistance from 5GS to support data pre-processing, model downloading, learning, training and reporting for various distributed learning operations (e.g. Federated Learning).

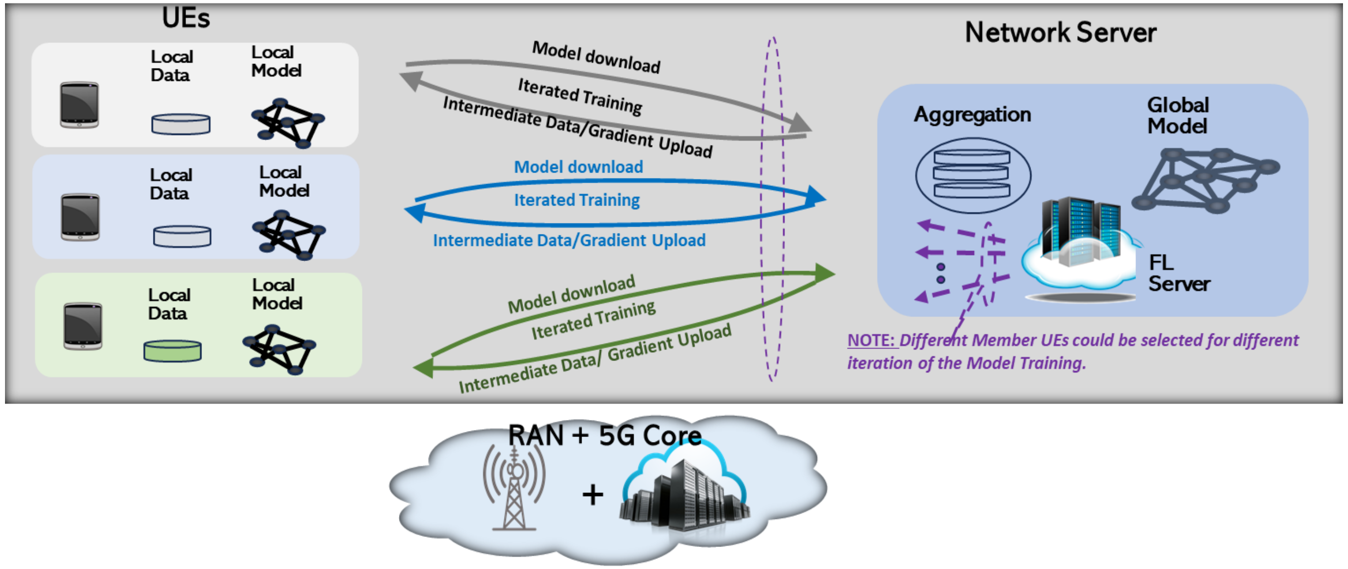
1) **QoS enhancement, AF influence traffic routing & policy control for planned and/or on-demand QoS transport**

2) & 3) 5GC analytical information exposure extension to support application server AI/ML analytic operation, & AF service provisioning 5GC parameters for application AI/ML operation assistance

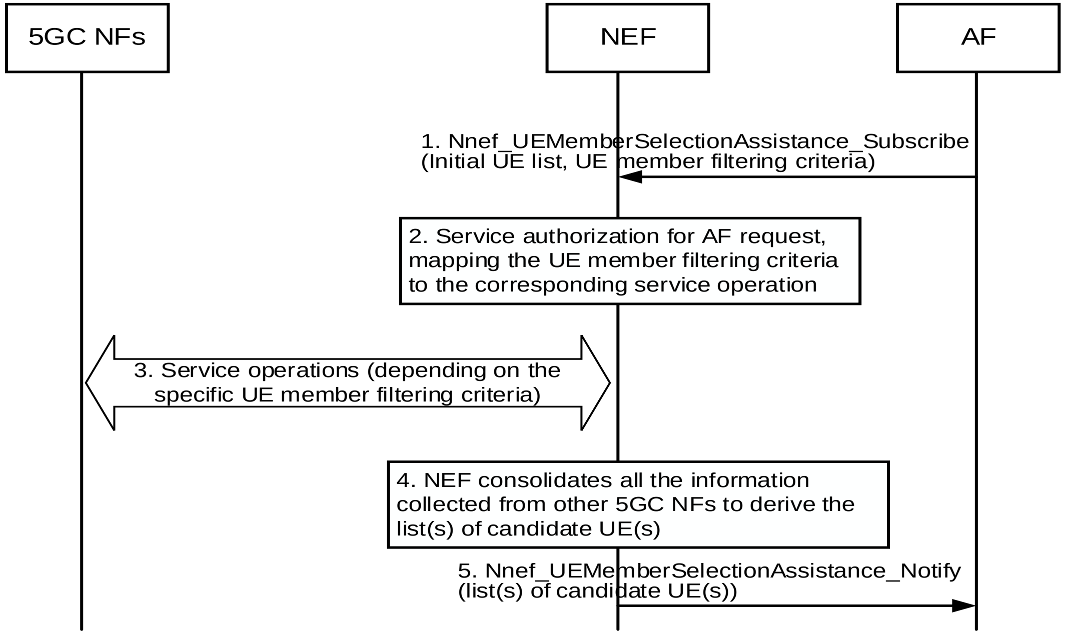
SA2 Architecture Enhancement for Application AI/ML Operation (AIMLsys) – Horizontal Federated Learning (HFL) Support

5G Core provides assistant to support Application layer Federated Learning operation, including

1) **Candidate FL member selection according to specific set of selection criteria (e.g. UE performance, location and trajectory, network resource availability etc.)**
2) **Real time Aggregated QoS monitoring to monitor the QoS usage for the FL task**
3) **Proper time window negotiation with required QoS in order to perform FL and other AIML model transfer service**
4) **KPI definitions for efficient transmission of FL model**



NOTE: Different Member UEs could be selected for different iteration of the Model Training.



1. Nnef_UEMemberSelectionAssistance_Subscribe (Initial UE list, UE member filtering criteria)

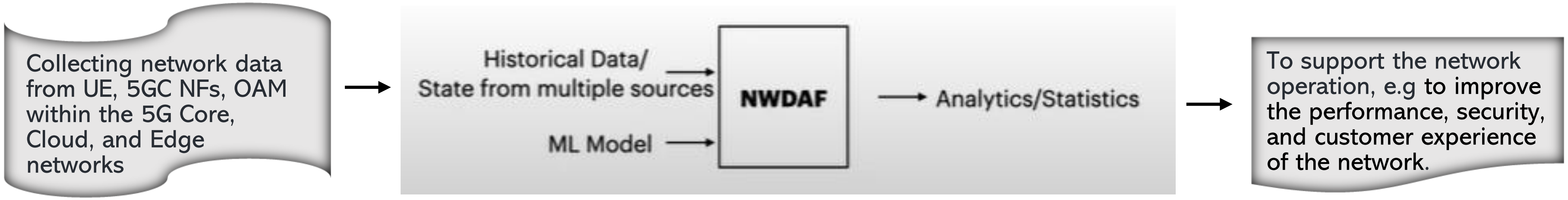
2. Service authorization for AF request, mapping the UE member filtering criteria to the corresponding service operation

3. Service operations (depending on the specific UE member filtering criteria)

4. NEF consolidates all the information collected from other 5GC NFs to derive the list(s) of candidate UE(s)

5. Nnef_UEMemberSelectionAssistance_Notify (list(s) of candidate UE(s))

SA2 architecture enhancement for network AI/ML operation (eNA)


Collecting network data from UE, 5GC NFs, OAM within the 5G Core, Cloud, and Edge networks

To support the network operation, e.g to improve the performance, security, and customer experience of the network.

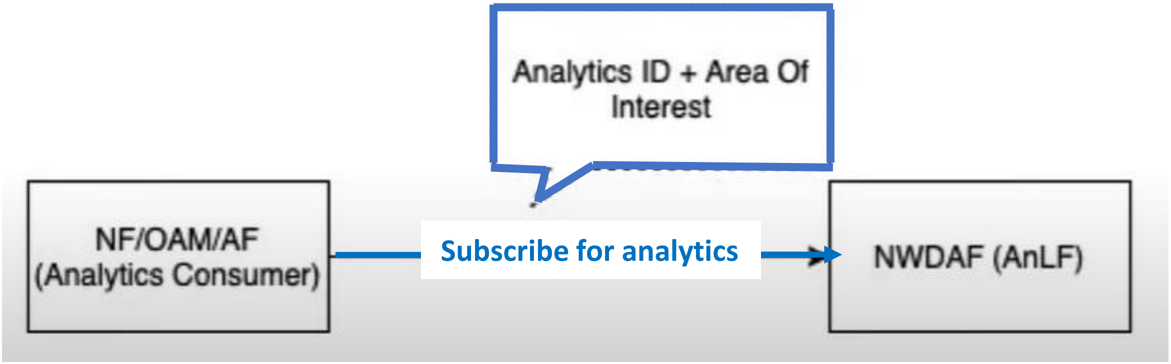
What Is Network Data Analytics Function (NWDAF)?

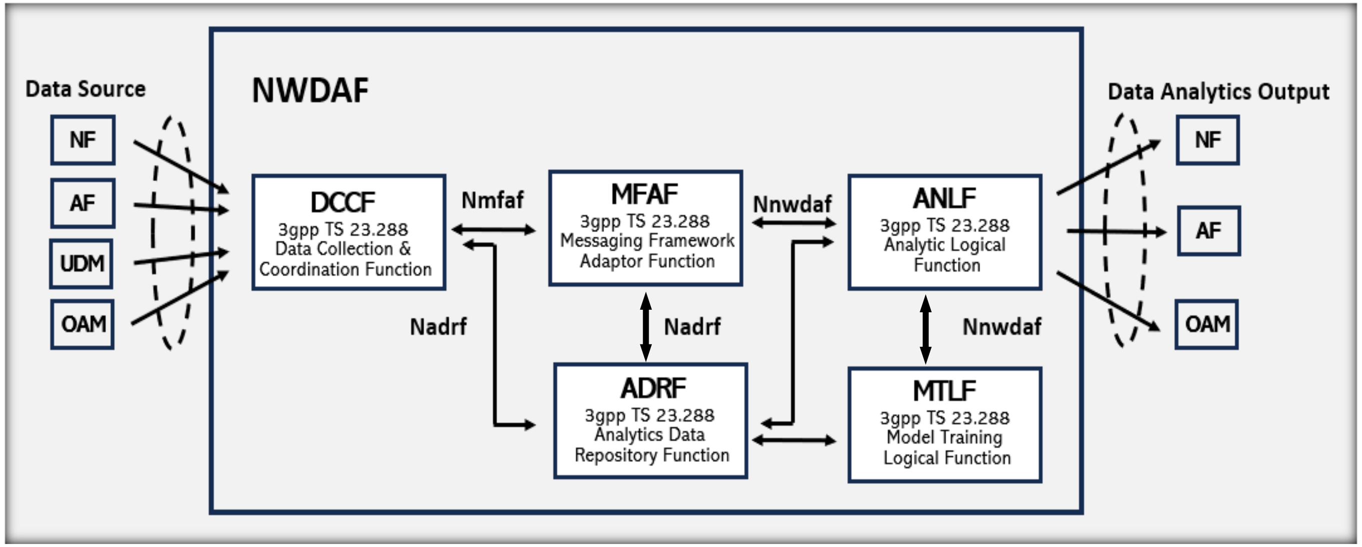
NWDAF as defined in 3GPP TSs 23.288 & 29.520 incorporates standard interfaces from the service-based architecture to collect data by subscription or request model from other network functions.

NWDAF defined in 3GPP TS 29.520 incorporates standard interfaces from the service-based architecture to collect data by subscription or request model from other NFs and similar procedures. This is to deliver analytics functions in the network for automation or reporting, solving major custom interface or format challenges.

Group of standard functions that defined by 3GPP for supporting data analytics to support 5G Network Operation:

- ❑ NWDAF-ANLF – Analytical Logical Function
- ❑ NWDAF-MTLF – Model Training Logical Function
- ❑ DCCF – Data Collection Coordination (& Delivery) Function
- ❑ ADRF – Analytical Data Repository Function
- ❑ MFAF – Messaging Framework Adaptor Function

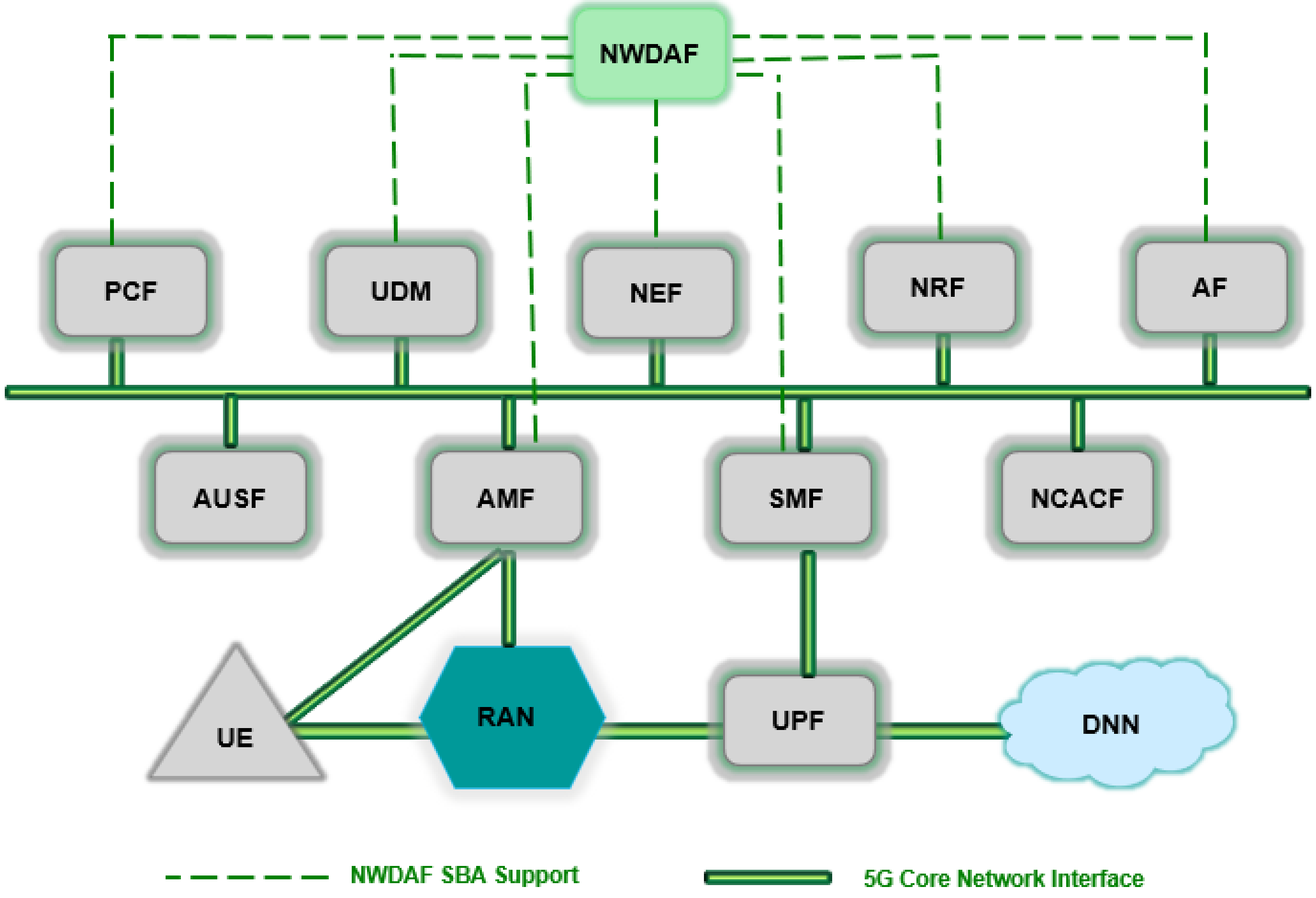




SA2 architecture enhancement for network AI/ML operation (eNA)

3GPP 5G Core SBA Architecture w.r.t NWDAF

Refer to 3GPP TS 23.288 for further details



What are the Key Functionalities of NWDAF?

- ✓ Support data collection from NFs and AFs.
- ✓ Support data collection from OAM.
- ✓ NWDAF service registration and metadata exposure to NFs and AFs.
- ✓ Support analytics information provisioning to NFs and AFs.
- ✓ Support Machine Learning (ML) model training and service provisioning to NWDAF-MTLF & MWDAF-AnLF

SA2 architecture enhancement for network AI/ML operation (eNA)

Referring to 3GPP TS 23.288, clause 5.3 ….

Federated learning among multiple NWDAFs is a machine learning technique in core network that trains an ML Model across multiple decentralized entities holding local data set, without exchanging/sharing local data set. This approach stands in contrast to traditional centralized machine learning techniques where all the local datasets are uploaded to one server, thus allowing to address critical issues such as data privacy, data security, data access rights.

NOTE 1: Horizontal Federated Learning is supported among multiple NWDAFs, which means the local data set in different FL client NWDAFs have the same feature space for different samples (e.g. UE IDs).

For Federated Learning supported by multiple NWDAFs containing MTLF, there is one NWDAF containing MTLF acting as FL server (called FL server NWDAF for short) and multiple NWDAFs containing MTLF acting as FL client (called FL client NWDAF for short), the main functionality includes:

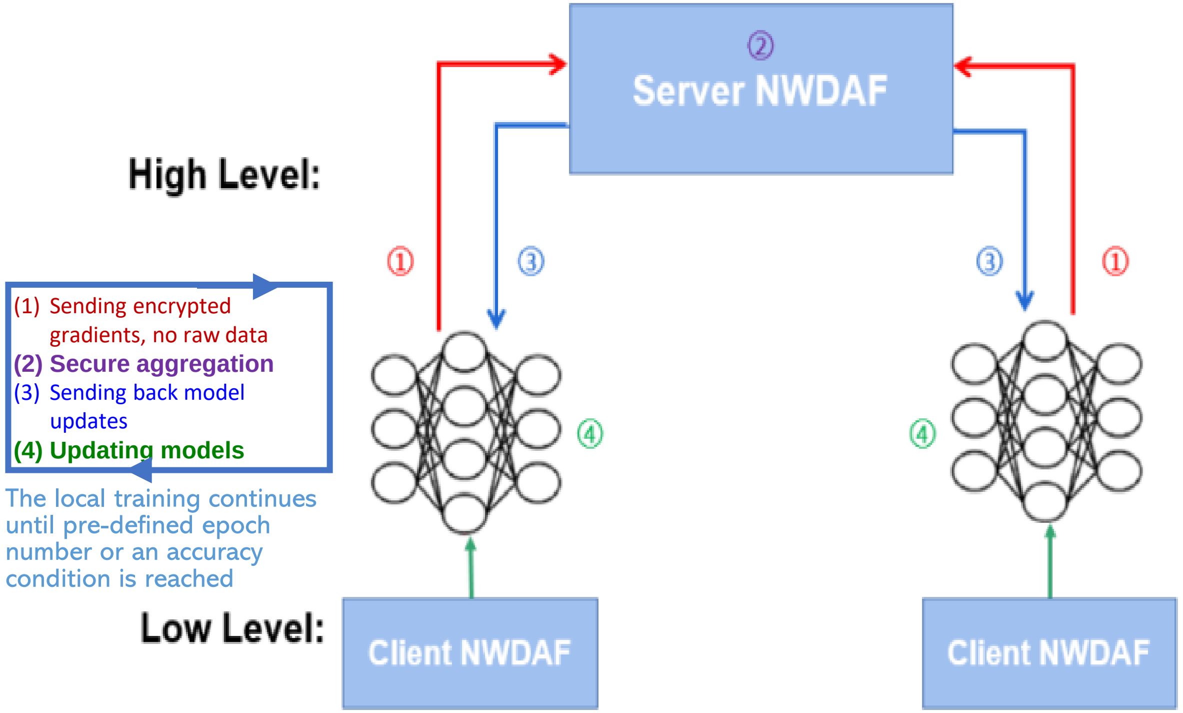
FL server NWDAF:

- discovers and selects FL client NWDAFs to participant in an FL procedure
- requests FL client NWDAFs to do local model training and to report local model information.
- generates global ML model by aggregating local model information from FL client NWDAFs.
- sends the global ML model back to FL client NWDAFs and repeats training iteration if needed.

FL client NWDAF:

- locally trains ML model that tasked by the FL server NWDAF with the available local data set, which includes the data that is not allowed to share with others due to e.g. data privacy, data security, data access rights.
- reports the trained local ML model information to the FL server NWDAF.
- receives the global ML model feedback from FL server NWDAF and repeats training iteration if needed.

FL server NWDAF or FL client NWDAF register to NRF with their FL capability information as described in clause 5.2.



(1) Sending encrypted gradients, no raw data
(2) Secure aggregation
(3) Sending back model updates
(4) Updating models

The local training continues until pre-defined epoch number or an accuracy condition is reached

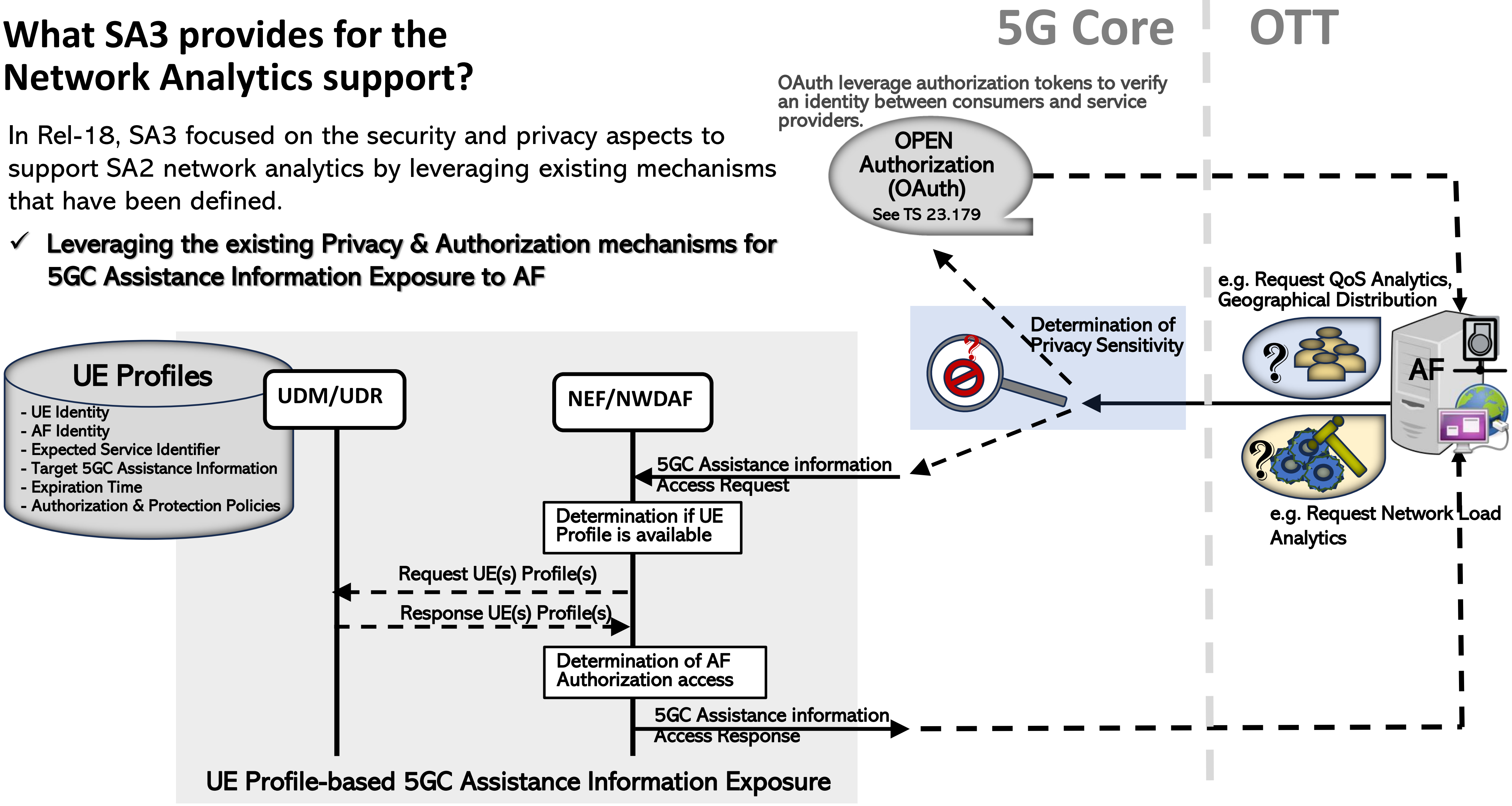
Basic Architecture Framework for Federated Learning is supported in TODAY 5G Core

SA3 Security & Privacy support for Network Analytics

What SA3 provides for the Network Analytics support?

In Rel-18, SA3 focused on the security and privacy aspects to support SA2 network analytics by leveraging existing mechanisms that have been defined.

✓ **Leveraging the existing Privacy & Authorization mechanisms for 5GC Assistance Information Exposure to AF**



5G Core

OTT

OAuth leverage authorization tokens to verify an identity between consumers and service providers.

OPEN Authorization (OAuth)

See TS 23.179

Determination of Privacy Sensitivity

e.g. Request QoS Analytics, Geographical Distribution

AF

e.g. Request Network Load Analytics

UE Profiles

- UE Identity
- AF Identity
- Expected Service Identifier
- Target 5GC Assistance Information
- Expiration Time
- Authorization & Protection Policies

UDM/UDR

NEF/NWDAF

5GC Assistance information Access Request

Determination if UE Profile is available

Request UE(s) Profile(s)

Response UE(s) Profile(s)

Determination of AF Authorization access

5GC Assistance information Access Response

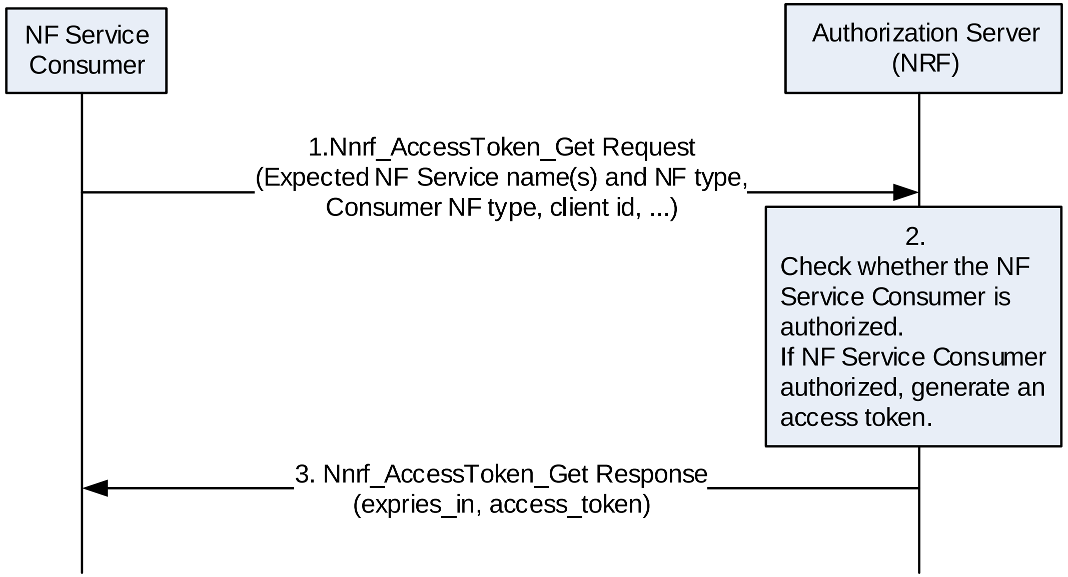
UE Profile-based 5GC Assistance Information Exposure

SA3 Security & Privacy support for network AI/ML operation （eNA_Sec)

□ In Rel-18 eNA_Sec, SA3 has identified and provided security requirements and procedures for the Network Automation features. mainly including:

- Authorization of NF Service Consumers for data access via DCCF;
- Authorization of NF Service Consumers for data access via DCCF when notification sent via MFAF;
- Security protection of data via messaging framework;
- Protection of data transferred between AF and NWDAF;
- Protection of UE data in transit between NFs;
- User consent requirements

Note: The feature for enablers for Network Automation by 5GS is described in 3GPP TS23.501 and 3GPP TS23.288



NF Service Consumer

Authorization Server (NRF)

1.Nnrf_AccessToken_Get Request (Expected NF Service name(s) and NF type, Consumer NF type, client id, ...)

2. Check whether the NF Service Consumer is authorized. If NF Service Consumer authorized, generate an access token.

3. Nnrf_AccessToken_Get Response (expries_in, access_token)

□ In Rel-18

➢ AIMLsys_Sec, SA3 had a study on AIML application only study and no normative work was pursued.

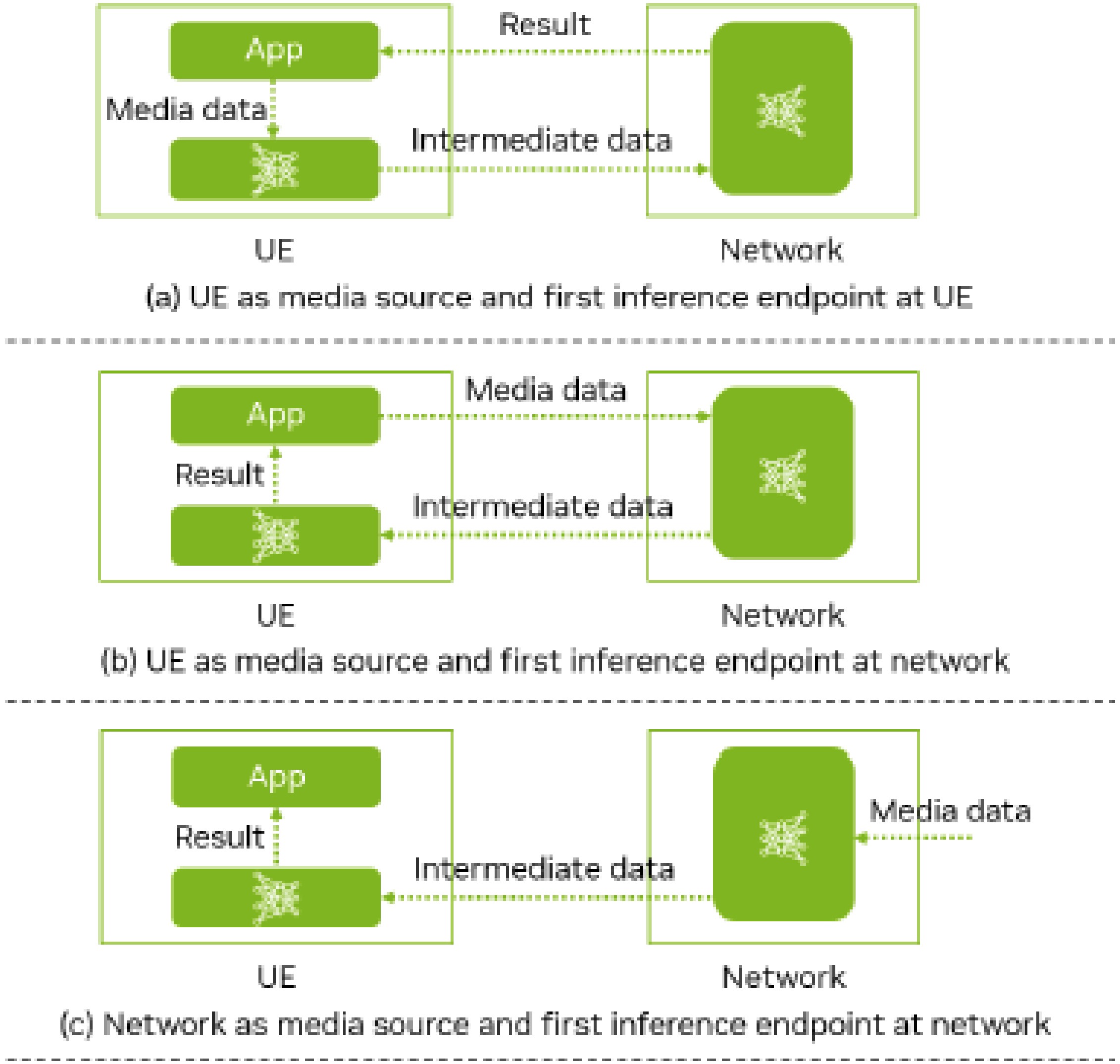
➢ SA3 also focused on the security and privacy aspects to support the RAN3 Rel-18 AI/ML Framework (see later slide on RAN3 reporting) with the study “Study on the security aspects of Artificial Intelligence (AI)/Machine Learning (ML) for the NG-RAN. The study was concluded with no pursued normative work

SA4 AI/ML for Multi-media

Main Objectives – Defining media service architecture for AI/ML and relevant service flows; in addition, determining the data formats and protocols for various types of data components for AI/ML-based media services, traffic characteristics of the data components delivered over 5G and the respective KPIs.

When applying AI/ML for media, one main consideration is ***the splitting the AI/ML inference between network and UE***. Split points can depend on a number of factors including UE capabilities (e.g., memory, compute, energy consumption, and inference latency), network conditions (e.g., capacity, load, and latency), model characteristics, and user/task specific requirements (e.g., delay and privacy)

Illustration of different orders of operations & corresponding media flows for splitting AI/ML inference operations between network and UE



(a) UE as media source and first inference endpoint at UE

(b) UE as media source and first inference endpoint at network

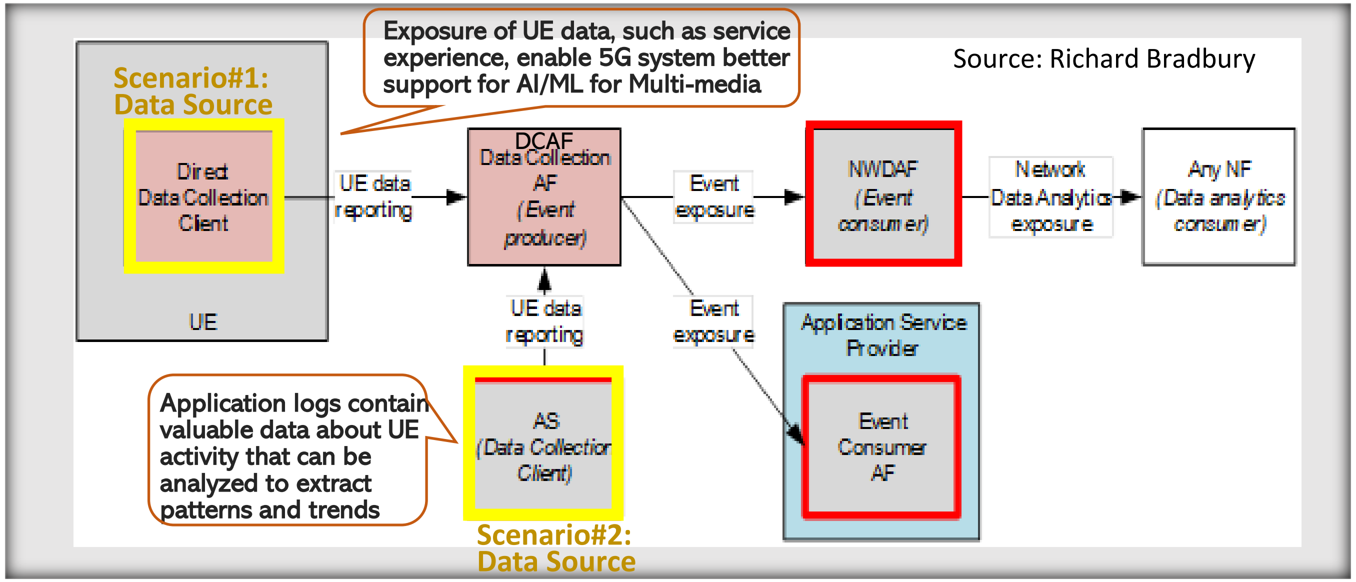
(c) Network as media source and first inference endpoint at network

Source: Artificial Intelligence in 3GPP 5G-Advanced: A survey Xingqin Lin

SA4 Supporting UE Data Collection, Reporting & Exposure

Data collection is essential to support AI/ML operation. SA4 defines the **Data Collection AF (DCAF)** and the related architecture for **UE data collection**, **reporting** and **exposure** to assist AI/ML operation for 5G system as well as for the Application Service Provider (ASP).

In order to support UE Data Collection over 5G for Multi-media services, additional data protection mechanisms were defined by SA4. When the collection of UE data is provisioned by an ASP at the DCAF, a number of **data processing instructions** can be specified to limit the UE data exposed to event consumers. These instructions are expressed in the form of **Data Access Profiles** as follows:



Exposure of UE data, such as service experience, enable 5G system better support for AI/ML for Multi-media

Application logs contain valuable data about UE activity that can be analyzed to extract patterns and trends

Source: Richard Bradbury

- For a particular event type, the exact parameters to be collected can be limited by each Data Access Profile. This permits compliance with one of the key principles of data protection legislation that **only data necessary** for specific purposes should be collected.
- In addition, each metric of collected UE data can be summarized along the axes of time, user and/or location using an **aggregation function**. For example, rather than exposing events detailing the service experienced by individual UEs, a particular Data Access Profile may expose only maximum, minimum and mean average values aggregated over five-minute intervals.
- Multiple Data Access Profiles can be provisioned for a given event type to vary the data restrictions imposed on different event consumers. When more than one Data Access Profile is provisioned, the Data Collection AF selects one based on local policy when it receives a new subscription request from an event consumer.
- As part of the authorization procedure for event consumers, the Data Collection AF may also collaborate with an external Authorization AS, following a similar message exchange pattern to OAuth.

SA5 AI/ML Management In 5G System

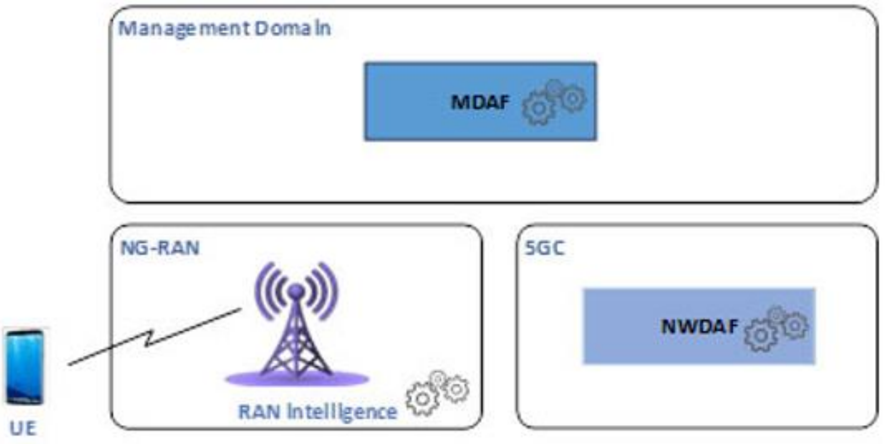
SA5 AI/ML Management is developed to support and facilitate the efficient deployment and operation of AI/ML capabilities/features over 5GS and to manage entire AI/ML lifecycle.

SA5 started the **Management Data Analytics (MDA)** since Rel-17 and continues the AI/ML management specifications development in Rel-18 on the concepts and operational workflows, as well as to address a wide range of use cases (for MDA capabilities) along with the corresponding potential requirements and solutions for the management capabilities and services required for AI/ML **training** & **inference** phases.

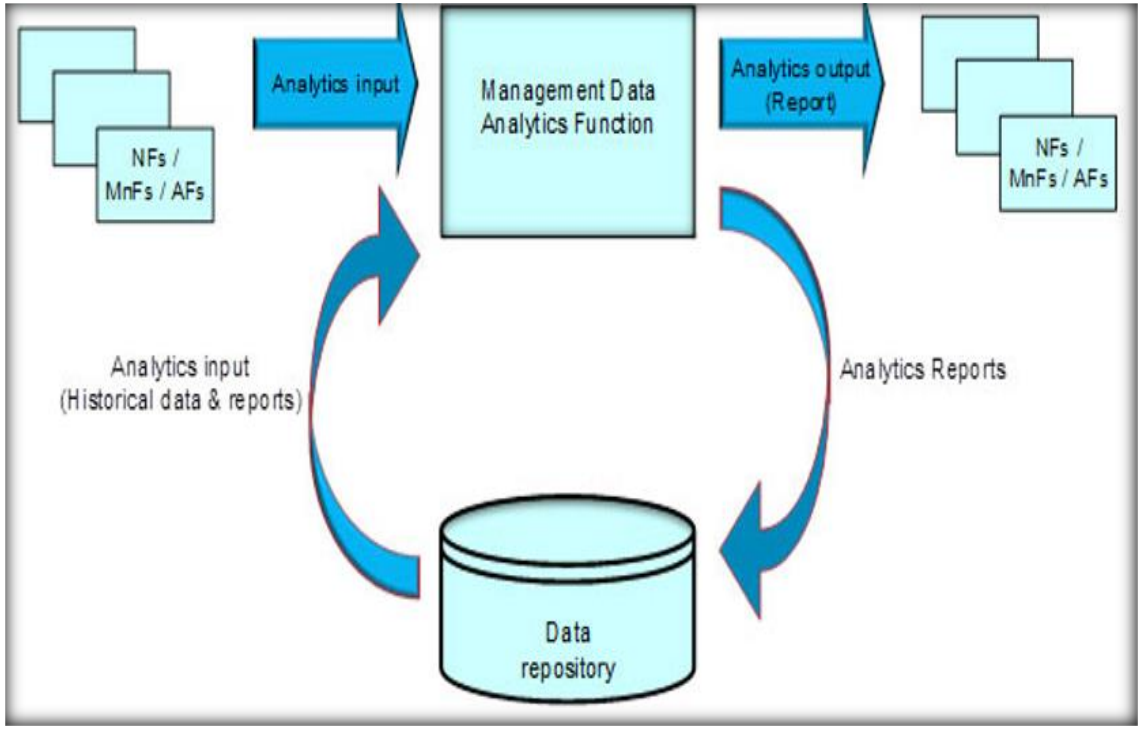
The MDA, in the context of the 3GPP-defined Service Based Management Architecture (SBMA) offers a management service (MnS), usually referred to as MDA MnS or MDAS, allowing any authorized consumer (MDA MnS consumer, e.g. MDAF, NFs, NWDAF, SON, operators etc.) to request and receive analytics.

MDAF may also play the role of MDA MnS producer by leveraging current and historical data from 3GPP cross-domain, e.g. RAN, CN, OAM system as well as data from external entities including non-3GP management system (e.g., MANO, verticals). The data includes e.g.,

- Performance Measurements,
- Trace data including MDA/RF/RCEF,
- QoE and service experience data,
- Analytics data from CN NWDAF,
- Alarm information and notifications,
- Configuration Management information and notifications,
- UE location information,
- MDA reports from other MDA MnS producers,
- Management data from non-3GPP systems.



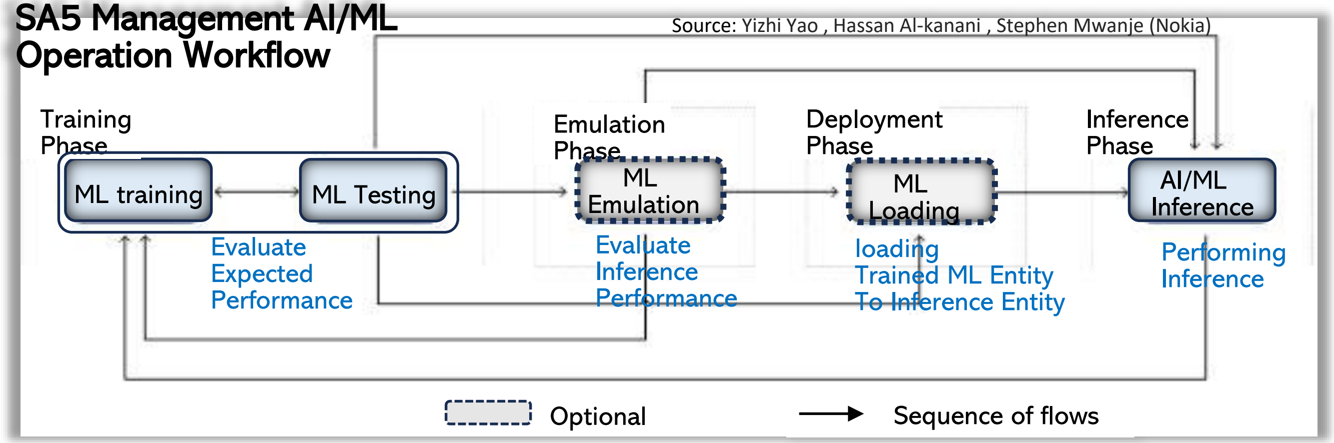
Source: Hassan Al-Kanani, NEC and Yizhi Yao, Intel


NFs / MnFs / AFs
Analytics input
Management Data Analytics Function
Analytics output (Report)
NFs / MnFs / AFs
Analytics input (Historical data & reports)
Analytics Reports
Data repository

Overview of MDA Functionality

SA5 AI/ML Management Operations

SA5 Management AI/ML Operation Workflow



Source: Yizhi Yao , Hassan Al-kanani , Stephen Mwanje (Nokia)

Category	Use cases
Management Capabilities for ML training phase	
Event data for ML training	Pre-processed event data for ML training
ML entity validation	ML entity validation performance reporting
ML entity testing	Consumer-requested ML entity testing
	Control of ML entity testing
	Multiple ML entities joint testing
ML entity re-training	Producer-initiated threshold-based ML Retraining
	Efficient ML entity re-training
	ML entities updating initiated by producer
ML entity joint training	Support for ML entity modularity – joint training of ML entities
Training data effectiveness	Training data effectiveness reporting
	Training data effectiveness analytics
	Measurement data correlation analytics for ML training
ML context management	ML context monitoring and reporting
	Mobility of ML Context
	Standby mode for ML entity
ML entity capability discovery and mapping	Identifying capabilities of ML entities
	Mapping of the capabilities of ML entities
Performance evaluation for ML training	Performance indicator selection for ML model training
	Monitoring and control of AI/ML behavior
	ML entity performance indicators query and selection for ML training
	ML entity performance indicators selection based on MnS consumer policy for ML training
Configuration management for ML training	Control of producer-initiated ML training
ML Knowledge Transfer Learning	Discovering sharable Knowledge
	Knowledge sharing and transfer learning

Category	Use cases
Management Capabilities for AI/ML inference phase	
AI/ML Inference History	Tracking AI/ML inference decisions and context
Orchestrating AI/ML Inference	Knowledge sharing on executed actions
	Knowledge sharing on impacts of executed actions
	Abstract information on impacts of executed actions
	Triggering execution of AI/ML inference functions or ML entities
	Orchestrating decisions of AI/ML inference functions or ML entities
Coordination between the ML capabilities	Alignment of the ML capability between 5GC/RAN and 3GPP management system
Performance evaluation for AI/ML inference	AI/ML performance evaluation in inference phase
	ML entity performance indicators query and selection for AI/ML inference
	ML entity performance indicators selection based on MnS consumer policy for AI/ML inference
	AI/ML abstract performance
Configuration management for AI/ML inference	ML entity configuration for RAN domain ES initiated by consumer
	ML entity configuration for RAN domain ES initiated by producer
	Partial activation of AI/ML inference capabilities
	Configuration for AI/ML inference initiated by MnS consumer
	Configuration for AI/ML inference initiated by producer
	Enabling policy-based activation of AI/ML capabilities
AI/ML update control	Availability of new capabilities or ML entities
	Triggering ML entity update
Common management capabilities for ML training and AI/ML inference phase	
Trustworthy Machine Learning	AI/ML trustworthiness indicators
	AI/ML data trustworthiness
	ML training trustworthiness
	AI/ML inference trustworthiness
	Assessment of AI/ML trustworthiness

Trustworthiness is identified as a **common** management capability for both the training phase and the inference phase.

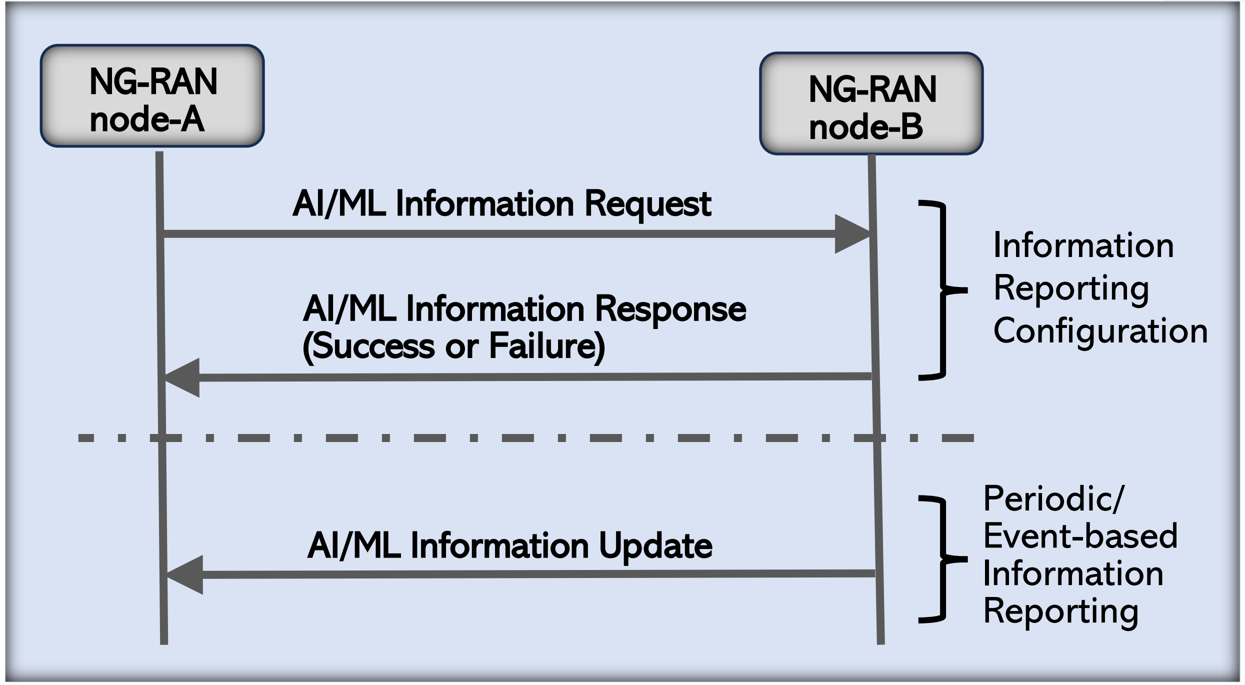
- ➢ Trustworthiness = AI/ML models {robust, explainable, and fair}.
- ➢ Trustworthiness Indicator – configurable and be monitored/evaluated according to Risk & Use Case(s).
  - ▪ Preprocessing of training/testing/inference data may be needed according to the desired trustworthiness measure of the corresponding AI/ML model.
- ➢ The AI/ML MnS should equip the consumer with the trustworthiness capability of data processing requirement to the producer as well as enabling the producer to expose the supported trustworthiness data processing capabilities.
- ➢ AI/ML MnS consumer should be able to query the AI/ML training producer, inference producer, and/or assessment producer about the supported trustworthiness capabilities and request the configuration, measurement, and reporting of a selected set of trustworthiness characteristics.

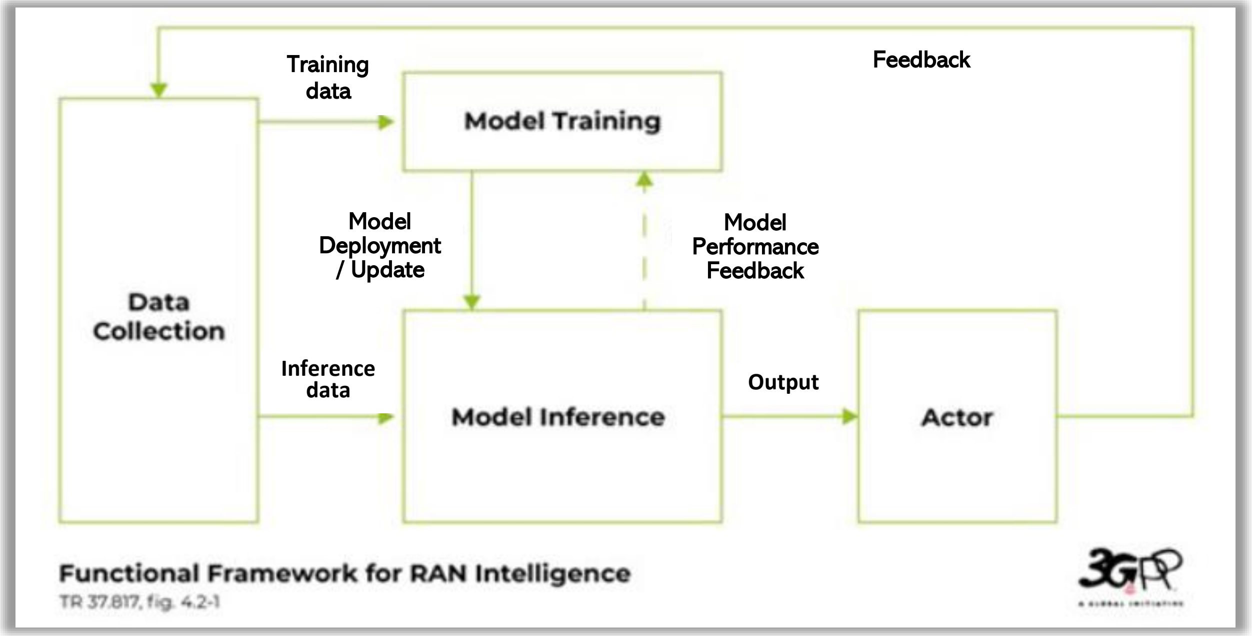
RAN3 AI/ML-enabled NG-RAN

Objective: Improving network performance and user experience, through analyzing the data collected and autonomously process by the NG-RAN with signaling support for: (1) AI/ML based network energy saving, (2) Load Balancing, and (3) Mobility Optimization.

Principles:

- ❑ The AI/ML function requires inputs from neighbor NG-RAN nodes over Xn (e.g. predicted information such as cell-granularity UE trajectory, number of active UEs, RRC connections and radio resources, feedback information such as UE's UL/DL throughput performance, packet delay, PER, measurements such as energy efficiency metric etc.)
- ❑ Signaling procedures used for the exchange of AI/ML related information are use case and data type agnostic and not dependent on the input, output and feedback
- ❑ AI/ML algorithm and models as well as required performance are out of 3GPP scope
- ❑ Deployment options for RAN AI intelligence could be:
  - ➢ AI/ML model training is located in OAM and inference in gNB, or
  - ➢ both can be located in gNB


NG-RAN node-A
NG-RAN node-B
AI/ML Information Request
AI/ML Information Response (Success or Failure)
Information Reporting Configuration
AI/ML Information Update
Periodic/ Event-based Information Reporting



Functional Framework for RAN Intelligence

TR 37.817, fig. 4.2-1

RAN1&2 AI/ML for Air Interface (pave the way to 6G)

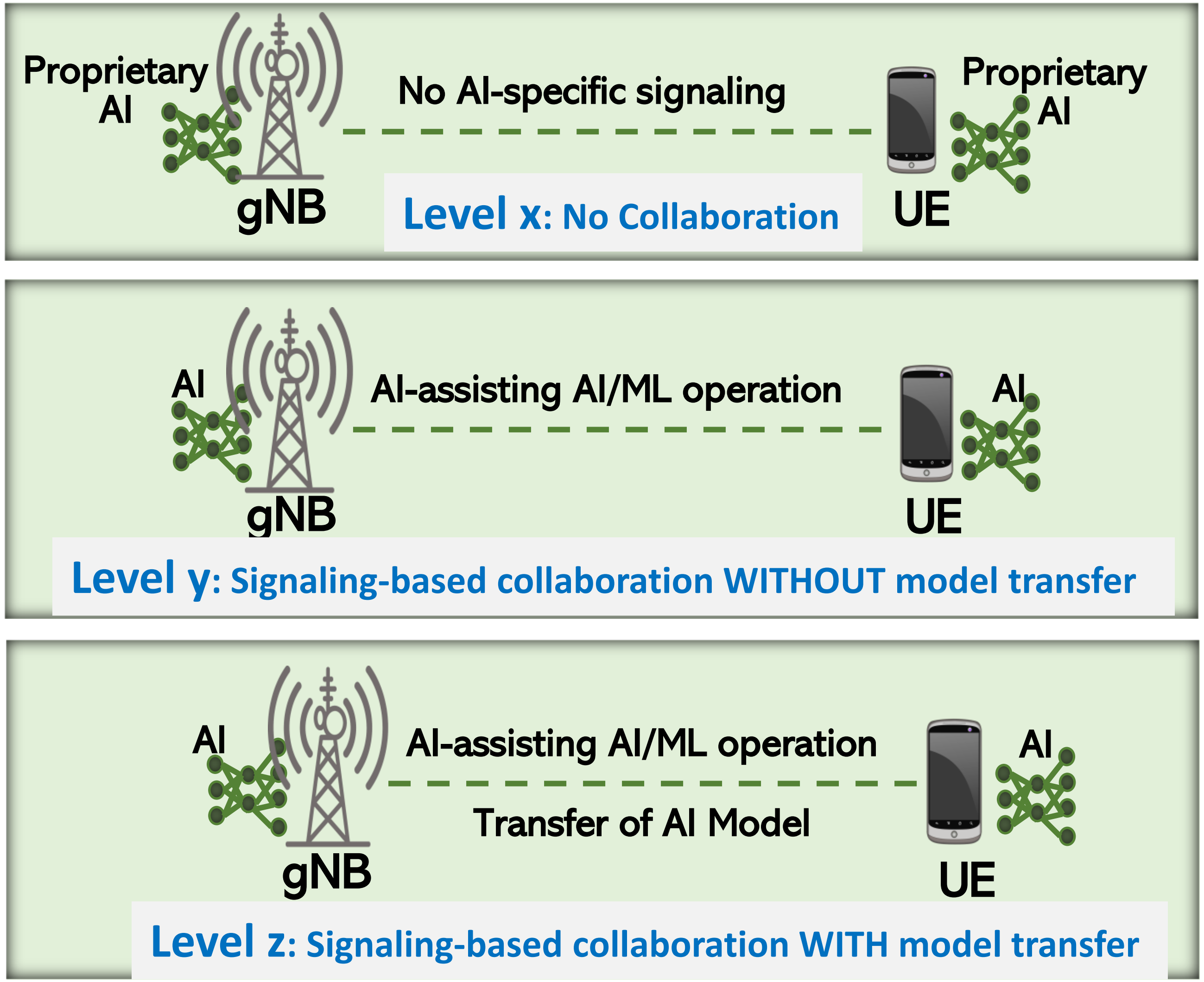
Objective: Establishing a general framework for enhancing the air interface using AI/ML – stages of AI/ML algorithms, collaboration levels between gNB and UE, required datasets for AI/ML model training, validation and testing, and life cycle management of AI/ML models.

Three training collaboration models under investigated:

- □ **Level x:** No collaboration
- □ **Level y:** Signaling-based collaboration without model transfer
- □ **Level z:** Signaling-based collaboration with model transfer

Focusing on 3 use cases:

- □ **Channel state information (CSI) feedback Enhancement** – leveraging AI/ML techniques to improve CSI compression which includes an AI/ML-based CSI encoder at the UE and decoder at the gNB as well CSI Prediction.
- □ **Beam management** – leveraging AI/ML techniques to reduce beam management overhead and latency, as well as improving beam selection accuracy via spatial & temporal prediction.
- □ **Positioning** – leveraging AI/ML techniques to improve Direct AI/ML and AI/ML assisted positioning accuracy for different scenarios including those with heavy Non-line-of-sign (NLOS).

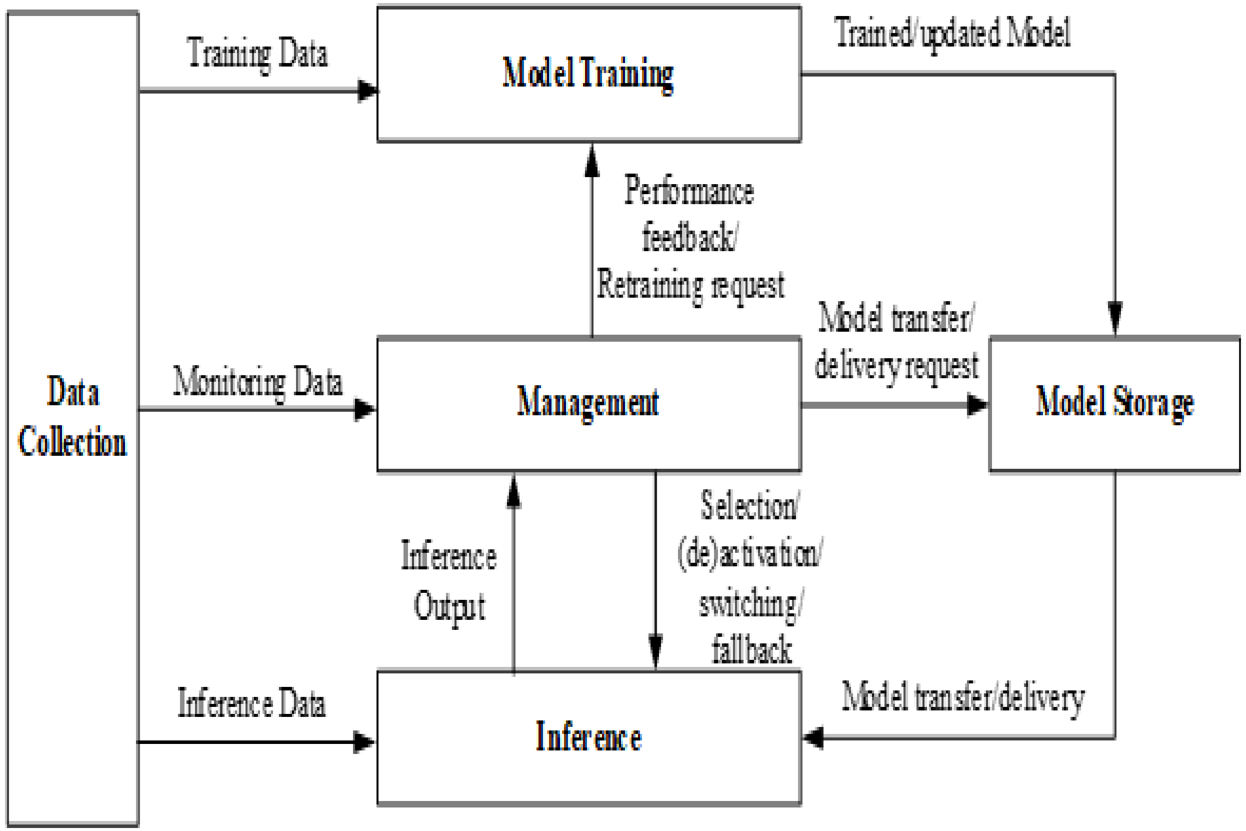

Proprietary AI
No AI-specific signaling
Proprietary AI
gNB
Level x: No Collaboration
UE
AI
AI-assisting AI/ML operation
AI
gNB
UE
Level y: Signaling-based collaboration WITHOUT model transfer
AI
AI-assisting AI/ML operation
AI
Transfer of AI Model
gNB
UE
Level z: Signaling-based collaboration WITH model transfer

RAN1&2 AI/ML for Air Interface (LCM Framework)

Life Cycle Management (LCM) for Air Interface: Establishing a general framework for LCM referred as Functional-based LCM which is considered Model-based LCM as the subset. The key functional elements for Air-interface LCM are:

The list of main components of the Air Interface LCM are as follows:

- ❑ Data collection
  - Note: This also includes associated assistance information, if applicable.
- ❑ Model training
- ❑ Functionality/model identification
- ❑ Model transfer
- ❑ Model inference operation
- ❑ Functionality/model selection, activation, deactivation, switching, and fallback operation.
- ❑ Including: Decision by the network (either network initiated or UE-initiated and requested to the network), decision by the UE (event-triggered as configured by the network, UE's decision reported to the network, or UE-autonomous either with UE's decision reported to the network or without it)
- ❑ Functionality/model monitoring
- ❑ Model update
- ❑ UE capability



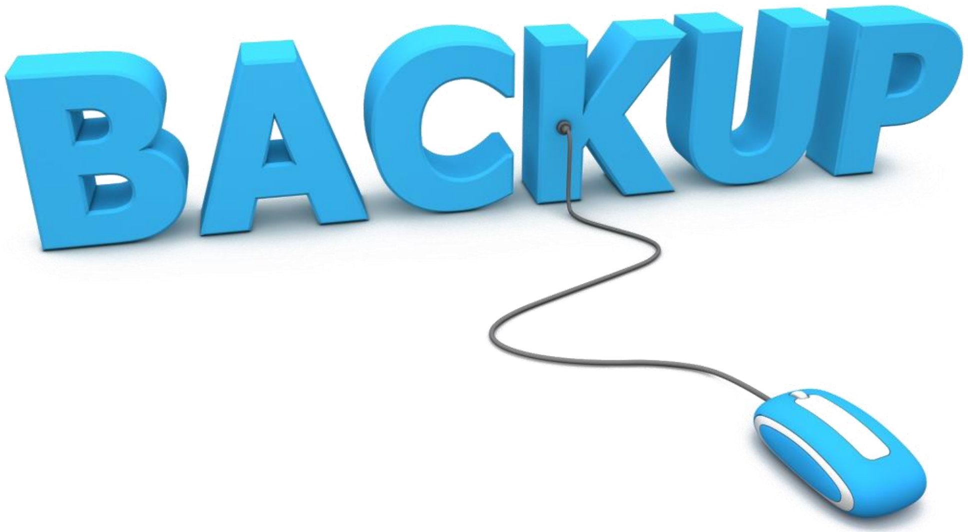
Notes: Some aspects may not have specification impact.

3GPP Rel-18 AI/ML Related Study/Work Items

3GPP Rel-18 AI/ML Related Study/Work Items	Working Group	SID/WID Descriptions
AI/ML model transfer in 5GS	SA1	SP-210520
Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface	RAN1/2	RP-221348
Artificial Intelligence (AI)/Machine Learning (ML)	SA2	SP-230095
Study on Security and Privacy of AI/ML-based Services and Applications in 5G	SA3	SP-220687
Study on the security aspects of Artificial Intelligence (AI)/Machine Learning (ML) for the NG-RAN	SA3	SP-220529
Artificial Intelligence (AI)/Machine Learning (ML) for NG-RAN	RAN3	RP-220635
AI/ML management	SA5	SP-230335
Study on Artificial Intelligence (AI) and Machine Learning (ML) for Media	SA4	SP-220328
CT3 aspects of AIML (CT aspects of System Support for AI/ML-based Services)	CT3	CP-230329
CT4 aspects of AIML (CT aspects of System Support for AI/ML-based Services)	CT4	CP-230329

NOTE: The table above is just to reflect the list of 3GPP projects that are related to AI/ML in Rel-18 and not all require normative work.

Backup Slide



SA2 5G Core Architecture enhancement for network AI/ML operation (eNA)

How NWDAF supports Service Provisioning to assist 5G Network Operation?

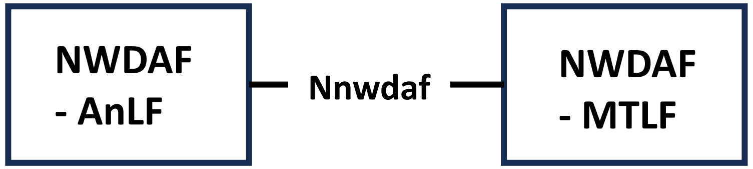
The 5G System architecture allows NWDAF-AnLF to use trained ML model provisioning services from another NWDAF-MTLF.

NWDAF-AnLF performs inference, derives analytics information (i.e. derives statistics and/or predictions based on Analytics Consumer request) and exposes analytics service i.e. Nnwdaf_AnalyticsSubscription or Nnwdaf_AnalyticsInfo.

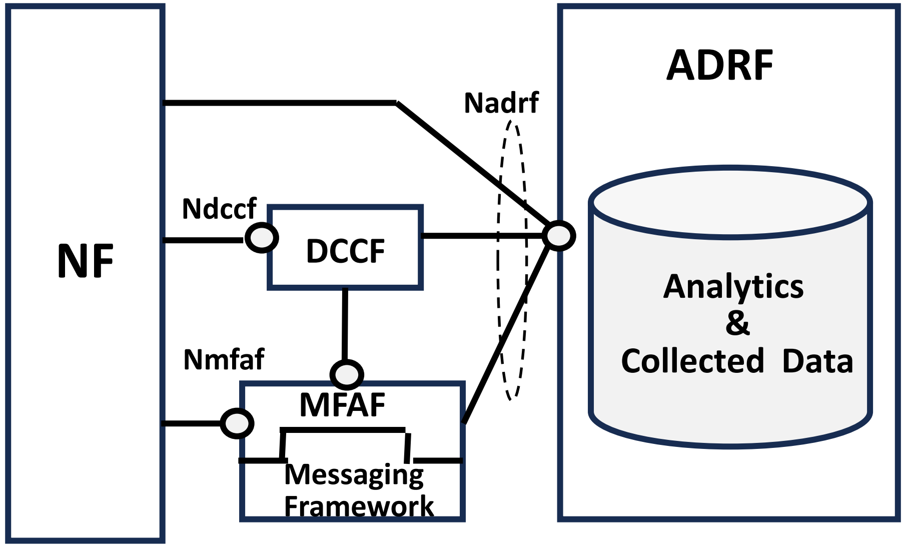
NWDAF-MTLF trains Machine Learning (models and exposes new training services (e.g. providing trained ML model)

The 5G System architecture allows ADRF to store and retrieve the collected data and analytics.

- ADRF exposes the Nadrf service for storage and retrieval of data by other 5GC NFs (e.g. NWDAF) which access the data using Nadrf services.
- Based on the NF request or configuration on the DCCF, the DCCF may determine the ADRF and interact directly or indirectly with the ADRF to request or store data.
- The ADRF stores data received in a Nadrf_DataManagement_Storage Request sent directly from an NF, or data received in an Ndccf_DataManagement_Notify / Nmfaf_3caDataManagement_Notify or
- The ADRF checks if the Data Consumer is authorized to access ADRF services.



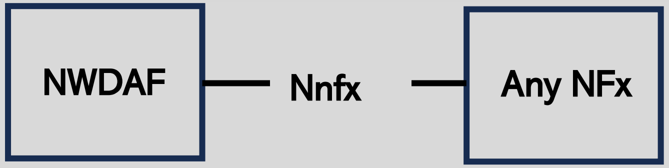
Trained ML Model Provisioning Support



Data Storage Support for Analytics & Data Collection

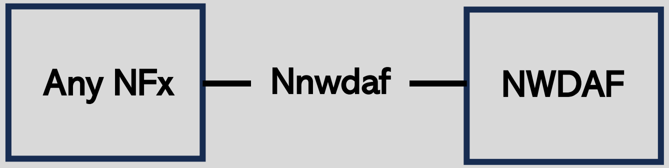
SA2 5G Core Architecture Enhancement for network AI/ML operation (eNA)

How NWDAF support Data Collections to assist 5G Network Operation?



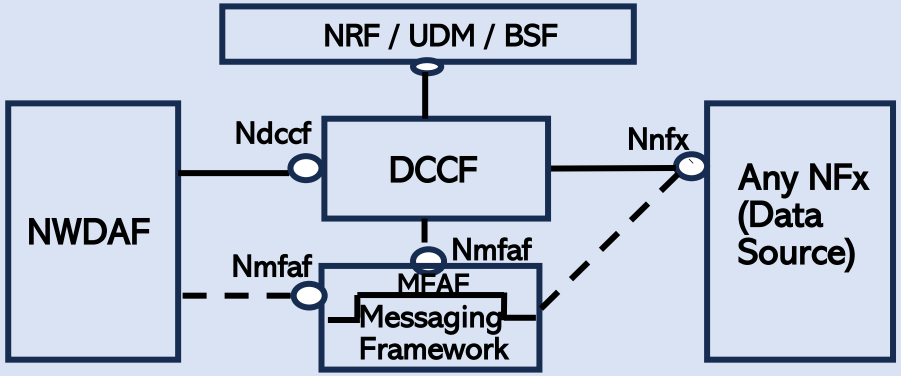
Data Collection Architecture from any 5GC NF

The NWDAF belongs to the same PLMN as the 5GC NF that provides the data



Network Data Analytics Exposure Architecture

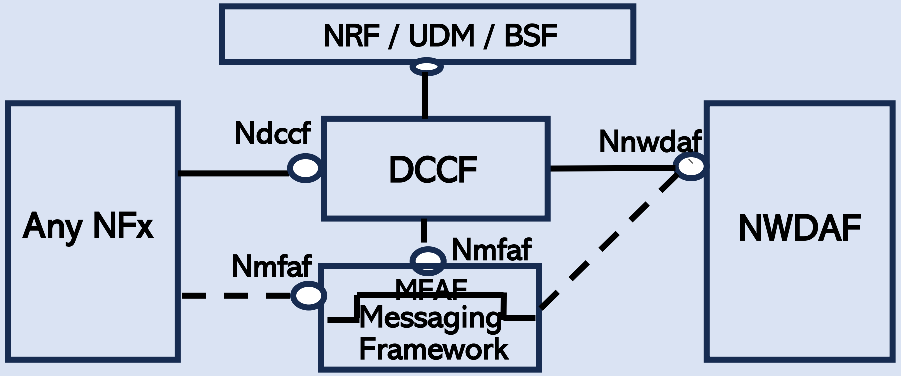
The NWDAF belongs to the same PLMN as the 5GC NF that consumes the analytics information.



Data Collection Architecture using Data Collection Coordination

The Ndccf interface is defined for the NWDAF to subscribe/unsubscribe for data delivery and to request a specific report of data.

Data collection and coordination function (DCCF) coordinates the collection and distribution of data requested by NF consumers. It prevents data sources from handling multiple subscriptions for the same data and sending multiple notifications containing the same information due to uncoordinated requests from data consumers



Network Data Analytics Exposure Architecture using Data Collection Coordination

The Ndccf interface is defined for any NF to subscribe/unsubscribe and to request a specific report of network analytics. If the analytics is not already being collected, the DCCF requests the analytics from the NWDAF using Nnwdaf services. The DCCF may collect the analytics and deliver them to the NF, or the DCCF may rely on a messaging framework to collect analytics and deliver it to the NF.

спасибо 谢谢

GRACIAS

THANK YOU

ありがとうございました MERCI

DANKE धन्यवाद

شُكراً OBRIGADO