September 2020

OPTIMIZATION OF USER PLANE CONNECTION RESTORATION

Sateesh Reddy Mommidi

Ravi Shankar Mantha

Sandeep Dasgupta

OPTIMIZATION OF USER PLANE CONNECTION RESTORATION

AUTHORS:
Sateesh Reddy Mommidi
Ravi Shankar Mantha
Sandeep Dasgupta

ABSTRACT

During user plane General Packet Radio Service (GPRS) Tunneling Protocol (GTP) GTP-U connection failures in a Third Generation Partnership Project (3GPP) Fifth Generation (5G) network environment that arise from an error indication, it is desirable to restore a connection in as short a period of time as possible (i.e., with minimal latency so that services can be restored to a user as early as possible) with as efficient use of signaling as possible. Techniques are presented that achieve those objectives.

DETAILED DESCRIPTION

As described in Section 5.3.3 of 3GPP Technical Specification (TS) 23.527, the current user plane restoration procedure includes a number of steps. Those steps may be described with reference to Figures 1 and 2, below.

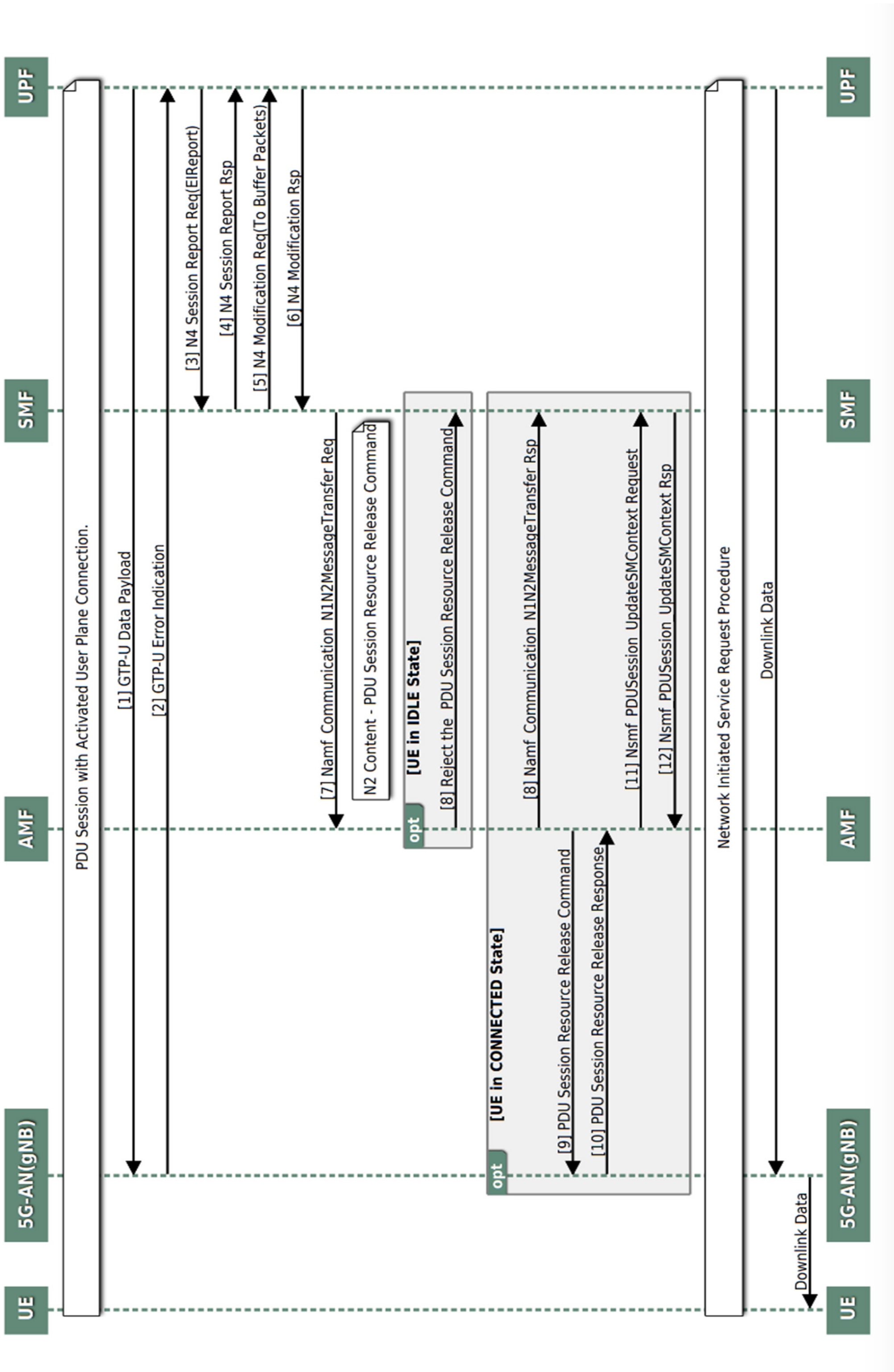

Figure 1: Current GTP-U Restoration Procedure

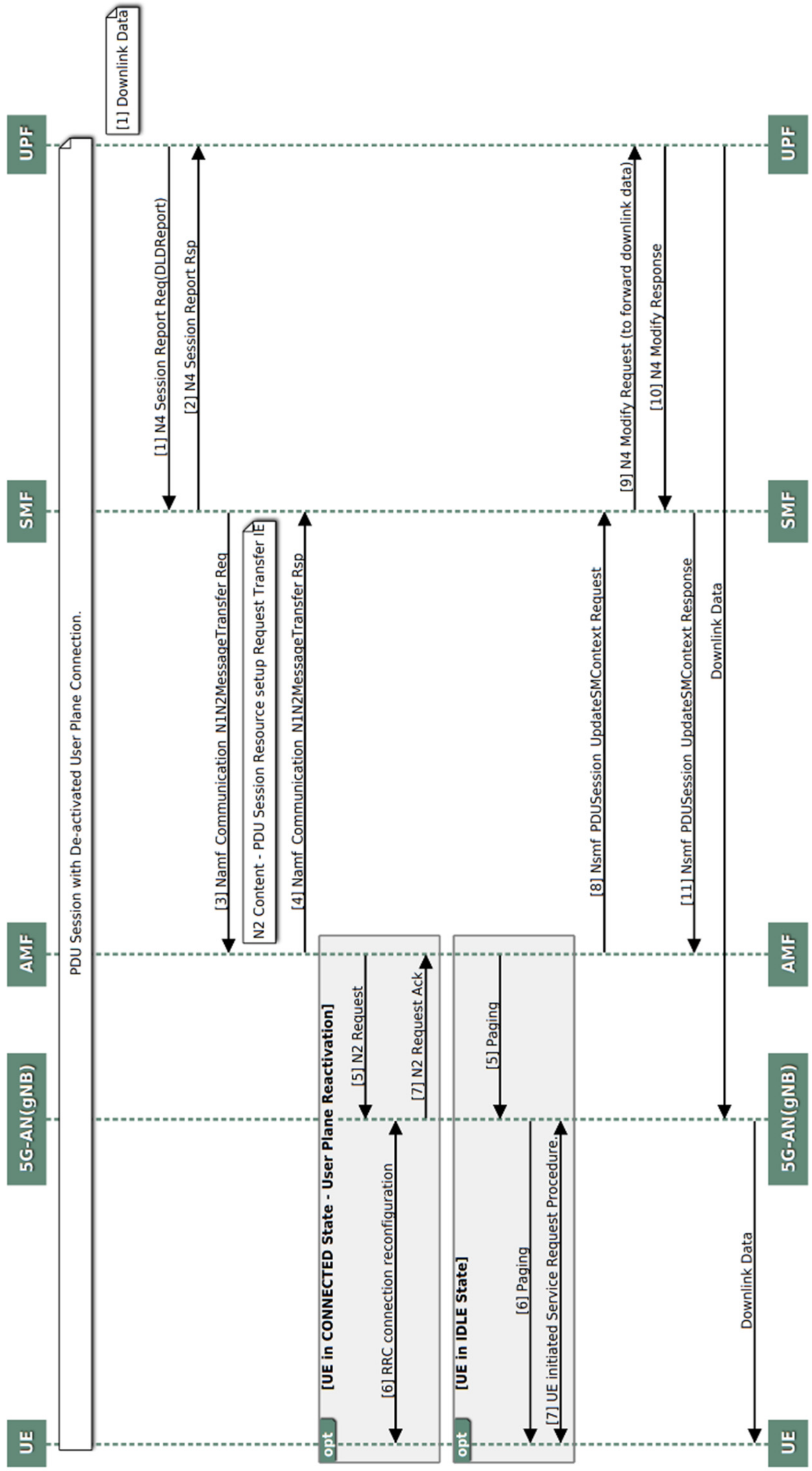

Figure 2: Current Network Initiated Service Request Procedure

As depicted in the above figures:

1. On receiving downlink packet, User Plan Function (UPF) will forward the packet to gNB, gNB sends GTP-U error indication due to lack of data session.

On receiving a GTP-U error indication the User Plane Function (UPF) identifies the Packet Forwarding Control Protocol (PFCP) session and conveys it to a Session Management Function (SMF).

2. The SMF is to modify the PFCP session to instruct the UPF to buffer downlink (DL) packets.

3. If the user plane connection of the Protocol Data Unit (PDU) session is seen as activated by the SMF, the SMF is to initiate a `Namf_Communication_N1N2MessageTransfer` request operation to request the 5G access network (5G-AN) to release the PDU session's resources.

4. Upon receipt of a `Namf_Communication_N1N2MessageTransfer` request to transfer the PDU Session Resource Release Command, the Access and Mobility Management Function (AMF) is to:

a. Proceed with the request if the User Equipment (UE) is in a CM-CONNECTED state for the access network type associated with the PDU session;

b. Otherwise, reject the request with an error indicating that the UE is in a CM-IDLE state for the access network type associated with the PDU session.

5. If the AMF sends a PDU Session Resource Release Command to the 5G-AN, the PDU session's resource release is acknowledged to the SMF.

6. The SMF initiates the network triggered service request procedure to re-activate the user plane connection of the PDU session.

As described above, the restoration procedure involves multiple steps, including:

1. Deactivation of a N4 Tunnel (in buffer mode),
2. Release of a user plane session/connection, and
3. Re-establishment of the user plane session.

Each step of the procedure entails multiple message exchanges between UPF, SMF, AMF and 5G-AN (gNB) nodes. That volume of messaging may be reduced, thus optimizing the restoration procedure, through the techniques described herein whereby the signaling between the different nodes is decreased. The signaling optimization achieved through this solution is not possible with the current 3GPP Specification-driven approach.

It is important to note that a downlink data notification will almost certainly occur after the GTP-U error indicating trigger, necessitating the same being communicated/conveyed to a user.

Figure 3, below, illustrates features of the techniques herein, which provide an efficient mechanism to restore a user plane connection by reducing the signaling between various network nodes.

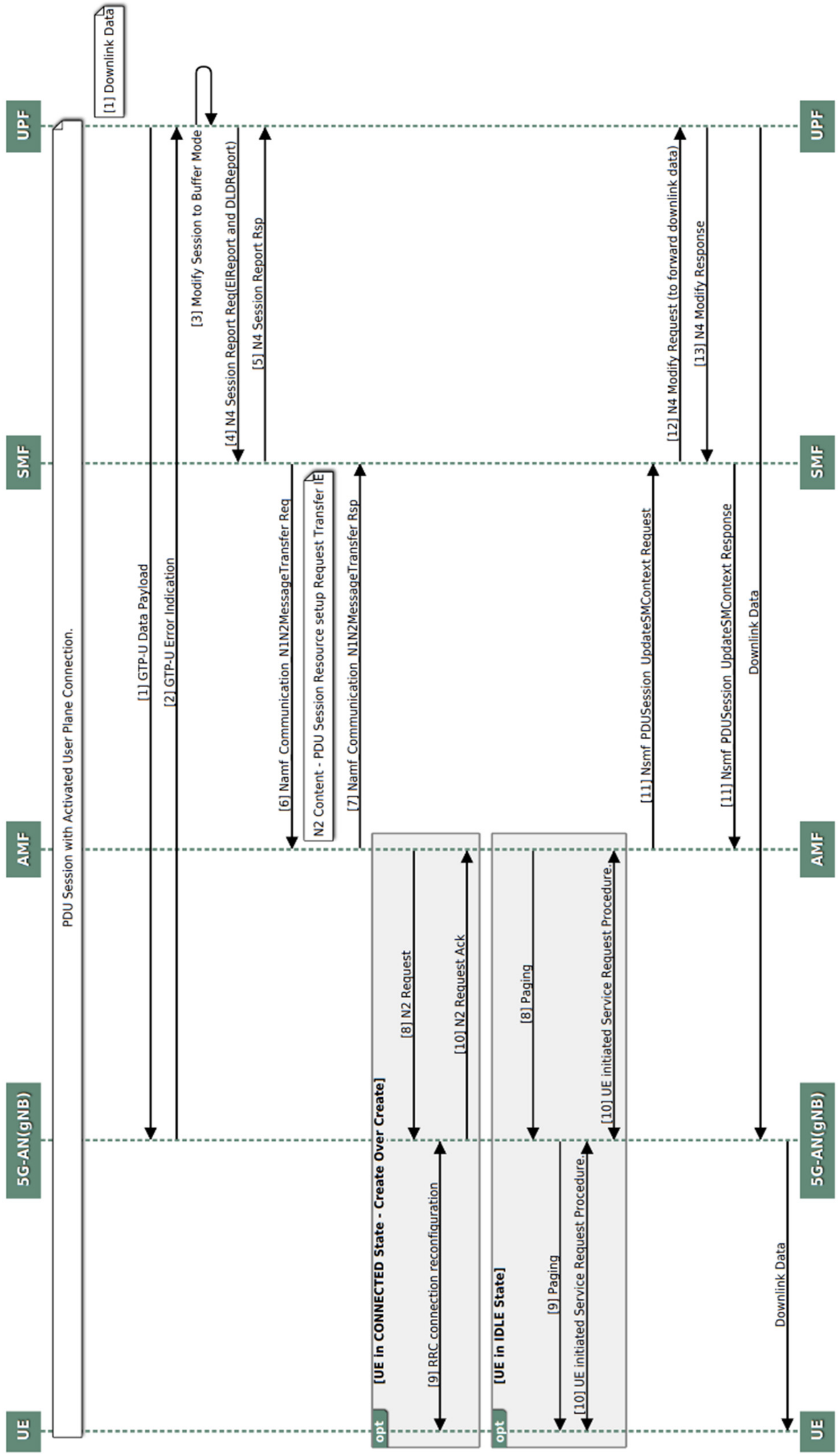

Figure 3: Optimized User Plane Connection Restoration

One aspect of the techniques described herein provides for reducing the signaling between the UPF and SMF nodes following receipt of a GTP-U error indication. This may be achieved by:

1. The UPF modifying the PFCP session to buffer downlink packets on receiving a GTP-U error indication, without requiring the N4 modification request from a SMF.

2. Conveying both a GTP-U error indication and a Down Link Data Notification in a single session report as depicted in Steps 3, 4, and 5. Note, combining multiple session reports in a single message is provided for by various 3GPP Specifications/Standards.

This will avoid the following messages between a SMF and a UPF:

a. A N4 modify request and a N4 modify response to put the session into buffering mode.

b. A N4 session report for a Downlink Data Report (DLDR) and a N4 session report response.

The N4 optimizations can be achieved with existing standards in which the configuration is independent of PFCP standards. However, in some instances, the N4 optimizations may be based on signaling enhancements rather than configuration. For example, in the Packet Detection Rule (PDR) Information Element (IE), a new GTPU-Error-FAR identifier (ID) can be introduced and the corresponding Forwarding Action Rule (FAR) IE can include actions such as buffer or drop. A local policy or configuration on the SMF can control whether packets need to be dropped or buffered.

Another aspect of the techniques described herein provides for reducing the signaling between the SMF, AMF, and 5G-AN (gNB) nodes. This may be achieved by:

1. On receiving a session report with a GTP-U error indication and a DLDR notification, the SMF is to mark the user plane connection as IDLE and start the network-initiated service request procedure (doing away with the N1N2TransferMessage request to release the user plane connection).

This avoids additional signaling between the SMF and AMF nodes and also between the AMF and 5G-AN (gNB) nodes during release of the user plane connection, as described below.

For example, upon receipt of a `Namf_Communication_N1N2MessageTransfer` request to transfer the PDU session resource setup request (as part of the network-initiated service request procedure), the AMF is to proceed with the request if the UE is in a CM-CONNECTED state for the access network type associated with the PDU session. In particular, if an access network finds the PDU session already present, it will abort/delete the old session, honor the creation of a new session, and respond back with a PDU session resource setup response (e.g., 'Create Over Create' as depicted in Figure 3 above).

It is important to note that the particular optimized sequence just described – i.e., the deletion of an old session and the creation of a new session – is one new mechanism within the techniques described herein that achieve a reduction in signaling.

Through the steps described above that involve encompassing a launch of the network-initiated service request procedure by skipping the user plane release procedure, the volume of signaling is reduced between a SMF and an AMF and between an AMF and a 5G-AN (gNB).

As depicted in Steps 6, 7, 8, 9, and 10 of Figure 3:

1. If a UE is in a CM-IDLE state, the paging procedure is initiated to reactivate the user plane connection.

2. The AMF acknowledges the PDU session resource setup response back to a SMF (thus, following receipt of a 5G-AN response, accounting for restoration when a UE is in a CM-CONNECTED mode (as described above) or a paging procedure when a UE is in a CM-IDLE mode).

3. The SMF is to modify the PFCP session to instruct the UPF to forward the downlink packets to a 5G-AN (gNB) instead of taking a buffering action. (The SMF conveys the AN Node (gNB) Internet Protocol (IP) address and tunnel endpoint identifier (TEID) details as part of the modify message).

It is noted that the mechanism to send multiple report types (e.g., a GTP-U error indication and a Downlink Data notification) is already supported by the PFCP protocol. Additionally, the network-initiated service request is a 3GPP defined procedure.

Mechanisms for a configuration push to the UPF to automatically program/modify the session to buffer action on a GTP-U error indication and to combine the GTP-U error indication and DLDR session reports can be achieved through:

1. Configuration, which can be read during UPF initialization/start to apply this action for specific SMFs. A SMF will apply this action after observing a session report with a GTP-U error indication and a Downlink Data notification.

2. The creation of a Buffering Action Rule (BAR) IE, introducing a custom IE to convey the actions (e.g., buffer action and combining of reports) during the PFCP session establishment (from the SMF to the UPF).

Aspects of the techniques described herein do not add any new requirements for, or necessitate any changes to the roles and responsibilities of, the SMF and the AMF. The SMF remains the entity which makes session management decisions. At the AMF, existing functionality based on UE state knowledge (e.g., IDLE or CONNECTED) may be leveraged – e.g., the AMF decides whether to page the UE to initiate the UE-initiated service request for an IDLE UE or forward the N2 payload to the 5G-AN for a CONNECTED UE.

Further, even when the combined error indication and DLDR report is received, the SMF can still choose to release the session based on local configuration. When the decision is not to release the session, re-attachment will be more costly and user unfriendly in such instances. Rather, it is a better approach to seamlessly/transparently restore the connection with less disruption to services. The session will not be stale. If a UE has moved, paging and subsequent handover procedures should address such issues.

Under the techniques described herein, only three nodes require optimization:

1. The UPF needs to combine error indication and DLDR reports.

2. The SMF needs to avoid the N2 release signaling sequence and instead directly initiate a N2 setup signaling procedure, as a configuration option.

3. The gNB needs to recreate the N2 session when a N2 setup request is received for an already present session (a 'Create Over Create' session for the case where only the user plane of the gNB has lost the session but the control plane of the gNB still retains it).

Per-session messaging will be reduced, rather than creating a cascading effect. This will also aid in reducing the latency to restore the data connection so that a user can enjoy a better user experience. Consider, for example, the case where the user plane part of a gNB has reloaded and all of the data path connections are lost. In the absence of the techniques described herein the entire sequence that was presented above would need to be repeated for all the sessions associated with that gNB resulting in a significant signaling surge between a SMF, a UPF and a gNB control plane. The gain in signaling in such a case, even on the N4 interface, is significant.

As discussed herein, the optimizations may be provided for GTP-U failures, during which time the gNB user plane reloads. In a typical cloud Radio Access Network (RAN) involving split control plane-user plane (CP-UP) deployments, it is likely that the UPF will receive an error indication on all DL traffic for active sessions. In such instances, the gNB-CP may not have detected the gNB-UP failure yet, so an access network release may take time. In such a deployment, savings may be significant but conditional.

Resiliency is one of the prime differentiators for network functions (NFs), that is when peer NFs are reloaded or performing in a degrading fashion, an NF has to reduce its signaling as well as processing burden so that it can return to normal operations sooner.

Such optimizations can also be seen in 3GPP standards, such as, for Final Unit Indication on Charging (e.g., Gy/CHF) interfaces in which the Charging Function (CHF) or Online Charging System (OCS) provides an action to the SMF upon final quota grant. Thus, signaling between UPF/SMF/CHF after the quota is exhausted can be avoided.

In summary, during GTP-U connection failures due to an error indication, techniques herein provide for the ability to restore a connection in a short period of time and with minimal latency (so that services can be restored to a user as early as possible) through efficient signaling, which facilitates a reduction in the volume of messaging between various network nodes.