October 2020

CONVERGING SESSIONS OR DATA PATHS ON A USER PLANE

Sanjeev Panem Jaya

Ameo Ghosh

Suyog Belsare

Shailesh Goel

Aaseem Sufi

See next page for additional authors

Inventor(s)
Sanjeev Panem Jaya, Ameo Ghosh, Suyog Belsare, Shailesh Goel, Aaseem Sufi, and Dishant Parikh

CONVERGING SESSIONS OR DATA PATHS ON A USER PLANE

AUTHORS:
Sanjeev Panem Jaya
Ameo Ghosh
Suyog Belsare
Shailesh Goel
Aaseem Sufi
Dishant Parikh

ABSTRACT

As mobile operators look to deploy Third Generation Partnership Project (3GPP) Fifth Generation (5G) services (either standalone (SA) or non-standalone (NSA)) to their subscribers, the process of migrating from Fourth Generation (4G) Evolved Packet Core (EPC) deployments to 5G core (5GC) is going to take time. During such a migration process the principal focus will be on providing higher throughputs and lower data path latency. It is likely that operators will have both EPC and 5GC deployments in their network for some time as they deploy a converged core – ‘ AnyG’ – and roll out nation-wide 5G coverage. This implies that a substantial number of subscribers may still be using the Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access Network (EUTRAN) and operators will see a large number of 5G to 4G handovers. With dual cores deployed in the network, the total cost of ownership for the operators will be high. Techniques are presented herein to address these challenges by enabling an operator to use a common user plane for Control and User Plane Separation of EPC nodes (CUPS), 5G NSA, and 5G SA deployments thus helping to reduce capital expenditures and operating expenses.

DETAILED DESCRIPTION

In a 3GPP 5GC and 4GC EPC interworking scenario, the User Equipment (UE) has a user plane serving gateway (SGW-U) session and a User Plane Function (UPF)/user plane Packet Data Network (PDN) gateway (PGW-U) session separately created on the UPFs. This introduces an additional hop in the subscriber's data path. The additional hop can be avoided if a control plane serving gateway (SGW-C) and a Session Management Function (SMF)/control plane PDN gateway (PGW-C) can select the same UPF.

Even though solutions are already known for enabling a SGW-C and a SMF/PGW-C to select the same UPF, such a selection removes only the physical hop. Since there are two distinct sessions on the UPF – a Sxa session and a N4/Sxb sessions – the packet processing will happen separately, once for the Sxa session and then for the N4/Sxb session. This is because packet processing for Sxa is done at the user plane General Packet Radio Service (GPRS) Tunneling Protocol (GTP-U) layer and packet processing for N4/Sxb is done at an inner Internet Protocol (IP) and payload level. Thus even though the physical hop is avoided the separate processing introduces an additional delay.

Moreover, a number of other inefficiencies arise, including:

- At scale, the additional sessions being hosted on a user plane will impact capacity.
- Throughput at a UPF will degrade as packets are processed twice (as noted above, first for a Sxa session and then for a N4/Sxb session).

Techniques are presented herein that address these inefficiencies.

A high level depiction of the 5GC network architecture as defined by the 3GPP is presented in Figure 1, below.

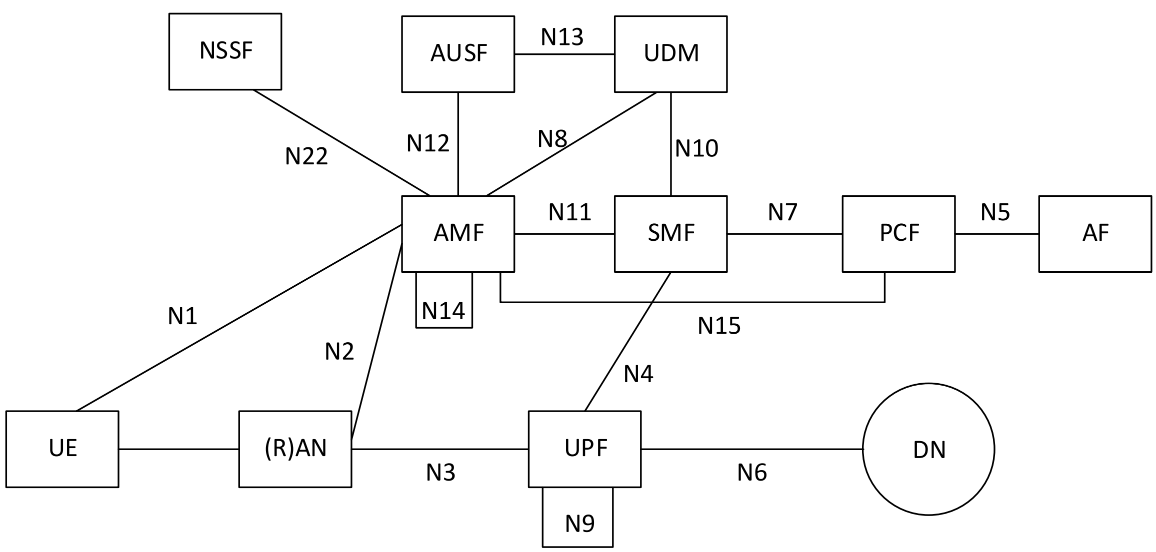

Figure 1: 5GC Network Architecture

As illustrated in the above figure, the data path setup for the subscriber (UE) consists of only a UPF from the core network. But when the UE moves into a 4G coverage area (e.g., under EPC/EUTRAN), an additional node – a SGW-U – is introduced into the data path for the subscriber as part of the 3GPP EPC interworking procedures (further described below).

As noted previously, there are solutions defined to enable selection of the same UPF by a SGW-C and a SMF/PGW-C. Such a selection addresses some of the inefficiencies noted above by co-locating both the Sxa session and N4/Sxb session on the same physical node.

Even though the same UPF is selected by a SGW-C and a SMF/PGW-C, thus removing the physical hop, given the two distinct sessions (i.e., Sxa and N4/Sxb) on the UPF the packet processing will happen separately (once for a Sxa session and then for a N4/Sxb session). As noted previously, this is because packet processing for Sxa is done at the GTP-U layer and for N4/Sxb is done at an inner IP and payload level. Consequently even though the physical hop is avoided the separate processing introduces an additional delay. With today's applications being largely distributed, the issue becomes more pronounced as the packet hops from one instance to another as part of the packet processing.

Techniques are presented herein that provide a mechanism to merge the two distinct Sxa and N4/Sxb sessions on the UPF to create a collapsed data path.

There are three principal parts to the techniques presented herein:

- Identifying the separate Sxa and N4/Sxb sessions of the same PDN using GTP-U Fully Qualified Tunnel Endpoint Identifiers (F-TEIDs) – e.g., Tunnel Endpoint Identifier (TEID) + GTP-U node IP address.
- Anchoring the two distinct sessions on the same instance/Pod in a distributed system architecture.
- Merging Sxa and N4/Sxb sessions and creating an optimized data path (wherein a packet is processed only once) while emulating separate sessions and state machines to the external peer entities.

Each part will be described below.

A first part of the techniques described herein in which separate Sxa and N4/Sxb sessions of the same PDN are identified using GTP-U F-TEIDs, may be explicated with reference to Figure 2, below.

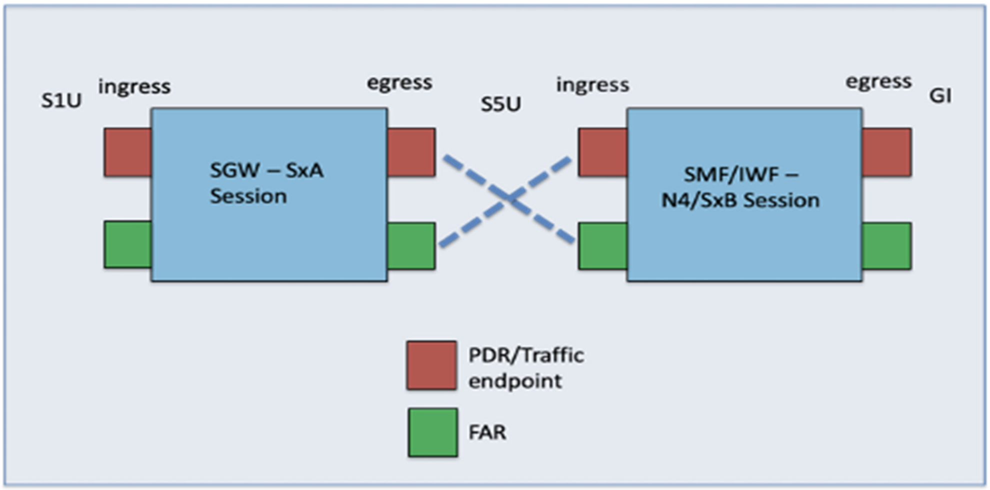

Figure 2: Sxa and N4/Sxb Session Correlation

As shown in the above figure, two separate sessions are established by a SGW-C (Sxa) and a SMF/PGW-C (N4/Sxb) on the same UPF instance. The following logic may be employed to confirm that these two sessions belong to the same PDN:

- For an Uplink Packet: A SxA session's egress Forwarding Action Rule's (FAR's) F-TEID should match to a N4/SxB session's ingress Packet Detection Rule's (PDR's) F-TEID.
- For a Downlink Packet: A N4/SxB session's ingress FAR's F-TEID should match with a SxA session's egress PDR's F-TEID.

Other approaches for confirming that two sessions belong to the same PDN are possible, but such approaches entail various deficiencies. For example:

- One could involve using a private GTP Information Element (IE), including a SGW-U's S5 F-TEID, as part of a Create Session Request from a SGW-C to a SMF (PGW-C) to communicate the co-location information. A SMF could include the S5-U IP addresses as a 'hint' to the Network Repository Function (NRF) to select a UPF that is co-located with the SGW-U. However, this approach would involve changes at a SGW-C and at a SMF (PGW-C) to support such a private IE.

- One could employ a User ID IE, as defined by 3GPP, on a Sxa and a N4/Sxb interface. A International Mobile Subscriber Identity (IMSI)/ Mobile Station International Subscriber Directory Number (MSISDN) present in such a User ID IE could be sent by a SGW-C and a SMF (PGW-C) during session establishment. However, a User ID is an optional IE on a Sxa and a N4/Sxb and hence it would require that both a SGW-C and a SMF (PGW-C) always include such an IE.

There are singular advantages to employing F-TEIDs for confirming that two sessions belong to the same PDN. Such employment, as found within the techniques that are described herein, does not deviate from any standards and is not dependent on the implementation or support of a SGW-C and a SMF (PGW-C). Hence it can, if required, also operate with a third party SGW-C and SMF/PGW-C. It only requires that the UPF has the ability to anchor both SGW-U and SMF/PGW-U sessions and that operator policies/criteria enable a SGW-C and a SMF/PGW-C to select the same UPF instance.

A second part of the techniques described herein, involving anchoring the two distinct sessions on the same instance/Pod in a distributed system architecture, may be explicated with reference to Figure 3, below.

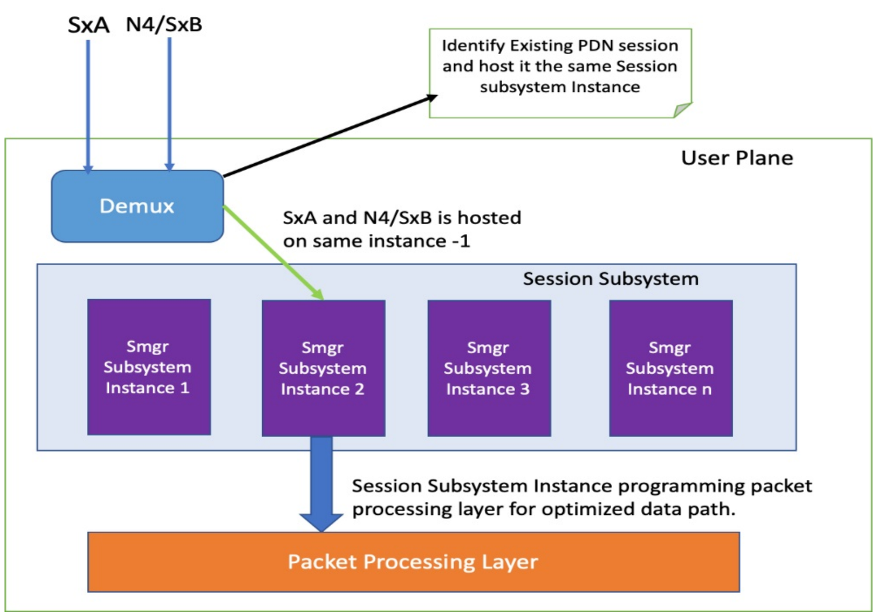

Figure 3: Enhanced Session Management

As illustrated in the above figure, at the UPF, when a Packet Forwarding Control Protocol (PFCP) Session Establishment request is received, a Demux subsystem may be used to parse the message and extract the F-TEID values as mentioned above. A mapping may be created between an extracted F-TEID and the Session Management Subsystem on which the session is anchored so that:

- When a UE is registering, the same Session Management Subsystem instance is selected when a N4/Sxb PFCP Session Establishment Request is received.
- When a UE moves from a 5GC environment to a EPC environment, the same Session Management Subsystem instance is selected when a Sxa PFCP Session Establishment Request is received
- During SGW relocation, the same Session Management Subsystem instance is selected when a Sxa PFCP Session Establishment Request is received

After a Session Management Subsystem instance is identified, the Demux subsystem may pass the unique Session Identifier of the existing Session (which could be either a Sx session or a N4/Sxb session) along with the PFCP Session Establishment Request to the Session Management Subsystem instance.

A third part of the techniques described herein involves merging Sxa and N4/Sxb sessions and creating an optimized data path. As part of a PFCP Session Establishment Request, the Session Management Subsystem instance may query the existing session with the unique Session Identifier received from the Demux subsystem and update the session state/parameters, creating a merged session on the UPF.

Once the session is merged, the packet processing sequence is modified (if it already exists) to apply the combined policy so that:

- For Uplink, packets received on S1-U are forwarded to Sgi/N6 directly.
- For Downlink, packets received on Sgi/N6 are forwarded to S1-U directly.

An S5-U interface is abstracted in the packet processing path in both the directions, thus creating an optimized data path for the UE.

Aspects of the techniques that are described herein may be explicated through a series of illustrative call flows, which are presented below.

Note that the call flows that are presented below assume the use of Option 3 for the identification and correlation of a Sxa session and a N4/Sxb session belonging to the UE.

A first illustrative call flow, depicting aspects of a 5G UE initial attach in a 4G coverage area with sessions collapsed at a UPF, may be described with reference to Figure 4, below.

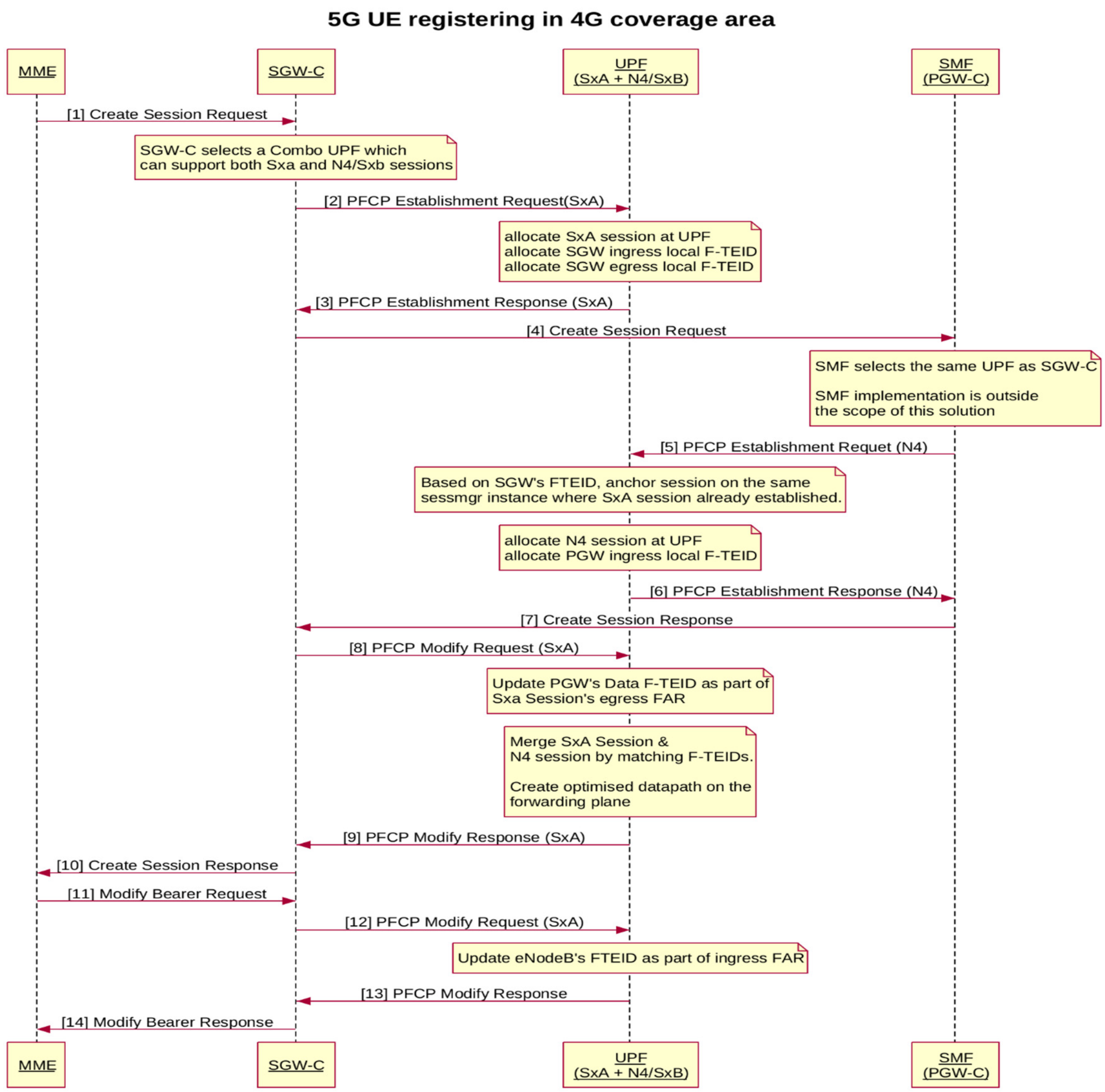

Figure 4: 5G UE Registering in a 4G Coverage Area

In the illustrative call flow that was depicted above, various steps may include:

- Step 1. A Create Session Request message is received on a SGW-C. The SGW-C selects one of the UPFs which supports a combined Sxa+N4 interface based on predefined criteria.
- Step 2. A PFCP Establishment Request message is sent towards the selected UPF instance. At the UPF a new session is created for a SGW-C.

- Step 3. A PFCP Establishment Response is received from the UPF with an allocated F-TEID for SGW ingress and SGW egress endpoints.
- Step 4. The SGW-C initiates a Create Session Request message toward a SMF/PGW-C.
- Steps 5 and 6. The SMF/PGW-C selects the same UPF as the SGW-C based on predefined criteria for UPF selection. The SMF initiates a PFCP Establishment Request message towards the UPF. The UPF creates a new session for the SMF. The UPF responds with a PFCP Establishment Response message to the SMF with the PGW ingress F-TEID.
    - In a distributed system architecture, the UPF selects the same subsystem in which the SGW-C's Sxa session is already anchored so that the sessions can be merged
- Step 7. The SMF/PGW-C responds with a Create Session Response message, including its local F-TEID.
- Step 8. The SGW-C updates the remote F-TEID to the UPF on the egress side FAR using a PFCP Session Modification Request message. The Sxa and N4 sessions of the same PDU session are merged. An optimized data path is created as described above.
- Steps 9 and 10. The UPF responds with a PFCP Modification Response to the SGW-C. The SGW-C responds with a Create Session Response to the Mobility Management Entity (MME).
- Steps 11 and 12. The MME initiates a Modify Bearer Request message towards the SGW-C with the Evolved Node B (eNodeB) F-TEID. The SGW-C updates the eNodeB F-TEID to the UPF on a Sxa interface for the respective FAR. The optimized data path is updated with the eNodeB's address for downlink packets.
- Steps 13 and 14. The UPF responds with a PFCP Session Modification Response message to the SGW-C. The SGW-C responds with a Modify Bearer Response to the MME.

A second illustrative call flow, depicting aspects of a 5G to 4G handover (encompassing interworking with EPC) with sessions collapsed at a UPF, may be described with reference to Figure 5, below.

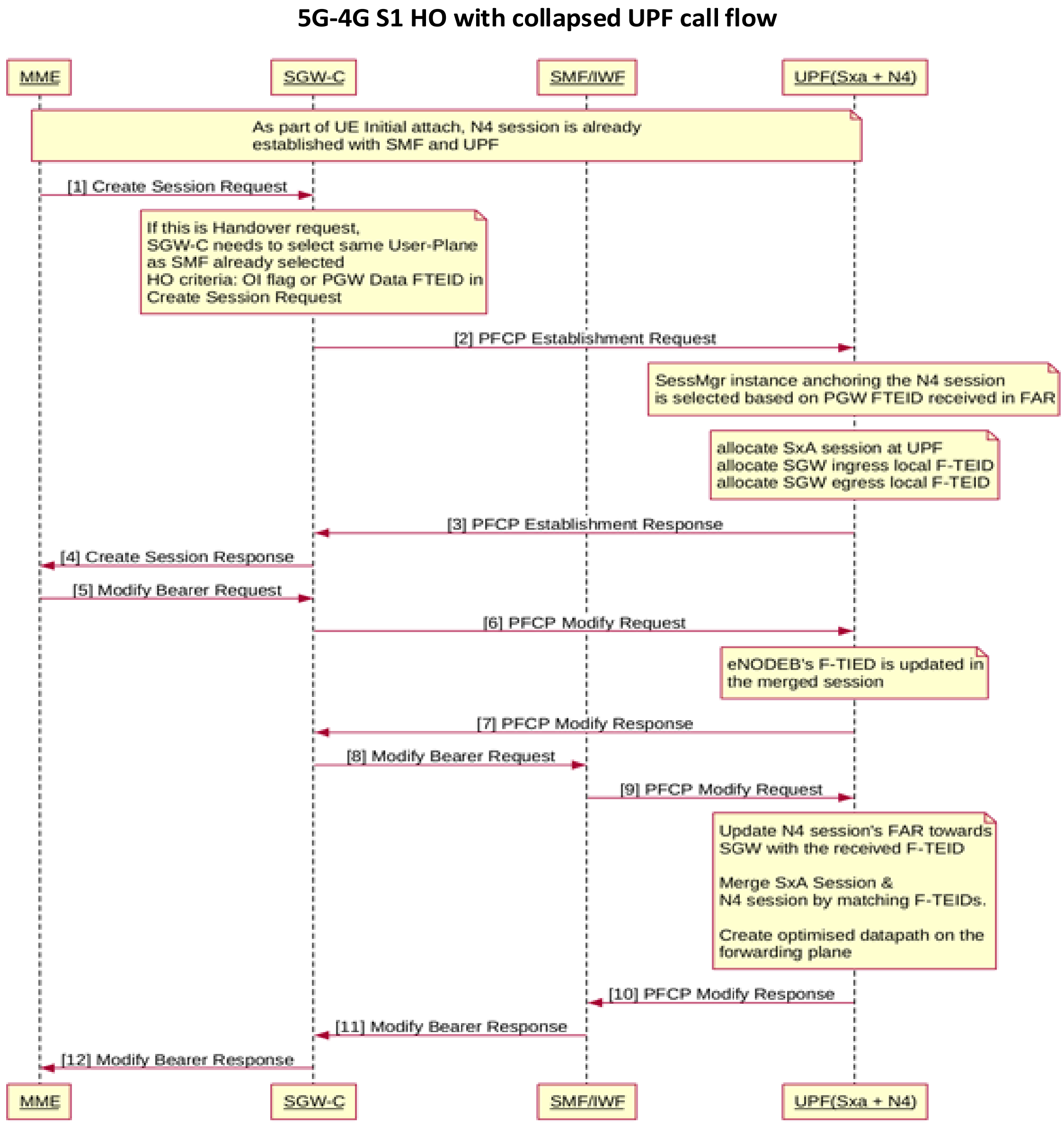

Figure 5: 5G to 4G Handover with Collapsed UPF Flow

In the illustrative call flow as depicted above, various steps may include:

- Step 1. A Create Session Request message is received on a SGW-C. The SGW-C selects one of the UPFs which supports a combined Sxa+N4 interface based on predefined criteria.

- Step 2. The SGW-C initiates a PFCP Establishment Request message towards the UPF instance. The same Session Subsystem instance anchoring the N4 session is selected.
- Step 3. The UPF responds with a PFCP Establishment Response with the allocated F-TEID for SGW ingress and SGW egress endpoints.
- Step 4. The SGW-C sends a Create Session response to the MME.
- Step 5. The SGW-C receives a Modify Bearer Request from the MME with the eNodeB's TEID.
- Steps 6 and 7. The SGW-C updates the eNodeB's TEID at the UPF.
- Step 8. The SGW-C sends a Modify bearer Request indicating a path switch to the SMF/PGW-C.
- Steps 9 and 10. The SMF/PGW-C updates the UPF with the remote F-TEID for the downlink path.
- Steps 11 and 12. A Modify Bearer response is sent from the SMF/PGW-C to the SGW-C, and from the SGW-C to the MME.

A third illustrative call flow, depicting aspects of SGW relocation with a destination SGW-C selecting the same user plane instance, may be described with reference to Figure 6, below.

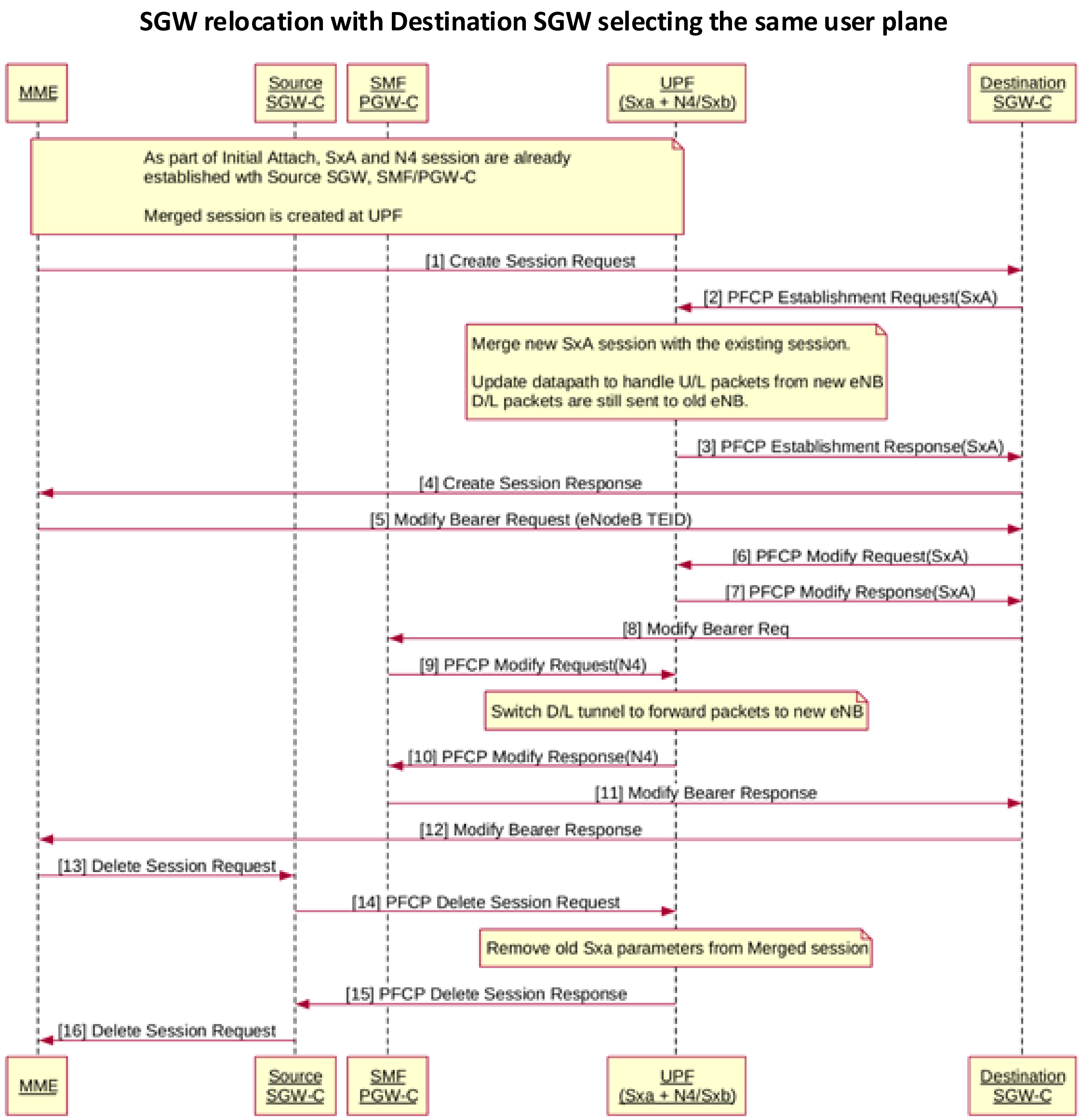

Figure 6: SGW Relocation with Destination SGW Selecting the Same User Plane

In the illustrative call flow that was depicted above:

- A merged session with optimized data path exists on the UPF.
- A UE moves from a Source SGW-C to a Destination SGW-C service area.
- Steps 1 and 2. The Destination SGW-C receives a Create Session Request. The same UPF is selected and a PFCP Session Establishment Request is initiated towards the UPF.
- Existing PDN sessions on the UPF are identified and updated with the new Sxa parameters.

- Steps 3 and 4. Upon receiving a PFCP Session Establishment Response from the UPF, the SGW-C sends a Create Session Response to the MME.
- Steps 5 through 7. The MME updates the SGW-C with eNodeB's F-TEID which in turn is updated on the UPF.
- Steps 8 through 11. The Destination SGW-C informs the SMF/PGW-C of a path switch. The SMF/PGW-C updates the N4 FAR on the UPF with Sxa's F-TEID. The UPF switches the downlink path to the new eNodeB's address.
- Steps 13 through 16. After receiving a Modify Bearer Response from the Destination SGW-C, the MME initiates a Session deletion operation on the old SGW-C.

As described and illustrated above, the techniques presented herein provide significant benefits to an operator looking to deploy 5G services for their subscribers. Such benefits include, for example:

- An ability to merge sessions on a user plane independent of the control plane capabilities (e.g., CUPS EPC and 5GC control planes).
- An ability to create an optimized data path for a UE, by removing additional hops in such a path, thus helping to reduce latency and improve throughout, improve a UPF's performance, and enhance an end user's experience.
- Optimal utilization of user plane resources, due to the merging of sessions and the creation of an optimized data path, resulting in reduced capital expenditures and operating expenses.
- Providing a collapsed user plane (SAEGW-U) independent of control plane nodes (e.g., EPC and 5GC).
- Realizing the benefits of a collapsed user plane even in multi-vendor control plane deployments – e.g., a CUPS EPC and a 5GC can be from separate vendors. An existing CUPS EPC deployment need not be replaced, further reducing capital expenditures.
- Providing significant advantages in the case of remote CUPS/ Multi-Access Edge Computing (MEC) use cases for CUPS EPC and 5GC deployments.