CLAIMS

1. A method for wireless communication, comprising:

switching, by a first network node, a quality of service (QoS) flow from a source network node to a target network node by:

sending, by the first network node, a first request to the target network node, the first network node comprising a control plane (CP) of a central unit (CU) of a new generation radio access network node (gNB);

receiving, by the first network node, a first response to the first request from the target network node;

sending, by the first network node, a second request to the source network node, the source network node comprising a first user plane (UP) of the CU of the gNB and the target network node comprising a second UP of the CU of the gNB;

receiving, by the first network node, a second response to the second request from the source network node; and

sending, by the first network node, a radio resource control (RRC) reconfiguration complete indication to the source network node by:

sending, by the first network node, a third bearer context modification request to the source network node, the third bearer context modification request comprising the RRC reconfiguration complete indication, and

receiving, by the first network node, a third bearer context modification response; or

sending, by the first network node, the RRC reconfiguration complete indication to the source network node, wherein, upon receiving the RRC

reconfiguration complete indication, the source network node does not send a response to the first network node.

2. The method according to claim 1, further comprising:

prior to sending the RRC reconfiguration complete indication to the source network node:

sending, by the first network node, an RRC reconfiguration to a user equipment; and

receiving, by the first network node, an RRC reconfiguration complete from the user equipment, wherein the RRC reconfiguration complete indicates a completion of the RRC reconfiguration.

3. The method according to claim 1, wherein:

the first request comprises a bearer context setup request comprising a list of to-be-setup data radio bearers (DRBs);

the first response comprising a bearer context setup response;

the second request comprises a bearer context release request comprising a list of to-be-released DRBs;

the second response comprises a bearer context release response.

4. The method according to claim 1, wherein:

the first request comprises a bearer context setup request comprising a list of DRBs;

the first response comprises a bearer context setup response;

the second request comprises a bearer context modification request comprising a list of to-be-modified DRBs; and

the second response comprises a bearer context modification response.

5. The method according to claim 1, wherein:

the first request comprises a first bearer context modification request comprises a list of to-be-modified DRBs, wherein the list of to-be-modified DRBs is configured to be modified according to the QoS flow;

the first response comprising a first bearer context modification response;

the second request comprises a second bearer context modification request; and

the second response comprises a second bearer context modification response.

6. The method according to claim 1, wherein:

the first request comprises a first bearer context modification request comprises a list of to-be-modified DRBs, wherein the list of to-be-modified DRBs is configured to be modified according to the QoS flow;

the first response comprising a first bearer context modification response;

the second request comprises a bearer context release request; and

the second response comprises a bearer context release response.

7. The method according to any of claims 3-6, wherein:

the first response comprises data-forwarding information; and

the second request comprises the data-forwarding information;

wherein the data-forwarding information comprises general packet radio service (GPRS) tunneling protocol (GTP) tunnel information.

8. A method for wireless communication, comprising:

transmitting user data between a user equipment and a core network by:

switching a quality of service (QoS) flow by a first network node from a source network node to a target network node, the first network node comprising a control plane (CP) of a central unit (CU) of a new generation radio access network node (gNB), the source network node comprising a first user plane (UP) of the CU of the gNB, and the target network node comprising a second UP of the CU of the gNB;

continuously receiving first uplink user data by the target network node from the user equipment;

receiving a first uplink end-marker by the source network node from the user equipment;

upon receiving the first uplink end-marker, determining by the source network node whether the source network node finishes sending second uplink user data to the core network, wherein the second uplink user data is received by the source network node from the user equipment;

in response to the determination that the source network node finishes sending the second uplink user data to the core network, sending a second uplink end-marker by the source network node to the target network node; and

upon receiving the second uplink end-marker, sending the first uplink user data by the target network node to the core network.

9. The method according to claim 8, wherein:

the first uplink end-marker comprises a service data adaptation protocol (SDAP) end-marker.

10. The method according to claim 8, wherein:

the second uplink end-marker comprises a general packet radio service (GPRS) tunneling protocol-user plane (GTP-U) end-marker.

11. The method according to claim 8, wherein:

the second uplink end-marker comprises a message indicating that the source network node completes sending all second uplink user data to a user plan function (UPF).

12. The method according to claim 8, further comprising:

receiving downlink user data by the target network node from the source network node;

receiving a first downlink end-marker by the source network node from the core network;

receiving a second downlink end-marker by the target network node from the source network; and

upon receiving the second downlink end-marker, sending the downlink user data from the target network node to the user equipment.

13. The method according to claim 12, wherein:

the first downlink end-marker comprises a GTP-U end-marker.

14. The method according to claim 12, wherein:

the second uplink end-marker comprises a message indicating that the target network node begins sending the downlink user data to the user equipment.

15. An apparatus comprising:

a memory storing instructions; and

a processor in communication with the memory, wherein, when the processor executes the instructions, the processor is configured to cause the apparatus to perform the steps according to any of claims 1-7.

16. An apparatus comprising:

a memory storing instructions; and

a processor in communication with the memory, wherein, when the processor executes the instructions, the processor is configured to cause the apparatus to perform the steps according to any of claims 8-14.