Qualcomm

RP-212043

Views on XR enhancements for Rel-18

Source: Qualcomm

Agenda item: 9.0.2

Document for: Discussion

Introduction

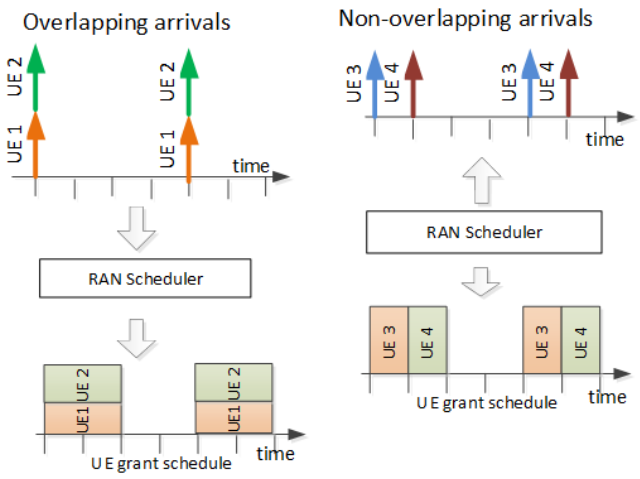
- Email discussions on enhancements for XR were conducted from Aug 30th to Sep 3rd, and their summary is presented by the moderator in **RP-211655** (Moderator's summary for discussion [RAN93e-R18Prep-05] Enhancements for XR, Nokia).
- In this document, we discuss the following aspects.
  - Starting from SI vs. WI in Rel-18?
  - Motivations and proposals on various identified topics:

Starting from SI vs. WI in Rel-18?

- The main objectives of Rel-17 XR SI, “study on evaluations for NR” are performance evaluation.
- Main objective of the Rel-17 XR study was baseline evaluation results
- In addition, evaluation results showing performance improvement via new schemes not currently supported by the spec are expected to be discussed and potentially captured in the TR
  - Such enhancement schemes have not been discussed at length and are not expected to be much discussed as part of the ongoing Rel-17 XR study item
- Therefore, a choice between the two below options need to be adopted for Rel-18:
  - Option 1: Work item from the beginning of the Rel-18
  - Option 2: Start with a Study Item (e.g., for 6 months), followed by a Work Item (e.g., for 12 months)

RAN Awareness @ Application

- Motivation
  - Improve Application awareness of RAN status for better user experience management
  - RAN and application coordination for traffic arrival pattern that improves capacity and power
- **Expose RAN information to the Application to enable codec adaptation**
  - **RAN reports** - RAN performance metrics including supported Burst-rate and notification of temporary degradation
  - **5G aware traffic generation** - RAN recommendation to application on traffic timing to stagger traffic burst offset among UEs



Overlapping arrivals among UEs:

Detrimental to XR **Capacity** & UE **Power** Consumption

RAN reports - QoS information

QoS notification of Temporary Degradation Event

- **Proposal**: Augment QoS notification to convey "temporary degradation" of service
  - Convey duration of degradation
  - Convey "Reduced GFBR", "higher PDB", "higher PER" that can be handled during degradation
  - Predict degradation event time
- Temporary degradation events:
  - Handover events
  - High power dissipation
- **Expected Action** from application:
  - Temporary reduction in application bit-rates, traffic periodicity
  - Trade-off user experience for service continuity

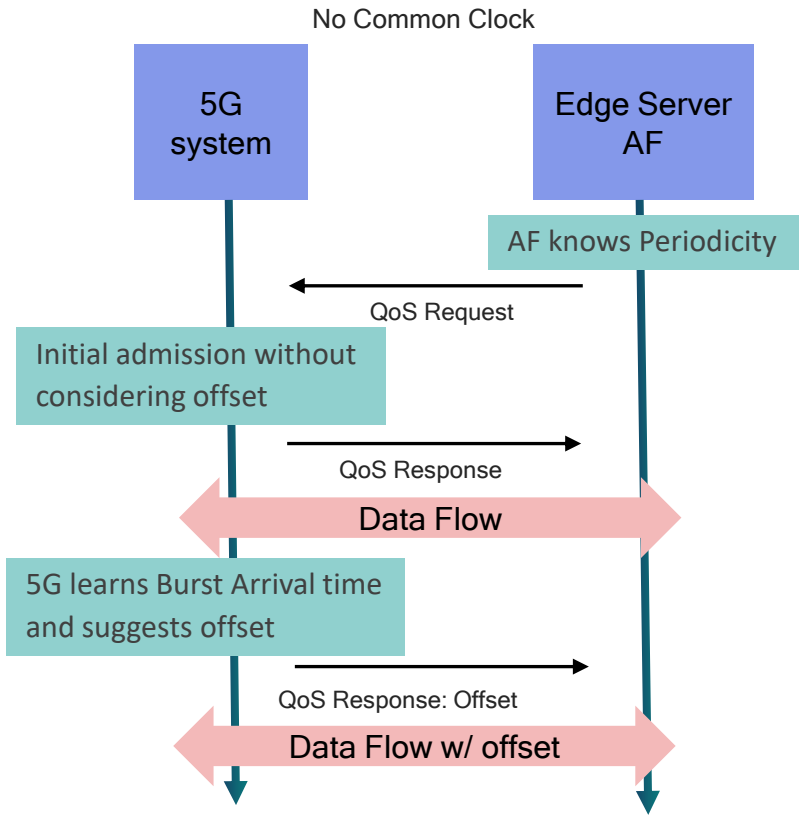
QoS notification of Supportable Burst-rate

- Currently supported in 5GS
  - QoS notification for supportable GFBR when requested GFBR cannot be supported
- **Proposal**: Augment QoS notification to convey supportable max burst-rate
- **Expected Action** from application:
  - Change peak to average of the generated traffic to reduce required burst-rate

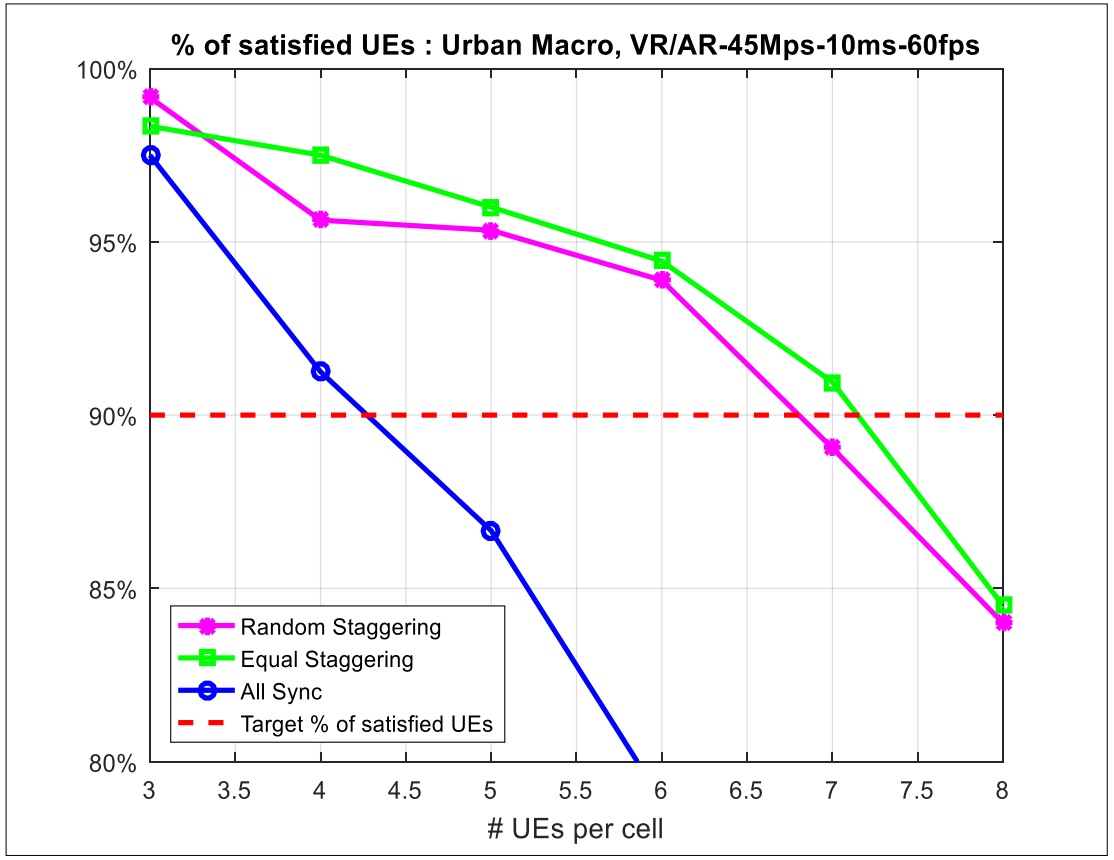
5G Aware Traffic Generation

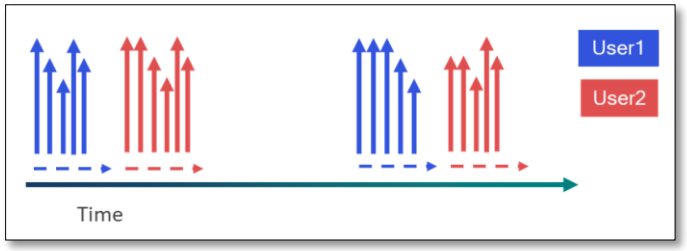
Proposal

- RAN may receive XR burst periodicity information from SMF during user plane resources activation.
  - Statically configured or provided by XR app via NEF
- RAN determines optimal time offset from arriving XR traffic burst for subsequent bursts
- RAN provides this information to CN (SMF) which provides to XR App via NEF
- XR app expected to accommodate subsequent XR traffic bursts based on received time offset
- No clock synchronization required



5G Aware Traffic Generation: FR1 Performance

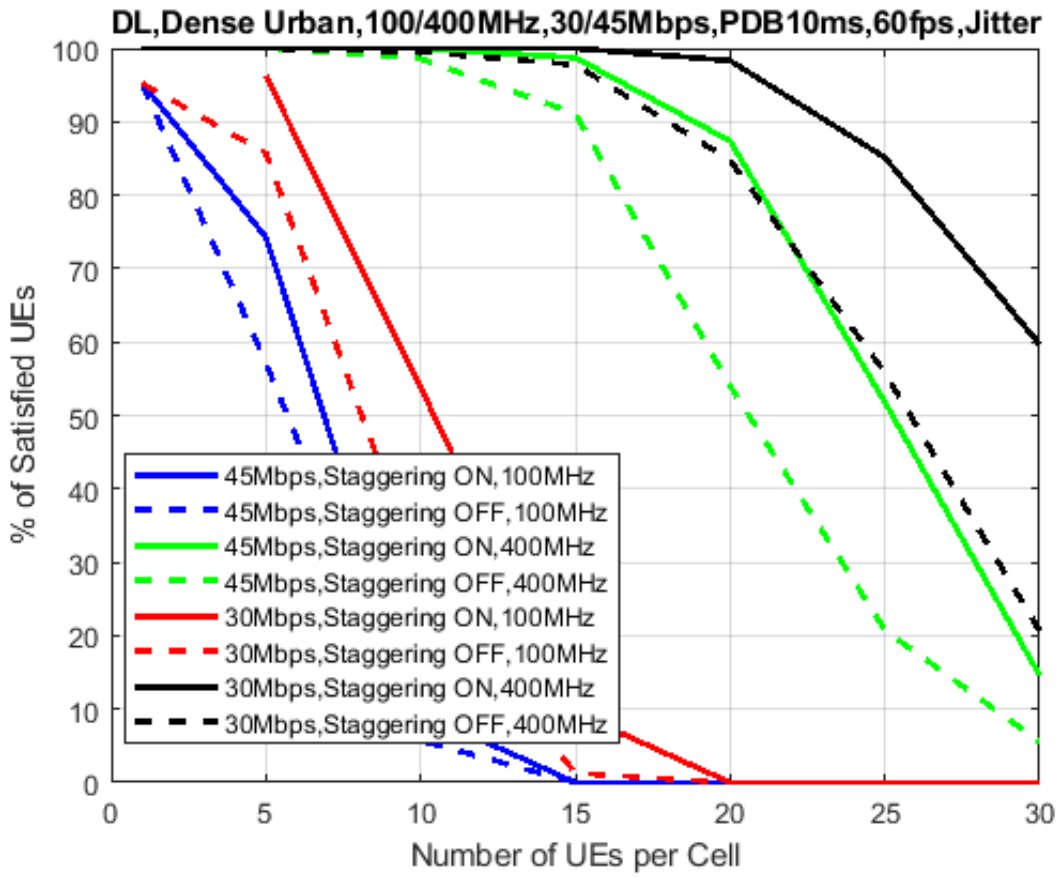




- Simulation assumptions:
  - Single-cell
  - Wideband SU-MIMO scheduling
  - Traffic: 45 Mbps, 10 ms PDB, 60 frames/s
- Frame offset modes compared:
  - *All sync:* All UEs share same frame offset
  - *Equal staggering:* The frame-arrival offset of the UEs are equally spaced within a frame period
  - *Random staggering:* The frame-arrival offset of the UEs are uniformly distributed
- Appropriate staggering across UEs within one cell could increase XR capacity

Ref: RAN1-106e Tdoc R1-2108251 (Qualcomm)

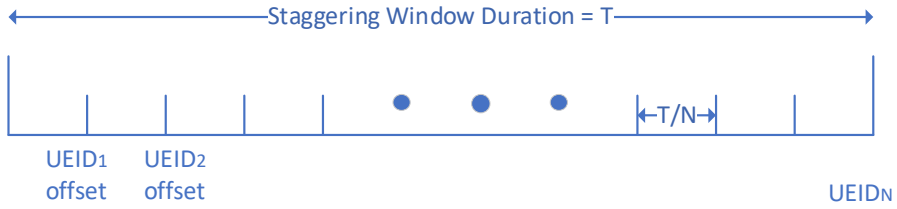
5G Aware Traffic Generation: FR2 Performance



Ref: RAN1-106e Tdoc R1-2108251 (Qualcomm)

Modes:

- **Staggering OFF:** All UEs' initial and subsequent packet arrival times are aligned.
- **Staggering ON:**



- A staggering window size is divided to uniform offsets based on the number of UEs in the network. For each user, the first packet transmission offset is selected based on the UE ID. This model ensures UEs across the entire network (and the cell) do not have the same packet arrival time.

Simulation Assumptions:

- **Staggering window size, T:** 16ms
- **Network:** 21cells with wraparound, ISD = 200m,
- **Numbers of UEs in network, N =** 21* Number of UEs per Cell
- **Channel model:** UMa Channel
- **Traffic:** Data rates = 35Mbps and 40Mbps, PDB=10ms, PER =1%, Frame Rates = 60 frame/s.
- Other traffic and systems parameters used for this evaluation can be found in [REF]. These parameters are based on Release 17 XR RAN1 SI agreed parameters for FR2 XR evaluations.

Conclusion:

- Staggering UEs' packet arrival times results in an increase in system capacity in Dense Urban deployment scenarios. The increase in capacity increases with bandwidth.

Application Awareness

We agree with the two majors selected for study and normative work which have been captured by R18Prep email discussion conclusion:

1. Identify the XR traffic characteristics and application layer attributes beneficial/feasible for the gNB to be aware of, e.g., the QoS flow association, frame-level QoS, ADU-based QoS, XR specific QoS etc. that was concluded from KPI and QoS
2. Application layer information (e.g., frame rate, delay, packet importance, etc.) to aid XR-specific handling, e.g., scheduling, radio bearer handling, etc; One potential mechanism for this is to introduce UE assistance information;

Our opinion is that #2 item may be broaden to include areas such as beam management and measurement configurations. Specifically,

- The RAN and/or UE may leverage XR application information such as XR viewport or pose information to predict the channel and UE/gNB beam changes so that necessary beam management updates or measurement configurations may be triggered. For example, if the pose information indicates an increase in speed or large movements at the UE, the UE may want to request a decrease in measurement periodicity or increase in the number of CSI measurements from the RAN. The RAN may also trigger the measurement configuration updates.

- Therefore, we recommend that #2 item under “Potential Areas of Work for Application Awareness” should be modified to clearly include beam management and measurement configuration:
- “2. Application layer information (e.g., frame rate, delay, packet importance, pose information etc.) to aid XR-specific handling, e.g., scheduling, radio bearer handling, beam management, measurement configuration etc; One potential mechanism for this is to introduce UE assistance information;”

Power Saving

- **The following has been captured by R18Prep email discussion conclusion:**
  - XR-optimized C-DRX for aligning C-DRX with XR service periodicity and jitter, XR-specific multi-flow aspects. Enhancements to PDCCH monitoring and WUS can also be considered.
- **Proposal:** Consider following areas further for XR power saving:
  - Further **enhancements for SPS/CG**
    - Dynamic config update for SPS/CG, etc
  - Techniques for **UL/DL alignment**
    - UL/DL alignment could reduce # of UE wake-up/On time
    - CG/SR restriction to DL or DRX Active Time
  - **UE assistance information** for helping gNB configuration for efficient resource allocation
    - UL XR Traffic pattern information, e.g., preferred UL pose periodicity, offset, data size, etc
  - Frequent **UL pose tx**
    - Very frequent UL tx makes it hard for UE to enter sleep state due to frequent DRX timer extension.
    - Decoupling/reducing DRX retx timer update and UL pose transmission (aka UL only tx) could reduce UE awake time.
  - Timer based **BWP/CA adaptation**
    - Periodic nature of XR traffic could be fully utilized by timer-based BWP/CA switching for faster switching and power saving.

Ref: RAN1-106e Tdoc R1-2107377 (Qualcomm)

Capacity

- **The following has been captured by R18Prep email discussion conclusion:**
  - XR-specific resource allocation and scheduling enhancement, SPS and CG enhancements including dynamic SPS/CG parameter update, multi-TB SPS/CG, HARQ-ACK enhancement etc... to address XR-specific non-integer periodicity and jitter.

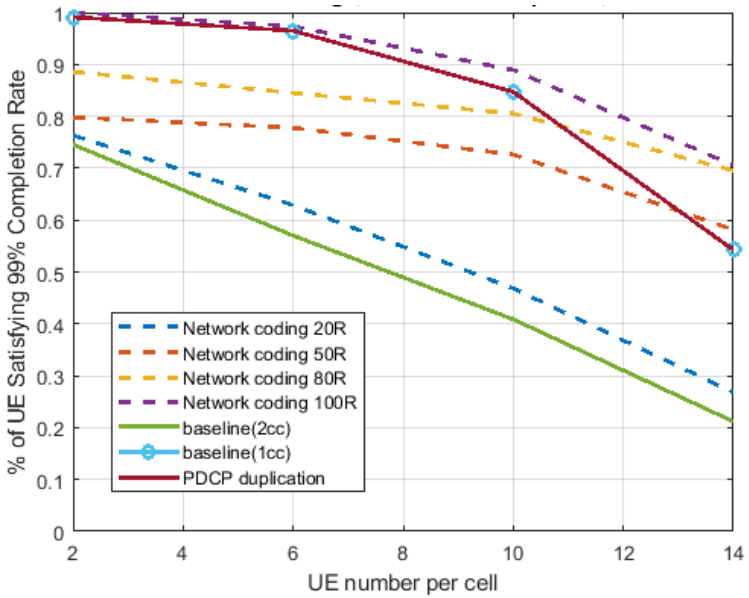
- **Proposal: Consider following areas further for XR capacity**
  - **SPS/CG enhancements**
    - Multi-SPS/CG configurations with dynamic parameter adjustment and SPS/CG occasions skipping based on the XR traffic characteristics.
    - UE assistance information for helping gNB configuration for SPS/CG parameters
      - UL XR Traffic pattern information, e.g., preferred UL pose periodicity, offset, data size, etc
  - **Network coding** (in next slides)

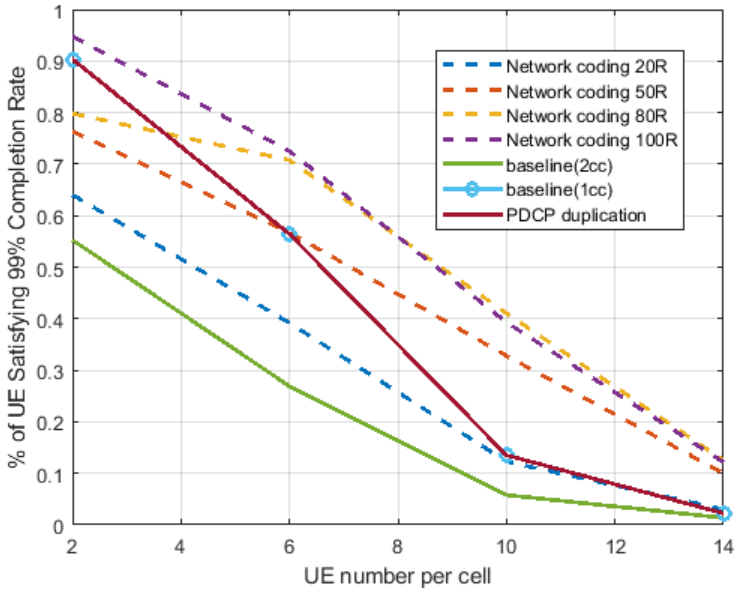
Network Coding/Outer Coding for XR Traffic

- Network Coding (NC)/Outer Coding (OC) inclusion into the RAN protocol stack could improve XR capacity
- XR traffic characteristics include relatively high data rate, stringent latency bound, and reliability requirements
  - Given these requirements, NC/OC could offer performance benefits over other schemes, e.g., baseline HARQ/RLC and PDCP duplication
    - Redundancy added upfront in case of NC/OC could help XR traffic fulfil latency and reliability requirements without having to resort to HARQ/RLC retransmissions that would increase the delay of packet reception
    - In case of blocking of one of the links (e.g., one of the CCs or TRPs), due to the added redundancy, the XR application frames could be successfully received only from the data transferred over the link that is not blocked
      - In the absence of NC/OC, the packets on the blocked link would have to go through multiple HARQ/RLC retransmissions, resulting in potentially significant delays/packet drop rate
- Compared to the PDCP duplication, NC/OC can offer
  - Adaptive redundancy – more efficient operation by adapting to the current traffic load and reliability/latency requirements
    - With a constant redundancy (one RLC packet copy per carrier), PDCP duplication in many cases results in excessive system load, which may stall XR traffic and reduce capacity
  - More efficient coding – PDCP duplication is just a repetition coding

Network Coding/Outer Coding for XR Traffic

- The initial findings suggest that utilizing the NC/OC in the RAN protocol stack could provide performance benefits
  - Smaller delays, smaller packet loss rate, larger capacity - compared to the baseline HARQ/PDCP duplication schemes for the variety of XR traffic requirements, such as traffic bit rate, and number of users in the system





XR capacity analysis for 2CCs with a blocking pattern

- 45 Mbps (left) & 100 Mbps (right) traffic
- NC redundancy parameterized (20%, 50%, 80%, and 100%)
- NC performed at RLC layer

Qualcomm

Thank you!

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

©2018 Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark of Qualcomm Incorporated, registered in the United States and other countries. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes Qualcomm’s licensing business, QTL, and the vast majority of its patent portfolio. Qualcomm Technologies, Inc., a wholly-owned subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of Qualcomm’s engineering, research and development functions, and substantially all of its product and services businesses, including its semiconductor business, QCT.