**3GPP SA4 RTC #130 S4-241902**

**Orlando, FL, US, November 18-22, 2024 Revision of S4aR240087**

|  |
| --- |
| *CR-Form-v12.0* |
| **PSEUDO CHANGE REQUEST** |
|  |
|  | **26.822** | **CR** | pseudo | **rev** | **-** | **Current version:** | **1.0.0** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **x** | Radio Access Network | **x** | Core Network | **x** |

|  |
| --- |
|  |
| ***Title:***  | **[FS\_5G\_RTP\_Ph2] Traffic Pattern Prediction for Real-Time Video Communication** |
|  |  |
| ***Source to WG:*** | Qualcomm Incorporated |
| ***Source to TSG:*** |  |
|  |  |
| ***Work item code:*** | FS\_5G\_RTP\_Ph2 |  | ***Date:*** | 11/18/2024 |
|  |  |  |  |  |
| ***Category:*** | **B** |  | ***Release:*** | Rel-19  |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-10 (Release 10)Rel-11 (Release 11)Rel-12 (Release 12)**Rel-13 (Release 13)Rel-14 (Release 14)Rel-15 (Release 15)Rel-16 (Release 16)* *Rel-17 (Release 17)* *Rel-18 (Release 18)* |
|  |  |
| ***Reason for change:*** | This addresses Key Issue #12: Enhancements of Data Burst MarkingIt is important to investigate whether it is feasible to predict the traffic patterns (delay and data burst size) for real-time multimedia applications.To address comments at the RTP Ad Hoc Meeting on Oct 23, 2024:* Rufael: If everything is periodic why is it not predictable? What am I missing?
* Liangping: Time to next frame is periodic, but this is the size of the next video frame. It is a different metric.
* Serhan: Sent commented version over reflector (30 min ago). Suggest to take offline.
 |
|  |  |
| ***Summary of change:*** | Analyzed the traffic pattern prediction problem for both conversational video and XR split rendering. Addressed the comments. |
|  |  |
| ***Consequences if not approved:*** | Any potential solution for traffic pattern prediction may not be justified. |
|  |  |
| ***Clauses affected:*** |  |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  |  |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  |  |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  |  |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

\* \* \* \* 1st change (all new)\* \* \* \*

## 6.x Traffic pattern prediction for real-time video communication

### 6.x.1 Key issue mapping

This maps to Key Issue #12.

### 6.x.2 Description

### 6.x.2.1 The need for traffic pattern indication

The discussion focuses on real-time video traffic. There are two categories: conversational video and XR video (including real time gaming and augumented reality). For conversational video, the delay requirement is 150ms, while for XR video the delay requirement is 50ms and 10 ms, as shown in the table 6.x-1 extracted from [3].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5QIValue | Resource Type | Default Priority Level | Packet Delay Budget(NOTE 3) | Packet ErrorRate  | Default Maximum Data Burst Volume(NOTE 2) | DefaultAveraging Window | Example Services |
| 2 | (NOTE 1) | 40 | 150 ms(NOTE 11,NOTE 13) | 10-3 | N/A | 2000 ms | **Conversational Video** (Live Streaming) |
| 3 |  | 30 | 50 ms(NOTE 11,NOTE 13) | 10-3 | N/A | 2000 ms | **Real Time Gaming**, V2X messages (see TS 23.287 [121]).Electricity distribution – medium voltage, Process automation monitoring |
| 80 |  | 68 | 10 ms(NOTE 5,NOTE 10) | 10-6 | N/A | N/A | Low Latency eMBB applications **Augmented Reality** |

Table 6.x-1 QoS characteristics for real-time video communications

Given that the delay requirement is 150ms for conversational video, the penalty of letting a video frame wait for the next DRX one period at the RAN is likely negligible, whereas this is not the case for XR video, due to the much more stringent delay requirement.

**Observation 1:** For conversational video, the benefit of indicating the delay to the next frame is neglibible.

### 6.x.2.2 Time to the next data burst (TTNB)

For simplicity, we consider a data burst that consists of a video frame. For the more general case of data burst, it will be enven more difficult to predict the TTNB. For conversational video, two factors affect the TTNB:

* **Video encoding time:** The video encoding time depends on the complexity of the scene. For the same target frame size, typically the more complex the scene is, the long it takes to complete the encoding. It may be difficult for the sender to predict the scene and the video encoding time.
* **Rate adaptation:** The frame rate is part of the rate adaptation, e.g., triggered by congestion control. With rate adaptation, the delay to the next frame may change. If a rate reduction is requested after the time to the next frame was predicted and sent in the current PDU Set, the prediction may be become obsolete. This is illustrated in Figure 6.x-2. This issue can be avoided if the sender delays the rate adaptation until the predicted frame is transmitted.



Figure 6.x-2 Time to the next data burst is obsoleted by a rate reduction request triggered by congestion control.

For XR split-rendering video, TTNB is typically fixed based on TR26.926. It is more efficient to use control-plane signaling.

**Conclusion 1:** For XR split-rendering video, it is beneficial to indicate the time to the next burst, and control-plane approach is more efficient.

#### 6.x.2.3 Burst size

When a data burst consists of a video frame, as noted in Solution #16, if a packager generates all packets of the burst at once, there is no delay in knowing the burst size and there is no error in the data burst size. However, a data burst may consist of multiple PDU Sets according to TS 23.501 [3]. For example, when a data burst consists of multiple application data units (ADUs) such as an audio frame and video frame, to aovid buffering which introduces latency, the sender needs to predict the size of the ADUs that come after the first ADU. This is illustrated in Figure 6.x-3.

NOTE: TS 23.501 [3] allows a data burst to include multiple PDU Sets. The benefit of such a data burst for low-latency communication is FFS.

NOTE: The PDU Set is not necessarily the minimum unit of traffic. A data burst does not have to be composed of PDU Sets.



Figure 6.x-3 The sender needs to predict the size of PDU Set #(N+1) when generating the burst size.

Therefore, there is a need for predicting the size of the next ADU. When doing prediction, one way is to use the past frame sizes of a particular media stream to predict the size of the next frame of the same media stream. *This reduces the burst size prediction problem to one of predicting the frame size.* However, it needs to be verified that such prediction is feasible. To test it, an experiment was carried out. The setup is shown in, the HMD is connected to a split rendering server that performs split rendering via Wi-Fi 6 (IEEE 802.11ax). The content is Steam VR with complex graphics. The video codec is a commercial hardware HEVC codec with the IPPP GOP structure. The video frame rate is driven by the display refresh rate, which is 90 FPS. The implementation of the split rendering server is proprietary, not based on WebRTC.



Figure 6.x-4 Experimental setup for XR video.

The video frame size for the left video stream as a function of the frame number is shown in Figure 6.x-5.

Figure 6.x-5 The frame size of the left video.

Some prediction algorithms are tested, and the results are shown in Figure 6.x-6:

* EWMA (exponentially weighted moving average). The weight is selected to achieve the smallest prediction error.
* MA (moving average): the prediction is set to the average of the frame sizes in a sliding window of most recent frames. The size of the window is 5 and is selected to achieve the smallest prediction error.
* Past observation: the last frame size is used as the prediction of the current frame size.



Figure 6.x-6 The frame size predictions of the left video: actual frame size (blue), EWMA prediction (red), MA (cyan), and past observation prediction (green).

The absolute value of the average prediction error (normalized by the average frame size) is: 4.76% for EWMA, 4.75% for MA, and 5.28% for the past observation method.

The distribution of the prediction error for the best performing method – MA – is show in Figure 6.x-7. It is seen that although the average error is 4.75%, the error varies a lot, as much as 30%.



**Figure 6.x-7 CDF of the prediction error (relative to the mean frame size) for the MA prediction method.**

**Observation 1:** for XR video, the prediction error for the next frame size can be significant.

**Observation 2**: Given the large prediction error, it is not clear what the RAN will do with the predicted size.

**Conclusion 2:** for XR video, if the prediction of the next video frame size is used in the data plane, the prediction accuracy (e.g., the 99% confidence interval) needs to be indicated along with the prediction.

\* \* \* \* end of first change \* \* \* \*