Flows

DD MD Owned by David Cherney DISH •••
Last updated: Jul 03, 2023 • Add Workflow

Contents

The flow of user plane data between UE and DN is divided up into flows of three nested kinds in 5G.

1. SDFs (service data flows) that are tunnels through the 5G core with tunnel endpoints (addresses)
2. PDU sessions (protocol data unit sessions)
3. QoS flows (quality of service flows)

Before presenting them, it will help to have the following.

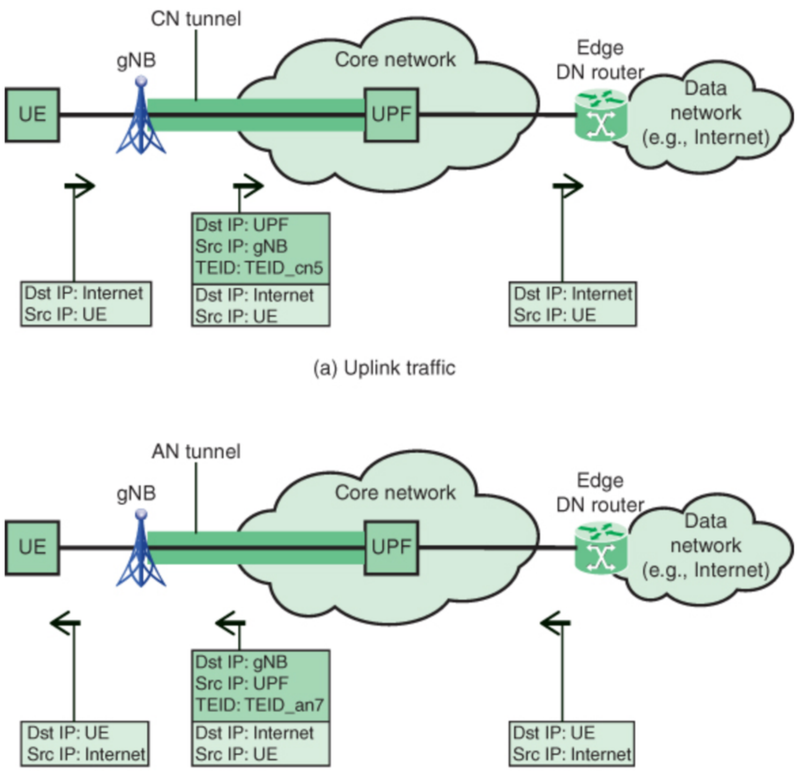
Definition: A **flow** is a sequence of packets traversing a network that share a set of packet header values.

1) service data flows (SDFs)

For a UE to send a packet of data to the internet or any other DN, that packet must first pass through the core network. That is, a core network tunnel (CN tunnel) must be created. This is accomplished in parts as shown in the figure (a) below.

- The relevant parties have addresses that appear in the header as source and destination addresses;
  - the UE has an address to receive packets through this flow
  - a device/application/service in the DN also has an address.
- The UE sends a packet with a destination IP and source IP in the header of the packet to a radio tower (typically over radio waves). That packet then travels to a DU, then to its CU. Recall that a gNB is made up of several DUs and one CU.
- Since the gNB does not have a direct connection to the internet (or other data network) but does have a (typically fiberoptic) connection to the core, the UE wraps this packet with a new header with an addresses for the gNB as source and an address for the UPF as a destination. This is called tunneling, and the addresses of the gNB and UPF are called tunnel endpoints.
  - Since this tunnel is **into** the core network, it is called a CN tunnel.
  - The wrapping is for a protocol called GTP for GPRS tunneling protocol,

    - GPRS is for general packet radio service, since any kind of packet can be wrapped. The “radio“ part is no legacy; GPRS has been used since 1G.
  - The wrapping treats the original packet, including the original header as the (typically encrypted) body for a new packet. The new packet facilitates routing to appropriate points in the 5G core because the new header includes
    - the destination UPF instance’s IP address,
    - the source gNB’s IP address (which will change if the UE moves), and
    - a tunnel endpoint ID (TEID), the mechanism by which the core keeps track of the UE that is the source of the wrapped packet.
- When the wrapped packet reaches the UPF, the wrapping is removed and the original packet is sent to the internet through a router on the edge of the data network.
- The data network then sends the packet to its destination.



(a) Uplink traffic

(b) Downlink traffic

A core network tunnel is uni-directional. We just went through the uplink direction. In the reverse direction, the downlink direction, a unidirectional tunnel takes packets from the UPF to the gNB. In this direction,

- since the tunnel is **into** the AN it is called an AN tunnel (as opposed to CN tunnel for uplink).
- the UPF does the work of wrapping packets in a header and the wrapping is removed by the gNB.
- multiple tunnels for multiple UEs terminate in this gNB, and the TEID allows the gNB to know which UE should be given the packet.
- has the characters AN for “access network”.
- The original packet from the DN is sent by radio waves to the UE.

Tunneling is, then, the passing of data packets from UE to DNs through the 5G core via wrapping of packets with headers that include a tunnel ID. A single UE can have multiple tunnels going at once for multiple different applications the UE accesses through a DN, or several DNs.

Note that a single UE can have different IP addresses for different SDFs.

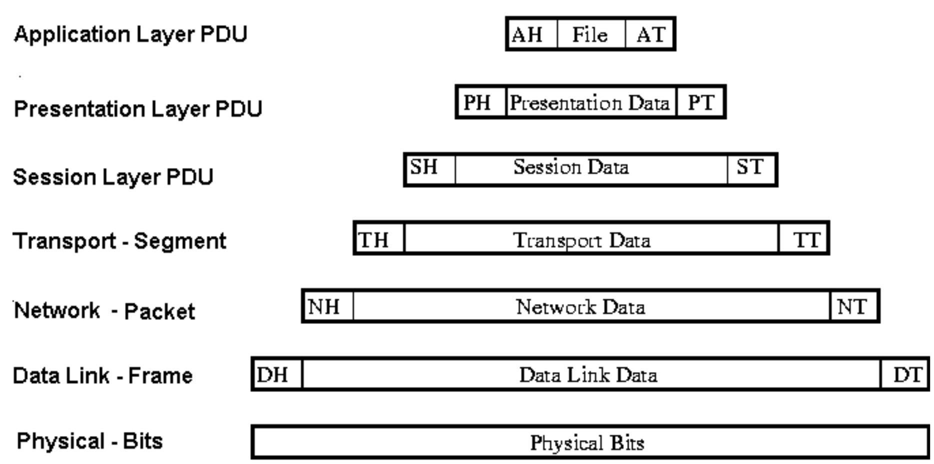
Definition: Each unique combination of UE IP address and application IP address in a DN together with that combination's CN and AN tunnels is called a service data flow (SDF).

2) Protocol Data Unit (PDU) Session

Packets get wrapped for their passage through the core, as described above. But the core also needs to make preparations for this passage; the network functions have to be made aware that packets will be coming through so that the functions know how to route the packets within the core. To present this, it will help to have some terminology.

Definition: A protocol data unit (PDU) is a unit of information exchanged between peers on a computer network in some protocol.

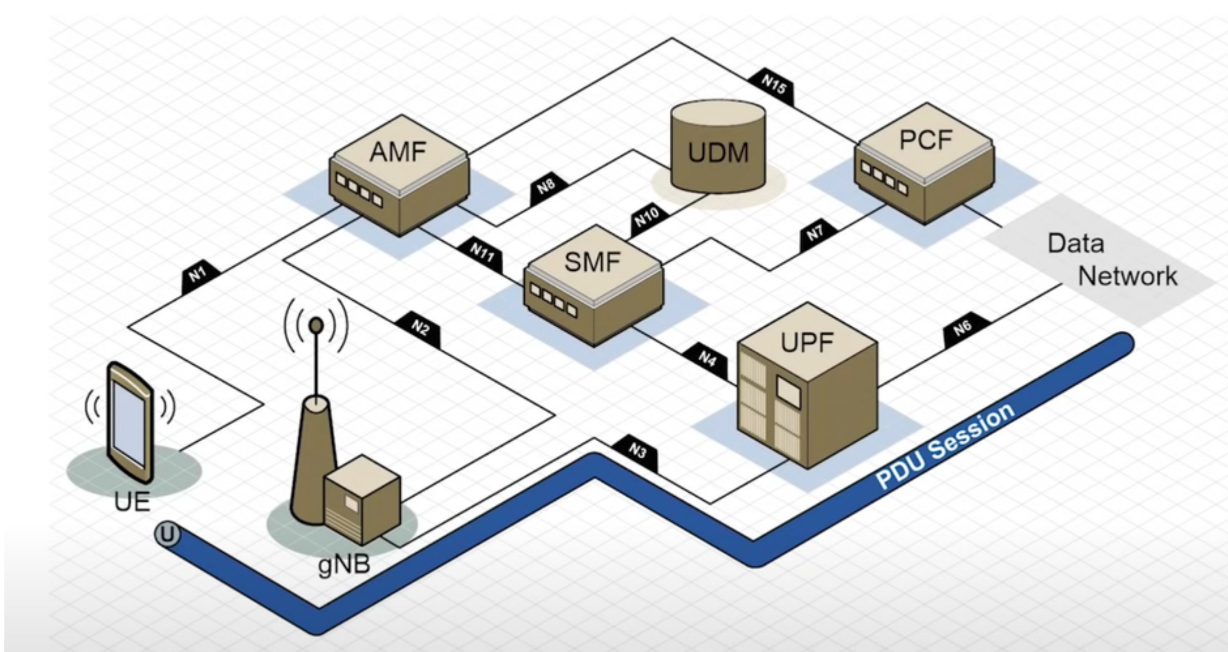
Different transmission protocols have different PDUs. For some examples, a PDU in the data layer with protocol Bluetooth is called a bit, but a PDU in the network layer with protocol IP is called a packet. PDUs often contain protocol-specific control information in headers and the data to be transmitted as their body.



Protocol Data Units at Different OSI Layers

Definition: A PDU connectivity service is a service that provides for the exchange of PDUs for a particular protocol.

Definition (version 1.0): A PDU session is a connection between a UE and a DN through a CN such that the connection provides a PDU connectivity service.



5G supports 5 kinds of PDU sessions for 5 kinds of protocols;

1. IPv4
2. IPv6
3. IPv4v6 (a hybrid wherein network address translation is applicable)
4. Ethernet
5. Unspecified

Thus, a single UE might have simultaneously

- an IPv4 session with 3 SDFs each to different addresses on the DN, and
- an IPv6 session with 2 SDFs both to different addresses on the DN.

Putting SDFs for different protocols into different PDU sessions is one way that 5G facilitates giving different service to different data flows. Another way is an intermediate level of grouping of SDFs called QoS flows.

3) QoS Flows

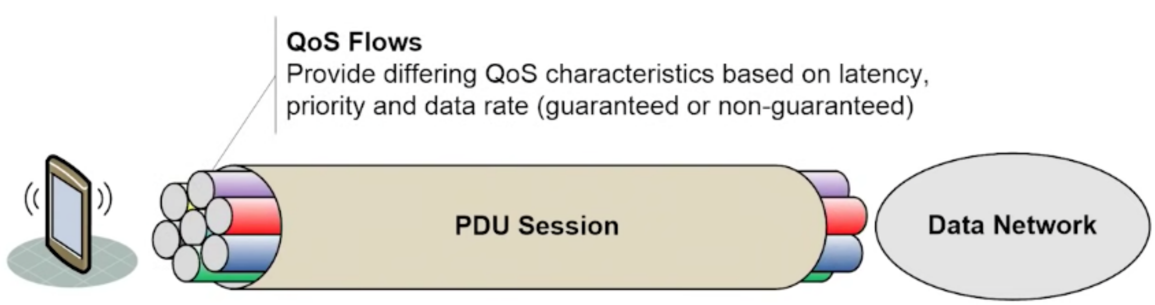
As presented above, service data flows are flows from a address on a UE to a app with an address on a DN. Packets from those flows are wrapped in GTP headers for routing in the core. Part of the information in that header is a quality of service flow identifier, QFI. Data packets can have different tunneling endpoint IDs (TEIDs) but the same QoS identifier. All packets with the same QFI are treated with the same traffic forwarding treatment.

Definition: A QoS flow is a bundle of SDFs in a PDU that have the same QFI.

e.g. Your UE might have

- one PDU session carrying two QoS flows A and B, with
  - QoS flow A carrying 2 SDFs, in particular
    - both SDFs carry QFI 7
  - QoS flow B carrying 3 SDFs, and in particular
    - all three SDF carry QFI 9.

so that your UE is connected to 5 apps in 1 data network, since PDU sessions have one DN.



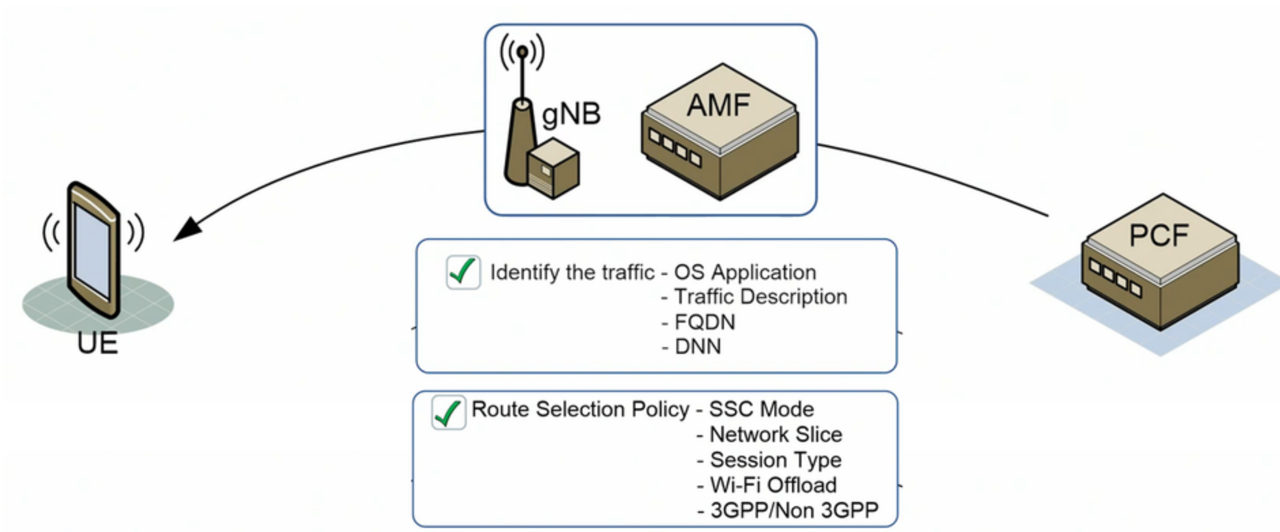
Thus, QoS flows bundle together SDFs that can be treated with the same policies in the core. As a result, a QoS flow is the lowest granularity of a traffic flow control in a 5G core. That means that calculations for charging (as in billing) can be applied differently only to different QoS flows. It also means that differentiated treatment of packets, priority for some packets being forwarded so that other packets are placed in routing buffer in times of congestion, can be applied differently only to different QoS flows.

QoS Flow Binding

Say a UE starts a new application, and thus a new SDF. The UE needs to make decisions about

- which existing QoS flow can host the SDF
- if a new QoS flow is needed because
  - e.g. the new SDF has QoS parameters unlike any existing QoS flow.
- a new PDU session is needed
  - e.g. because a new data network is being accessed or
  - e.g. because the guaranteed flow bit rate of the new SDF is larger that the difference between the PDU session's aggregate maximum bit rate and the currently reserved guaranteed flow bit rates is too small.

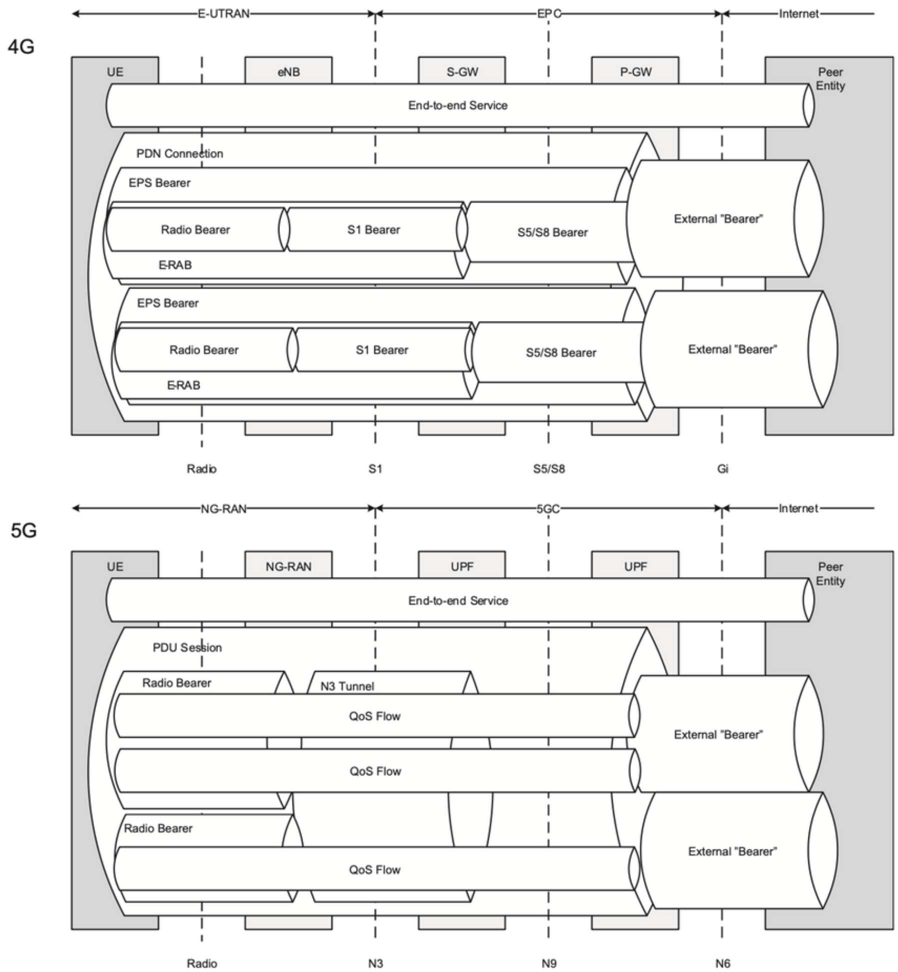
These decisions are made by using a policy called a UE route selection policy (URSP). This policy is not the decision of the UE, but the decision of the network; URSPs are given to the UE by the policy control function, PCF. As shown below, a URSP policy takes into account the nature of the application the UE is running (including the fully qualified domain name (FQDN) of the app on the data network and the data network name DNN) as well as the PDU session information like service and session continuity (SSC) mode, slice, PDU session type, and access network used.



Once the UE has URSP, it can make choices about which SDFs to put in which QoS flows and which PDU sessions. When the UE makes that decision, the SMF is informed (with AMF as communication proxy) and the SMF creates the relevant structures in the UPF. The information given to the UPF by the SMF in this process is what allows the SDF flow to be in a QoS flow in the core, and thus is called QoS flow binding.

New to 5G

To say it again, QoS flows are the mechanism by which 5G facilitates differentiated service between data flows. QoS flows are new in name, but not in concept; 4G had differenced service mechanisms too. One distinct change from 4G advanced to 5G is that in 5G multiple QoS flows can be passed on one radio channel (data radio bearer, DRB). The following diagram tries to make this comparison visual; in 4G the analog of QoS flows were EPS bearers. (Evolved packet system is the general name for 4G system). Each EPS bearer had to be transmitted by a separate radio bearer. This was sometimes a waste of radio resources; each of three EPS bearers for a UE might be carrying a small amount of traffic but still occupying three radio bearers. Since in 5G multiple QoS flows can be put on the same radio bearer, 5G is more efficient in its use of radio spectrum, and require less energy consumption. This is one of the reasons that, although mobile traffic has increased by a factor of 4 since the start of 5G in 2018, user experience has not degraded.



All Together

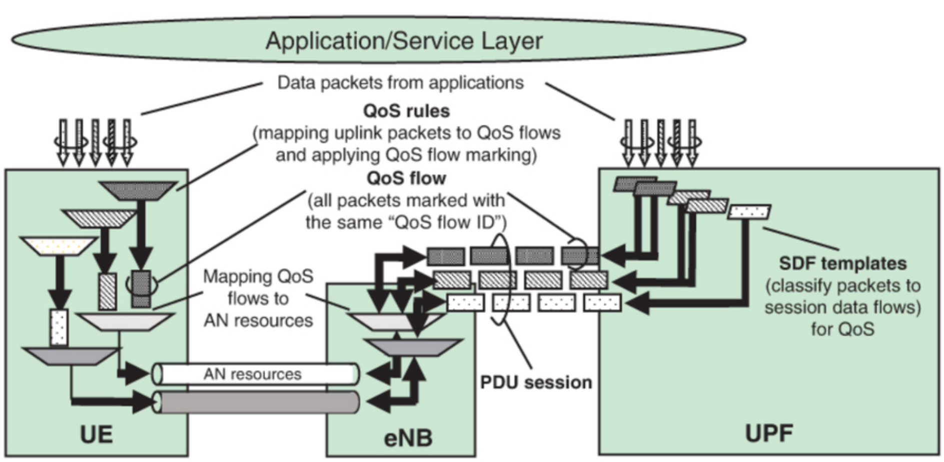


FIGURE 9.11 The Principles for Classification, User Plane Marking, and Differentiation in 5G

In the direction UE to DN, after a PDU session has been established so that SMF has informed the UE and UDM (the network function called unified data management) of the TEIDs for all SDFs, QoS flow IDs (QFI) for each QoS flow, and QoS profiles/rules

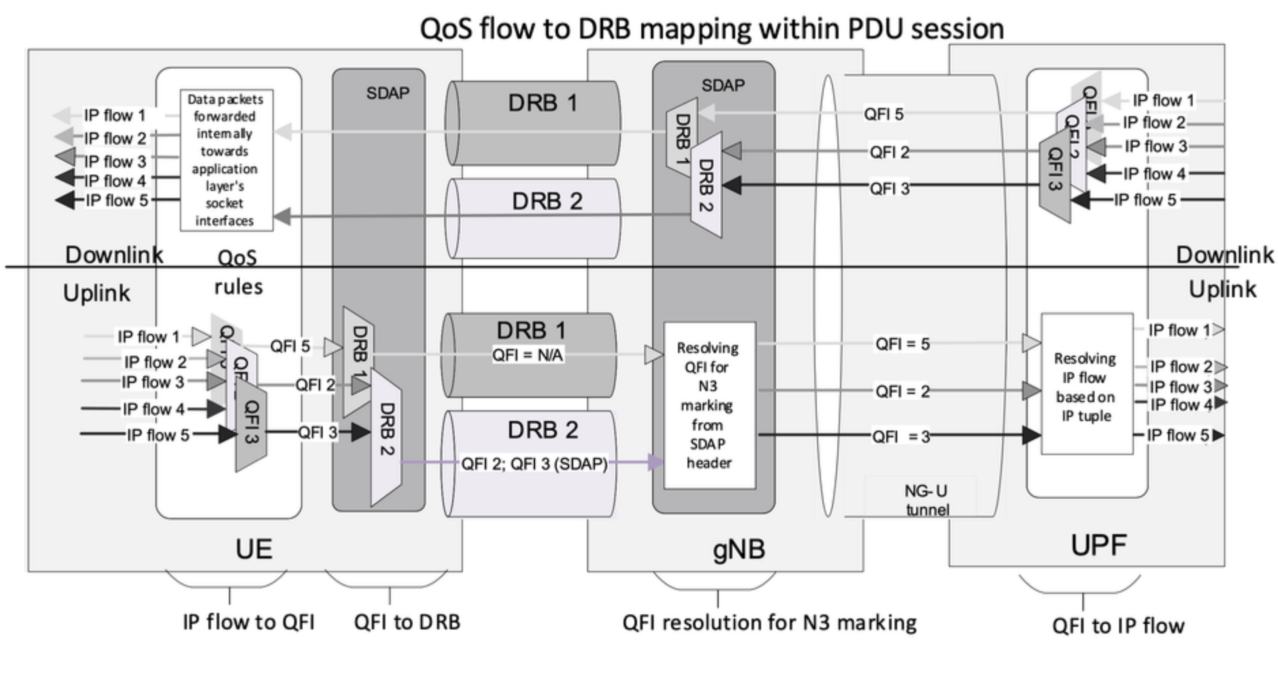
- The UE uses SDF templates to gather packets into SDFs, one SDF for each application. Each SDF has a unique TEID. The SDF templates are applied in an order in case multiple templates apply to the same packet.
- The UE applies QoS rules to bundle SDFs into QoS flows. Each QoS flow has a unique QoS flow ID.
- The UE bundles together QoS flows into a PDU session. (This is not shown explicitly in the diagram.)
- The UE bundles QoS flows (that are within the PDU session) into radio bearers of radio transmission to the gNB. (The eNB in the diagram is a typo.) 5G introduces the possibility of having multiple QoS flows in the same radio bearer.
- The UE sends the packets for all SDFs in the QoS flows in the PDU session to the gNB via radio bearers, and the gNB receives the packets. This is transmission over the Uu reference point. (Uu is the 3GPP name of the reference point between UE and gNB.)
- The gNB unbundles the radio bearers into QoS flows.
- The gNB bundles the QoS flows into a PDU Session transmitted with GTP headers, and sends (via fiber) to the 5G core.
- Packets for the PDU Session are routed through a series of physical routers that underly the virtual networking between the gNB and UPF. There are many physical connections but one virtual connection. This is transmission through the N3 reference point.
- The UPF unbundles the PDU session into QoS flows, and the QoS flows into SDFs.
- Each SDF is routed to the appropriate address on a data network. This is communication via the N6 reference point.

The DN to UE direction is similar in concept with a few possible complications.

The similarity is enhanced if the PDU session uses a property called reflective QoS, which applies the same QoS profile to both uplink and downlink packets in the PDU session. You can imagine, then, that the QoS profiles can be different in the up and down link directions.

The diagram above details a session, but does not show how that sessions is managed. Session management will be presented in another section.

The diagram below shows the same facts but with a few more acronyms like data radio bearer (DRB) and Service Data Adaptation Protocol (SDAP).



Lingo

Bringing SDFs, QoS flows, and PDU sessions together, here is a description of how packets of data are passed through the 5G core using 3 new terms.

Definition: Traffic classification is the grouping of traffic into classes based on user defined QoS values.

Assigning a QFI to a SDF is traffic classification.

Definition: User plane marking is the marking of packets to indicate each packet's traffic classification.

Writing a QFI to a GTP header around a packet from an SDF is user plane marking.

Definition: QoS differentiation is the use of different QoS values or different categories of traffic.

The UPF treats one packet differently than another because the two packets carry different QFIs is QoS differentiation.

Consider why this three level division (SDF, QoS flow, PDU session) was introduced in 5G. As time goes on, an ever widening range of apps involve wireless transmission, and the QoS of those apps are ever widening too;

- some IoT devices will only need to send a few bytes every few days, while some will need to send gigabytes per second, and
- some apps will need less than 1ms of latency (e.g. remote surgery) while others can tolerate seconds of latency.

Previous mobile technologies did not support the kind of differentiation of service that these future applications need if they are to be enabled by the same core network. Who wants to pay for 10gB/s capacity when all that is needed is 1B/day?

Definition: A service level agreement is an agreement between a customer and a provider of a service that specifies the level of availability, serviceability, performance, operation, or other attributes of the service.

Service Level Agreements (SLAs) between mobile network operators and their customers need to match QoS requirements to QoS guarantees. MNOs can deliver on those SLAs using the tool of QoS flows. The systems that control QoS flows are discussed under session management.

UP Session Connection State in 5G VS 4G

The analog to PDU sessions in 4G is called PDN sessions (packet data network). Just as a single UE might have several PDU sessions, a single device might have several PDN sessions. For both kinds of sessions, when the UE goes into connection management state `idle` the sessions still exist, but the tunneling endpoint information is not kept. A session in this state is called an inactive session. A session that has tunneling endpoints assigned is called an active session. This terminology applies to both 5G PDU sessions and 4G PDN sessions.

So what differences are there between 4G and 5G user plane sessions, and why do those difference exist?

In 4G, when a UE switched from cm state `idle` to cm state `connected` all PDN sessions were assigned tunneling endpoints; all PDN sessions for the UE went from inactive to active. In 5G, when a UE switches from cm state `idle` to cm state `connected` it is possible for zero, one or any number of the PDU sessions to regain their tunnel endpoint information; a subset of the PDU sessions becomes active.

The motivation for this change is subtle but important. It involves the concept of slicing, covered later in this introduction.

A goal of 5G was to provide native support for network slicing, a concept which is new in name but not new in concept; 4G had a feature similar to slicing called dedicated cores (DECOR). In network slicing the behavior of the UE over the various slices is as independent as possible; the UE might behave as a low data rate IoT device over one slice and use high resolution video streaming over another slice. Slicing should allow these different behaviors of the same UE to be treated as though handled independently by independent networks. However, in the 4G setup, when a single PDN sessions went from inactive to active, all other PDN sessions had to go from inactive to active, which is a dependency between slices. In 5G, PDU sessions are more independent since they can become active independently.

Another way to think of this is in terms of (connection management) CM states; it would be nice to have the capability for a UE to be in cm state `idle` in some slices and cm state `connected` in others. But CM state is a property of the UE, not a property of a session. To get as close as possible, 5G specifications allow some 5G PDU sessions to be active while others are inactive.

Recap

UE get UE route selection policies from the core. A URSP allows the UE and core to coordinate is packaging packets into service data flows (SDFs), package SDFs into QoS flows, and package QoS flows into radio bearers and PDU sessions. The smallest level of flow, an SDF, is a single connection between a UE and an app. The intermediate level, QoS flows, facilitates differentiated treatment of SDFs. The largest level, PDU sessions, represent a connection between a UE and a data network.