**Agenda item:** 10.6

**Source:** Qualcomm Inc.

**Title: [5G\_RTP\_Ph2] The need for redefining Time to the Next Data Burst based on experimental results**

**Document for** Agreement

# Introduction

Time to next data burst (TTNB) may be beneficial for RAN to achieve UE power saving, but the definition should be also considered from the operation’s perspective – whether there is a way to signal it so that the benefit materializes.

The term of TTNB originated from SA4. Since then, it has been defined in SA2 in TS23.501 as follows:

*The time to next data burst, … is the interval between the transmission of the last PDU in the current data burst and the first PDU of the next data burst*

Regarding the accuracy of TTNB indication, in R2-2407733 [1], RAN2 stated:

***Regarding Question 2 on time to next burst****, RAN2 thinks that time to next burst may be useful for the network scheduling for downlink, if it is provided in advance and is reliable and accurate at RAN.*

In the 5G\_RTP\_Ph2 WID, SA4 agreed to indicate TTNB in the RTP header extension for dynamically changing traffic characteristics on the timeliness and accuracy of the indication:

1. *Conduct normative work on burst size, time to next burst, data boosting indication and the definition of data burst.*

*NOTE: RAN2 has indicated that TTNB may be useful if provided in time and is reliable. SA4 needs further evaluation before proceeding with normative work.*

The timeliness requirement obliges the RTP sender to indicate the TTNB before the next data burst is generated, and this calls for prediction.

There was disagreement within SA4 about the meeting of “reliable and accurate”. SA4 sent an LS to RAN2 in SA4 #131-bis-e meeting to RAN2 for clarification and received the reply in R2-2504812 [2]. In the rest of the paper, we look at the consequence (in terms of prediction errors) if the SA2 definition of TTNB were adopted, using the experimental data from S4-241835 [3] and described in TR 26.822 [5], and propose to change the definition of TTNB or use a new term altogether in order to meet the accuracy requirement from RAN2.

# RAN2 requirement on TTNB accuracy

In R2-2504812 [2] RAN2 provided the following feedback on the requirement of the TTNB indication accuracy:

RAN2 thanks SA4 for the question on the preferred accuracy of the TTNB indication.

The TTNB can be used by RAN in its scheduling and configurations to achieve benefits such as UE power savings. Therefore, the more accurate the indicated TTNB is (up to a practical limit and considering the impact of factors such as network jitters), the more benefits (e.g. UE power savings) can be achieved. More specifically, from RAN2’s perspective, the desired accuracy of the TTNB indication equals the shortest PDCCH skipping duration, which is 0.125 msec (i.e. one slot for 120 KHz SCS). If this level of accuracy is not feasible with SA4, then the most accurate indication that SA4 can practically support is acceptable to RAN2.

RAN2 clearly expects the TTBN indication to be as accurate as possible, up to 0.125msec.

# Prediction results

We present prediction results for the packet traces provided in S4-241835 [3] for conversational video and in Clause 6.24.2.3 of TR 26.822 [5] for XR split rendering if the SA2 definition of TTNB were adopted. The prediction uses exponential weighted moving algorithm (EWMA) with the optimal weight found through exhaustive search. The prediction for the kth TTNB is calculated as

$$p\_{k}=αp\_{k-1}+(1-α)t\_{k-1}$$

where $p\_{k-1}$ is the previous predicted TTNB, and $t\_{k-1}$is the previous measured TTNB, and $α$ is the weight in the range from 0 to 1.

The Python script in [3] produces two TTNB sequences, saved as a\_b\_bursts\_windows\_duplex.csv and b\_a\_bursts\_windows\_duplex.csv. In what follows, we present the predictions for them separately.

**3.1 Conversational video: The first TTNB sequence (a\_b\_bursts\_windows\_duplex.csv)**

The optimal weight of EWMA predictor is found by searching for the weight that minimizes the mean absolute prediction error over the entire first TTN sequence. The optimal EWMA is 0.82, as shown in Figure 1.



Figure 1 Mean absolute prediction error vs the WEMA weight for the first TTNB sequence

The measured TTNB (i.e., actual TTNB) vs the predicted TTNB with the optimal predictor for the first TTNB sequence is shown in Figure 2. It is seen that although the predicted TTNB well tracks the trends of the measured TTNB, it often over-estimates or under-estimates the measured TTNB with a significant prediction error.



Figure 2 The measured TTNBs and the predicted TTNBs for the first TTNB sequence

The cumulative distribution function (CDF) of the prediction error, i.e., the predicted TTNB minus the measured TTNB, is shown in Figure 3. It is seen that the prediction over-estimates by 14.7ms 10% of the time, and under-estimates by 19ms 30% of the time.



Figure 3 CDF of the prediction error for the first TTNB sequence

**3.2 Conversational video: The second TTN sequence (b\_a\_bursts\_windows\_duplex.csv)**

For the second TTN sequence, the optimal EWMA weight is 0.99, as shown in Figure 4. The measured TTNB and the predicted TTNB is shown in Figure 5. Figures show that the prediction over-estimates by 12.4ms 20% of the time, and under-estimates by 13.4ms 20% of the time.



Figure 4 Mean absolute prediction error vs the EWMA weight for the second TTNB sequence



Figure 5 The measured TTNBs and the predicted TTNBs for the second TTNB sequence



Figure 6 CDF of the prediction error for the second TTNB sequence

**3.3 XR split rendering traffic**

The optimal weight of EWMA is 0.99, as shown in Figure 7, and the corresponding prediction error and the CDF are shown in Figures 8 and 9.

****

Figure 7 Mean absolute prediction error vs the EWMA weight for the third TTNB sequence

****

Figure 8 The measured TTNBs and the predicted TTNBs for the third TTNB sequence

****

Figure 9 CDF of the prediction error for the third TTNB sequence

Given that RAN2 expects the error to be 1/8 msec, and the errors from the prediction observed here, we have the following observation:

**Observation:** The definition TTNB in TS23.501 could result in prediction errors that are orders of magnitude larger than the desired error expected from RAN2.

# Conclusion

Based on the observation in clause 3, we have the following proposal:

**Proposal 1:** SA4 redefines TTNB or replace it with a new term with the associated definition and inform other working groups including SA2, RAN2 and CT.

In S4-250356 [4], the idle period was proposed to replace TTNB. The idle period can be considered as the TTNB currently defined in TS23.501 minus the (magnitude of) the prediction error.

There are two alternatives to address the signalling accuracy issue:

**Alternative 1**: Replace the TTBN definition currently in TS23.501 with the following new definition:

**TTNB** is the interval between the transmission of the last PDU in the current data burst and a time instant that is earlier than or equal to the transmission of the first PDU of the next data burst.

**NOTE:** The definition uses “a time instant that is earlier than or equal to the transmission of the first PDU of the next data burst” rather than “the transmission of the first PDU of the next data burst” because, at the time TTNB is indicated, the RTP sender may not yet know exactly when the first PDU of the next data burst will be transmitted. The uncertainty arises from factors such as variable and not fully predictable video encoding delays. Despite this uncertainty, no PDU is transmitted by the RTP sender during the indicated TTNB interval.

**Alternative 2**: keep the TTBN definition currently in TS23.501 but replace TTNB from the RTP header extension for dynamically changing traffic characteristics in TS26.522 with the idle period defined as follows:

**Idle period** is the duration of a time interval from the transmission of the last packet of the current data burst to a time before the transmission of the first packet of the next data burst during which the traffic source is guaranteed not to send delay-sensitive packets.

Because Alternative 1 minimizes the spec changes especially in WG’s other than SA4, we have the following proposal.

**Proposal 2:** agree the following change to the definition of TTNB and send an LS to other WGs to inform them of the change.

**TTNB** is the interval between the transmission of the last PDU in the current data burst and a time instant that is earlier than or equal to the transmission of the first PDU of the next data burst.

**NOTE:** The definition uses “a time instant that is earlier than or equal to the transmission of the first PDU of the next data burst” rather than “the transmission of the first PDU of the next data burst” because, at the time TTNB is indicated, the RTP sender may not yet know exactly when the first PDU of the next data burst will be transmitted. The uncertainty arises from factors such as variable and not fully predictable video encoding delays. Despite this uncertainty, no PDU is transmitted by the RTP sender during the indicated TTNB interval.

# References

[1] S4-250606, LS on Indicating Time to the Next Data Burst (TTNB), 3GPP TSG SA WG4 #131-bis-e, online, April 14-18, 2025.

[2] R2-2504812, Reply to LS on Indicating Time to the Next Data Burst (TTNB), 3GPP TSG RAN WG2 #130, Malta, 19~23 May 2025.

[3] S4-241835, Additional statistics for RTP Senders and Receivers in Solution #13, KI#12 for dynamic traffic, SA WG4#130, November 2024.

[4] S4-250356, On the ending time of the Time To the Next Data Burst (TTNB), 3GPP SA4 #131 Meeting, Geneva, Switzerland, 17-21 February 2025.

[5] TR26.822, Study on 5G Real-time Transport Protocol Configurations, V1.2.0, Rel-19, Nov 2024.