SUPPORT OF VERTICAL FEDERATED LEARNING

TECHNICAL FIELD

[0001] The present disclosure relates to wireless communications, and more specifically to network architectures and methods that support vertical federated learning.

BACKGROUND

[0002] A wireless communications system may include one or multiple network communication devices, such as base stations, which may support wireless communications for one or multiple user communication devices, which may be otherwise known as user equipment (UE), or other suitable terminology. The wireless communications system may support wireless communications with one or multiple user communication devices by utilizing resources of the wireless communication system (e.g., time resources (e.g., symbols, slots, subframes, frames, or the like) or frequency resources (e.g., subcarriers, carriers, or the like)). Additionally, the wireless communications system may support wireless communications across various radio access technologies including third generation (3G) radio access technology, fourth generation (4G) radio access technology, fifth generation (5G) radio access technology, among other suitable radio access technologies beyond 5G (e.g., sixth generation (6G)).

SUMMARY

[0003] An article "a" before an element is unrestricted and understood to refer to "at least one" of those elements or "one or more" of those elements. The terms "a," "at least one," "one or more," and "at least one of one or more" may be interchangeable. As used herein, including in the claims, "or" as used in a list of items (e.g., a list of items prefaced by a phrase such as "at least one of" or "one or more of" or "one or both of") indicates an inclusive list such that, for example, a list of at least one of A, B, or C means A or B or C or AB or AC or BC or ABC (i.e., A and B and C). Also, as used herein, the phrase "based on" shall not be construed as a reference to a closed set of conditions. For example, an example step that is described as "based on condition A" may be based on both a condition A and a condition B without departing from the scope of the present disclosure. In other words, as used herein,

the phrase “based on” shall be construed in the same manner as the phrase “based at least in part on”. Further, as used herein, including in the claims, a “set” may include one or more elements.

[0004] Some implementations of the method and apparatuses described herein may include a network entity of a wireless communication system, the network entity comprising at least one memory and at least one processor coupled with the at least one memory and configured to cause the network entity to receive an application layer request from a consumer entity for supporting a machine learning enabled application service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier, and determine a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request. Based on the first requirement, a second requirement is determined, wherein the second requirement comprises a data set requirement for the machine learning enabled service.

[0005] The processor may be configured to cause the network entity to discover a plurality of entities acting as a candidate federated learning clients based on the first and/or second requirement, wherein the discovery comprises obtaining the capabilities of the plurality of the entities.

[0006] The processor may be configured to cause the network entity to configure at least one alignment parameter based on the second requirement and the capabilities of the discovered entities acting as candidate federated learning clients.

[0007] The processor may be configured to cause the network entity to select at least one entity from the plurality of the candidate federated learning clients to serve as a federated learning client for the machine learning enabled application service.

[0008] The processor may be configured to cause the network entity to transmit the at least one alignment parameter to the or each entity acting as a federated learning client.

[0009] The data set requirement may include one or more of:

a set of identifiers corresponding to a machine learning model related event or to an analytics event;

a sample range requirement, for example identifying a plurality of data sources such as UEs, a geographical range, and/or a time range; and

an identification of one or more statistics.

[0010] The network entity may be configured for use as an Application Data Analytics Enablement Server, ADAES, or other Artificial Intelligence enabler in an application layer of the wireless communication system.

[0011] The consumer entity may be an entity of a Vertical Application Layer, VAL, or an entity of a wireless core network.

[0012] The processor may be configured to identify at least one entity acting as a candidate federated learning client by performing a federated learning client discovery procedure with a Network Repository Function or other registry.

[0013] The processor, and/or one or more further processors, may be configured to receive federated learning training responses from the or each federated learning client and to perform a global model aggregation to obtain machine learning model parameters consistent with the request. The or each processor may be configured to:

send the obtained machine learning model parameters to the consumer entity; or

use the obtained machine learning model parameters to obtain analytics data and send the obtained analytics data to the consumer entity.

[0014] The federated machine learning client or clients may be located within one or more of:

an application enabler layer of the wireless communication system;

a core network;

an Edge data network; regional data network; a user equipment; a vertical application;

an enterprise network; and

an external cloud.

[0015] A data set requirement may be used to facilitate a consistent collection of data across the or each candidate federated learning client. The data set requirement may be a data sample requirement. The data set requirement may be a feature selection requirement. The data sample requirement may comprise information on the data required to train the first model. The data sample requirement may comprise an indication of the data needed to be collected to train the first model. The data sample requirement may comprise statistics

corresponding to the data needed to be collected to train the first model. The data sample requirement may comprise features corresponding to the data needed to be collected to train the first model. The data sample requirement may comprise one or more sample range requirement. The one or more sample range requirement may be identified by one or more of: one or more user equipment identifiers; one or more location areas; one or more time ranges. The one or more user equipment identifiers may include one or more of: a Subscription Permanent Identifier (SUPI), Global Public Subscriber Identity (GPSI) or user application identifier. The data sample requirement may comprise one or more event identification(s).

[0016] Some implementations of the method and apparatuses described herein may include a method performed by a network entity of a wireless communication system. The method comprises receiving an application layer request from a consumer entity for supporting a machine learning enabled application service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier, and determining a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request. The method further comprises determining a second requirement based on the first requirement, wherein the second requirement comprises a data set and/or feature selection requirement for the machine learning enabled service.

[0017] The method may comprise discovering a plurality of entities acting as a candidate federated learning clients based on the first and/or second requirement, wherein the discovery comprises obtaining the capabilities of the plurality of the entities.

[0018] The method may comprise selecting at least one entity from the plurality of the candidate federated learning clients to serve as a federated learning client for the machine learning enabled application service.

[0019] The method may comprise receiving federated learning training responses from the or each federated learning client and performing a global model aggregation to obtain machine learning model parameters consistent with the request.

[0020] The method may comprise sending the obtained machine learning model parameters to the consumer entity, or using the obtained machine learning model parameters to obtain analytics data and send the obtained analytics data to the consumer entity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Figure 1 illustrates a current architecture and approach to deriving and providing analytics to Analytics Consumers NFs.

[0022] Figure 2 illustrates a NWDAF architecture for analytics generation based on trained models.

[0023] Figure 3A illustrates a horizontal federated learning architecture.

[0024] Figure 3B illustrates a vertical federated learning architecture.

[0025] Figure 4 illustrates impacted entities for vertical federated learning and data alignment between ADAES and 5G core network according to an embodiment.

[0026] Figure 5 illustrates a procedure for cross-domain vertical federated learning according to an embodiment.

[0027] Figure 6 illustrates a procedure for app layer vertical federated learning according to an embodiment.

[0028] Figure 7 illustrates a procedure for cross-domain vertical federated learning according to an embodiment.

[0029] Figure 8 illustrates an example of a wireless communications system in accordance with aspects of the present disclosure.

[0030] Figure 9 illustrates an example of a network equipment (NE) 200 in accordance with aspects of the present disclosure.

[0031] Figure 10 illustrate a flowchart of a method performed by a NE in accordance with aspects of the present disclosure.

DETAILED DESCRIPTION

[0032] The 3rd Generation Partnership Project (3GPP) is an umbrella term for a number of standards organizations which develop protocols for mobile telecommunications technologies, including radio access, core network and service capabilities, which provide a complete system description for mobile telecommunications. The 3GPP TSG SA WG6 (SA6) is the application enablement and critical communication applications group for

vertical markets. The main objective of SA6 is to provide application layer architecture specifications for verticals, including architecture requirements and functional architecture for supporting the integration of verticals to 3GPP systems. In this context, SA6 functionality serves as a middleware, offering a toolbox of Platform as a Service (PaaS) capabilities for easing the interactions between the application providers and the network layer, while providing application layer support.

[0033] 3GPP SA6 in particular provides the following key enablers:

Common API Framework [TS 23.222] provides a unified and harmonized API framework across several 3GPP API specifications. Some of the key features are: Onboarding of API invokers, Publication and discovery of service APIs, API security.

Services Enabler Architecture Layer (SEAL) [TS 23.434] is a middleware layer with core capabilities that are needed for multiple industry vertical applications and it includes group management, configuration management, location management, key management, network resource management, slice enablement, analytics enablement, data delivery etc.

Application Architecture for Edge Applications [TS 23.558] enables the applications to be hosted on the Edge of the 3GPP network. Some of the key requirements considered are: minimal impacts to the applications on UE for use on Edge, service differentiation (enable / disable edge compute capabilities), flexible deployments, service continuity.

[0034] Network Data Analytics Function (NWDAF) is a 5G 3GPP standard method used to collect data from user equipment (UEs), network functions (NFs), operations, administration, and maintenance (OAM) systems, etc. from the 5G Core (5GC), Cloud, and Edge networks that can be used for analytics. The NWDAF is located within the 5G core network domain. The NWDAF may provide data analytics to allow communication service providers to improve customer experiences and to increase network efficiency and generate new sources of revenue. Via open APIs the NWDAF may also be exposed to external users such as streaming services, financial institutions including banks, etc.

[0035] Typically, the NWDAF makes use of one or more Machine Learning (ML) models the building of which is an iterative process. A ML model can be identified by its ID and can be provided by the network or operator, a telco vendor, edge provider or an application provider and can be applicable for mobile devices in 5G system, e.g., image recognition, localization, performance, speech recognition, and video processing, as well as

for optimizing performance of the network elements / communication aspects. A ML model can be trained for example for deriving the UE location pattern for a given time and area, the performance for a given cell/ network access or for a UE or the load of a network entity. ML models at the network side are stored at network functions serving as ML model repositories using the ML model ID. For external ML models, there is no provisioning of how the ID is configured and where the ML models are stored.

[0036] In the current 3GPP architecture (up to Release 18) the NWDAF provides analytic output to one or more Analytics Consumer NFs based on data collected from one or more Data Producer NFs as shown in Figure 1. Analytics Consumer NFs subscribe to the NWDAF to receive analytics data therefrom.

[0037] In Release 17 the NWDAF is split into an Analytics Logical Function (AnLF) and a Model Training Function (MTLF) as shown in Figure 2. The AnLF is responsible for collecting analytics requests from consumers and for returning responses. The MTLF is responsible for training ML models based on data received from the Data Producer NFs (or DCCF). According to this architecture, a single AnLF may act as an "aggregator", calculating an aggregate model (where a model may be defined by a set of model parameters) from model data received from the multiple MTLFs. The model aggregator provides updated model parameters to each MTLF that each MTLF uses to re-train its own model thus allowing every local model training function to have a trained model using data from multiple sources.

[0038] This type of shared ML model training is known as "federated learning" or FL. Release 18 defines two types of federated learning, namely horizontal and vertical federated learning. Horizontal federated learning (HFL), or sample-based federated learning, is introduced where data sets share the same feature space but have different samples. This is illustrated in Figure 3A and might arise where different MTLFs create models for the same feature set but for a different group of users, e.g. UEs. Vertical federated learning (VFL) is illustrated in Figure 3B and might arise where different UEs are creating models based on different feature sets for the same or different groups of users.

[0039] Two types VFL, the non-split and split VFL types, are available. In non-split VFL each participant owns a full model, whereas in split VFL each participant owns part of the model. There are different approaches on how to train a model with VFL depending on whether non-split or split VFL is used. In non-split VFL, VFL is either supported through

the use of a Coordinator VFL where each party exchange intermediate results and computes gradients for the model which are sent to a coordinator and the coordinator updates each model and scenarios where no coordinator is used and each party autonomously trains their model based on exchanging intermediate results and gradients. In split VFL the model is split between several parties. One party own the top model (global model) and other parties own one or more bottom models (that are used to train the top model). The most probable scenario for supporting VFL within 5G Systems is non-split VFL.

[0040] Examples of how a model aggregator can generate aggregated models for both types of federated learning are presented in:

1. Concept and Applications, ACM Transactions on Intelligent Systems and Technology (TIST) Volume 10 Issue 2, Article No. 12, January 2019; and
2. 5GPPP, AI and ML – Enablers for Beyond 5G Networks.

[0041] 3GPP SA6 has also specified in 3GPP TS 23.436 the application layer architecture to enable data analytics as a new Service Enabler Architecture Layer (SEAL) service, also known as Application Data Analytics Enablement Server (ADAES). This architecture provides an application layer analytics framework which offers generic analytics exposure and value-added services for verticals and Application Service Providers (ASPs). It also includes application layer analytics related to end-to-end application performance, edge load, service API availability, location accuracy and slice-related performance and fault analysis. In Release 19, the ADAES functional architecture is expected to study enhancements to further improve and enhance functionality in 3GPP and in particular to improve support for analytics using AI/ML methods.

[0042] An analytics event corresponds to an ADAE layer analytics event is specified in 3GPP TS 23.436. Such an event in certain embodiments can be an NWDAF provided analytics event/ID as specified in 3GPP TS 23.288.

[0043] A currently open question is how best to support the registration, discovery and notification handling/subscriptions for entities which are expected to serve as FL members in the AI/ML model pipeline (FL clients, FL server, FL server/aggregator, data collection coordinator). Here, an AI/ML model pipeline is considered to be a means to codify and automate the workflow it takes to produce a machine learning model. An AI/ML learning

pipeline may include multiple sequential steps that do everything from data extraction and preprocessing to model training and deployment.

[0044] The AI/ML pipeline includes building blocks, e.g. ML model training, inference, data preparation and collection, which can be distributed over different entities in the application layer as well as in the 5G network side. For the case of FL, the ML model pipeline includes multiple entities which are operating in parallel for some of these blocks (e.g. training, inference). This presents some additional complexity when allowing different entities to be registered and discovered as FL members since it includes diverse entities from different domains and possibly from different stakeholders.

[0045] Solutions to support vertical federated learning, enabling the 5G System (5GS) to assist in collaborative AI/ML operations involving the ADAE layer and possibly core network and AFs, is currently being studied for Release 19. An example use case is a scenario where a third party application provider wishes to combine data available from the application with data available from a network operator, as well as from Edge/Cloud providers, in order to train a ML model. Taking this example further, a third party application/third party service provider (e.g., a banking application) may require a model with which to map a user's purchasing behaviour against the user's mobility trends. The user can be identifiable by an application ID or by a 3GPP subscription. However, the data (features) available for the same user will be different in the third party and operator network since the third party will have transaction details for the user whereas the core network will have network information, e.g., mobility / location information, for the same user. Vertical Federated Learning can support model training when different domains have the same sample data, e.g., for the same users, but have different feature space.

[0046] The following issues will be considered further in order to support cross domain vertical federated learning:

- How an Enabler/AF can ensure that cross domain FL clients (which may be at the server as well as at the UE side) have aligned their sample ranges, e.g., the same users, to support vertical federated learning (Data Alignment between NWDAF and AF)?
- How an Enabler/AF can discover what features are available in different domains (for the same sample range) in order to support vertical federated learning (Feature Alignment between ADAES, Vertical Application Layer (VAL) and the 5GC)?

[0047] In the context of this discussion a ML enabled application service can be an application layer analytics service as defined in ADAES specifications (e.g. VAL performance analytics) which is expected to used ML techniques for deriving the analytics outputs. Such ML enabled analytics service in certain embodiments can be also a service provided by an AI/ML application (e.g. VAL server) like an automation services at the vertical/ASP premises or an application service (e.g. gaming service, XR service) using ML techniques which is requiring the assistance from the network entity to act as FL client.

[0048] Embodiment 1: SA6 functionality (ADAES, AI enabler) performs alignment and triggers the 5GC to act as FL client.

[0049] Figure 4 illustrates the architecture for supporting vertical FL, the architecture comprising the relevant analytics entities in the 5G core (NWDAF MTLFs), the NRF, the data producing entities and the enabler layer (in particular the SEAL ADAE layer or an AI enabler) which sits on top of 5G core as AFs. In this embodiment the VFL process is initiated from the SA6 functionality (which can be ADAES or another AI enabler). Initiation is triggered by the VAL server for ADAE layer analytics (e.g. the VAL server may be external to the 5G system e.g. a bank server). However it is also possible that this is triggered by a VAL server (towards ADAES or another AI enabler) for training an ML model. Specifically, the VFL process involves the following steps as illustrated in Figure 5:

1. The VAL server sends a request to the ADAES (or equivalent enabler server e.g. AI enabler) to subscribe for analytics or to subscribe for requesting assistance from the 3GPP system (including Enabler layer/AF) to train a ML model. The request includes an analytics and/or ML model ID.

2. The ADAES agrees to support FL for the indicated model and identifies the data set requirement based on the VAL server request. The ADAES may also determine at least some of the FL clients to be used (if known) both at the 5GC and at enablement layer (ADAEC, edge ADAES).

3. The ADAES instantiates the FL server capability (if not already instantiated) and discovers the (possibly additional) entities which are available to act as FL clients from both the 5GC and the enablement layer. This can be done via the Network Repository

Function (NRF) for 5GC FL clients, and other registry e.g., CAPIF Core Function, for external FL clients.

4. The ADAES FL server determines which of the candidate FL clients can be part of the FL process according to the Service Area and time interval for supporting FL and Data Set requirements and features.

5. The ADAES FL server sends an FL preparation request to each of the candidate FL clients, both in the 5GC and outside the 5GC.

6. The candidate FL clients checks the Model Id, Analytics Id (or equivalent), model interoperability and the data availability, time schedule, and confirm or reject their participation.

7. The candidate FL clients (that confirm participation) send an FL preparation response to the ADAES FL server with a positive acknowledgement and with the statistics description of the available data set and supported features.

8. The ADAES FL server initiates the FL process.

9. The ADAES FL server sends a ML model training request to all selected FL clients.

10. The FL clients optionally perform data collection and train locally the ML model, and then provide the output to FL server, e.g. as a set of model parameters.

11. The ADAES FL server performs aggregation of the locally trained ML models and optionally, if the task is to provide FL-enabled analytics at ADAES, derives analytics based on the aggregated trained model.

12. The ADAES FL server sends a response to the VAL server providing the aggregated ML trained model or the analytics output (in case of analytics at the ADAES).

[0050] Embodiment 2: SA6 functionality (ADAES, AI enabler) triggers other applications (Edge, Cloud, UE) to act as FL clients.

[0051] In this embodiment the VFL process is initiated from the SA6 functionality (which can be ADAES or another AI enabler), and the FL operation involves only entities at the application side (either Edge/Cloud platform or AI apps at the UE side). This is triggered

by the VAL server for ADAE layer analytics. However it is also possible that this is triggered by the VAL server (towards ADAES or another AI enabler) for training an ML model. One of the new functionalities in this embodiment for SA6 (Enabler/AF) is how the AF determines its available sample range for the FL process.

[0052] The embodiment has the following high-level steps as illustrated in Figure 6:

1. The VAL server sends a request to the ADAES (or equivalent enabler server e.g. AI enabler) to subscribe for analytics or to subscribe for requesting assistance from the 3GPP system (including Enabler layer/AF) for training of the ML model. The request includes an analytics and/or ML model ID.

2. The ADAES determines to support FL for the identified model and identifies the data set requirement based on the VAL server request. The ADAES may also determine the FL clients to be used (if known) at enablement layer (SEAL, EEL) or at VAL layer / EAS.

3. The ADAES instantiates FL server capability (if not already instantiated) and discovers the entities which are available to act as FL clients from the enablement layer and/or VAL layer / EAS or external system (e.g. MEC Services, O-RAN RIC). This can be done via the NRF (if supported in the 5GC) or other registry, e.g., CAPIF Core Function or other service registry for external FL clients (where "external" here means outside of the PLMN trusted domain).

4. The ADAES determines which of the candidate FL clients can be part of the FL process according to the Service Area and time interval for supporting FL and Data Set requirements and features.

5. The ADAES sends an FL preparation requests to each of the candidate FL clients both at the AF/Enablement layer and externally (e.g. third party servers, EAS/VAL), as well as to FL clients at the UE side. The format of these requests, and the parameters that they contain, may be different based on the interface and the type of FL clients. For example:

    - For FL clients at the Enabler layer or VAL layer or edge: the FL preparation request may include the ML model information, the data set requirement, the

feature requirements or the feature selection method (supervised, semi-supervised, embedded), the configuration of the FL operation and the reporting to the FL server, time of validity and service area, and possibly tools/libraries for undertaking the new capability as FL clients.

- For the FL clients at the VAL / UE side: the FL preparation request may include additionally a status report request for capturing possible changes at the availability/capability and conditions of the respective VAL / UE as well as the indication of expected unavailability or communication limitations and optionally performance measurements / statistics and mobility information.

6. The candidate FL clients check Model Id and/or Analytics Id (or equivalent), model interoperability and the data availability, time schedule and confirm or reject their participation.

7. The candidate FL clients that confirm, send an FL preparation response to the ADAES with a positive acknowledgement and with the statistics description of the available data set and supported features.

8. The ADAES initiates the FL process.

9. The ADAES sends an ML model training request to all selected FL clients.

10. The FL clients optionally perform data collection and train locally the ML model, and then provide the output to ADAES.

11. The ADAES performs aggregation of the locally trained ML models and optionally, if the task is to provide FL-enabled analytics at ADAES, it also derives analytics based on the aggregated trained model.

12. The ADAES sends a response to VAL server providing the aggregated ML trained model or the analytics output (in case of analytics at ADAES).

[0053] Embodiment 3: 5GC triggers ADAES/SA6 functionality to act as FL client.

[0054] In this embodiment the VFL process is initiated from the 5GC to ADAES to include also the VAL layer FL clients (e.g., external and untrusted AFs). This embodiment is used for scenarios when the 5GC requires from external entities to undertake the FL client

operations, but where such FL clients have not registered to the NRF and there is no direct agreement between the Mobile Network Operator (MNO) and the third party FL clients. This scenario is for example for cases when the ADAES/Enabler/AF is provided as PaaS at the Edge Platform provider (e.g. Lenovo) and the third party FL client resides at the vertical or enterprise/IT domain (e.g. AWS). In this case, the Edge platform provider has an agreement with the MNO for the communication aspects, and has also a service agreement with the vertical/enterprise application provider. The embodiment has the following high-level steps as illustrated in Figure 7:

0. The 5GC triggers the FL preparation which includes identifying the FL client as being external to the core network entity (AF / ADAES). The 5GC discovers the ADAES (assuming that ADAES is registered as FL client AF to the NRF).
1. The Network Exposure Function (NEF) sends an ML model training request to the ADAES: this may also include the permission to delegate some ML model training to further application entities (which are in agreement with ADAES provider e.g., Edge provider). This request includes the data set requirements and the feature requirements / feature selection methods to be used.
2. The ADAES determines to delegate/offload some training to external applications, due to various reasons such as energy or load constraints or data limitations / unavailability. It then discovers VAL server entities which can serve as candidate FL clients for the given ML model. This can be done via CAPIF or other service registry at the Edge platform.
3. The ADAES sends to the discovered VAL servers acting as FL clients, ML model training requests, including the data set requirements.
4. The VAL servers confirm and, after fetching the ML model (e.g. from an external ML model repository or from ADRF), they send to the ADAES a response including the trained ML model.
5. The ADAES may perform local ML model aggregation for the plurality of FL clients at the third party domain. Local training at ADAES itself is also possible, based on step 2.
6. The ADAES sends an ML model training response to the NWDAF via the NEF, including the trained ML model which can be also locally aggregated.

[0055] Aspects of the present disclosure are described in the context of a wireless communications system including a core network such as a 5GC or equivalent.

[0056] Figure 8 illustrates an example of a wireless communications system 100 in accordance with aspects of the present disclosure. The wireless communications system 100 may include one or more NE 102, one or more UEs 104, and a core network (CN) 106. The wireless communications system 100 may support various radio access technologies. In some implementations, the wireless communications system 100 may be a 4G network, such as an LTE network or an LTE-Advanced (LTE-A) network. In some other implementations, the wireless communications system 100 may be a NR network, such as a 5G network, a 5G-Advanced (5G-A) network, or a 5G ultra-wideband (5G-UWB) network. In other implementations, the wireless communications system 100 may be a combination of a 4G network and a 5G network, or other suitable radio access technology including Institute of Electrical and Electronics Engineers (IEEE) 802.11 (Wi-Fi), IEEE 802.16 (WiMAX), IEEE 802.20. The wireless communications system 100 may support radio access technologies beyond 5G, for example, 6G. Additionally, the wireless communications system 100 may support technologies, such as time division multiple access (TDMA), frequency division multiple access (FDMA), or code division multiple access (CDMA), etc.

[0057] The one or more NE 102 may be dispersed throughout a geographic region to form the wireless communications system 100. One or more of the NE 102 described herein may be or include or may be referred to as a network node, a base station, a network element, a network function, a network entity, a radio access network (RAN), a NodeB, an eNodeB (eNB), a next-generation NodeB (gNB), or other suitable terminology. An NE 102 and a UE 104 may communicate via a communication link, which may be a wireless or wired connection. For example, an NE 102 and a UE 104 may perform wireless communication (e.g., receive signaling, transmit signaling) over a Uu interface.

[0058] An NE 102 may provide a geographic coverage area for which the NE 102 may support services for one or more UEs 104 within the geographic coverage area. For example, an NE 102 and a UE 104 may support wireless communication of signals related to services (e.g., voice, video, packet data, messaging, broadcast, etc.) according to one or multiple radio access technologies. In some implementations, an NE 102 may be moveable, for example, a

satellite associated with a non-terrestrial network (NTN). In some implementations, different geographic coverage areas 112 associated with the same or different radio access technologies may overlap, but the different geographic coverage areas may be associated with different NE 102.

[0059] The one or more UEs 104 may be dispersed throughout a geographic region of the wireless communications system 100. A UE 104 may include or may be referred to as a remote unit, a mobile device, a wireless device, a remote device, a subscriber device, a transmitter device, a receiver device, or some other suitable terminology. In some implementations, the UE 104 may be referred to as a unit, a station, a terminal, or a client, among other examples. Additionally, or alternatively, the UE 104 may be referred to as an Internet-of-Things (IoT) device, an Internet-of-Everything (IoE) device, or machine-type communication (MTC) device, among other examples.

[0060] A UE 104 may be able to support wireless communication directly with other UEs 104 over a communication link. For example, a UE 104 may support wireless communication directly with another UE 104 over a device-to-device (D2D) communication link. In some implementations, such as vehicle-to-vehicle (V2V) deployments, vehicle-to-everything (V2X) deployments, or cellular-V2X deployments, the communication link 114 may be referred to as a sidelink. For example, a UE 104 may support wireless communication directly with another UE 104 over a PC5 interface.

[0061] An NE 102 may support communications with the CN 106, or with another NE 102, or both. For example, an NE 102 may interface with other NE 102 or the CN 106 through one or more backhaul links (e.g., S1, N2, N3, or network interface). In some implementations, the NE 102 may communicate with each other directly. In some other implementations, the NE 102 may communicate with each other or indirectly (e.g., via the CN 106. In some implementations, one or more NE 102 may include subcomponents, such as an access network entity, which may be an example of an access node controller (ANC). An ANC may communicate with the one or more UEs 104 through one or more other access network transmission entities, which may be referred to as a radio heads, smart radio heads, or transmission-reception points (TRPs).

[0062] The CN 106 may support user authentication, access authorization, tracking, connectivity, and other access, routing, or mobility functions. The CN 106 may be an evolved packet core (EPC), or a 5G core (5GC), which may include a control plane entity that manages access and mobility (e.g., a mobility management entity (MME), an access and mobility management functions (AMF)) and a user plane entity that routes packets or interconnects to external networks (e.g., a serving gateway (S-GW), a Packet Data Network (PDN) gateway (P-GW), or a user plane function (UPF)). In some implementations, the control plane entity may manage non-access stratum (NAS) functions, such as mobility, authentication, and bearer management (e.g., data bearers, signal bearers, etc.) for the one or more UEs 104 served by the one or more NE 102 associated with the CN 106.

[0063] The CN 106 may communicate with a packet data network over one or more backhaul links (e.g., via an S1, N2, N3, or another network interface). The packet data network may include an application server. In some implementations, one or more UEs 104 may communicate with the application server. A UE 104 may establish a session (e.g., a protocol data unit (PDU) session, or the like) with the CN 106 via an NE 102. The CN 106 may route traffic (e.g., control information, data, and the like) between the UE 104 and the application server using the established session (e.g., the established PDU session). The PDU session may be an example of a logical connection between the UE 104 and the CN 106 (e.g., one or more network functions of the CN 106).

[0064] In the wireless communications system 100, the NEs 102 and the UEs 104 may use resources of the wireless communications system 100 (e.g., time resources (e.g., symbols, slots, subframes, frames, or the like) or frequency resources (e.g., subcarriers, carriers)) to perform various operations (e.g., wireless communications). In some implementations, the NEs 102 and the UEs 104 may support different resource structures. For example, the NEs 102 and the UEs 104 may support different frame structures. In some implementations, such as in 4G, the NEs 102 and the UEs 104 may support a single frame structure. In some other implementations, such as in 5G and among other suitable radio access technologies, the NEs 102 and the UEs 104 may support various frame structures (i.e., multiple frame structures). The NEs 102 and the UEs 104 may support various frame structures based on one or more numerologies.

[0065] One or more numerologies may be supported in the wireless communications system 100, and a numerology may include a subcarrier spacing and a cyclic prefix. A first numerology (e.g., μ=0) may be associated with a first subcarrier spacing (e.g., 15 kHz) and a normal cyclic prefix. In some implementations, the first numerology (e.g., μ=0) associated with the first subcarrier spacing (e.g., 15 kHz) may utilize one slot per subframe. A second numerology (e.g., μ=1) may be associated with a second subcarrier spacing (e.g., 30 kHz) and a normal cyclic prefix. A third numerology (e.g., μ=2) may be associated with a third subcarrier spacing (e.g., 60 kHz) and a normal cyclic prefix or an extended cyclic prefix. A fourth numerology (e.g., μ=3) may be associated with a fourth subcarrier spacing (e.g., 120 kHz) and a normal cyclic prefix. A fifth numerology (e.g., μ=4) may be associated with a fifth subcarrier spacing (e.g., 240 kHz) and a normal cyclic prefix.

[0066] A time interval of a resource (e.g., a communication resource) may be organized according to frames (also referred to as radio frames). Each frame may have a duration, for example, a 10 millisecond (ms) duration. In some implementations, each frame may include multiple subframes. For example, each frame may include 10 subframes, and each subframe may have a duration, for example, a 1 ms duration. In some implementations, each frame may have the same duration. In some implementations, each subframe of a frame may have the same duration.

[0067] Additionally or alternatively, a time interval of a resource (e.g., a communication resource) may be organized according to slots. For example, a subframe may include a number (e.g., quantity) of slots. The number of slots in each subframe may also depend on the one or more numerologies supported in the wireless communications system 100. For instance, the first, second, third, fourth, and fifth numerologies (i.e., μ=0, μ=1, μ=2, μ=3, μ=4) associated with respective subcarrier spacings of 15 kHz, 30 kHz, 60 kHz, 120 kHz, and 240 kHz may utilize a single slot per subframe, two slots per subframe, four slots per subframe, eight slots per subframe, and 16 slots per subframe, respectively. Each slot may include a number (e.g., quantity) of symbols (e.g., OFDM symbols). In some implementations, the number (e.g., quantity) of slots for a subframe may depend on a numerology. For a normal cyclic prefix, a slot may include 14 symbols. For an extended cyclic prefix (e.g., applicable for 60 kHz subcarrier spacing), a slot may include 12 symbols.

The relationship between the number of symbols per slot, the number of slots per subframe, and the number of slots per frame for a normal cyclic prefix and an extended cyclic prefix may depend on a numerology. It should be understood that reference to a first numerology (e.g., μ =0) associated with a first subcarrier spacing (e.g., 15 kHz) may be used interchangeably between subframes and slots.

[0068] In the wireless communications system 100, an electromagnetic (EM) spectrum may be split, based on frequency or wavelength, into various classes, frequency bands, frequency channels, etc. By way of example, the wireless communications system 100 may support one or multiple operating frequency bands, such as frequency range designations FR1 (410 MHz – 7.125 GHz), FR2 (24.25 GHz – 52.6 GHz), FR3 (7.125 GHz – 24.25 GHz), FR4 (52.6 GHz – 114.25 GHz), FR4a or FR4-1 (52.6 GHz – 71 GHz), and FR5 (114.25 GHz – 300 GHz). In some implementations, the NEs 102 and the UEs 104 may perform wireless communications over one or more of the operating frequency bands. In some implementations, FR1 may be used by the NEs 102 and the UEs 104, among other equipment or devices for cellular communications traffic (e.g., control information, data). In some implementations, FR2 may be used by the NEs 102 and the UEs 104, among other equipment or devices for short-range, high data rate capabilities.

[0069] FR1 may be associated with one or multiple numerologies (e.g., at least three numerologies). For example, FR1 may be associated with a first numerology (e.g., μ=0), which includes 15 kHz subcarrier spacing; a second numerology (e.g., μ=1), which includes 30 kHz subcarrier spacing; and a third numerology (e.g., μ =2), which includes 60 kHz subcarrier spacing. FR2 may be associated with one or multiple numerologies (e.g., at least 2 numerologies). For example, FR2 may be associated with a third numerology (e.g., μ=2), which includes 60 kHz subcarrier spacing; and a fourth numerology (e.g., μ =3), which includes 120 kHz subcarrier spacing.

[0070] Figure 9 illustrates an example of a NE 200 in accordance with aspects of the present disclosure. The NE 200 may include a processor 202, a memory 204, a controller 206, and a transceiver 208. The processor 202, the memory 204, the controller 206, or the transceiver 208, or various combinations thereof or various components thereof may be examples of means for performing various aspects of the present disclosure as described

herein. These components may be coupled (e.g., operatively, communicatively, functionally, electronically, electrically) via one or more interfaces.

[0071] The processor 202, the memory 204, the controller 206, or the transceiver 208, or various combinations or components thereof may be implemented in hardware (e.g., circuitry). The hardware may include a processor, a digital signal processor (DSP), an application-specific integrated circuit (ASIC), or other programmable logic device, or any combination thereof configured as or otherwise supporting a means for performing the functions described in the present disclosure.

[0072] The processor 202 may include an intelligent hardware device (e.g., a general-purpose processor, a DSP, a CPU, an ASIC, an FPGA, or any combination thereof). In some implementations, the processor 202 may be configured to operate the memory 204. In some other implementations, the memory 204 may be integrated into the processor 202. The processor 202 may be configured to execute computer-readable instructions stored in the memory 204 to cause the NE 200 to perform various functions of the present disclosure.

[0073] The memory 204 may include volatile or non-volatile memory. The memory 204 may store computer-readable, computer-executable code including instructions when executed by the processor 202 cause the NE 200 to perform various functions described herein. The code may be stored in a non-transitory computer-readable medium such as the memory 204 or another type of memory. Computer-readable media includes both non-transitory computer storage media and communication media including any medium that facilitates transfer of a computer program from one place to another. A non-transitory storage medium may be any available medium that may be accessed by a general-purpose or special-purpose computer.

[0074] In some implementations, the processor 202 and the memory 204 coupled with the processor 202 may be configured to cause the NE 200 to perform one or more of the functions described herein (e.g., executing, by the processor 202, instructions stored in the memory 204). For example, the processor 202 may support wireless communication at the NE 200 in accordance with examples as disclosed herein.

[0075] The NE 200 may be configured to support a means for receiving an application layer request from a consumer entity for supporting a machine learning enabled application

service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier, determining a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request, and determining a second requirement based on the first requirement, wherein the second requirement comprises a data set requirement for the machine learning enabled service.

[0076] The controller 206 may manage input and output signals for the NE 200. The controller 206 may also manage peripherals not integrated into the NE 200. In some implementations, the controller 206 may utilize an operating system such as iOS®, ANDROID®, WINDOWS®, or other operating systems. In some implementations, the controller 206 may be implemented as part of the processor 202.

[0077] In some implementations, the NE 200 may include at least one transceiver 208. In some other implementations, the NE 200 may have more than one transceiver 208. The transceiver 208 may represent a wireless transceiver. The transceiver 208 may include one or more receiver chains 210, one or more transmitter chains 212, or a combination thereof.

[0078] A receiver chain 210 may be configured to receive signals (e.g., control information, data, packets) over a wireless medium. For example, the receiver chain 210 may include one or more antennas for receiving the signal over the air or wireless medium. The receiver chain 210 may include at least one amplifier (e.g., a low-noise amplifier (LNA)) configured to amplify the received signal. The receiver chain 210 may include at least one demodulator configured to demodulate the received signal and obtain the transmitted data by reversing the modulation technique applied during transmission of the signal. The receiver chain 210 may include at least one decoder for decoding the demodulated signal to receive the transmitted data.

[0079] A transmitter chain 212 may be configured to generate and transmit signals (e.g., control information, data, packets). The transmitter chain 212 may include at least one modulator for modulating data onto a carrier signal, preparing the signal for transmission over a wireless medium. The at least one modulator may be configured to support one or more techniques such as amplitude modulation (AM), frequency modulation (FM), or digital modulation schemes like phase-shift keying (PSK) or quadrature amplitude modulation

(QAM). The transmitter chain 212 may also include at least one power amplifier configured to amplify the modulated signal to an appropriate power level suitable for transmission over the wireless medium. The transmitter chain 212 may also include one or more antennas for transmitting the amplified signal into the air or wireless medium.

[0080] Figure 10 illustrates a flowchart of a method in accordance with aspects of the present disclosure. The operations of the method may be implemented by a NE as described herein. In some implementations, the NE may execute a set of instructions to control the function elements of the NE to perform the described functions.

[0081] At 302, the method may include receiving an application layer request from a consumer entity for supporting a machine learning enabled application service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier. The operations of 302 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 302 may be performed by a NE as described with reference to Figure 9.

[0082] At 304, the method may include determining a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request. The operations of 304 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 304 may be performed by a NE as described with reference to Figure 9.

[0083] At 306, the method may include determining a second requirement based on the first requirement, wherein the second requirement comprises a data set and/or feature selection requirement for the machine learning enabled service. The operations of 306 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 306 may be performed a NE as described with reference to Figure 9.

[0084] It should be noted that the method described herein describes A possible implementation, and that the operations and the steps may be rearranged or otherwise modified and that other implementations are possible.

[0085] The description herein is provided to enable a person having ordinary skill in the art to make or use the disclosure. Various modifications to the disclosure will be apparent to a person having ordinary skill in the art, and the generic principles defined herein may be applied to other variations without departing from the scope of the disclosure. Thus, the disclosure is not limited to the examples and designs described herein but is to be accorded the broadest scope consistent with the principles and novel features disclosed herein.

CLAIMS

1. A network entity of a wireless communication system, the network entity comprising:
at least one memory; and
at least one processor coupled with the at least one memory and configured to cause the network entity to:

receive an application layer request from a consumer entity for supporting a machine learning enabled application service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier;

determine a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request; and

determine a second requirement based on the first requirement, wherein the second requirement comprises a data set requirement for the machine learning enabled service.

2. The network entity of claim 1, the processor being configured to cause the network entity to discover a plurality of entities acting as candidate federated learning clients based on the first and/or second requirement, wherein the discovery comprises obtaining the capabilities of the plurality of the entities.

3. The network entity of claim 2, the processor being configured to cause the network entity to configure at least one alignment parameter based on the second requirement and the capabilities of the discovered entities acting as candidate federated learning clients.

4. The network entity of claim 3, the processor being configured to cause the network entity to select at least one entity from the plurality of the candidate federated learning clients to serve as a federated learning client for the machine learning enabled application service.

5. The network entity of claim 4, the processor being configured to cause the network entity to transmit the at least one alignment parameter to the or each entity acting as a federated learning client.

6. The network entity of any one of the preceding claims, wherein the data set requirement includes one or more of:

a set of identifiers corresponding to a machine learning model related event or to an analytics event;

a sample range requirement, for example identifying a plurality of data sources such as UEs, a geographical range, and/or a time range; and

an identification of one or more statistics.

7. The network entity of any preceding claim and being configured for use as an Application Data Analytics Enablement Server, ADAES, or other Artificial Intelligence enabler in an application layer of the wireless communication system.

8. The network entity of any preceding claim, wherein the consumer entity is an entity of a Vertical Application Layer, VAL, or an entity of a wireless core network.

9. The network entity of any preceding claim, wherein the processor is configured to identify at least one entity acting as a candidate federated learning client by performing a federated learning client discovery procedure with a Network Repository Function or other registry.

10. The network entity of any preceding claim, wherein the processor, and/or one or more further processors, is or are configured to receive federated learning training responses from the or each federated learning client and to perform a global model aggregation to obtain machine learning model parameters consistent with the request.

11. The network entity of claim 10, the or each processor being configured to:

send the obtained machine learning model parameters to the consumer entity; or

use the obtained machine learning model parameters to obtain analytics data and send the obtained analytics data to the consumer entity.

12. The network entity of any one of the preceding claims, wherein the federated machine learning client or clients is or are located within one or more of:

an application enabler layer of the wireless communication system;

a core network;

an Edge data network; regional data network; a user equipment; a vertical application;

an enterprise network; and

an external cloud.

13. A method performed by a network entity of a wireless communication system, the method comprising:

receiving an application layer request from a consumer entity for supporting a machine learning enabled application service, wherein the application layer request comprises a machine learning model identifier and/or an analytics event identifier;

determining a first requirement for providing vertical federated learning training for the machine learning enabled application service, based on the application layer request; and

determining a second requirement based on the first requirement, wherein the second requirement comprises a data set and/or feature selection requirement for the machine learning enabled service.

14. The method of claim 13 and comprising:

discovering a plurality of entities acting as a candidate federated learning clients based on the first and/or second requirement, wherein the discovery comprises obtaining the capabilities of the plurality of the entities.

15. The method of claim 14 and comprising:

configuring at least one alignment parameter based on the second requirement and the capabilities of the discovered entities acting as candidate federated learning clients.

16. The method of claim 15 and comprising:

selecting at least one entity from the plurality of the candidate federated learning clients to serve as a federated learning client for the machine learning enabled application service.

17. The method of any of claims 13 to 16 and comprising receiving federated learning training responses from the or each federated learning client and performing a global model aggregation to obtain machine learning model parameters consistent with the request.

18. The method of claim 17 and comprising:

sending the obtained machine learning model parameters to the consumer entity; or using the obtained machine learning model parameters to obtain analytics data and send the obtained analytics data to the consumer entity.