COMMUNICATION METHOD AND APPARATUS

[0001] This application claims priority to Chinese Patent Application No. 202410327695.8, filed with the China National Intellectual Property Administration on March 20, 2024 and entitled "COMMUNICATION METHOD AND APPARATUS", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This application relates to the field of communication technologies, and in particular, to a communication method and apparatus.

BACKGROUND

[0003] In a communication system, a radio access network (radio access network, RAN) node may obtain a quality of service (quality of service, QoS) profile (QoS profile) based on a current network status, and indicate the QoS profile to a terminal and/or an application server. Subsequently, the terminal and/or the application server may determine a task scheduling result corresponding to the QoS profile, and execute a task based on the task scheduling result. In the foregoing process, because the QoS profile is related to the task scheduling result, quality of service provided by a current network may meet a QoS requirement of the task, to ensure user experience.

[0004] It may be understood that, in actual application, a channel state in the communication system constantly changes. Therefore, after obtaining the QoS profile, the RAN node needs to indicate the QoS profile to the terminal and/or the application server in time, so that the terminal and the application server determine the task scheduling result in time, thereby avoiding a case in which user experience is affected because the quality of service provided by the current network cannot meet the QoS requirement of the task. Therefore, how the RAN node indicates the QoS profile to the terminal and/or the application server in time is critical.

SUMMARY

[0005] This application provides a communication method and apparatus, so that a RAN node can indicate a QoS profile to a terminal and/or an application server in time.

[0006] To achieve the foregoing objective, the following technical solutions are used in this application.

[0007] According to a first aspect, a communication method is provided. The method may be performed by a second RAN node. The second RAN node herein may be the second RAN node, or may be a processor, a circuit, a module, a logical node, a chip, a chip system, or the like that is in the second RAN node and implements the method.

[0008] The method includes: obtaining first indication information that indicates a transmission mode of second indication information, and sending the first indication information to a first RAN node. The second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information includes: sending, through a user plane path, the second indication information to a first node associated with a first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service. The first RAN node is a RAN node to which the second node is to be handed over or is potentially to be handed over, or the first RAN node is a secondary RAN node that provides dual connectivity for the second node.

[0009] According to the method provided in the first aspect, the second RAN node may indicate the transmission mode of the second indication information to the RAN node to which the second node is to be handed over or is potentially to be handed over, or the secondary RAN node that provides dual connectivity for the second node, so that these RAN nodes may send the second indication information to the first node/second node in the mode indicated by the second RAN node. The second indication information is sent to the first node through the user plane path, so that the first node can quickly obtain the second indication information. The second indication information is sent to the second node through the air interface signaling, so that the second node can quickly obtain the second indication information.

[0010] In a possible implementation, the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

[0011] Based on the foregoing possible implementation, the second RAN node may send the first indication information to the first RAN node in a handover procedure, a secondary base station addition procedure, a secondary base station change procedure, or an inter-base station anchor handover procedure.

[0012] In a possible implementation, obtaining the first indication information includes: receiving the first indication information from a first network element in a core network.

[0013] Based on the foregoing possible implementation, the second RAN node may obtain the

first indication information from the core network (for example, a session management network element in the core network), and then send the first indication information to the first RAN node.

[0014] In a possible implementation, the method further includes: receiving third indication information, where the third indication information indicates that the first RAN node does not support sending of the second indication information through the user plane path, or indicates that the first RAN node does not send the second indication information through the user plane path.

[0015] Based on the foregoing possible implementation, the second RAN node may determine that the first RAN node does not send the second indication information through the user plane path. The second RAN node may further indicate, to the first node and/or the second node, that the first RAN node does not send the second indication information through the user plane path, so that the first node and/or the second node determines that there is a delay in receiving the second indication information.

[0016] In a possible implementation, sending the first indication information to the first RAN node includes: sending the first indication information to the first RAN node via a second network element in the core network.

[0017] Based on the foregoing possible implementation, the second RAN node may forward the first indication information to the first RAN node over the core network (for example, via an access and mobility management network element in the core network). For example, if the second RAN node and the first RAN node are not located in an area managed by a same access and mobility management network element; or if the second RAN node and the first RAN node are located in the area managed by the same access and mobility management network element, but no Xn interface exists between the two RAN nodes, the second RAN node may send the first indication information to the first RAN node via the access and mobility management network element.

[0018] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[0019] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be indicated by using one or more pieces of the foregoing information.

[0020] In a possible implementation, the first node is a user plane network element or an application server, and the second node is a terminal.

[0021] Based on the foregoing possible implementation, the second RAN node may indicate the first RAN node to send the second indication information to the user plane network element or

the application server through the user plane path, or send the second indication information to the terminal through the air interface signaling.

[0022] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

[0023] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be quality of service configuration information at a session granularity, or quality of service configuration information at a quality of service flow granularity.

[0024] According to a second aspect, a communication method is provided. The method may be performed by a first RAN node. The first RAN node is a RAN node to which a second node associated with a first service is to be handed over or is potentially to be handed over, or the first RAN node is a secondary RAN node that provides dual connectivity for the second node. The first RAN node herein may be the first RAN node, or may be a processor, a circuit, a module, a logical node, a chip, a chip system, or the like that is in the first RAN node and implements the method.

[0025] The method includes: receiving first indication information, determining second indication information based on at least one piece of alternative quality of service configuration information, and sending the second indication information based on the first indication information. The first indication information indicates a transmission mode of the second indication information. The second indication information indicates quality of service configuration information associated with the first service; and the transmission mode of the second indication information includes sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending the second indication information to the second node through air interface signaling.

[0026] According to the method provided in the second aspect, the first RAN node may send the second indication information to the first node through the user plane path, so that the first node can quickly obtain the second indication information; and/or the first RAN node may send the second indication information to the second node through the air interface signaling, so that the second node can quickly obtain the second indication information.

[0027] In a possible implementation, the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

[0028] Based on the foregoing possible implementation, the first RAN node may receive the

first indication information in a handover procedure, a secondary base station addition procedure, a secondary base station change procedure, or an inter-base station anchor handover procedure.

[0029] In a possible implementation, receiving the first indication information includes: receiving the first indication information from a second network element in a core network; or receiving the first indication information from a second RAN node.

[0030] Based on the foregoing possible implementation, the first RAN node may obtain the first indication information from the core network (for example, a session management network element in the core network), or obtain the first indication information from the second RAN node.

[0031] In a possible implementation, the transmission mode of the second indication information includes: sending the second indication information to the first node through the user plane path, and sending the second indication information based on the first indication information includes: sending the second indication information to a user plane network element through the user plane path.

[0032] Based on the foregoing possible implementation, the first RAN node may send the second indication information to the user plane network element through the user plane path, so that the user plane network element sends the second indication information to an application server.

[0033] In a possible implementation, the method further includes: when determining not to send the second indication information to the first node through the user plane path, sending third indication information, where the third indication information indicates that sending of the second indication information through the user plane path is not supported, or indicates that the second indication information is not sent through the user plane path.

[0034] Based on the foregoing possible implementation, a node that receives the third indication information, for example, the second RAN node, may determine that the first RAN node does not send the second indication information to the first node through the user plane path.

[0035] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[0036] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be indicated by using one or more pieces of the foregoing information.

[0037] In a possible implementation, the first node is the user plane network element or the application server, and the second node is a terminal.

[0038] Based on the foregoing possible implementation, the first RAN node may send the

second indication information to the user plane network element or the application server through the user plane path, or send the second indication information to the terminal through the air interface signaling.

[0039] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

[0040] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be quality of service configuration information at a session granularity, or quality of service configuration information at a quality of service flow granularity.

[0041] According to a third aspect, a communication method is provided. The method may be performed by an access and mobility management network element. The access and mobility management network element herein may be the access and mobility management network element, or may be a processor, a circuit, a module, a logical node, a chip, a chip system, or the like that is in the access and mobility management network element and implements the method.

[0042] The method includes: receiving first indication information from a second RAN node, and sending the first indication information to a first RAN node. The first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information includes sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service. The first RAN node is a RAN node to which the second node is to be handed over.

[0043] According to the method provided in the third aspect, the access and mobility management network element may forward, to the first RAN node, the first indication information sent by the second RAN node, so that the first RAN node may send the second indication information to the first node/second node in the mode indicated by the second RAN node. The second indication information is sent to the first node through the user plane path, so that the first node can quickly obtain the second indication information. The second indication information is sent to the second node through the air interface signaling, so that the second node can quickly obtain the second indication information.

[0044] In a possible implementation, the first indication information is carried in a handover request message.

[0045] Based on the foregoing possible implementation, the access and mobility management network element may forward, to the first RAN node through a handover procedure, the first indication information sent by the second RAN node.

[0046] In a possible implementation, the method further includes: receiving third indication information from the first RAN node, where the third indication information indicates that the first RAN node does not support sending of the second indication information through the user plane path, or indicates that the first RAN node does not send the second indication information through the user plane path.

[0047] Based on the foregoing possible implementation, the access and mobility management network element may determine that the first RAN node does not send the second indication information through the user plane path.

[0048] In a possible implementation, the method further includes: sending the third indication information to the second RAN node.

[0049] Based on the foregoing possible implementation, the access and mobility management network element may further send the third indication information to the second RAN node, so that the second RAN node indicates, to the first node and/or the second node, that the first RAN node does not send the second indication information through the user plane path. In this case, the first node and/or the second node determine/determines that there is a delay in receiving the second indication information.

[0050] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[0051] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be indicated by using one or more pieces of the foregoing information.

[0052] In a possible implementation, the first node is a user plane network element or an application server, and the second node is a terminal.

[0053] Based on the foregoing possible implementation, the first RAN node may send the second indication information to the user plane network element or the application server through the user plane path, or send the second indication information to the terminal through the air interface signaling.

[0054] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of

the first service.

[0055] Based on the foregoing possible implementation, the quality of service configuration information associated with the first service may be quality of service configuration information at a session granularity, or quality of service configuration information at a quality of service flow granularity.

[0056] According to a fourth aspect, a communication apparatus is provided, to implement the foregoing method. The communication apparatus may be the second RAN node in the first aspect, the communication apparatus may be the first RAN node in the second aspect, or the communication apparatus may be the access and mobility management network element in the third aspect. The communication apparatus includes a corresponding module, unit, or means (means) for implementing the foregoing method. The module, unit, or means may be implemented by using hardware, software, or hardware executing corresponding software. The hardware or the software includes one or more modules or units corresponding to the foregoing functions.

[0057] In a possible implementation, the communication apparatus may include a processing module and an interface module. The processing module may be configured to implement a processing function in any one of the foregoing aspects and the possible implementations of the foregoing aspects. The processing module may be, for example, a processor. The interface module may also be referred to as an interface unit, and is configured to implement a sending function and/or a receiving function in any one of the foregoing aspects and the possible implementations of the foregoing aspects. The interface module may include an interface circuit, a transceiver machine, a transceiver, or a communication interface.

[0058] In a possible implementation, the interface module includes a sending module and a receiving module, respectively configured to implement sending and receiving functions in any one of the foregoing aspects and the possible implementations of the foregoing aspects.

[0059] According to a fifth aspect, a communication apparatus is provided, including: a processor, configured to execute a computer program (or computer-executable instructions) stored in a memory, and/or a logic circuit, to cause the communication apparatus to perform the method in any one of the foregoing aspects. The communication apparatus may be the second RAN node in the first aspect, the communication apparatus may be the first RAN node in the second aspect, or the communication apparatus may be the access and mobility management network element in the third aspect. Optionally, there may be one or more processors.

[0060] In a possible implementation, the communication apparatus further includes the memory.

[0061] In a possible implementation, the processor and the memory are integrated together; or the memory is independent of the processor.

[0062] In a possible implementation, the communication apparatus further includes a communication interface. The communication interface is configured for the communication apparatus to communicate with another device, for example, send or receive data and/or a signal. For example, the communication interface may be a transceiver, a circuit, a bus, a module, or another type of communication interface.

[0063] In a possible implementation, the processor and/or the memory further include an artificial intelligence (artificial intelligence, AI) module, configured to implement an AI-related function. The AI module may implement an AI function by using software, hardware, or a combination of software and hardware. For example, the AI module includes a radio access network (radio access network, RAN) intelligent controller (RAN intelligent controller, RIC) module. For example, the AI module may be a near-real-time RIC or a non-real-time RIC.

[0064] In a possible implementation, the communication apparatus is a chip or a chip system. Optionally, when the communication apparatus is a chip system, the communication apparatus may include a chip, or may include a chip and another discrete device.

[0065] According to a sixth aspect, a communication apparatus is provided, including a processor and an interface circuit. The interface circuit is configured to: receive a computer program or instructions, and transmit the computer program or the instructions to the processor. The processor is configured to execute the computer program or the instructions, to cause the communication apparatus to perform the method in any one of the foregoing aspects. The communication apparatus may be the second RAN node in the first aspect, the communication apparatus may be the first RAN node in the second aspect, or the communication apparatus may be the access and mobility management network element in the third aspect. Optionally, there may be one or more processors.

[0066] In a possible implementation, the processor further includes an AI module, configured to implement an AI-related function. The AI module may implement an AI function by using software, hardware, or a combination of software and hardware. For example, the AI module includes an RIC module. For example, the AI module may be a near-real-time RIC or a non-real-time RIC.

[0067] In a possible implementation, the communication apparatus is a chip or a chip system. Optionally, when the communication apparatus is a chip system, the communication apparatus may include a chip, or may include a chip and another discrete device.

[0068] According to a seventh aspect, a computer-readable storage medium is provided. The computer-readable storage medium stores instructions. When the instructions are run on a computer, the computer is caused to perform the method in any one of the foregoing aspects.

[0069] According to an eighth aspect, a computer program product including instructions is

provided. When the computer program product is run on a computer, the computer is caused to perform the method in any one of the foregoing aspects.

[0070] According to a ninth aspect, a communication system is provided. The communication system includes a second RAN node configured to perform the method in the first aspect and a first RAN node configured to perform the method in the second aspect.

[0071] In a possible implementation, the communication system further includes an access and mobility management network element configured to perform the method in the third aspect.

[0072] For technical effects achieved by any one of the possible implementations of the third aspect to the ninth aspect, refer to technical effects achieved by the first aspect or the second aspect or different possible implementations of the first aspect or the second aspect. Details are not described herein again.

[0073] It may be understood that the solutions in the foregoing aspects may be combined if the solutions are not contradictory.

BRIEF DESCRIPTION OF DRAWINGS

[0074] FIG. 1A is a diagram of a virtual reality (virtual reality, VR) task according to this application;

[0075] FIG. 1B is a diagram of a deep neural network (deep neural network, DNN) model according to this application;

[0076] FIG. 1C is a schematic flowchart in which a RAN node indicates a task scheduling result or a service traffic mode to a network node according to an alternative quality of service mechanism;

[0077] FIG. 2 is a diagram of an architecture of a communication system according to this application;

[0078] FIG. 3 is a diagram of a hardware structure of a communication apparatus according to this application;

[0079] FIG. 4 is a schematic flowchart 1 of a communication method according to this application;

[0080] FIG. 5 is a schematic flowchart 2 of a communication method according to this application;

[0081] FIG. 6 is a schematic flowchart 3 of a communication method according to this application;

[0082] FIG. 7 is a schematic flowchart 4 of a communication method according to this application; and

[0083] FIG. 8 is a diagram of a structure of a communication apparatus according to this application.

DESCRIPTION OF EMBODIMENTS

[0084] Before technical solutions of this application are described, related technical terms in this application are described. It may be understood that the descriptions are intended to make this application easier to understand, but should not be construed as a limitation on the protection scope claimed in this application.

[0085] 1. Terminal

[0086] In this application, the terminal is a device with wireless receiving and sending functions. The terminal may be deployed on land, including indoor, outdoor, handheld, or vehicle-mounted; may be deployed on water (for example, on a ship); or may be deployed in the air (for example, on an airplane, a balloon, and a satellite). The terminal may also be referred to as a terminal device. The terminal device may be user equipment (user equipment, UE), a mobile station (mobile station, MS), a mobile terminal (mobile terminal, MT), or the like, or may be a device configured to provide voice or data connectivity for a user. The UE includes a handheld device with a wireless communication function, a vehicle-mounted device (for example, a car, a bicycle, an electric vehicle, an airplane, a ship, a train, or a high-speed train), a wearable device (for example, a smartwatch, a smart band, or a pedometer), or a compute device. For example, the UE may be a mobile phone (mobile phone), a tablet computer, a notebook computer, a palmtop computer, a mobile internet device (mobile internet device, MID), a satellite terminal, or a computer with wireless receiving and sending functions. The UE may alternatively be a VR terminal device, an augmented reality (augmented reality, AR) terminal device, a wireless modem (modem), a smart point of sale (point of sale, POS) machine, customer-premises equipment (customer-premises equipment, CPE), a smart robot, a robot arm, a workshop device, a smart home device (for example, a refrigerator, a television, an air conditioner, or an electric meter), a wireless terminal in industrial control, a wireless terminal in self-driving, a wireless terminal in telemedicine, a wireless terminal in a smart grid (smart grid), a wireless terminal in transportation security, a wireless terminal in smart transportation, a wireless terminal in a smart city (smart city), a wireless terminal in a smart home (smart home), a vehicle-mounted terminal, a road side unit (road side unit, RSU) with a terminal function, a flight device (for example, a smart robot, a hot air balloon, an uncrewed aerial vehicle, or an airplane), or the like. The terminal may alternatively be another device with a terminal function. For example, the terminal may alternatively be a device that functions as a terminal in device to device (device to device, D2D) communication.

[0087] By way of example and not limitation, the terminal in this application may be a wearable device. The wearable device may also be referred to as a wearable intelligent device, and is a general term of wearable devices, such as glasses, gloves, watches, clothes, and shoes, that are developed by applying wearable technologies to intelligent designs of daily wear. The wearable device is a portable device that can be directly worn on the body or integrated into clothes or an accessory of a user. For example, the wearable device is not only a hardware device, but also a device that implements a powerful function through software support, data exchange, and cloud interaction. In a broad sense, wearable intelligent devices include full-featured and large-sized devices that can implement all or some functions without depending on smartphones, for example, smartwatches or smart glasses, and include devices that are dedicated to only one type of application function and need to collaboratively work with other devices such as smartphones, for example, various smart bands or smart jewelry for monitoring physical signs.

[0088] In this application, the terminal may alternatively be a terminal in an internet of things (internet of things, IoT) system. IoT is an important part of future information technology development. A main technical feature of the IoT is to connect objects to networks using communication technologies, to implement intelligent networks for interconnection between persons and machines and between things. The terminal in this application may be a terminal in machine type communication (machine type communication, MTC). The terminal in this application may be a vehicle-mounted module, a vehicle-mounted component, an onboard component, a vehicle-mounted chip, or a vehicle-mounted unit that is built in a vehicle as one or more components or units. The vehicle uses the vehicle-mounted module, the vehicle-mounted component, the onboard component, the vehicle-mounted chip, or the vehicle-mounted unit that is built in the vehicle, to implement the method in this application.

[0089] 2. RAN node

[0090] The RAN node in this application may be any device with wireless receiving and sending functions, and may provide a wireless access service for a terminal. The RAN node may include but is not limited to an evolved NodeB (NodeB or eNB or e-NodeB, evolved NodeB) in long term evolution (long term evolution, LTE), an evolved NodeB (next generation eNB, ng-eNB) in next generation LTE, a gNodeB (gNodeB or gNB) or a transmission reception point (transmission receiving point/transmission reception point, TRP) in new radio (new radio, NR), a future evolved base station of the 3rd generation partnership project (3rd generation partnership project, 3GPP), an access node in a wireless fidelity (wireless fidelity, Wi-Fi) system, a wireless relay node, a wireless backhaul node, and the like. A base station may be a macro base station, a micro base station, a pico base station, a small cell, a relay station, a balloon station, or the like. A plurality of base stations may support networks using a same technology mentioned above, or may

support networks using different technologies mentioned above. The base station may include one or more co-site or non-co-site TRPs. Alternatively, the RAN node may be a radio controller in a cloud radio access network (cloud radio access network, CRAN) scenario. Alternatively, the RAN node may be a central unit (central unit, CU), a distributed unit (distributed unit, DU), a CU-control plane (control plane, CP), a CU-user plane (user plane, UP), a radio unit (radio unit, RU), a road side unit (road side unit, RSU) with a base station function, a wired access gateway, a core network element, or the like. Alternatively, the RAN node may be a server, a wearable device, a machine communication device, a vehicle-mounted device, or the like. For example, an access network device in a vehicle to everything (vehicle to everything, V2X) technology may be an RSU. An example in which the RAN node is a base station is used below for description. The plurality of RAN nodes may be base stations of a same type, or may be base stations of different types. The base station may communicate with a terminal, or may communicate with the terminal via a relay station. The terminal may communicate with a plurality of base stations using different technologies. For example, the terminal may communicate with a base station that supports an LTE network, may communicate with a base station that supports a 5th generation (5th generation, 5G) network, may support dual connectivity to a base station in an LTE network and a base station in a 5G network, or may support dual connectivity to a base station in a 6G network and a base station in a 5G network. Alternatively, the terminal may communicate with a plurality of base stations using a same technology. For example, the terminal may support dual connectivity to base stations in an LTE network, dual connectivity to base stations in a 5G network (which may be referred to as NR-DC (new radio dual connectivity)), or dual connectivity to base stations in a 6G network.

[0091] In this application, the CU and the DU may be separately disposed, or may be included in a same network element, for example, a baseband unit (baseband unit, BBU). The RU may be included in a radio frequency device or a radio frequency unit, for example, included in a remote radio unit (remote radio unit, RRU), an active antenna unit (active antenna unit, AAU), or a remote radio head (remote radio head, RRH). It may be understood that the CU may be classified as a network device in an access network, or the CU may be classified as a network device in a core network. This is not limited herein.

[0092] In different systems, the CU (or the CU-CP and the CU-UP), the DU, or the RU may also have different names, but persons skilled in the art may understand meanings thereof. For example, in an open radio access network (open radio access network, ORAN) system, the CU may alternatively be referred to as an O-CU (open CU), the DU may alternatively be referred to as an O-DU, the CU-CP may alternatively be referred to as an O-CU-CP, the CU-UP may alternatively be referred to as an O-CU-UP, and the RU may alternatively be referred to as an O-

RU. For ease of description, the CU, the CU-CP, the CU-UP, the DU, and the RU are used as examples for description in this application. Any one of the CU (or the CU-CP or the CU-UP), the DU, and the RU in this application may be implemented by using a software module, a hardware module, or a combination of a software module and a hardware module.

[0093] It may be understood that, in some scenarios, roles of the RAN node and the terminal are relative. For example, a helicopter or an uncrewed aerial vehicle that is usually configured as a terminal may be alternatively configured as a mobile base station, and a device that accesses a RAN via the helicopter or the uncrewed aerial vehicle is configured as a terminal.

[0094] 3. Extended reality (extended reality, XR)

[0095] The XR may be various environments that combine reality and virtuality and are generated by using computing technologies and wearable devices, and human-computer interaction. The XR has advantages such as a plurality of views and strong interactivity, can provide brand-new experience for users, and has great application value and commercial potential. The XR mainly includes virtual-reality interaction technologies such as VR, AR, and mixed reality (mixed reality, MR), and can be widely used in many fields such as entertainment, gaming, medical care, advertising, industry, online education, and engineering. The VR technology, the AR technology, and the MR technology are separately described below.

[0096] The VR technology integrates a plurality of technologies such as computer graphics and multimedia, to simulate functions of human sensory organs, for example, a visual sense, an auditory sense, and a tactile sense. In this way, people immerse in a virtual world generated by a computer, feeling as if they were there, and can perform real-time communication by using a language, a gesture, or the like, thereby enhancing a sense of immersion. By using the VR technology, people feel reality of the real world, and can further break through restrictions of time, space, and the like, to have wonderful experience of entering the virtual world. In the VR technology, users usually need to wear XR terminals (for example, head-mounted devices), so that visual senses, auditory senses, and/or tactile senses are simulated for the users. The VR technology may be further used to perform motion tracking on the users, to update simulated visual content, auditory content, and/or tactile content in time. For example, in the VR technology, status information of the users (for example, position information and posture information of the users) may be processed, to display, on the XR terminals, scene content corresponding to the status information of the users.

[0097] In the AR technology, virtual information may be superposed on the real world by using a computer technology, and displayed through a device, for example, a mobile phone, a tablet computer, or glasses, to be perceived by people. In this way, reality and virtuality are merged, to enrich the real world. In short, more information is given to a real object, to enhance a stereoscopic

sense, and enhance visual effect and interactive experience. For example, in the AR technology, perceived visual information (the visual information usually includes depth information) may be processed, to merge virtual information with the real world and enable the virtual information to be perceived by a user. In this way, the real world is "enhanced".

[0098] In the MR technology, the real world and a virtual world may be mixed to generate a new visual environment. The environment includes both a physical entity and virtual information, and content seen in the visual environment is "real-time" content.

[0099] 4. Multi-node collaboration

[00100] In recent years, video rendering services represented by cloud gaming and VR, or AI services represented by terminal visual cognition, AR, and MR, have increasingly high requirements on transmission bandwidth of a network and computational power of a terminal. However, the transmission bandwidth of the network and the computational power of the terminal are limited, and cannot meet the requirements of these services.

[00101] The video rendering service is used as an example. If video rendering is performed on the terminal, the computational power of the terminal is insufficient to implement fully local high-quality rendering of a video. If video rendering is performed on a cloud, a low-delay requirement cannot be met, and a picture may be subject to black-border effect, distortion, or the like due to a limited transmission capability of the network.

[00102] The AI service is used as an example. If the AI service is performed on the terminal, the computational power and battery power of the terminal are insufficient to support local AI inference. If the AI service is performed on the cloud, high uplink bandwidth is needed for AI inference on the cloud when a low-delay requirement is met, resulting in limited uplink coverage and quantity of users.

[00103] Therefore, to balance the computational power of the terminal and the transmission capability of the network, a multi-node collaborative task processing scheme is provided. The task may be referred to as work of processing data through a plurality of steps (step). The plurality of steps may be parallel steps, serial steps, or hybrid parallel and serial steps. For ease of describing a plurality of steps included in one task, one step or a plurality of associated steps in the plurality of steps may be considered as one subtask, that is, one task may include a plurality of subtasks.

[00104] In this application, one service, for example, one video rendering service or one AI service, may include one or more tasks. If one service includes one task, the terminal and the cloud may separately execute a part of subtasks, to reduce a requirement of the task on the terminal and the computational power, and a requirement on the transmission bandwidth of the network. If one service includes a plurality of tasks, the terminal and the cloud may jointly process all or a part of the tasks in the plurality of tasks. For example, for each of all or a part of the tasks, the terminal

and the cloud may separately execute a part of subtasks, to reduce a requirement of the plurality of tasks on the terminal and the computational power, and a requirement on the transmission bandwidth of the network.

[00105] It may be understood that the terminal or the cloud in this application may be collectively referred to as a network node. The network node may be any device with a computing capability and a communication capability. In addition to the terminal and the cloud, the network node may alternatively be a RAN node, a functional module (for example, a CU, a DU, or an RIC) of a RAN node, a core network element, a server, an application server (APP server), a cloud server, a cloud platform (cloud platform), a mobile edge computing (mobile edge computing, MEC) platform, a computing execution entity (computing execution entity, CEF), or the like. This is not limited. For descriptions of the terminal and the RAN node, refer to the foregoing descriptions of the technical terms in this application. The core network element is, for example, one or more of the following network elements: a user plane function (user plane function, UPF) network element, an access management function (access management function, AMF) network element, a session management function (session management function, SMF) network element, a policy control function (policy control function, PCF) network element, or an application function (application function, AF) network element. The cloud platform may be located in a data network behind the UPF network element, and may exchange user plane application layer data with a 5th generation system (5th generation system, 5GS) through an N6 interface, to provide a computing service.

[00106] In conclusion, one task may be split into a plurality of subtasks, and the plurality of subtasks may be executed by a plurality of network nodes. It should be understood that a quantity of network nodes that execute one task is not limited in this application. For example, one task may be split into three subtasks, which are respectively executed by three network nodes; or may be split into two subtasks, which are respectively executed by two network nodes.

[00107] The following uses the terminal and the cloud as an example to describe a specific process in which a plurality of network nodes execute a task.

[00108] Refer to FIG. 1A. A VR task may be split into a subtask 101 and a subtask 102. The terminal may obtain initial data (for example, image information), input the initial data to the subtask 101 to obtain intermediate data, and send the intermediate data to the cloud. After receiving the intermediate data, the cloud may input the intermediate data to the subtask 102 to obtain target data.

[00109] It may be understood that the task may be split in a plurality of split manners, and different split manners may correspond to different communication requirements.

[00110] For example, a video rendering task is used as an example. The video rendering task may be split by separating a foreground from a background. For example, a foreground part of a

picture is usually unpredictable with a small amount of rendering, and may be locally rendered by the terminal, and a background part of the picture is predictable with a large amount of rendering, and is rendered by the cloud. Alternatively, the video rendering task may be split based on a fixation point of a user. For example, a picture at a center of the fixation point may be locally rendered by the terminal, and a picture around the fixation point may be rendered by the cloud. Alternatively, the video rendering task may be split based on an object. For example, different nodes are selected, for rendering, based on rendering computation amounts and data volumes of different objects.

[00111] For the video rendering task, different split manners correspond to different volumes of video data to be transmitted. Therefore, different split manners correspond to different communication requirements. For example, in Table 1, an average bit rate of data transmission required by full cloud rendering is 4 Mbps, a peak value is 40 Mbps, and an allowed transmission delay is 25 ms. An average bit rate of data transmission required by device-cloud collaborative rendering is 0.49 Mbps to 1.8 Mbps, a peak value is 3.8 Mbps, and an allowed transmission delay is 50 ms. It can be learned that device-cloud collaborative processing of the video rendering task can reduce a requirement on a downlink transmission rate. Therefore, selecting a proper split manner for the video rendering task can increase a quantity of users of downlink transmission that access a network.

Table 1

	Bit rate of data transmission (Mbps)		Transmission delay (ms)
	Average value	Peak value	
Full cloud rendering	4	40	25
Device-cloud collaborative rendering	0.49 to 1.8	3.8	50

[00112] Then, an AI compute task is used as an example. A computing model of the AI compute task may be split into a plurality of submodels, and one or more submodels may be considered as one subtask. For example, the AI computing model is a neural network (neural network, NN) model or a DNN model. The NN model/DNN model may include a plurality of layers, for example, an input (input) layer, a convolutional (convolution) layer, a pooling (pooling) layer, and a fully connected (fully connected) layer. The layers may be split. For example, the input layer and the convolutional layer are deployed on the terminal, and the pooling layer and the fully connected layer are deployed on the cloud. In this way, the terminal may sense behavior data of a user (for example, position information, posture information, or voice information of the user) or data of an

ambient environment (for example, a video or an image of the ambient environment) through a sensing module, for example, a radar, a camera, a handle, or a microphone, sequentially input the data to the input layer and the convolutional layer to obtain intermediate data, and send the intermediate data to the cloud. Subsequently, after receiving the intermediate data, the cloud sequentially inputs the intermediate data to the pooling layer and the fully connected layer to obtain target data. Optionally, the cloud may further send the target data to the terminal.

[00113] It is found through further research that, when a task is split, if positions of task splitting (referred to as task split points below) are different, a data volume of the intermediate data may vary, and computation load in the terminal may also vary.

[00114] For example, a DNN model shown in FIG. 1B is used as an example. The DNN model includes an input layer, a convolutional layer 1, a pooling layer 1, a convolutional layer 2, a pooling layer 2, a convolutional layer 3, a pooling layer 3, a convolutional layer 4, a pooling layer 4, a convolutional layer 5, an activation function (for example, a rectified linear unit (rectified linear unit, ReLU)), a pooling layer 5, a fully connected layer 1, a fully connected layer 2, and an output layer. Initial data may be sequentially input to the foregoing layers to obtain target data. FIG. 1B further shows five candidate split points (candidate split points), which are respectively a candidate split point 0 to a candidate split point 4. The five candidate split points are at different positions in the DNN model. It can be understood that a layer preceding a candidate split point may be deployed on the terminal, and a layer following the candidate split point may be deployed on the cloud. The candidate split point 0 precedes the input layer. Therefore, the terminal sends the initial data to the cloud without executing a task, and the cloud executes all tasks. The candidate split point 1 follows the pooling layer 1. Therefore, the terminal executes a task preceding the pooling layer 1, and the cloud executes a task following the pooling layer 1. The candidate split point 2 follows the pooling layer 2. Therefore, the terminal executes a task preceding the pooling layer 2, and the cloud executes a task following the pooling layer 2. The candidate split point 3 follows the pooling layer 5. Therefore, the terminal executes a task preceding the pooling layer 5, and the cloud executes a task following the pooling layer 5. The candidate split point 4 follows the output layer. Therefore, the terminal executes all tasks, the cloud does not execute a task, and the terminal does not send the intermediate data to the cloud. An approximate output uplink data size (approximate output UL data size) (that is, the data volume of the intermediate data) and computation load in the terminal (computation load in UE) (required UL data rate) that correspond to each candidate split point vary. The candidate split point 1 and the candidate split point 2 are used as an example. In Table 2, an approximate output uplink data size corresponding to the candidate split point 1 is 120 Mbit/s, and computation load that is in the terminal and corresponds to the candidate split point 1 is low (low); and an approximate output uplink data size

corresponding to the candidate split point 2 is 24 Mbit/s, and computation load that is in the terminal and corresponds to the candidate split point 2 is high (high). A larger approximate output uplink data size indicates more data to be transmitted by the terminal, that is, a larger data volume of the intermediate data. Therefore, a computation amount of the terminal is smaller, that is, the computation load in the terminal is lower. On the contrary, a smaller approximate output uplink data size indicates less data to be transmitted by the terminal, and higher computation load in the terminal. Therefore, on the premise that a requirement on the computation load in the terminal is met, selecting a proper task split point can reduce a rate of data to be transmitted, thereby increasing a quantity of users who can access a network.

Table 2

Split point (split point)	Approximate output uplink data size (Mbit/s)	Computation load in a terminal
Candidate split point 1	120	Low
Candidate split point 2	24	High

[00115] In conclusion, a task scheduling result may affect a size of intermediate data, a communication requirement, computation load in a terminal, and the like. Therefore, how to determine the task scheduling result is critical. The task scheduling result may indicate a manner based on which at least one network node executes a task, in other words, indicate a specific subtask that is in the task and to be executed by the network node.

[00116] 5. Alternative quality of service (quality of service, QoS) mechanism

[00117] To reduce a risk of releasing quality of service flows (QoS flows) caused by limited RAN resources, a RAN node may indicate a task scheduling result or a service traffic mode to a network node according to the alternative quality of service mechanism. The service procedure mode indicates a bit rate, a code rate, a frame rate, a resolution, or the like. Specifically, different alternative quality of service configuration information (or alternative quality of service profiles (QoS profiles)) may be used to describe quality of service requirements of different task split points or different service traffic modes. In this way, the RAN node may select proper alternative quality of service configuration information based on a current communication status, and indicate the selected alternative quality of service configuration information to the network node, so that the network node determines the task scheduling result or the service traffic mode.

[00118] The following uses an example in which a network node includes a terminal and an application server to describe a specific process in which a RAN node indicates a task scheduling result or a service traffic mode to the network node according to an alternative quality of service

mechanism. As shown in FIG. 1C, the process may include the following steps.

[00119] S101: The RAN node obtains at least one piece of alternative quality of service configuration information from a core network.

[00120] S102: The RAN node determines target quality of service configuration information from the at least one piece of alternative quality of service configuration information based on a current network status.

[00121] It may be understood that, if the RAN node cannot meet quality of service required by current quality of service configuration information, for example, the RAN node cannot meet a guaranteed flow bit rate (guaranteed flow bit rate, GFBR), a packet delay budget (packet delay budget, PDB) (which may be an uplink PDB, a downlink PDB, or a sum of uplink and downlink PDBs), a packet error rate (packet error rate, PER), or the like required by the current quality of service configuration information, the RAN node may determine, one by one based on a priority sequence of the at least one piece of alternative quality of service configuration information, whether a corresponding quality of service requirement can be met. If the RAN node can meet a GFBR, a PDB, or a PER required by a specific piece of alternative quality of service configuration information, the RAN node determines to use the matched alternative quality of service configuration information, to provide a service for a corresponding quality of service flow. The matched alternative quality of service configuration information is the target quality of service configuration information.

[00122] S103: The RAN node sends, to the core network, indication information that indicates the target quality of service configuration information. Correspondingly, the core network receives the indication information from the RAN node.

[00123] For example, the RAN node sends the indication information to a session management function (session management function, SMF) network element in the core network.

[00124] S104: The core network sends the indication information to the terminal. Correspondingly, the terminal receives the indication information from the core network.

[00125] For example, the SMF network element sends the indication information to the terminal through non-access stratum (non-access stratum, NAS) signaling, so that the terminal determines, based on the indication information, that the current quality of service configuration information changes, or determines the task scheduling result or the service traffic mode.

[00126] S105: The core network sends the indication information to the application server. Correspondingly, the application server receives the indication information from the core network.

[00127] For example, the SMF network element sends the indication information to the application server, so that the application server determines, based on the indication information, that the current quality of service configuration information changes, or determines the task

scheduling result or the service traffic mode.

[00128] It may be understood that the core network may perform S104 before S105, perform S105 before S104, or simultaneously perform S104 and S105. This is not limited.

[00129] It may be understood that, when the network status is restored and the RAN can meet the previous quality of service configuration information, the RAN node may notify the terminal and the application server over the core network, so that the terminal and the application server use previous task splitting to execute a task, or use a previous service traffic mode to process service data.

[00130] In the foregoing process, because the quality of service configuration information is related to the task scheduling result or the service traffic mode, quality of service provided by the current network may meet the quality of service requirement of the task, to ensure user experience. However, in actual application, a channel state constantly changes. Therefore, after determining the target quality of service configuration information, the RAN node needs to indicate the target quality of service configuration information to the terminal and/or the application server in time, so that the terminal and the application server determine the task scheduling result or the service traffic mode in time, thereby avoiding a case in which user experience is affected because quality of service provided by the current network cannot meet the quality of service requirement of the task. Therefore, how the RAN node indicates the target quality of service configuration information to the terminal and/or the application server in time is critical.

[00131] To resolve the foregoing problem, this application provides a communication method. In the method, a second RAN node may obtain first indication information, and send the first indication information to a first RAN node. The first indication information indicates a transmission mode of second indication information, and the second indication information indicates quality of service configuration information associated with a first service. The transmission mode of the second indication information includes: sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service. The first RAN node is a RAN node to which the second node is to be handed over or is potentially to be handed over, or the first RAN node is a secondary RAN node that provides dual connectivity for the second node.

[00132] In the foregoing process, the second RAN node may indicate the transmission mode of the second indication information to the first RAN node, so that the first RAN node may send the second indication information to the first node/second node based on the transmission mode indicated by the second RAN node. In this way, the first node/second node quickly obtains the second indication information.

[00133] The following describes implementations of this application in detail with reference to accompanying drawings.

[00134] The method provided in this application may be applied to various communication systems. For example, the communication system may be a universal mobile telecommunications system (universal mobile telecommunications system, UMTS), an LTE system, a 5G communication system, a wireless fidelity (wireless fidelity, Wi-Fi) system, a 3GPP-related communication system, a post-5G evolved communication system (for example, a sixth generation (6th generation, 6G) communication system), or a system integrating a plurality of systems. This is not limited. 5G may also be referred to as new radio (new radio, NR). The method provided in this application is described below by using a communication system 1000 shown in FIG. 2 as an example. FIG. 2 is merely a diagram, and does not constitute a limitation on an application scenario of the technical solutions provided in this application.

[00135] FIG. 2 is a diagram of an architecture of a communication system 1000 according to this application. In FIG. 2, the communication system 1000 includes a RAN 100 and a core network (core network, CN) 200. The RAN 100 includes at least one RAN node (for example, 110a and 110b in FIG. 2, which are collectively referred to as 110) and at least one terminal (for example, 120a to 120j in FIG. 2, which are collectively referred to as 120). The RAN 100 may further include another RAN node, for example, a wireless relay device and/or a wireless backhaul device (not shown in FIG. 2). The terminal 120 is connected to the RAN node 110 in a wireless manner. The RAN node 110 is connected to the core network 200 in a wireless or wired manner. A core network device in the core network 200 and the RAN node 110 in the RAN 100 may be different physical devices, or may be a same physical device that integrates a core network logical function and a radio access network logical function.

[00136] The RAN 100 may be a 3GPP-related cellular system, for example, a 4G or 5G mobile communication system, or a future-oriented evolved system (for example, a 6G mobile communication system). The RAN 100 may alternatively be an open radio access network (open RAN, O-RAN or ORAN), a cloud radio access network (cloud radio access network, CRAN), or a Wi-Fi system. The RAN 100 may alternatively be a communication system that integrates the foregoing two or more systems.

[00137] The RAN node 110 may also be referred to as an access network device, a RAN entity, an access node, or the like sometimes, and forms a part of the communication system, to help the terminal implement radio access. A plurality of RAN nodes 110 in the communication system 1000 may be nodes of a same type, or may be nodes of different types. For other descriptions of the RAN node and the terminal, refer to the foregoing corresponding descriptions.

[00138] The core network 200 may include network elements configured to implement various

core network functions, for example, include a session management function (session management function, SMF) network element, an access and mobility management function (access and mobility management function, AMF) network element, a user plane function (user plane function, UPF) network element, and the like.

[00139] Optionally, the communication system 1000 further includes an application server 300. The application server 300 may communicate with the core network 200. For example, the application server 300 may receive information from the RAN 100 over the core network 200. Optionally, the application server is located in a data network after the user plane function in the core network, and performs exchanging of user plane application layer data through an N6 interface, to provide computing and data services.

[00140] It may be understood that the communication system 1000 shown in FIG. 2 is merely used as an example, and is not intended to limit the technical solutions in this application. Persons skilled in the art should understand that, in a specific implementation process, the communication system 1000 may further include another device, and a quantity of RAN nodes and a quantity of terminals may alternatively be determined based on a specific requirement. This is not limited.

[00141] Optionally, each network element or device (for example, the RAN node, the terminal, or the access and mobility management network element) in FIG. 2 of this application may also be referred to as a communication apparatus, and may be a general-purpose device or a dedicated device. This is not specifically limited in this application.

[00142] Optionally, related functions of each network element or device (for example, the RAN node, the terminal, or the access and mobility management network element) in FIG. 2 in this application may be implemented by one device, may be jointly implemented by a plurality of devices, or may be implemented by one or more functional modules in one device. This is not specifically limited in this application. It may be understood that the foregoing function may be a network element in a hardware device, a software function running on dedicated hardware, a combination of hardware and software, or a virtualization function instantiated on a platform (for example, a cloud platform).

[00143] During specific implementation, each network element or device (for example, the RAN node, the terminal, or the access and mobility management network element) in FIG. 2 of this application may use a composition structure shown in FIG. 3, or include components shown in FIG. 3. FIG. 3 is a diagram of a hardware structure of a communication apparatus to which this application is applicable. It may be understood that the communication apparatus 30 includes a means in a necessary form, for example, a module, a unit, an element, a circuit, or an interface, which is properly configured together to implement the solutions provided in this application. For example, the communication apparatus 30 includes one or more processors 301, configured to

implement the method provided in this application.

[00144] The processor 301 may be a general-purpose processor, a dedicated processor, or the like. For example, the processor 301 may be a baseband processor or a central processing unit (central processing unit, CPU). The baseband processor may be configured to process a communication protocol and communication data. The central processing unit may be configured to: control the communication apparatus 30 (for example, a RAN node, a terminal, an access and mobility management network element, or a chip), execute a software program, and process data of the software program. Optionally, in a design, the processor 301 may include a program 305 (which may sometimes also be referred to as code or instructions). The program 305 may be run on the processor 301, to cause the communication apparatus 30 to perform the method described in the following embodiments. In another possible design, the communication apparatus 30 includes a circuit (not shown in FIG. 3), and the circuit is configured to implement the method in the following embodiments.

[00145] Optionally, the communication apparatus 30 may include one or more memories 303. The memory 303 may be a read-only memory (read-only memory, ROM), another type of static storage device that can store static information and instructions, a random access memory (random access memory, RAM), a cache (cache), another type of dynamic storage device that can store information and instructions, an electrically erasable programmable read-only memory (electrically erasable programmable read-only memory, EEPROM), a compact disc read-only memory (compact disc read-only memory, CD-ROM), another compact disc storage medium, an optical disc storage medium (including a compact disc, a laser disc, an optical disc, a digital versatile disc, a Blu-ray disc, or the like), a magnetic disk storage medium, another magnetic storage device, or any other medium that can be configured to carry or store expected program code in a form of instructions or a data structure and is accessible by a computer, but is not limited thereto. Usually, the memory provided in this application may be non-volatile. Optionally, the memory 303 stores a program 307 (which may sometimes also be referred to as code or instructions). The program 307 may be run on the processor 301, to cause the communication apparatus 30 to perform the method described in the following method embodiments.

[00146] Optionally, the processor 301 may include an AI module 306, and/or the memory 303 may include an AI module 308. The AI module is configured to implement an AI-related function. The AI module may be implemented by using software, hardware, or a combination of software and hardware. For example, the AI module may include a RIC module. For example, the AI module may be a near-real-time RIC or a non-real-time RIC.

[00147] Optionally, the processor 301 and/or the memory 303 may further store data. The processor 301 and the memory 303 may be separately disposed, or may be integrated together.

[00148] Optionally, the communication apparatus 30 may further include a transceiver 302 and/or an antenna 304. The processor 301 may sometimes also be referred to as a processing unit, and controls the communication apparatus 30. The transceiver 302 may sometimes also be referred to as a transceiver unit, a transceiver machine, a transceiver circuit, a transceiver, or the like, and is configured to implement receiving and sending functions of the communication apparatus 30 through the antenna 304.

[00149] It may be understood that the composition structure shown in FIG. 3 does not constitute a limitation on the communication apparatus. In addition to the components shown in FIG. 3, the communication apparatus may include more or fewer components than those shown in the figure, some components may be combined, or different component arrangements may be used.

[00150] The following describes the method provided in this application with reference to the accompanying drawings. Network elements in the following embodiments may have the components shown in FIG. 3. Details are not described again.

[00151] It may be understood that the "connection" in this application may be a direct connection or an indirect connection. In addition, the "connection" may be an electrical connection or a communication connection. For example, that two electrical elements A and B are connected may mean that A and B are directly connected, or may mean that A and B are indirectly connected by using another electrical element or a connection medium, so that electrical signal transmission can be performed between A and B. For another example, that two devices A and B are connected may mean that A and B are directly connected, or may mean that A and B are indirectly connected via another communication device or a communication medium, so that A and B can communicate with each other.

[00152] It may be understood that names of messages between network elements, names of parameters in messages, or the like in the following embodiments of this application are merely examples, and there may alternatively be other names during specific implementation. This is not specifically limited in this application.

[00153] It may be understood that "sending information to... (for example, a terminal)" in this application may be understood as that a destination end of the information is the terminal, and may include directly or indirectly sending the information to the terminal. "Receiving information from... (for example, the terminal)" may be understood as that a source end of the information is the terminal, and may include directly or indirectly receiving the information from the terminal. Information may undergo necessary processing, for example, a format change, between the source end for sending the information and the destination end. However, the destination end may understand valid information from the source end. Similar descriptions in this application may be understood similarly. Details are not described herein again.

[00154] It may be understood that, in this application, "/" may indicate an "or" relationship between associated objects. For example, A/B may indicate A or B. "And/or" may be used to describe three relationships between associated objects. For example, A and/or B may indicate three cases: Only A exists, both A and B exist, and only B exists. A and B may be singular or plural. In addition, a representation similar to "at least one of A, B, and C" or "at least one of A, B, or C" is usually used to represent any one of the following: Only A exists; only B exists; only C exists; both A and B exist; both A and C exist; both B and C exist; and A, B, and C all exist. The foregoing uses three elements A, B, and C as an example to describe an optional item of the project. When there are more elements in the representation, a meaning of the representation may be obtained according to the foregoing rules.

[00155] For ease of description of the technical solutions of this application, in this application, the terms "first", "second", and the like may be used to distinguish between technical features with a same function or similar functions. Terms such as "first" and "second" do not limit a quantity and an execution order, and the terms such as "first" and "second" do not indicate a definite difference. In this application, terms such as "example" or "for example" represent giving an example, an illustration, or a description. Any embodiment or design scheme described with "example" or "for example" should not be explained as being more preferred or having more advantages than another embodiment or design scheme. Use of the terms such as "example" or "for example" is intended to present a related concept in a specific manner for ease of understanding.

[00156] It may be understood that an "embodiment" mentioned throughout this specification means that particular features, structures, or characteristics related to this embodiment are included in at least one embodiment of this application. Therefore, embodiments in the entire specification do not necessarily refer to a same embodiment. In addition, these particular features, structures, or characteristics may be combined in one or more embodiments by using any proper manner. It may be understood that sequence numbers of processes do not mean an execution order in various embodiments of this application. The execution order of the processes should be determined based on functions and internal logic of the processes, and should not be construed as any limitation on implementation processes of this application.

[00157] It may be understood that in this application, "when ...", "in a case of ...", "provided that", and "if" all mean that corresponding processing is performed in an objective case, and are not intended to limit a time, do not require a determining action during implementation, and do not mean that there is another limitation.

[00158] "Simultaneously" in this application may be understood as being at a same time point, may be understood as being in a time period, or may be understood as being in a same periodicity.

[00159] It may be understood that in some scenarios, some optional features in this application may be independently implemented without depending on other features, for example, a solution on which the optional features are currently based, to resolve a corresponding technical problem and achieve corresponding effects. Alternatively, in some scenarios, the optional features may be combined with other features based on a requirement. Correspondingly, an apparatus provided in this application may also correspondingly implement these features or functions. Details are not described herein.

[00160] It may be understood that, in this application, a same step, or steps or technical features that have a same function may be mutually referenced between different embodiments.

[00161] It may be understood that in this application, the first RAN node and/or the second RAN node and/or the access and mobility management network element may perform some or all of the steps in this application. These steps are merely examples. In this application, other steps or variations of various steps may be further performed. In addition, the steps may be performed in a sequence different from that presented in this application, and not all the steps in this application may be performed.

[00162] It may be understood that, in the following method provided in this application, an example in which the first RAN node, the second RAN node, and the access and mobility management network element are used as execution bodies of an interaction example is used to illustrate the method. However, the execution bodies of the interaction example are not limited in this application. For example, the first RAN node in the method provided in the following embodiments of this application may alternatively be a chip, a chip system, or a processor that supports the first RAN node in implementing the method, or may be a logical node, a logical module, or software that can implement all or some functions of a server. The second RAN node in the method provided below in this application may alternatively be a chip, a chip system, or a processor that supports the second RAN node in implementing the method, or may be a logical node, a logical module, or software that can implement all or some functions of the second RAN node. The access and mobility management network element in the following method provided in this application may alternatively be a chip, a chip system, or a processor that supports the access and mobility management network element in implementing the method, or may be a logical node, a logical module, or software that can implement all or some functions of the access and mobility management network element.

[00163] FIG. 4 shows a communication method according to this application. The method may include the following steps.

[00164] S401: A second RAN node obtains first indication information.

[00165] In this application, the second RAN node may be any RAN node in the communication

system 1000 shown in FIG. 2. The first indication information may indicate a transmission mode of second indication information. The transmission mode of the second indication information includes sending the second indication information to a first node through a user plane path, and/or sending the second indication information to a second node through air interface signaling. The first node and the second node are network nodes associated with a first service. For descriptions of the network nodes, refer to the foregoing corresponding descriptions. For example, the first node is an application server, and the second node is a terminal. In addition, because transmission of the second indication information may be performed through the user plane path, the first node may alternatively be a user plane network element (for example, a UPF network element). It may be understood that after receiving the second indication information, the user plane network element may send the second indication information to the application server.

[00166] Optionally, the air interface signaling includes at least one of the following: downlink control information (downlink control information, DCI), a medium access control control element (medium access control control element, MAC CE), a packet data convergence protocol (packet data convergence protocol, PDCP) control protocol data unit (protocol data unit, PDU) (PDCP control PDU) message, or a radio resource control (radio resource control, RRC) message.

[00167] In an example, the first indication information includes at least one bit, and a value of the at least one bit indicates the transmission mode of the second indication information. For example, the first indication information includes one bit. When a value of the one bit is "0" or "1", it indicates that the second indication information is sent to the first node through the user plane path, or indicates that the second indication information is sent to the second node through the air interface signaling. For another example, the first indication information includes two bits. When a value of the two bits is "11" or "00", it indicates that the second indication information is sent to the first node through the user plane path, and indicates that the second indication information is sent to the second node through the air interface signaling. When a value of the two bits is "01", it indicates that the second indication information is sent to the second node through the air interface signaling. When a value of the two bits is "10", it indicates that the second indication information is sent to the first node through the user plane path.

[00168] In another example, the first indication information includes an identifier of the transmission mode, and the identifier indicates the transmission mode of the second indication information. For example, when the first indication information includes "upf", "upfQoSnotification", or "up", it indicates that the second indication information is sent to the first node through the user plane path; when the first indication information includes "uu", "dci", "macce", "pdcppdu", or "rrc", it indicates that the second indication information is sent to the second node through the air interface signaling; or when the first indication information includes

one of "upf", "upfQoSnotification", or "up", and one of "uu", "dci", "macce", "pdcppdu", or "rrc", it indicates that the second indication information is sent to the first node through the user plane path, and indicates that the second indication information is sent to the second node through the air interface signaling.

[00169] In a possible design, the second indication information indicates quality of service configuration information associated with the first service. For example, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control (QoS notification control, QNC) information.

[00170] In this application, a quality of service profile associated with the first service is a quality of service profile determined according to the foregoing alternative quality of service mechanism. The quality of service profile includes the quality of service profile index information and the quality of service configuration parameter information. The quality of service profile index information may identify the quality of service profile. The quality of service configuration parameter information may include at least one of a GFBR, an uplink PDB, a downlink PDB, a sum of uplink and downlink PDBs, or a PER. The quality of service profile adjustment information and the quality of service notification control information may indicate that the quality of service profile changes or is adjusted. Therefore, after receiving the second indication information, the first node and the second node may determine that a network side updates the quality of service profile, and the first node and the second node may re-determine at least one of a task scheduling result of the first service or a service traffic mode of the first service.

[00171] It may be understood that the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service or quality of service configuration information of a quality of service flow of the first service. In other words, the quality of service configuration information associated with the first service may be at a session (for example, a PDU session) granularity or a quality of service flow granularity. This is not limited.

[00172] In a possible implementation, the second RAN node obtains the first indication information from a core network or a RAN node other than the second RAN node.

[00173] In an example, a first network element in the core network, for example, a session management network element, sends the first indication information to the second RAN node. Correspondingly, the second RAN node receives the first indication information from the first network element. For example, the first network element detects that the session of the first service is established, and sends the first indication information to the second RAN node.

[00174] In another example, a second network element in the core network, for example, an access and mobility management network element, sends the first indication information to the second RAN node. Correspondingly, the second RAN node receives the first indication information from the second network element. For example, a third RAN node different from the second RAN node sends the first indication information to the second RAN node via the second network element. The third RAN node is a RAN node accessed by the second node before the second RAN node is accessed.

[00175] In another example, a fourth RAN node sends the first indication information to the second RAN node. Correspondingly, the second RAN node receives the first indication information from the fourth RAN node. The fourth RAN node is a RAN node accessed by the second node before the second RAN node is accessed, or the fourth RAN node is a master radio access network node that provides dual connectivity for the second node. That is, the second RAN node and the fourth RAN node provide dual connectivity for the second node. For example, after obtaining the first indication information in any one of the foregoing manners, the fourth RAN node may send the first indication information to the second RAN node.

[00176] S402: The second RAN node sends the first indication information to a first RAN node. Correspondingly, the first RAN node receives the first indication information from the second RAN node.

[00177] In this application, the first RAN node may be a RAN node that is in the communication system 1000 shown in FIG. 2 and different from the second RAN node.

[00178] In a possible design, the first RAN node is a RAN node to which the second node is to be handed over or is potentially to be handed over. In other words, when determining that the second node is to be handed over to the first RAN node or may be handed over to the first RAN node, the second RAN node may send the first indication information to the first RAN node, so that after determining the second indication information, the first RAN node may send the second indication information to the first node and/or the second node in time. In this case, the first indication information may be carried in a handover request (handover request) message.

[00179] In another possible design, the first RAN node is a secondary RAN node that provides dual connectivity for the second node. It may be understood that, in this case, the second RAN node may be a master RAN node that provides dual connectivity for the second node. That is, the second RAN node and the first RAN node may provide dual connectivity for the second node. In this case, the first indication information may be carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

[00180] In conclusion, the communication method provided in this application is applicable to

a scenario in which the second node moves, so that the RAN node to which the second node is to be handed over or is potentially to be handed over, or the secondary RAN node that provides dual connectivity for the second node determines the transmission mode of the second indication information in time, thereby quickly sending the second indication information to the second node and/or the first node, and avoiding a case in which quality of service provided by a current network cannot meet a quality of service requirement of the first service. Therefore, user experience can be improved.

[00181] It may be understood that the second RAN node may directly send the first indication information to the first RAN node, or may send the first indication information to the first RAN node via the second network element. This is not limited. In this application, that the second RAN node sends the first indication information to the first RAN node via the second network element may be understood as: The second RAN node sends the first indication information to the second network element. After receiving the first indication information, the second network element directly sends the first indication information to the first RAN node; or processes the first indication information, for example, performs an operation like encapsulation, information addition, or information deletion, and sends processed first indication information to the first RAN node.

[00182] Optionally, after receiving the first indication information, the first RAN node may determine not to send the second indication information to the first node through the user plane path, or may determine to send the second indication information to the first node through the user plane path. For example, if the first RAN node does not support sending of the second indication information through the user plane path, or the first RAN node supports sending of the second indication information through a control plane path but does not support sending of the second indication information through the user plane path, the first RAN node may determine not to send the second indication information through the user plane path. Alternatively, for an objective of network optimization (for example, considering load balancing), the first RAN node may determine not to send the second indication information through the user plane path.

[00183] Optionally, when determining not to send the second indication information to the first node through the user plane path, the first RAN node may send third indication information to the second RAN node. Correspondingly, the second RAN node receives the third indication information from the first RAN node. The third indication information indicates that sending of the second indication information through the user plane path is not supported, or indicates that second indication information is not sent through the user plane path. For example, the third indication information includes one bit. When a value of the one bit is "0" or "1", it indicates that sending of the second indication information through the user plane path is not supported, or

indicates that the second indication information is not sent through the user plane path. Subsequently, the second RAN node may indicate, to the core network or the application server, that there is a delay in receiving the second indication information.

[00184] It may be understood that the first RAN node may directly send the third indication information to the second RAN node, or may send the third indication information to the second RAN node via the second network element. For example, if the first RAN node directly receives the first indication information from the second RAN node, the first RAN directly sends the second indication information to the second RAN node; or if the first RAN node receives the first indication information via the second network element, the first RAN sends the second indication information to the second RAN node via the second network element.

[00185] In this application, that the first RAN node sends the third indication information to the second RAN node via the second network element may be understood as: The first RAN node sends the third indication information to the second network element. After receiving the third indication information, the second network element directly sends the third indication information to the second RAN node; or processes the third indication information, for example, performs an operation like encapsulation, information addition, or information deletion, and sends processed third indication information to the second RAN node.

[00186] S403: The first RAN node determines the second indication information based on at least one piece of alternative quality of service configuration information.

[00187] In a possible implementation, the first RAN node may obtain the at least one piece of alternative quality of service configuration information from the core network or the second RAN node, and determine target quality of service configuration information (that is, the quality of service configuration information associated with the first service) from the at least one piece of alternative quality of service configuration information based on a current network status. For this process, refer to the descriptions of S102. Subsequently, the first RAN node may determine the second indication information based on the target quality of service configuration information.

[00188] S404: The first RAN node sends the second indication information based on the first indication information.

[00189] It may be understood that if the first indication information indicates to send the second indication information to the first node through the user plane path, the first RAN node sends the second indication information to the first node through the user plane path. For example, the first RAN node sends the second indication information to the user plane network element through the user plane path, and after receiving the second indication information, the user plane network element sends the second indication information to the application server. The second indication information sent by the first RAN node to the user plane network element may be included in a

header, of a user plane part of a tunneling protocol (for example, a user plane part of a general packet radio service (general packet radio service, GPRS) tunneling protocol (user plane part of GPRS tunneling protocol, GTP-U)), of one or more pieces of data transmitted through an NG3 interface. The second indication information sent by the user plane network element to the application server may be sent to the application server in a capability exposure manner or in a piggyback manner. The capability exposure manner may be understood as that the user plane network element sends the second indication information to the application server via a capability exposure network element or a local capability exposure network element through an application programming interface (application programming interface, API). The piggyback manner may be understood as that the second indication information is carried in a header of a real-time transport protocol (real-time transport protocol, RTP), an RTP control protocol (RTP control protocol, RTCP), or a quick user datagram protocol (user datagram protocol, UDP) internet connection (quick UDP internet connection, QUIC). Optionally, after receiving the second indication information, the application server may send the second indication information to the second node.

[00190] It may be understood that, if the first indication information indicates to send the second indication information to the second node through the air interface signaling, the first RAN node sends the second indication information to the second node through the air interface signaling. For example, the second RAN node sends the second indication information to the second node by using the DCI, the MAC CE, the PDCP control PDU message, or the RRC message. Optionally, after receiving the second indication information, the second node may send the second indication information to the application server.

[00191] It may be understood that if the first indication information indicates to send the second indication information to the first node through the user plane path, and indicates to send the second indication information to the second node through the air interface signaling, the first RAN node sends the second indication information to the first node through the user plane path, and sends the second indication information to the second node through the air interface signaling.

[00192] In a possible implementation, after receiving the second indication information, the first node or the second node may determine the task scheduling result of the first service or the service traffic mode of the first service based on the second indication information. In other words, the first node or the second node may adjust the task scheduling result of the first service or the service traffic mode of the first service in time.

[00193] The task scheduling result of the first service indicates a split point of a task associated with the first service. The task scheduling result of the first service further indicates at least one of the following: a quantity of tasks executed by the first node or a quantity of tasks executed by the second node. The service traffic mode of the first service indicates at least one of the following: a

service feature of the first service, a service rate of the first service, a bit rate corresponding to the first service, a code rate corresponding to the first service, a frame rate corresponding to the first service, a resolution corresponding to the first service, or enabling/disabling of a rendering function. The service feature of the first service may represent importance of data of the first service. Alternatively, when the first service is an AI service, the service feature of the first service may represent a feature map (feature map) of an AI model. In this application, different service features may be associated with different service QoS. Therefore, the first node or the second node may determine corresponding quality of service of the service based on the service feature of the first service.

[00194] It may be understood that actions of the first RAN node, the second RAN node, or the second network element in the foregoing steps may be performed by the processor 301 in the communication apparatus 30 shown in FIG. 3 by invoking application code stored in the memory 303. This is not limited in this application.

[00195] According to the method shown in FIG. 4, in the scenario in which the second node moves, the second RAN node may indicate the transmission mode of the second indication information to the first RAN node, so that after obtaining the second indication information, the first RAN node sends the second indication information to the first node and/or the second node in time based on the indication of the second RAN node. Because the second indication information is determined based on an air interface channel state, the second indication information is sent to the first node and/or the second node in time, so that the first node and/or the second node can adjust the task scheduling result of the first service or the service traffic mode of the first service in time based on the air interface channel state, thereby avoiding packet loss during transmission of the data of the first service.

[00196] It may be understood that, in the scenario in which the second node moves, a case of inter-base station handover of the second node occurs. The inter-base station handover may cause the second node to enjoy an uninterrupted service regardless of how the second node moves in a network coverage area. However, when the second node performs inter-base station handover, quality of service change is caused because quality of service between the second node and the target cell fluctuates significantly. According to the method shown in FIG. 4, the first RAN node to which a target cell belongs may quickly notify the first node and/or the second node of quality of service change, so that the first node and/or the second node adjust/adjusts the task scheduling result of the first service or the service traffic mode of the first service in time.

[00197] Optionally, the foregoing "handover" may be an RRC connected mode mobility management process. For example, the process includes a serving cell change process in a standalone (standalone, SA) mode, or includes a primary cell (primary cell, PCell) change process

or a primary secondary cell (primary secondary cell, PSCell) change process in a non-standalone (non-standalone, NSA) mode.

[00198] In the SA networking mode, handover can be classified into intra-base station handover, inter-base station Xn-based handover, and inter-base station NG-based handover depending on whether RAN nodes to which serving cells before and after handover belong are inter-base station or whether there is a corresponding Xn interface between the RAN nodes in an inter-base station scenario. The intra-base station handover in the SA networking means that the second node is handed over from one cell to another cell, and the two cells belong to a same RAN node. The inter-base station Xn-based handover in the SA networking means that the second node is handed over from one cell to another cell, and the two cells belong to different RAN nodes. However, the RAN nodes to which the two cells belong are located in an area managed by a same access and mobility management network element, and there is an Xn interface between the RAN nodes to which the two cells belong. The inter-base station NG-based handover in the SA networking means that the second node is handed over from one cell to another cell, and the RAN nodes to which the two cells belong are different. The RAN nodes to which the two cells belong are not located in an area managed by a same access and mobility management network element. Alternatively, the RAN nodes to which the two cells belong are different, the RAN nodes to which the two cells belong are located in an area managed by a same access and mobility management network element, but there is no Xn interface between the RAN nodes to which the two cells belong.

[00199] In an NSA networking mode, handover may include a secondary gNB addition (SgNB addition) process, a secondary gNB change (SgNB change) process, or an inter-node anchor handover (MeNB handover) process. In the NSA mode, a RAN node in an LTE network may be used as a master node (master node), and a RAN node in 5G NR may be used as a secondary node (secondary node). This is referred to as EN-DC (EUTRA-NR dual connectivity) or MR-DC (multi-RAT dual connectivity) in 3GPP.

[00200] To better understand the method provided in this application, the following specifically describes the method provided in this application with reference to a plurality of handover scenarios. The following method is described by using an example in which a first node is a user plane network element, a second node is a terminal, a first network element is a session management network element, and a second network element is an access and mobility management network element.

[00201] First, an inter-base station Xn-based handover process in SA networking is used as an example to describe the communication method provided in this application.

[00202] FIG. 5 shows another communication method according to this application. The communication method may include the following steps.

[00203] S501: A session management network element sends first indication information to a second RAN node. Correspondingly, the second RAN node receives the first indication information from the session management network element.

[00204] For example, the session management network element may send the first indication information to the second RAN node via an access and mobility management network element. Correspondingly, the second RAN node receives the first indication information from the session management network element via the access and mobility management network element.

[00205] S502: The second RAN node sends measurement configuration information to a terminal. Correspondingly, the terminal receives the measurement configuration information from the second RAN node.

[00206] It may be understood that the measurement configuration information is used to configure the terminal to perform measurement. For example, the measurement configuration information is used to configure one or more of a measurement object (measurement object), a measurement gap (measurement gap), a reporting configuration (reporting configuration), a triggering quantity (triggering quantity), or a measurement identity (measurement identity). The measurement object is an object on which the terminal performs measurement. For example, the measurement object includes a synchronization signal and physical broadcast channel (physical broadcast channel, PBCH) block (synchronization signal and PBCH block, SSB) frequency, an SSB subcarrier spacing, an SSB-based measurement timing configuration (SSB-based measurement timing configuration, SMTC), an allow-listed cell, or a forbid-listed cell. The measurement gap is a time period during which the terminal leaves a current frequency to perform measurement on another frequency. Therefore, the measurement gap can be configured when inter-frequency measurement or inter-RAT measurement is involved. The reporting configuration may be used to configure a standard for triggering the terminal to report a measurement report and a format of the measurement report. The triggering quantity is a policy for triggering event reporting. The measurement identity is used to link the measurement object and the reporting configuration as a set.

[00207] Optionally, the measurement configuration information may be carried in an RRC message, for example, an RRC reconfiguration (RRCReconfiguration) message.

[00208] Optionally, after receiving the RRC reconfiguration message, the terminal may send an RRC reconfiguration complete (RRCReconfigurationComplete) message to the second RAN node, to indicate that the measurement configuration information is configured.

[00209] S503: The terminal performs measurement based on the measurement configuration information.

[00210] S504: The terminal sends the measurement report to the second RAN node.

Correspondingly, the second RAN node receives the measurement report from the terminal.

[00211] For example, after an event condition is met, the terminal sends the measurement report to the second RAN node.

[00212] S505: The second RAN node sends a handover request message to a first RAN node. Correspondingly, the first RAN node receives the handover request message from the second RAN node.

[00213] For example, after receiving the measurement report, the second RAN node may determine, based on a measurement result indicated by the measurement report, whether to perform handover. If the second RAN node determines that the terminal is to be handed over to the first RAN node, the second RAN node sends the handover request message to the first RAN node. The handover request message may carry the first indication information. It may be understood that the handover request is an Xn-based handover request message.

[00214] S506: The first RAN node sends a handover request acknowledge (HANDOVER REQUEST ACKNOWLEDGE) message to the second RAN node. Correspondingly, the second RAN node receives the handover request acknowledge message from the first RAN node.

[00215] It may be understood that, after receiving the handover request message, the first RAN node may perform admission control, and send the handover request acknowledge message to the second RAN node when access by the terminal is allowed. The handover request acknowledge message includes a transmission resource allocated by the first RAN node to the terminal.

[00216] Optionally, if handover of a part of PDU sessions (PDU sessions) fails, the handover request acknowledge message may indicate the PDU sessions whose handover fails.

[00217] It may be understood that if the first RAN node determines not to send second indication information to a user plane network element through a user plane path, the handover request acknowledge message may further carry third indication information.

[00218] S507: The second RAN node sends a handover indication to the terminal. Correspondingly, the terminal receives the handover indication from the second RAN node.

[00219] The handover indication indicates the terminal to be handed over to a cell of the first RAN node. The handover indication may be carried in an RRC reconfiguration message.

[00220] Optionally, the second RAN node may further send a PDCP sequence number (sequence number, SN) to the first RAN node by using a sequence number status transfer (SN STATUS TRANSFER) message, so that the first RAN node determines a specific data packet that has been sent by the second RAN node to the terminal and a specific data packet that has not been sent to the terminal.

[00221] Optionally, the terminal initiates non-contention-based random access to the cell of the first RAN node, and may include a dedicated preamble (preamble). After receiving a random

access request, the first RAN node may send a path switch request (PATH SWITCH REQUEST) message to the access and mobility management network element, to indicate to notify that the serving cell has been changed for the terminal. The message may carry an identifier of a target cell to be accessed by the terminal, a list of transferred PDU sessions, and the like. After receiving the message, a core network may update a downlink GTP-U data plane path address, for example, change a GTP-U address of a RAN side to an address of the first RAN node. Subsequently, the first RAN node may send a terminal context release (UE CONTEXT RELEASE) message to the second RAN node, so that the second RAN node releases a context of the terminal.

[00222] S508: The first RAN node determines the second indication information based on at least one piece of alternative quality of service configuration information.

[00223] S509: The first RAN node sends the second indication information based on the first indication information.

[00224] It may be understood that after allocating the transmission resource to the terminal, the first RAN node may determine the second indication information based on the at least one piece of alternative quality of service configuration information. Therefore, S508 and S509 may be performed in a handover process, or may be performed after the handover. For example, S508 and S509 may be performed simultaneously with S506, or may be performed after S506. This is not limited. For other descriptions of S508 and S509, refer to the corresponding descriptions of S403 and S404. Details are not described again.

[00225] It may be understood that actions of the first RAN node or the second RAN node in the foregoing steps may be performed by the processor 301 in the communication apparatus 30 shown in FIG. 3 by invoking application code stored in the memory 303. This is not limited in this application.

[00226] According to the method shown in FIG. 5, in the inter-base station Xn-based handover process in SA networking, the second RAN node (that is, a source RAN node) may indicate a transmission mode of the second indication information to the first RAN node (that is, a target RAN node), so that after obtaining the second indication information, the first RAN node may send the second indication information to the terminal and/or the user plane network element in time based on the indication of the second RAN node. Subsequently, the user plane network element may send the second indication information to an application server. Because the second indication information is determined based on an air interface channel state, the second indication information is sent to the terminal and/or the application server in time, so that the terminal and/or the application server can adjust a task scheduling result of a first service or a service traffic mode of the first service in time based on the air interface channel state, thereby avoiding impact on user experience due to the inter-base station Xn-based handover process in SA networking.

[00227] The following uses an inter-base station NG-based handover process in SA networking as an example to describe the communication method provided in this application.

[00228] FIG. 6 shows another communication method according to this application. The communication method may include the following steps.

[00229] S601: A session management network element sends first indication information to a second RAN node. Correspondingly, the second RAN node receives the first indication information from the session management network element.

[00230] S602: The second RAN node sends measurement configuration information to a terminal. Correspondingly, the terminal receives the measurement configuration information from the second RAN node.

[00231] S603: The terminal performs measurement based on the measurement configuration information.

[00232] S604: The terminal sends a measurement report to the second RAN node. Correspondingly, the second RAN node receives the measurement report from the terminal.

[00233] A process of S601 to S604 is similar to a process of S501 to S504. For details, refer to corresponding descriptions of S501 to S504. Details are not described again.

[00234] S605: The second RAN node sends a handover request message to an access and mobility management network element. Correspondingly, the access and mobility management network element receives the handover request message from the second RAN node.

[00235] For example, after receiving the measurement report, the second RAN node may determine, based on a measurement result indicated by the measurement report, whether to perform handover. If the second RAN node determines that the terminal is to be handed over to a first RAN node, the second RAN node sends the handover request message to the access and mobility management network element, so that the access and mobility management network element forwards the handover request message to the first RAN node. The handover request message may carry the first indication information. It may be understood that the handover request is an NG-based handover request message. In other words, the second RAN node sends the handover request message to the access and mobility management network element over an NG link.

[00236] It may be understood that, in S601, if the first indication information sent by the session management network element to the second RAN node via the access and mobility management network element is stored by the access and mobility management network element, the handover request message in S605 may alternatively not carry the first indication information.

[00237] Optionally, the handover request message further includes an identifier of the first RAN node, a PDU session list for performing data forwarding, and the like.

[00238] S606: The access and mobility management network element sends the handover request message to the first RAN node. Correspondingly, the first RAN node receives the handover request message from the access and mobility management network element.

[00239] It may be understood that the handover request message in S606 carries the first indication information. If the handover request message in S605 does not carry the first indication information, the access and mobility management network element may add the first indication information to the received handover request message, and then send, to the first RAN node, the handover request message to which the first indication information is added.

[00240] S607: The first RAN node sends a handover request acknowledge message to the access and mobility management network element. Correspondingly, the access and mobility management network element receives the handover request acknowledge message from the first RAN node.

[00241] It may be understood that, after receiving the handover request message, the first RAN node may perform admission control, and send the handover request acknowledge message to the access and mobility management network element when access by the terminal is allowed. The handover request acknowledge message may include a transmission resource allocated by the first RAN node to the terminal.

[00242] Optionally, if handover of a part of PDU sessions fails, the handover request acknowledge message may indicate the PDU sessions whose handover fails.

[00243] It may be understood that if the first RAN node determines not to send second indication information to a user plane network element through a user plane path, the handover request acknowledge message may further carry third indication information, so that the access and mobility management network element sends the third indication information to the second RAN node by using a handover command in S608.

[00244] S608: The access and mobility management network element sends the handover command (HANDOVER COMMAND) to the second RAN node. Correspondingly, the second RAN node receives the handover command from the access and mobility management network element.

[00245] For example, the handover command may include an address of the first RAN node, a list of tunnel endpoint identifiers (tunnel endpoint identifiers, TEIDs) used for data forwarding, and a list of radio bearers that need to be released by the second RAN node.

[00246] S609: The second RAN node sends a handover indication to the terminal. Correspondingly, the terminal receives the handover indication from the second RAN node.

[00247] The handover indication indicates the terminal to be handed over to a cell of the first RAN node. The handover indication may be carried in an RRC reconfiguration message.

[00248] Optionally, the second RAN node may further send a PDCP SN to the access and mobility management network element by using an uplink RAN status transfer (UPLINK RAN STATUS TRANSFER) message, so that the access and mobility management network element sends the PDCP SN to the first RAN node by using a downlink RAN status transfer (DOWNLINK RAN STATUS TRANSFER) message.

[00249] Optionally, the terminal initiates non-contention-based random access to the cell of the first RAN node, and may include a dedicated preamble (preamble). After receiving a random access request, the first RAN node may send a path switch request (PATH SWITCH REQUEST) message to the access and mobility management network element, to indicate and notify that the serving cell has been changed for the terminal. The message may carry an identifier of a target cell to be accessed by the terminal, a list of transferred PDU sessions, and the like. After receiving the message, a core network may update a downlink GTP-U data plane path address, for example, change a GTP-U address of a RAN side to an address of the first RAN node. Subsequently, the access and mobility management network element may send a terminal context release (UE CONTEXT RELEASE) message to the second RAN node, so that the second RAN node releases a context of the terminal.

[00250] S610: The first RAN node determines the second indication information based on at least one piece of alternative quality of service configuration information.

[00251] S611: The first RAN node sends the second indication information based on the first indication information.

[00252] It may be understood that after allocating the transmission resource to the terminal, the first RAN node may determine the second indication information based on the at least one piece of alternative quality of service configuration information. Therefore, S610 and S611 may be performed in a handover process, or may be performed after the handover. For example, S610 and S611 may be performed simultaneously with S607, or may be performed after S607. This is not limited. For other descriptions of S610 and S611, refer to the corresponding descriptions of S403 and S404. Details are not described again.

[00253] It may be understood that actions of the first RAN node, the second RAN node, or the access and mobility management network element in the foregoing steps may be performed by the processor 301 in the communication apparatus 30 shown in FIG. 3 by invoking application code stored in the memory 303. This is not limited in this application.

[00254] According to the method shown in FIG. 6, in the inter-base station NG-based handover process in SA networking, the second RAN node (that is, a source RAN node) may indicate a transmission mode of the second indication information to the first RAN node (that is, a target RAN node) via the access and mobility management network element, so that after obtaining the

second indication information, the first RAN node may send the second indication information to the terminal and/or the user plane network element in time based on the indication of the second RAN node. Subsequently, the user plane network element may send the second indication information to an application server. Because the second indication information is determined based on an air interface channel state, the second indication information is sent to the terminal and/or the application server in time, so that the terminal and/or the application server can adjust a task scheduling result of a first service or a service traffic mode of the first service in time based on the air interface channel state, thereby avoiding impact on user experience due to the inter-base station NG-based handover process in SA networking.

[00255] The following uses an inter-base station conditional handover (conditional handover, CHO) process as an example to describe the communication method provided in this application.

[00256] FIG. 7 shows another communication method according to this application. The communication method may include the following steps.

[00257] S701: A session management network element sends first indication information to a second RAN node. Correspondingly, the second RAN node receives the first indication information from the session management network element.

[00258] S702: The second RAN node sends measurement configuration information to a terminal. Correspondingly, the terminal receives the measurement configuration information from the second RAN node.

[00259] S703: The terminal performs measurement based on the measurement configuration information.

[00260] S704: The terminal sends a measurement report to the second RAN node. Correspondingly, the second RAN node receives the measurement report from the terminal.

[00261] It may be understood that, in the method shown in FIG. 7, the terminal supports CHO. Therefore, before S702, the terminal may report a capability of the terminal to the second RAN node. For example, the terminal may send, to the second RAN node by using terminal capability information (UECapabilityInformation), a capability of the terminal to support intra-frequency CHO and inter-frequency CHO. In S702, the second RAN node may configure the measurement configuration information based on the capability previously reported by the terminal. For example, the second RAN node may configure a reference signal received power (reference signal received power, RSRP) threshold of a measurement event by using the measurement configuration information. For example, the second RAN node may configure the terminal to report the measurement report when the terminal detects, through measurement, that an RSRP of a neighboring cell is 1 dB higher than that of a source cell. An objective of configuring the threshold by the second RAN node is to cause the terminal to find, as soon as possible, a plurality of

candidate cells that meet the threshold. The plurality of candidate cells belong to at least one candidate RAN node.

[00262] For other descriptions of S701 to S704, refer to corresponding descriptions of S501 to S504. Details are not described again.

[00263] S705: The second RAN node sends a handover request message to the at least one candidate RAN node. Correspondingly, the at least one candidate RAN node receives the handover request message from the second RAN node.

[00264] It may be understood that the second RAN node may send the handover request message to each of the at least one candidate RAN node. The at least one RAN node includes a first RAN node. Any handover request message may be used to request to perform CHO handover. The handover request message further includes the first indication information.

[00265] S706: The at least one candidate RAN node sends a handover request acknowledge message to the second RAN node. Correspondingly, the second RAN node receives the handover request acknowledge message from the at least one candidate RAN node.

[00266] It may be understood that, after receiving the handover request message, the at least one candidate RAN node may perform admission control, and send the handover request acknowledge message to the second RAN node when access by the terminal is allowed. The handover request acknowledge message includes a transmission resource reserved by the at least one candidate RAN node for the terminal. It may be understood that before receiving a handover cancel message sent by the second RAN node, the at least one candidate RAN node always reserves the transmission resource for the terminal.

[00267] Optionally, if handover of a part of PDU sessions (PDU sessions) fails, the handover request acknowledge message may indicate the PDU sessions whose handover fails.

[00268] It may be understood that if a specific candidate RAN node determines not to send second indication information to a user plane network element through a user plane path, the handover request acknowledge message sent by the candidate RAN node may further carry third indication information.

[00269] S707: The second RAN node sends a CHO handover command to the terminal. Correspondingly, the terminal receives the CHO handover command from the second RAN node.

[00270] The CHO handover command instructs the terminal to perform CHO handover. The CHO handover command may be carried in an RRC reconfiguration message.

[00271] Optionally, the CHO handover command may include information about each candidate RAN node and a handover condition. For example, if the handover condition is that the RSRP of the neighboring cell is 3 dB higher than that of the source cell, the handover may be performed. In this case, after finding a cell that meets the condition in the candidate cells, the

terminal may perform handover to the cell autonomously.

[00272] Optionally, after receiving the CHO handover command, the terminal does not immediately initiate a handover action, but continues to maintain a connection and transmission with the second RAN node. The terminal may continuously determine whether there is a cell that meets the handover condition. When detecting that a candidate cell meets the handover condition, the terminal may not send a measurement report to the second RAN node, but may directly perform handover autonomously. If the terminal determines to perform handover to the first RAN node, the terminal may initiate random access and establish an RRC connection to the first RAN node, and the terminal may further release the connection to the second RAN node. For example, the first RAN node may send a handover success message to the second RAN node, and after receiving the message, the second RAN node releases the connection to the terminal.

[00273] Optionally, the second RAN node may further send a PDCP SN to the first RAN node by using a sequence number status transfer (SN STATUS TRANSFER) message, so that the first RAN node determines a specific data packet that has been sent by the second RAN node to the terminal and a specific data packet that has not been sent to the terminal.

[00274] It may be understood that the second RAN node may send a handover cancel message to RAN nodes among the at least one RAN node other than the first RAN node, to indicate these RAN nodes to release reserved resource and buffered data. The first RAN node may further perform path switch (Path Switch), to switch a core network connection from the second RAN node to the first RAN node, thereby completing conditional handover.

[00275] S708: The first RAN node determines the second indication information based on at least one piece of alternative quality of service configuration information.

[00276] S709: The first RAN node sends the second indication information based on the first indication information.

[00277] It may be understood that actions of the at least one candidate RAN node or the second RAN node in the foregoing steps may be performed by the processor 301 in the communication apparatus 30 shown in FIG. 3 by invoking application code stored in the memory 303. This is not limited in this application.

[00278] According to the method shown in FIG. 7, in the inter-base station conditional handover process, the second RAN node (that is, a source RAN node) may indicate a transmission mode of the second indication information to the at least one candidate RAN node. Therefore, when the terminal is handed over to the first RAN node in the at least one candidate RAN node, the first RAN node may send the second indication information to the terminal and/or the user plane network element in time based on the indication of the second RAN node. Subsequently, the user plane network element may send the second indication information to an application server.

Because the second indication information is determined based on an air interface channel state, the second indication information is sent to the terminal and/or the application server in time, so that the terminal and/or the application server can adjust a task scheduling result of a first service or a service traffic mode of the first service in time based on the air interface channel state, thereby avoiding impact on user experience due to the inter-base station conditional handover process.

[00279] It should be understood that the methods shown in FIG. 5 to FIG. 7 are merely examples of a handover scenario. In specific application, the method provided in this application may be further combined with another handover scenario. For example, the method provided in this application may be further applied to a process in which a master node initiates secondary base station addition, secondary base station change, or inter-base station anchor handover to a secondary node in an NSA networking or dual connectivity scenario. Specifically, the master node (for example, the second RAN node) may include the first indication information in the foregoing message sent to the secondary node (for example, the first RAN node), for example, the secondary base station addition request message, the secondary base station change request message, or the inter-base station anchor handover request message, so that the secondary node sends the second indication information to the terminal and/or the application server in time based on the indication of the master node. For another example, the secondary node (for example, the second RAN node) may include the first indication information in a message sent to the master node (for example, the first RAN node), for example, a secondary base station adjustment request message or a secondary base station update request message, so that the master node sends the second indication information to the terminal and/or the application server in time based on the indication of the secondary node.

[00280] It may be understood that the foregoing dual connectivity scenario may include dual connectivity between the terminal and base stations in different technologies, for example, EN-DC or MR-DC, or include dual connectivity between the terminal and base stations in a same technology, for example, NR-DC. For details, refer to the foregoing descriptions of technical terms in this application. It should be understood that the dual connectivity scenario may alternatively be dual connectivity in a CP-UP split scenario. In this scenario, the master node may be configured to perform transmission of a control plane message, and the secondary node may be configured to perform transmission of a user plane message.

[00281] Embodiments mentioned above in this application may be combined when the solutions do not conflict. This is not limited.

[00282] The foregoing mainly describes the solutions provided in this application from the perspective of interaction between network elements. Correspondingly, this application further provides a communication apparatus. The communication apparatus may be the first RAN node

in the foregoing method embodiments, an apparatus including the first RAN node, or a component that can be used in the first RAN node. Alternatively, the communication apparatus may be the second RAN node in the foregoing method embodiments, an apparatus including the second RAN node, or a component that can be used in the second RAN node. Alternatively, the communication apparatus may be the access and mobility management network element in the foregoing method embodiments, an apparatus including the foregoing access and mobility management network element, or a component that can be used in the access and mobility management network element. It may be understood that, to implement the foregoing functions, the first RAN node, the second RAN node, the access and mobility management network element, or the like includes corresponding hardware structures and/or software modules for performing the functions. Persons skilled in the art should be easily aware that, in combination with units and algorithm operations of the examples described in embodiments disclosed in this specification, this application can be implemented by hardware or a combination of hardware and computer software. Whether a function is performed by hardware or hardware driven by computer software depends on particular applications and design constraints of the technical solutions. Persons skilled in the art may use different methods to implement the described functions for each particular application, but it should not be considered that the implementation goes beyond the scope of this application.

[00283] In this application, the first RAN node, the second RAN node, or the access and mobility management network element may be divided into functional modules based on the foregoing method examples. For example, each functional module may be obtained through division based on each corresponding function, or two or more functions may be integrated into one processing module. The integrated module may be implemented in a form of hardware, or may be implemented in a form of a software functional module. It may be understood that division into the modules in this application is an example, and is merely logical function division. In actual implementation, another division manner may be used.

[00284] For example, when functional modules are obtained through division in an integrated manner, FIG. 8 is a diagram of a structure of a communication apparatus 80. The communication apparatus 80 includes a processing module 801 and an interface module 802. The processing module 801 may also be referred to as a processing unit, and is configured to perform an operation other than receiving and sending operations. For example, the processing module may be a processing circuit or a processor. The interface module 802 may also be referred to as an interface unit, and is configured to perform sending and receiving operations. For example, the interface module may be an interface circuit, a transceiver machine, a transceiver, or a communication interface.

[00285] In some embodiments, the communication apparatus 80 may further include a storage

module (not shown in FIG. 8), configured to store program instructions and data.

[00286] In some embodiments, the communication apparatus 80 may further include an AI module (not shown in FIG. 8) configured to implement an AI-related function. The AI module may implement an AI function by using software, hardware, or a combination of software and hardware. For example, the AI module includes an RIC module. Optionally, the AI module and the storage module are integrated into one module, or the AI module and the processing module 801 are integrated into one module.

[00287] For example, the communication apparatus 80 is configured to implement functions of a second RAN node. The communication apparatus 80 is, for example, the second RAN node in the embodiments shown in FIG. 4 to FIG. 7.

[00288] The processing module 801 is configured to obtain first indication information. The first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information includes sending, through a user plane path, the second indication information to a first node associated with a first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service. For example, the processing module 801 may be configured to perform S401.

[00289] The interface module 802 is configured to send the first indication information to a first RAN node. The first RAN node is a RAN node to which the second node is to be handed over or is potentially to be handed over, or the first RAN node is a secondary RAN node that provides dual connectivity for the second node. For example, the interface module 802 may be configured to perform S402.

[00290] In a possible implementation, the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

[00291] In a possible implementation, the processing module 801 is specifically configured to receive the first indication information from a first network element in a core network via the interface module 802.

[00292] In a possible implementation, the interface module 802 is further configured to receive third indication information, where the third indication information indicates that the first RAN node does not support sending of the second indication information through the user plane path, or indicates that the first RAN node does not send the second indication information through the user plane path.

[00293] In a possible implementation, the interface module 802 is specifically configured to send the first indication information to the first RAN node via a second network element in the core network.

[00294] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[00295] In a possible implementation, the first node is a user plane network element or an application server, and the second node is a terminal.

[00296] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

[00297] When the communication apparatus 80 is configured to implement the functions of the second RAN node, for other functions that can be implemented by the communication apparatus 80, refer to related descriptions of the embodiments shown in FIG. 4 to FIG. 7. Details are not described again.

[00298] Alternatively, for example, the communication apparatus 80 is configured to implement functions of a first RAN node. The communication apparatus 80 is, for example, the first RAN node in the embodiments shown in FIG. 4 to FIG. 7.

[00299] The interface module 802 is configured to receive first indication information. The first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information includes sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending the second indication information to a second node through air interface signaling. For example, the interface module 802 may be configured to perform S402.

[00300] The processing module 801 is configured to determine the second indication information based on at least one piece of alternative quality of service configuration information. For example, the processing module 801 may be configured to perform S403.

[00301] The interface module 802 is further configured to send the second indication information based on the first indication information. For example, the interface module 802 may be configured to perform S404.

[00302] In a possible implementation, the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition

request message, a secondary base station change request message, or an inter-base station anchor handover request message.

[00303] In a possible implementation, the interface module 802 is specifically configured to receive the first indication information from a second network element in a core network; or the interface module 802 is specifically configured to receive the first indication information from a second RAN node.

[00304] In a possible implementation, the transmission mode of the second indication information includes sending the second indication information to the first node through the user plane path, and the interface module 802 is specifically configured to send the second indication information to a user plane network element through the user plane path.

[00305] In a possible implementation, the interface module 802 is further configured to: when determining not to send the second indication information to the first node through the user plane path, send third indication information, where the third indication information indicates that sending of the second indication information through the user plane path is not supported, or indicates that the second indication information is not sent through the user plane path.

[00306] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[00307] In a possible implementation, the first node is the user plane network element or an application server, and the second node is a terminal.

[00308] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

[00309] When the communication apparatus 80 is configured to implement the functions of the first RAN node, for other functions that can be implemented by the communication apparatus 80, refer to related descriptions of the embodiments shown in FIG. 4 to FIG. 7. Details are not described again.

[00310] Alternatively, for example, the communication apparatus 80 is configured to implement functions of an access and mobility management network element. The communication apparatus 80 is, for example, the access and mobility management network element in the embodiment shown in FIG. 4 or the embodiment shown in FIG. 6.

[00311] The interface module 802 is configured to receive first indication information from a second RAN node. The first indication information indicates a transmission mode of second

indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information includes sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service.

[00312] The interface module 802 is further configured to send the first indication information to a first RAN node. The first RAN node is a RAN node to which the second node is to be handed over.

[00313] In a possible implementation, the first indication information is carried in a handover request message.

[00314] In a possible implementation, the interface module 802 is further configured to receive third indication information from the first RAN node, where the third indication information indicates that the first RAN node does not support sending of the second indication information through the user plane path, or indicates that the first RAN node does not send the second indication information through the user plane path.

[00315] In a possible implementation, the interface module 802 is further configured to send the third indication information to the second RAN node.

[00316] In a possible implementation, the second indication information includes at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

[00317] In a possible implementation, the first node is a user plane network element or an application server, and the second node is a terminal.

[00318] In a possible implementation, the quality of service configuration information associated with the first service includes quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

[00319] When the communication apparatus 80 is configured to implement the functions of the access and mobility management network element, for other functions that can be implemented by the communication apparatus 80, refer to related descriptions of the embodiment shown in FIG. 4 or the embodiment shown in FIG. 6. Details are not described again.

[00320] In a simple embodiment, persons skilled in the art may figure out that the communication apparatus 80 may be in a form shown in FIG. 3. For example, the processor 301 in FIG. 3 may invoke computer-executable instructions stored in the memory 303, to cause the communication apparatus 80 to perform the methods in the foregoing method embodiments.

[00321] For example, functions/implementation processes of the processing module 801 and the interface module 802 in FIG. 8 may be implemented by the processor 301 in FIG. 3 by invoking computer-executable instructions stored in the memory 303. Alternatively, a function/an implementation process of the processing module 801 in FIG. 8 may be implemented by the processor 301 in FIG. 3 by invoking computer-executable instructions stored in the memory 303, and a function/an implementation process of the interface module 802 in FIG. 8 may be implemented by the transceiver 302 in FIG. 3.

[00322] It may be understood that one or more of the foregoing modules or units may be implemented by using software, hardware, or a combination thereof. When any one of the foregoing modules or units is implemented by software, the software exists in a form of computer program instructions, and is stored in a memory. The processor may be configured to execute the program instructions and implement the foregoing method procedures. The processor may be built in a system-on-a-chip (system-on-a-chip, SoC) or an ASIC, or may be an independent semiconductor chip. In addition to a core configured to execute software instructions to perform an operation or processing, the processor may further include a necessary hardware accelerator, for example, a field programmable gate array (field programmable gate array, FPGA), a programmable logic device (programmable logic device, PLD), or a logic circuit that implements a dedicated logic operation.

[00323] When the foregoing modules or units are implemented by using hardware, the hardware may be any one or any combination of a CPU, a microprocessor, a digital signal processing (digital signal processing, DSP) chip, a microcontroller unit (microcontroller unit, MCU), an artificial intelligence processor, an ASIC, a SoC, an FPGA, a PLD, a dedicated digital circuit, a hardware accelerator, or a non-integrated discrete device, and the hardware may run necessary software or does not depend on software to perform the foregoing method procedures.

[00324] Optionally, this application further provides a chip system, including at least one processor and an interface. The at least one processor is coupled to a memory through the interface, and when the at least one processor executes a computer program or instructions in the memory, the method in any one of the foregoing method embodiments is performed. In a possible implementation, the chip system further includes the memory. Optionally, the chip system may include a chip, or may include a chip and another discrete device. This is not specifically limited in this application.

[00325] Optionally, this application further provides a computer-readable storage medium. All or some of the procedures in the foregoing method embodiments may be implemented by a computer program instructing related hardware. The program may be stored in the computer-readable storage medium. When the program is executed, the procedures in the foregoing method

embodiments may be included. The computer-readable storage medium may be an internal storage unit of the communication apparatus in any one of the foregoing embodiments, for example, a hard disk or a memory of the communication apparatus. Alternatively, the computer-readable storage medium may be an external storage device of the communication apparatus, for example, a plug-in hard disk, a smart media card (smart media card, SMC), a secure digital (secure digital, SD) card, or a flash card (flash card) that is configured on the communication apparatus. Further, the computer-readable storage medium may include both an internal storage unit and an external storage device of the communication apparatus. The computer-readable storage medium is configured to store the computer program and store other programs and data that are needed by the communication apparatus. The computer-readable storage medium may be further configured to temporarily store data that has been output or is to be output.

[00326] Optionally, this application further provides a computer program product. All or some of the procedures in the foregoing method embodiments may be implemented by a computer program instructing related hardware. The program may be stored in the computer program product. When the program is executed, the procedures in the foregoing method embodiments may be included.

[00327] Optionally, this application further provides computer instructions. All or a part of the procedures in the foregoing method embodiments may be completed by the computer instructions instructing related hardware (for example, a computer, a processor, a terminal, or a RAN node). The program may be stored in the computer-readable storage medium or the computer program product.

[00328] Optionally, this application further provides a communication system, including the first RAN node and the second RAN node in the foregoing embodiments. Optionally, the communication system further includes the access and mobility management network element in the foregoing embodiments.

[00329] The foregoing descriptions about implementations allow persons skilled in the art to understand that, for the purpose of convenient and brief description, division of the foregoing functional modules is taken as an example for illustration. In actual application, the foregoing functions can be allocated to different functional modules and implemented based on a requirement, that is, an inner structure of an apparatus is divided into different functional modules to implement all or some of the functions described above.

[00330] In the several embodiments provided in this application, it should be understood that the disclosed apparatus and method may be implemented in other manners. For example, the described apparatus embodiment is merely an example. For example, the module or division into the units is merely logical function division and may be other division in actual implementation.

For example, a plurality of units or components may be combined or integrated into another apparatus, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings, direct couplings, or communication connections may be implemented by using some interfaces. The indirect couplings or communication connections between the apparatuses or units may be implemented in electronic, mechanical, or other forms.

[00331] The units described as separate parts may or may not be physically separate, and parts displayed as units may be one or more physical units, may be located in one place, or may be distributed on a plurality of different places. Some or all of the units may be selected based on actual requirements to achieve the objectives of the solutions of embodiments.

[00332] In addition, functional units in embodiments of this application may be integrated into one processing unit, each of the units may exist alone physically, or two or more units are integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of a software functional unit.

[00333] The foregoing descriptions are merely specific implementations of this application, but are not intended to limit the protection scope of this application. Any variation or replacement within the technical scope disclosed in this application shall fall within the protection scope of this application. Therefore, the protection scope of this application shall be subject to the protection scope of the claims.

CLAIMS

1. A communication method, wherein the method comprises:

obtaining first indication information, wherein the first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information comprises sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service; and

sending the first indication information to a first radio access network node, wherein the first radio access network node is a radio access network node to which the second node is to be handed over or is potentially to be handed over, or the first radio access network node is a secondary radio access network node that provides dual connectivity for the second node.

2. The method according to claim 1, wherein the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

3. The method according to claim 1 or 2, wherein obtaining the first indication information comprises:

receiving the first indication information from a first network element in a core network.

4. The method according to any one of claims 1 to 3, wherein the method further comprises:

receiving third indication information, wherein the third indication information indicates that the first radio access network node does not support sending of the second indication information through the user plane path, or indicates that the first radio access network node does not send the second indication information through the user plane path.

5. The method according to any one of claims 1 to 4, wherein sending the first indication information to the first radio access network node comprises:

sending the first indication information to the first radio access network node via a second network element in the core network.

6. The method according to any one of claims 1 to 5, wherein the second indication information comprises at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

7. The method according to any one of claims 1 to 6, wherein the first node is a user plane network element or an application server, and the second node is a terminal.

8. The method according to any one of claims 1 to 7, wherein the quality of service configuration information associated with the first service comprises quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

9. A communication method, wherein the method is applied to a first radio access network node or a chip in the first radio access network node, and the first radio access network node is a radio access network node to which a second node associated with a first service is to be handed over or is potentially to be handed over, or the first radio access network node is a secondary radio access network node that provides dual connectivity for the second node, and the method comprises:

receiving first indication information, wherein the first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with the first service; and the transmission mode of the second indication information comprises sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending the second indication information to the second node through air interface signaling;

determining the second indication information based on at least one piece of alternative quality of service configuration information; and

sending the second indication information based on the first indication information.

10. The method according to claim 9, wherein the first indication information is carried in a handover request message, or the first indication information is carried in a secondary base station addition request message, a secondary base station change request message, or an inter-base station anchor handover request message.

11. The method according to claim 9 or 10, wherein receiving the first indication information comprises:

receiving the first indication information from a second network element in a core network; or

receiving the first indication information from a second radio access network node.

12. The method according to any one of claims 9 to 11, wherein the transmission mode of the second indication information comprises sending the second indication information to the first node through the user plane path, and sending the second indication information based on the first indication information comprises:

sending the second indication information to a user plane network element through the user

plane path.

13. The method according to claim 12, wherein the method further comprises:

when determining not to send the second indication information to the first node through the user plane path, sending third indication information, wherein the third indication information indicates that sending of the second indication information through the user plane path is not supported, or indicates that the second indication information is not sent through the user plane path.

14. The method according to any one of claims 9 to 13, wherein the second indication information comprises at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

15. The method according to any one of claims 9 to 14, wherein the first node is the user plane network element or an application server, and the second node is a terminal.

16. The method according to any one of claims 9 to 15, wherein the quality of service configuration information associated with the first service comprises quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

17. A communication method, wherein the method comprises:

receiving first indication information from a second radio access network node, wherein the first indication information indicates a transmission mode of second indication information; the second indication information indicates quality of service configuration information associated with a first service; and the transmission mode of the second indication information comprises sending, through a user plane path, the second indication information to a first node associated with the first service, and/or sending, through air interface signaling, the second indication information to a second node associated with the first service; and

sending the first indication information to a first radio access network node, wherein the first radio access network node is a radio access network node to which the second node is to be handed over.

18. The method according to claim 17, wherein the first indication information is carried in a handover request message.

19. The method according to claim 17 or 18, wherein the method further comprises:

receiving third indication information from the first radio access network node, wherein the third indication information indicates that the first radio access network node does not support sending of the second indication information through the user plane path, or indicates that the first

radio access network node does not send the second indication information through the user plane path.

20. The method according to claim 19, wherein the method further comprises:

sending the third indication information to the second radio access network node.

21. The method according to any one of claims 17 to 20, wherein the second indication information comprises at least one of the following: quality of service profile index information associated with the first service, quality of service configuration parameter information associated with the first service, quality of service profile adjustment information, or quality of service notification control information.

22. The method according to any one of claims 17 to 21, wherein the first node is a user plane network element or an application server, and the second node is a terminal.

23. The method according to any one of claims 17 to 22, wherein the quality of service configuration information associated with the first service comprises quality of service configuration information of a session of the first service, or quality of service configuration information of a quality of service flow of the first service.

24. A communication apparatus, comprising a unit or a module configured to perform the method according to any one of claims 1 to 8, comprising a unit or a module configured to perform the method according to any one of claims 9 to 16, or comprising a unit or a module configured to perform the method according to any one of claims 17 to 23.

25. A communication apparatus, comprising a processor, wherein the processor is coupled to a memory, the memory is configured to store a program or instructions, and when the program or the instructions are executed by the processor, the apparatus is caused to perform the method according to any one of claims 1 to 8, perform the method according to any one of claims 9 to 16, or perform the method according to any one of claims 17 to 23.

26. A computer-readable storage medium, wherein the computer-readable storage medium stores a computer program or instructions, and when the computer program or the instructions are executed, a computer is caused to perform the method according to any one of claims 1 to 8, the method according to any one of claims 9 to 16, or the method according to any one of claims 17 to 23.

27. A computer program product, wherein the computer program product comprises computer program code, and when the computer program code is run on a computer, the computer is caused to implement the method according to any one of claims 1 to 8, implement the method according to any one of claims 9 to 16, or implement the method according to any one of claims 17 to 23.