3GPP TSG-SA WG2#162 S2-2405097

Changsha, China, 15-19 April 2024 (was S2-2403973, S2-2404728, S2-2404502)

**Source: Qualcomm Incorporated, Apple, OPPO**

**Title: Addressing open issues for Solution 1**

**Document for: Discussion/Approval**

**Agenda Item: 19.3**

**Work Item / Release: FS\_XRM\_Ph2 / Rel-19**

*Abstract of the contribution: This contribution addresses various open issues for Solution 1.*

# 1. Text proposal

It is proposed to agree the following changes to TR 23.700-70:

>>>>BEGINNING OF CHANGES<<<<

## 6.1 Solution #1: PDU Set content ratio awareness at RAN

### 6.1.1 Key Issue mapping

This solution addresses key issue #1.

### 6.1.2 Description

#### 6.1.2.1 Introduction

SA WG4 recently confirmed that "Commercial XR split rendering and cloud gaming services use Application Layer Forward Error Correction (FEC)." as documented clause 5.7.4 of TR 26.926 [12]. SA WG4 also illustrated the principles underpinning the use of AL-FEC by XR applications in clause 5.7.4 of TR 26.926 [12]:

- Applications send Application Data Units (ADUs) consisting of source symbols, which contain for instance a video frame, and in addition repair symbols.

- If the code that is used is maximum distance separable (MDS), e.g. in case of RaptorQ or Reed-Solomon codes, then the source and repair symbols are distributed across N packets such that the receiver can reconstruct the actual content (e.g. the video frame) if any K out of N packets (with K < N) are received.

In other words, from receiver perspective, it is sufficient to receive K packets of the ADU to be able to reconstruct the actual content. This also implies that once the receiver has successfully received K out of the N packets that the ADU consists of, transmitting the remaining N-K packets of the ADU to the receiver does not add any value because the receiver can already reconstruct the original content based on the first K packets.

According to [13], the overhead of AL-FEC schemes ranges from 10-50% with a typical value of 30%. For AL-FEC based XR content over 5G, this presents a significant optimization opportunity: If NG-RAN successfully delivered the first K PDUs of a PDU Set to the UE, then NG-RAN can refrain from sending the remaining PDUs to the UE because they anyhow do not provide any additional value. (We refer to these PDUs as obsolete PDUs hereafter.) Given that the typical overhead of AL-FEC schemes is 30%, this allows for significant savings in air interface resources.

Additional benefits lie in energy savings for the UE since unnecessary DL transmissions can be avoided and RAN may send the device back to sleep earlier.Editor's note: How RAN determines K packets (i.e. UDP packets) are successfully delivered over an unacknowledged mode data bearer will be decided based on RAN2 feedback.

This solution targets XR applications, e.g., commercial XR split rendering and cloud gaming services as identified by SA4 in clause 5.7.4 of TR 26.926 [12], that are using AL-FEC schemes regardless of the access technology that is used for the application's traffic. In other words, this solution targets applications that are not relying on specific reliability mechanisms that may be offered by a specific access technology but that always apply AL-FEC schemes.

Therefore this solution proposes to make NG-RAN aware of the ratio of PDUs of a PDU Set that are needed at the UE to be able to reconstruct the original content so that NG-RAN can discard the remaining, obsolete PDUs.

It is worth noting that this approach works regardless of whether the AL-FEC encoded traffic is encrypted or not.

One additional key aspect is the perspective of the application. If NG-RAN discards obsolete PDUs, then the application may observe that the UE constantly receives just enough PDUs to be able to reconstruct the original content. It is important to ensure that the application does not - based on this observation - increase the amount of AL-FEC information. This solution proposes to ensure this by informing the application server if a network supports active discarding of obsolete PDUs. Based on this, the application can refrain from increasing the rate of AL-FEC information as long as enough PDUs to reconstruct the original content are received.

An additional aspect that needs discussion is whether the application needs to react to discarded obsolete PDUs by reducing its sending rate to ensure fairness for other competing flows served by the same NG-RAN node. For this discussion, two cases need to be distinguished:

- **Case 1:** While NG-RAN **only** discards obsolete AL-FEC PDUs, i.e., if NG-RAN can still meet the admitted QoS characteristics for other QoS flows in the cell, there is obviously no fairness issue because no other flows are being impacted. Therefore, in this case there is no need for the source to reduce its sending rate.

- **Case 2:** If NG-RAN cannot meet the QoS characteristics of other QoS flow in the cell, e.g., if NG-RAN needs to discard PDUs other than just obsolete AL-FEC PDUs, then fairness may need to be ensured, i.e., also the QoS flows with content ratio information may need to reduce their sending rate.

This leads to the question how an application can distinguish these two cases.

This solution proposes to reuse the existing ECN marking for L4S mechanism to enable applications to distinguish these two cases and to determine when to reduce their sending rate. The key idea is as follows:

- AF indicates support of ECN marking for L4S when providing content ratio to 5GS.

- If the AF learns from 5GS that ECN marking for L4S is not supported (see TS 23.501 [3] clause 5.37.3.1), then the AF may decide to revoke the content ratio that it provided to 5GS by updating the QoS request without providing the content ratio.

- **Case 1 (no fairness issue, no need for sending rate reduction):**

- While NG-RAN only discards obsolete AL-FEC PDUs, i.e., if NG-RAN can still meet the admitted QoS characteristics for other QoS flows in the cell, NG-RAN does not need to from applying ECN markings for QoS flows for which it discards obsolete AL-FEC PDUs.

NOTE X: This aspect of not applying ECN marking in the case described above is assumed to only be captured as a Note in normative specifications.

- The AF does not reduce its sending rate since it does not receive any L4S feedback from the UE.

- **Case 2 (fairness may need to be ensured, potential need for sending rate reduction):**

- If NG-RAN cannot meet the QoS characteristics of other QoS flows in the cell, e.g., if NG-RAN needs to discard PDUs other than just obsolete AL-FEC PDUs for this or other UEs in the cell, then NG-RAN can apply ECN markings for QoS flows for which it discards obsolete AL-FEC PDUs to trigger sending rate reduction for this flow.

- If the UE/application client determines ECN markings for DL packets, the UE/application client provides L4S feedback to the AF following IETF RFC 9330 [14] as described in TS 23.501 [2] clause 5.37.3.1.

- The AF may decide to trigger rate reduction based on the L4S feedback received from the application client.

Editor's note: SA WG2 will reach out to SA WG4 to get feedback on this solution.

Editor's note: SA WG2 will reach out to RAN WG2 to get feedback on this solution.

#### 6.1.2.2 Definitions

The solution is based on the following definitions:

**- Definitions**

- **Content ratio:** The ratio of PDUs of a PDU Set that are needed at the UE to be able to reconstruct the original content.

- **Obsolete PDUs:** All PDUs of a PDU Set that have not been transmitted to the UE yet in a situation where already enough PDUs have been successfully transmitted to the UE according to the content ratio.

#### 6.1.2.3 Solution principles

The solution is based on the following principles:

**- General**

- When requesting or updating QoS for a DL flow, an AF may provide to PCF/NEF - together with PDU Set QoS parameters - the content ratio for the flow. An AF may provide either PSIHI or content ratio for a flow. The AF may additionally subscribe for receiving the indication of support/non-support of PDU Set content ratio awareness.

- How to deal with tethering scenarios:

- Option 1: It is up to the AF to determine whether to provide content ratio information to 5GS. The AF may take into account whether the application client is running on a UE, e.g., the AF may decide to only provide content ratio information to 5GS if the application client is running on a UE. The application client is assumed to be aware whether it is running on a UE and is assumed to be able to inform the AF using application layer means.

- Option 2: The UE may provide an indication of its ability to tolerate proactive dropping to NG-RAN (e.g., using RRC UE Assistance Information). NG-RAN will consider this additional information from the UE before deciding on discarding of obsolete PDUs.Editor's note: How the UE knows whether to send this indication to NG-RAN if FFS.

Editor's note: Whether Option 1 or Option 2 will be used, will be determined during the conclusion phase.

NOTE 2: Depending on the outcome of KI#8 (Enhancement for UE with tethered devices), other options to make the AF aware if the client application is running on a tethered device can be considered in addition.

- If a PCF receives the content ratio from an AF, then the PCF may include the content ratio in PCC rules that it provides to the SMF.

- The SMF provides content ratio to NG-RAN when establishing/modifying a QoS flow.

- If NG-RAN supports PDU Set content ratio awareness and has received content ratio information for a QoS flow, then NG-RAN may discard obsolete PDUs for this flow during congestion.

Editor's note: Whether NG-RAN may discard obsolete PDUs in cases other than during congestion (e.g., to reduce UE power consumption) is FFS and will be determined based on feedback from RAN2.

**- Rate adaptation at the source using ECN for L4S in combination with content ratio information**

- To assist the AF in determining when to reduce the sending rate for a flow for which the AF provides content ratio information to 5GS, ECN marking for L4S may be used as follows.

- When providing content ratio for a QoS flow, the AF may also provide the Indication of ECN marking for L4S to 5GC. If the AF learns from 5GS that ECN marking for L4S is not supported (see TS 23.501 [3] clause 5.37.3.1), then the AF may decide to revoke the content ratio information that it provided to 5GS by updating the QoS request without providing the content ratio.

- Based on operator policy, the PCF may reject a QoS request that includes content ratio information if the AF has not included the Indication of ECN marking for L4S.

- PCF indicates to SMF to enable ECN marking for L4S for that QoS flow as per TS 23.502 [3] clause 4.15.6.6. Based on this, SMF may either activate ECN marking for L4S in NG-RAN (see TS 23.501 [2] clause 5.37.3.2) or in PSA-UPF and, in case of ECN marking at PSA-UPF, also activate congestion reporting by NG-RAN for the QoS flow (see TS 23.501 [2] clause 5.37.3.3).

- If SMF activates ECN marking for L4S in NG-RAN, i.e., if NG-RAN has received an ECN marking for L4S indicator and content ratio information for a QoS flow, then RAN can apply ECN markings for this QoS flow as follows:

- While NG-RAN only discards obsolete AL-FEC PDUs, i.e., if NG-RAN can still meet the admitted QoS characteristics for other QoS flows in the cell, NG-RAN does not ned to apply ECN markings for this QoS flow.

NOTE X: This aspect of not applying ECN marking in the case described above is assumed to only be captured as a Note in normative specifications.

- If NG-RAN cannot meet the QoS characteristics of other QoS flows in the cell, e.g., if NG-RAN needs to discard PDUs other than just obsolete AL-FEC PDUs for this or other UEs in the cell, then NG-RAN can apply ECN markings for QoS flows for which it discards obsolete AL-FEC PDUs to trigger sending rate reduction for this flow.

- If SMF activates ECN marking for L4S in PSA-UPF, i.e., if NG-RAN has received a request from SMF to report congestion information (i.e. a percentage of packets that UPF uses for ECN marking for L4S) for the QoS Flow via GTP-U header extension to PSA UPF, then RAN can send the congestion information as follows:

- While NG-RAN only discards obsolete AL-FEC PDUs, i.e., if NG-RAN can still meet the admitted QoS characteristics for other QoS flows in the cell, NG-RAN does not need to indicate congestion for the QoS flow to the PSA-UPF.

NOTE X: This aspect of not indicating congestion information in the case described above is assumed to only be captured as a Note in normative specifications.

- If NG-RAN cannot meet the QoS characteristics of other QoS flows in the cell, e.g., if NG-RAN needs to discard PDUs other than obsolete AL-FEC PDUs for this or other UEs in the cell, then NG-RAN can indicate congestion for this QoS flow to the PSA-UPF.

- The UE application client provides L4S feedback to the AF following IETF RFC 9330 [14] as described in TS 23.501 [2] clause 5.37.3.1.

- The AF may decide to trigger rate reduction based on the L4S feedback received from the UE/application client.

**- Handling of supporting/non-supporting NG-RAN nodes**

- If an NG-RAN node supports PDU Set content ratio awareness:

- If the NG-RAN node receives content ratio information from SMF, then the NG-RAN node informs SMF in response that it supports PDU Set content ratio awareness.

- As part of Xn and N2 handovers, the target NG-RAN node indicates to SMF that NG-RAN supports PDU Set content ratio awareness.

- If the UE moves from a RAN node (e.g. an NG-RAN or E-UTRAN node) that does not support PDU Set content ratio awareness (referred to as non-supporting node hereafter) to an NG-RAN node that supports PDU Set content ratio awareness (referred to as supporting node hereafter), then SMF provides content ratio information (if available) to NG-RAN.

- If SMF has received content ratio information from PCF, and NG-RAN indicates that it supports PDU Set content ratio awareness, then SMF informs PCF, and subsequently PCF informs the AF, that PDU Set content ratio awareness is supported by NG-RAN.

- If the UE moves from a supporting to a non-supporting node, then SMF informs PCF, and subsequently PCF informs the AF, that PDU Set content ratio awareness is not supported by NG-RAN.

### 6.1.3 Procedures

Existing procedures are re-used and extended with content ratio information and the indication that PDU Set content ratio awareness is supported.

### 6.1.4 Impacts on services, entities and interfaces

**AF:**

- Provide content ratio to PCF

- Receive indication of support/non-support of PDU Set content ratio awareness from SMF

**PCF:**

- Receive content ratio from AF and provide content ratio to SMF as part of PCC rules

- Receive indication of support/non-support of PDU Set content ratio awareness from SMF and inform AF accordingly

**SMF:**

- Receive content ratio from PCF and provide content ratio to RAN

- Provide indication of support/non-support of PDU Set content ratio awareness to PCF

**NG-RAN:**

- Support receiving content ratio information for a QoS flow and support discarding obsolete PDUs for the QoS flow based on the content ratio.

- Indicate support of PDU Set content ratio awareness to SMF during QoS flow establishment and during Xn and N2 handovers.

**UE:**

(Only if Option 2 is chosen) - Indicates its ability to tolerate loss of PDUs in a PDU Set.

\* \* \* 2nd change\* \* \*

6.21 Solution #21: Enhancing PDU Set QoS Handling with Dynamic FEC Related Information Marking in GTP-U

6.21.1 Key Issue mapping

This solution addresses Key Issue #1, "Support of PDU set based QoS handling enhancement".

6.21.2 Description

The XRM traffic may undergo Forward Error Correction, where source data packets are used to generate additional data, called repair packets. Repair packets are generated according to the FEC scheme and transmitted with the source data packets. The repair packets help in the detection and correction of errors in the data stream. As explained in RFC 8627 [32], if the receiver successfully receives all the source packets, then the repair packets are not needed by the receiver. However, if the receiver does not successfully receive some of the source packets, then the receiver can use the repair packets to recover the information that was contained in the source packets that were not successfully received.

In some configurations, source and repair packets may be sent via different IP Port Numbers. In other configurations, source and repair packets may be sent via the same IP Port Numbers but different RTP streams.

- For example, Flex FEC, defined in RFC 8627 [32], is widely used where a number of FEC repair packets are generated from a set of source packets from one or more source RTP streams. These FEC repair packets are sent in a redundancy RTP stream separate from the source RTP stream(s) that carries the source packets. This (i.e. source packets and repair packets are transmitted in two RTP streams) actually provides a backward compatibility for the receivers that do not support Flex FEC. According to RFC 7656 [33], a redundancy RTP stream is an RTP stream that contains no original source data and only redundant data. Furthermore, as explained in S2-2210181 [34]:

 *"although some FEC codes allow for static redundancy ratio, the K/N ratio is not always static during a media delivery session. For example, Video usually relies on Flex-FEC configurations. In such a case, the application is expected to update the 5GS with any configuration change".*

The repair packets for each PDU Set may be dynamic due to the network conditions and the relative importance for real-time communication as shown following text from RFC 8627 [32].

*- It is RECOMMENDED that the amount and type (row, column, or both) of FEC protection is adjusted dynamically based on the packet loss rate and burst loss length observed by the applications.*

*- This would enable differential protection, i.e. application of FEC selectively to packets that require a higher level of reliability than the other packets in the source stream.*

This solution addresses how PDU set based QoS handling can be enhanced when the feature is applied to downlink user plane traffic that is protected via FEC. The solution does not apply to uplink data.

The solution describes two options which are applied to different FEC mechanisms. For example, Option 1 can be used in scenarios where source and repair bits are sent as different blocks of data packets and the UPF can detect whether each block of data packets carries source or repair data respectively. Option 2 can be used in scenarios where source and repair bits are mixed in all PDUs of the PDU Set and a ratio of PDUs at least of PDU Set are needed to use the PDU Set and it is possible for the UPF to dynamically detect the success ratio via inspection of a header inspection.

The principles of this solution are:

- The AF may provide the following information with Flow Descriptors:

- Option1: The Assistance Information that indicates whether traffic that matches the Flow Descriptor (e.g. SSRC) is a source or a repair packet and optionally indicate the information can be used to detect what source packet a repair packet is associated with.

- Option2: The protocol description which indicates that the RTP protocol can provide a success ratio and success ratio marking request.

NOTE: The success ratio is the ratio for PDUs of a PDU Set are needed at least for the usage of the PDU Set.

NOTE X: If the above information is provided by AF, the PSIHI will not be configured.

- The PCF may authorize and send the above Assistance Information to the SMF in PCC Rules.

- The SMF may:

- Option1: indicate in the Packet Detection Rules (PDR) that it sends to the UPF whether traffic that matches PDR is a source or a repair packet and request to mark the repair packet and the information for the source packet a repair packet is associated with.

- Option2: send the Protocol Description and success ratio marking request indication to the PSA UPF.

- indicate RAN to read the FEC related info included in the GTP-U header.

Editor's note: Whether this indication is needed is FFS.

- The UPF:

- Marks the following information in the GTP-U header in DL based on N4 rules and the protocol header of DL packet received from N6:

- Repair packet and optionally information for the source packet the repair packet is associated with.

- a success ratio.

- The AS may include a success ratio in the RTP header extension.

NOTE: Including a success ratio into the RTP header extension requires coordination with SA WG4.

- The RAN may use the information in GTP-U header from the UPF, e.g. to make packet discarding decisions in case of QoS flow congestion.

Editor's note: The overall impact on the sender adaptation and the resulting transported media quality from discarding repair packets is FFS.

6.21.3 Procedures

****

**Figure 6.21.1-1: Setting up with FEC Assistance Information**

1. The AF invokes Nnef\_AFsessionWithQoS\_Create to the NEF. The message includes Flow description(s).

 For each flow description, the AF may provide Assistance Information that

- indicates whether traffic that matches the Flow Descriptor (e.g. SSRC of source packets and the SSRC of repair packets) is a source or a repair packet and optionally indicate what information can be used detect what source packet a repair packet is associated with.

- alternatively, includes the protocol description which indicates that the RTP protocol can provide a success ratio and success ratio marking request.

2. The NEF authorizes the request from the AF.

3. The NEF sends the flow description and the Assistance Information (from step 1) to the PCF.

4. The PCF responds to the NEF.

5. The NEF responds to the AF.

6. The SMF Receives PCC Rules from the PCF. The PCC Rules include FEC Assistance Information.

7. The SMF responds to the PCF.

8. The SMF sends N4 Rules to the UPF.

 The PDRs of the N4 Rules can indicate whether traffic that matches the PDR is source or repair packets and requests to mark the repair packet and information for the source packet the repair packet is associated with. For example, the PDR can indicate the SSRC of source packets and the SSRC of repair packets.

 Alternatively, the N4 Rules include a success ratio marking request indication.

9. The PSA UPF responds to the SMF, and indicates RAN to read the FEC related information included in the GTP-U header.

10. The PSA UPF receives downlink data and uses the N4 Rules to detect whether the packet it is a source or repair packet and extract information from the FEC header to associate repair packets with source packets. The PSA UPF marks the above information in the GTP header in step 11.

 Alternatively, if the AS includes a success ratio in the PDU Set Information Header in the RTP header, the PSA UPF marks the success ratio for the PDU Set in the GTP-U header based on N4 rules in step 11.

11. The PSA UPF sends downlink data to the RAN.

- The PSA UPF includes an FEC source or FEC repair indication in the GTP-U header. When the packet is a repair packet, the PSA UPF also include information in the GTP-U header for the source packet the repair packet is associated with (e.g. information from the FEC header such as the Sequence Number (SN), the L/D offset, mask, etc.).

- Alternatively, the PSA UPF may include a success ratio for PDU Set in the GTP-U header based on information that was detected in the RTP header extension.

 The PSA UPF sends the traffic to the RAN. The RAN may use this information when making packet discarding decisions in the QoS Flow in case of congestion happens for the QoS flow.

Editor's note: SA WG2 will reach out to SA WG4 to get feedback on this solution.

Editor's note: SA WG2 will reach out to RAN WG2 to get feedback on this solution.

Editor's note: How RAN determines K packets (i.e. UDP packets) are successfully delivered over an unacknowledged mode data bearer, is FFS.

Editor's note: Whether the application needs to distinguish and if so how the application distinguishes RAN's intentionally dropped FEC packets from congestion related drops and if the application needs to react by reducing its send-rate to individual packet loss), is FFS.

Editor's note: How the removal of FEC data affects subsequent hops in DL/UL and consequently the end user experience.

Editor's note: How in this envisioned solution the e2e FEC relates to FEC introduced by the radio interfaces' channel coding and HARQ is FFS.

6.21.4 Impacts on services, entities and interfaces

**AF:**

- Provides Assistance Information to the NEF (or to the PCF directly).

**NEF:**

- Receives Assistance Information from the AF.

- Provides Assistance Information to the PCF.

**PCF:**

- Receives Assistance Information from the NEF (or directly from the AF).

- Creates PCC Rules that indicate include the Assistance Information.

**SMF:**

- Indicate RAN to read the FEC related information included in the GTP-U header.

- Creates N4 Rules that indicate

- Option1: whether traffic that matches a PDR is source or repair packet and requests to mark repair packet and information for the source packet the repair packet is associated with.

- Option2: success ratio marking request indication.

**AS (for option2):**

- Includes a success ratio in the RTP header extension.

**UPF:**

- Receives N4 Rules that indicate

- Option1: whether traffic that matches a PDR is source or repair packet and requests to mark the repair packet and information for the source packet the repair packet is associated with.

- Option2: success ratio marking request indication.

- Includes the following information in the GTP-U header based on received N4 rules:

- Option1: an FEC source or FEC repair indication in the GTP-U header and information for the source packet a repair packet is associated with.

- Option2: a success ratio in the GTP-U header.

**RAN:**

- May use the information in GTP-U header from the UPF, e.g. to make packet discarding decisions in the QoS Flow in case of congestion happens for the QoS flow.

**UE:**

- No impact.

>>>>END OF CHANGES<<<<