Coordinating Multi-access Edge Computing with TDM-PON based Mobile Franthaul for 5G

Wei Wang

BUPT Ph.d candidate & UC Davis visiting student
Email: weiw@bupt.edu.cn, waywang@ucdavis.edu

UCDAVIS

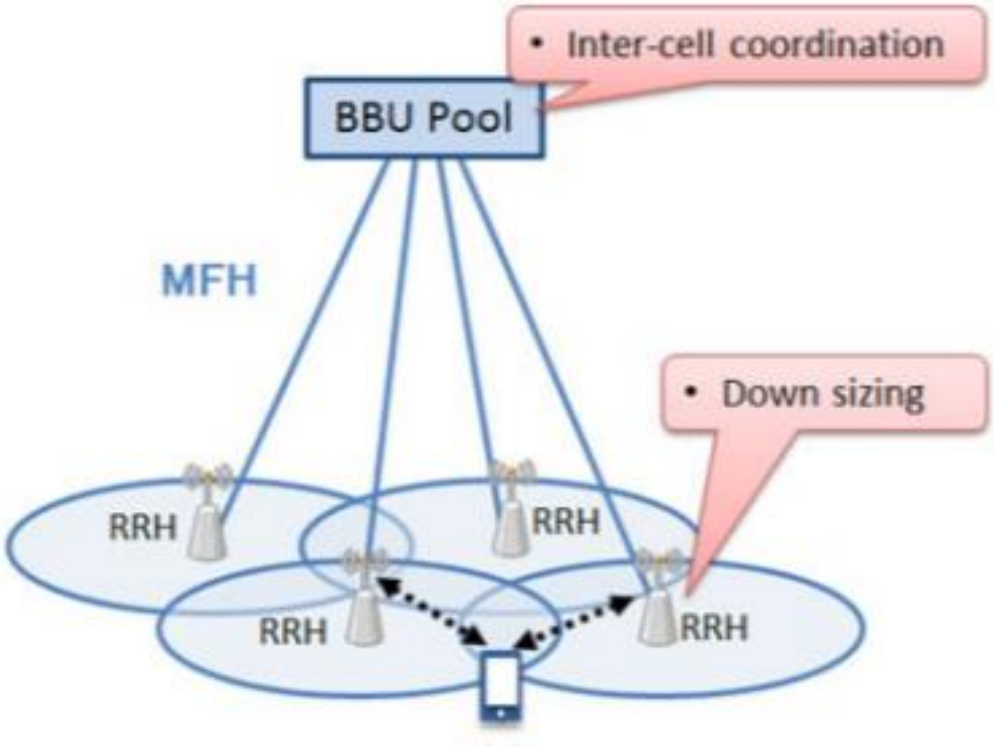
Group Meeting, Sep. 29, 2017

Contents

- Mobile Fronthaul (MFH) in C-RAN for 5G
- Issues for TDM-PON as MFH
- Architecture and latency in C-RAN with PON and MFH?
- Use cases for coordinating PON with MFH
- Summery

What is Mobile Fronthaul?

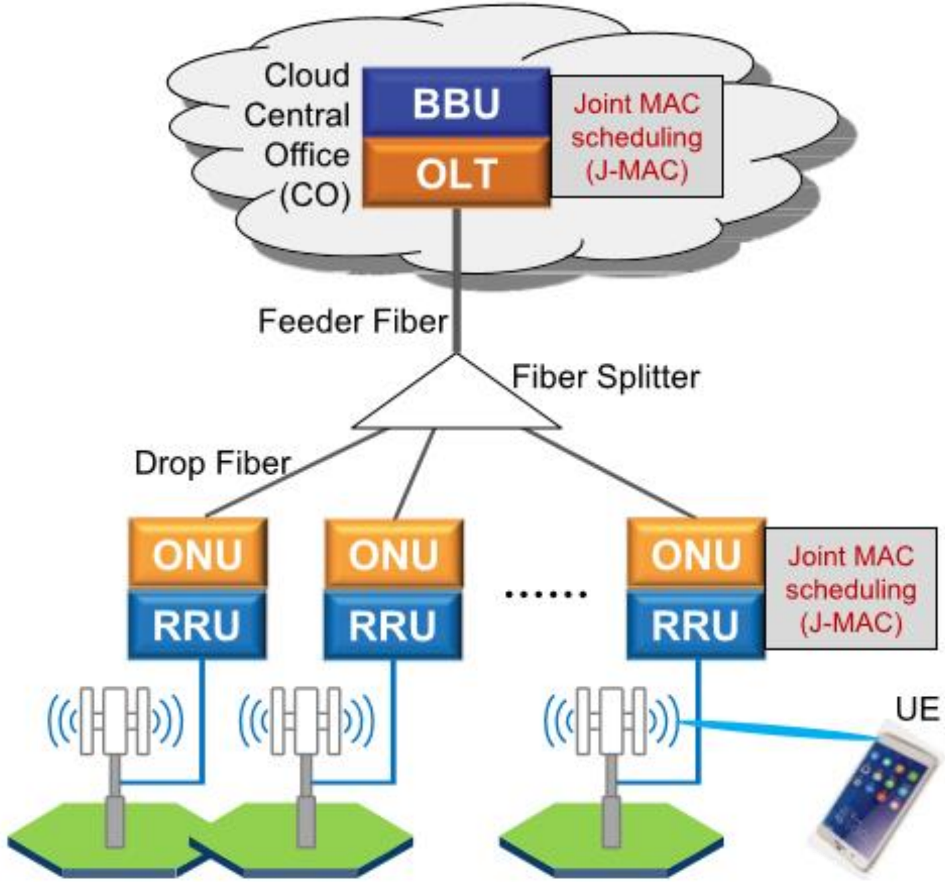
- Mobile Base-Station=Base Band Unit (BBU) + Radio Head (RH)
- Cloud-RAN(C-RAN): Centralized BBU pool + Remote Radio Heads (RRHs)
- RRH is the antennas sites with only RH functions
- Benefits offered by C-RAN: 1) inter-cell coordination, 2) down-sizing of antennas
- Mobile Fronthaul (MFH): intra-base station transport, used to connect BBU and RRH.



J. i. Kani, J. Terada, K. I. Suzuki and A. Otaka, "Solutions for Future Mobile Fronthaul and Access-Network Convergence," in *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 527-534, Feb.1, 1 2017.

TDM-PON for Fronthaul

- Why PON? (1) one to multi-point architecture; (2) cost-efficient; (3) reuse existing fibers
- https://www.nokia.com/en_int/news/releases/2017/06/20/nokia-bell-labs-first-to-show-use-of-ultra-low-latency-10g-pon-for-mobile-fronthaul
- Dynamic Bandwidth Assignment (DBA) in TDM-PON systems.
- Optical Line Terminal (OLT) assigns bandwidth grants to each Optical Network Unit (ONU) according to bandwidth requests from ONUs.
- Issue:
- The assignment procedures result in a delay of around 1 ms, which may exceed the latency threshold of 5G wireless communications.



Latency in TDM-PON

- With conventional DBA

- Control message latency
- 1) propagation latency of REPORT message
- 2) grant processing time
- 3) propagation latency of GATE message

- Data latency
- 4) propagation latency of data (distance)
- 5) transmission latency of data (b/w)

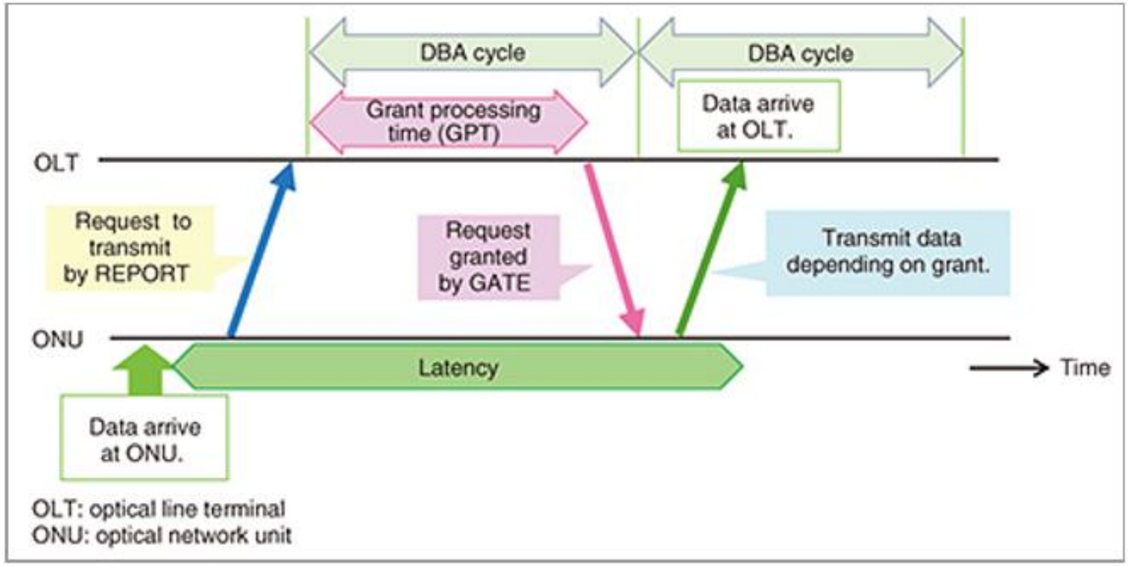
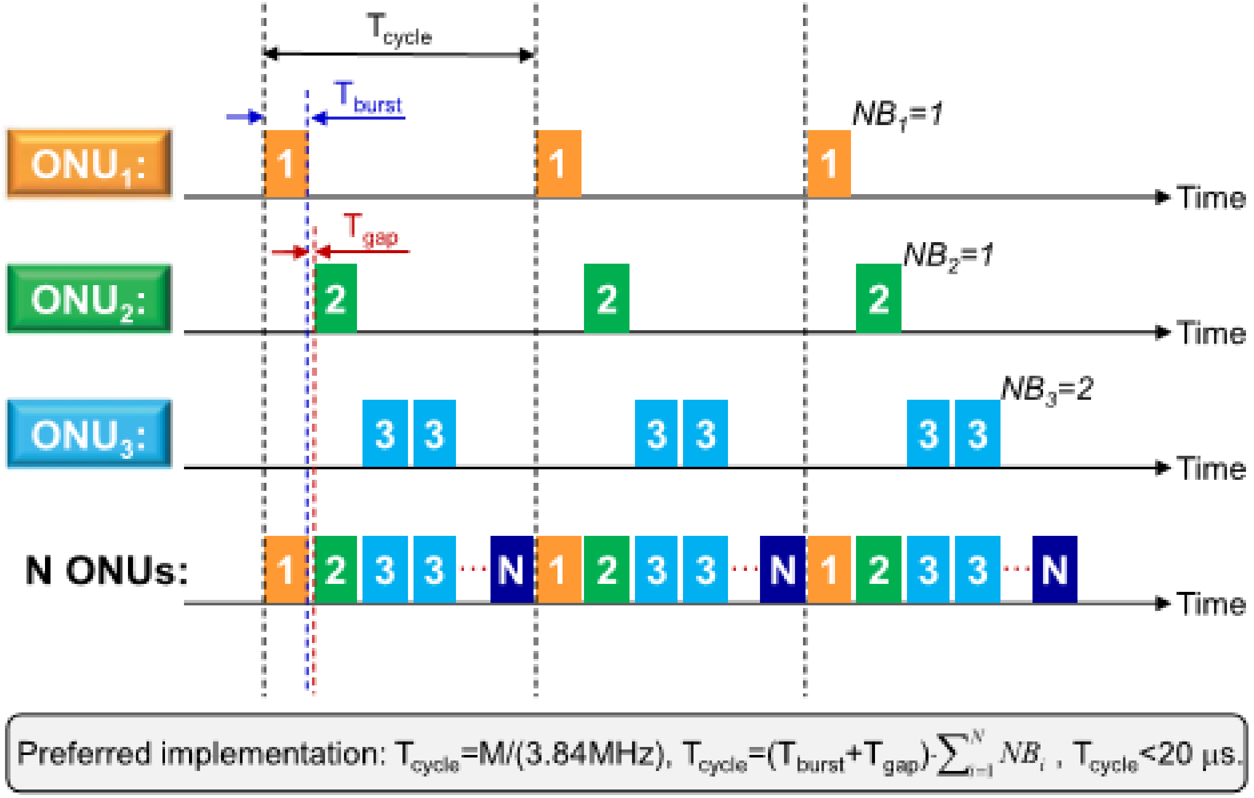


Fig. 2. Message exchange between ONU and OLT in SR-DBA.

Solutions for low latency TDM-PON

- Solution A: (Fixed scheduling)
- Accelerated burst scheduling of PON
- T_{cycle} : time for OLT to scan through all ONUs once (20μs preferred)
- T_{burst} : predetermined burst period for ONUs to transmit CPRI frames
- T_{gap} : to avoid implementation of imperfection-induced burst collision

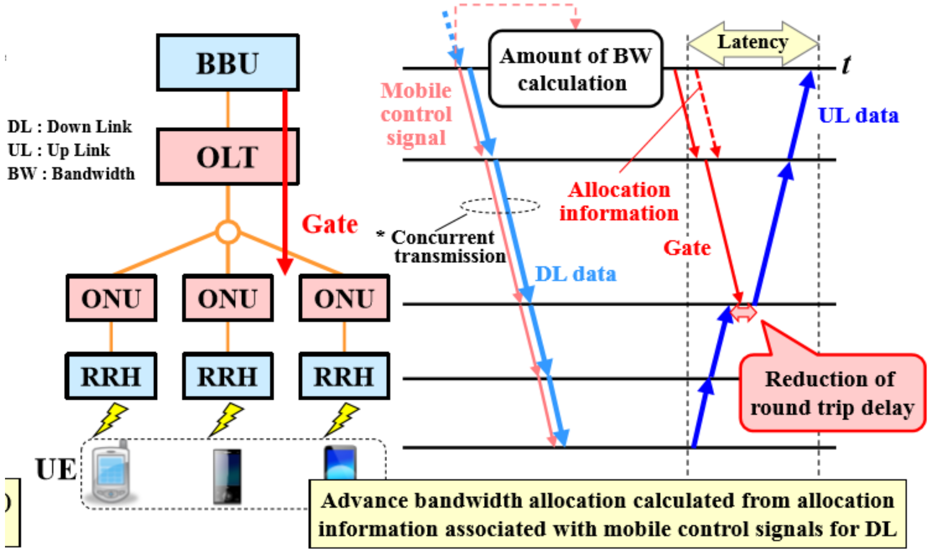
- Flexible bandwidth allocation can be realized by assigning each ONU a given number of bursts per cycle



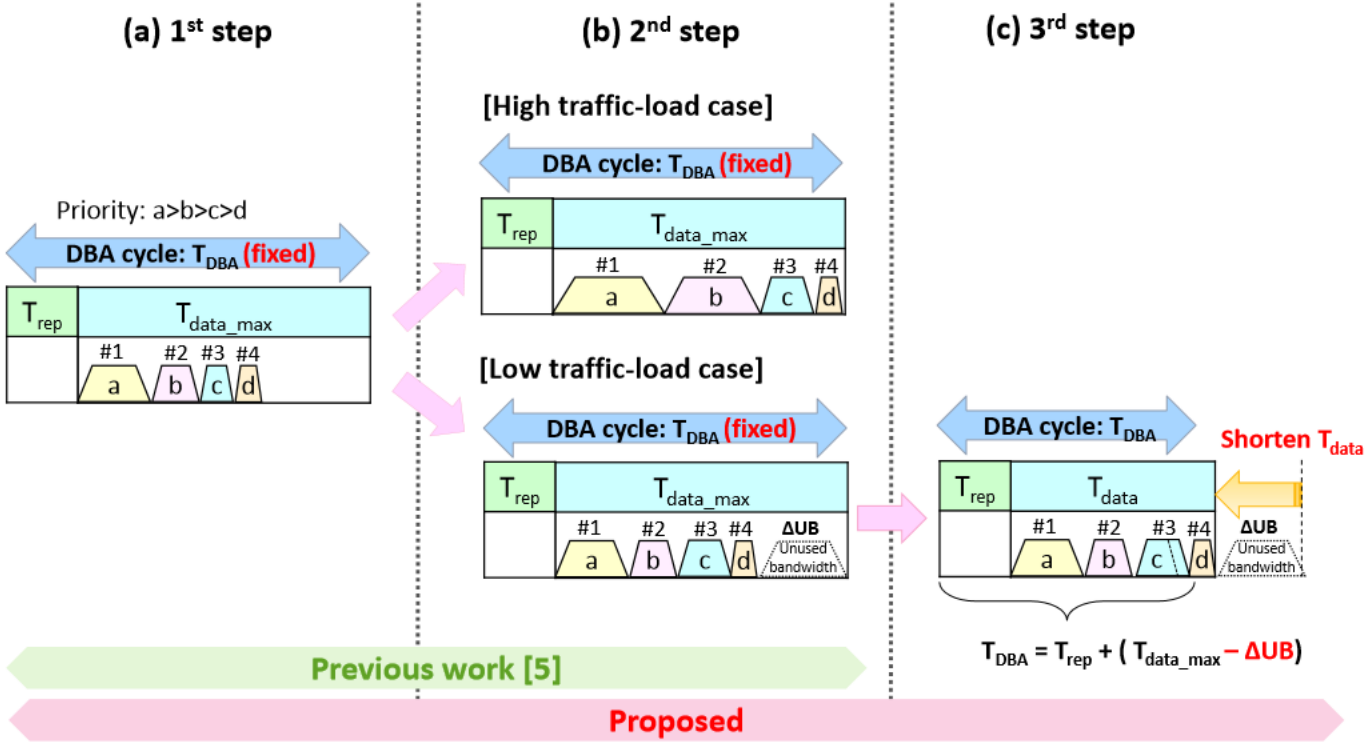
X. Liu and F. Effenberger, "Emerging optical access network technologies for 5G wireless [invited]," in IEEE/OSA Journal of Optical Communications and Networking, vol. 8, no. 12, pp. B70-B79, December 2016.

Solutions for low latency TDM-PON

- Solution B: (Advance Scheduling)
- Request Bandwidth(RB)
- Guaranteed Bandwidth(GB)
- Unused Bandwidth(EB)



(b) M-DBA

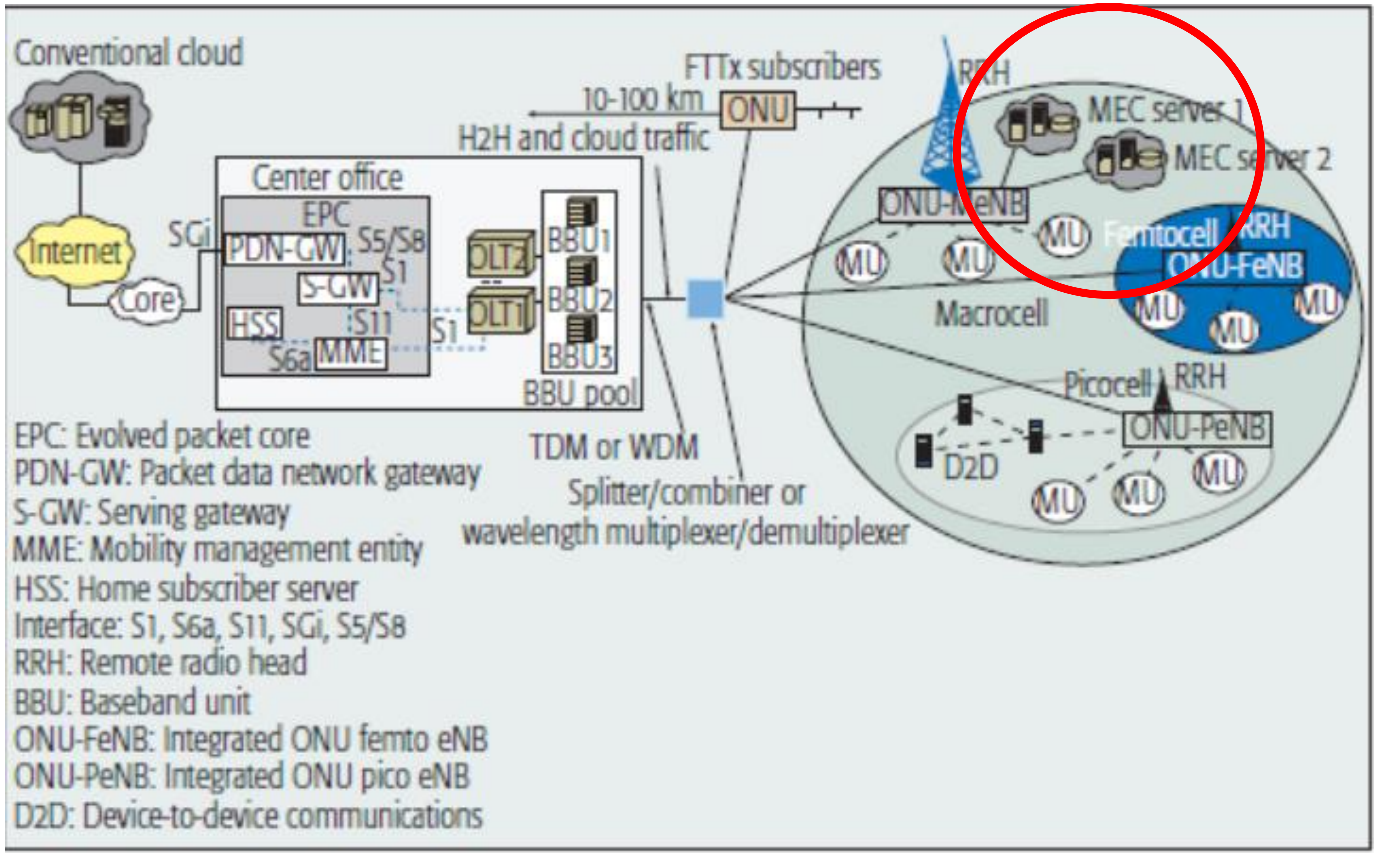


Where should MEC locate at in Fronthaul?

- MEC in PON-based Fronthaul

- MEC at ONU side?
- According to CPRI or other BBU-RRU split options, high layer(IP or above) protocols are not supported at ONU side.

- The most possible option is to deploy MEC servers at OLT (BBU) side

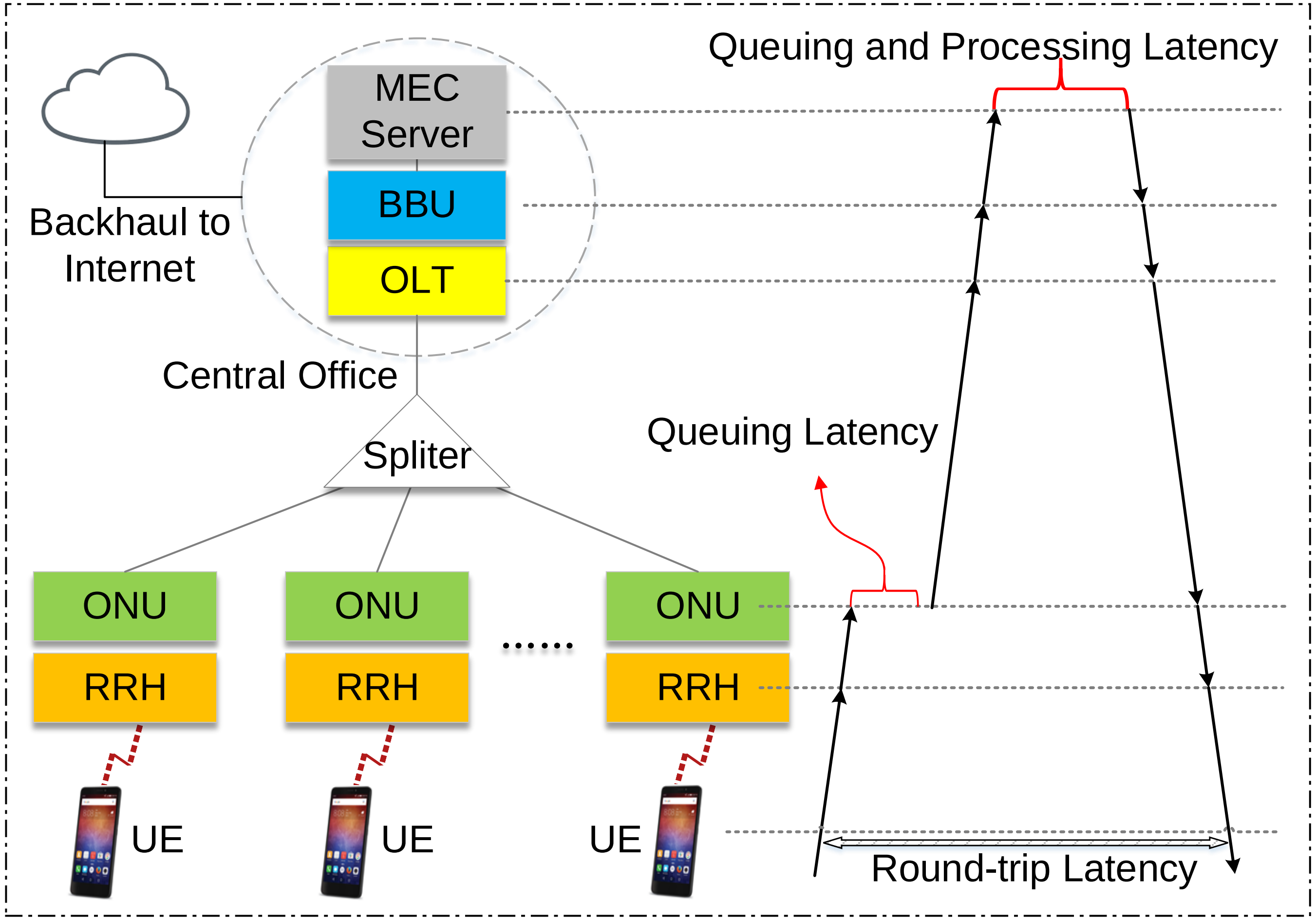


B. P. Rimal, D. P. Van and M. Maier, "Mobile Edge Computing Empowered Fiber-Wireless Access Networks in the 5G Era," in IEEE Communications Magazine, vol. 55, no. 2, pp. 192-200, February 2017.

C-RAN architecture with PON and MEC

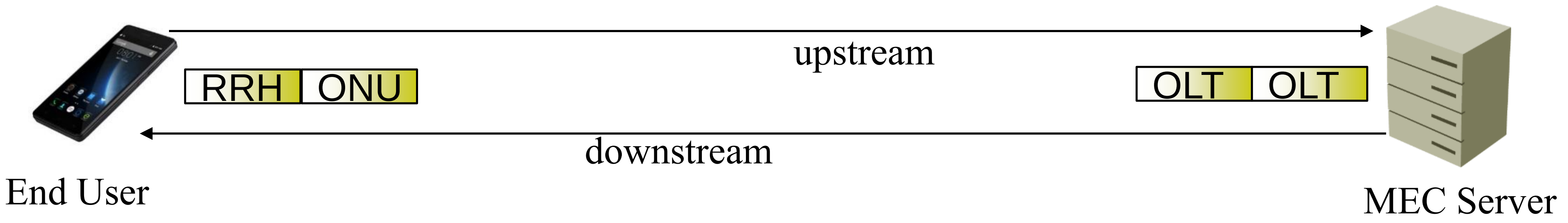
Architecture of C-RAN with PON and MEC:
1) MEC servers in central office.
2) PON as MFH

Latency components in C-RAN with PON and MEC:
1) Up and down stream propagation latency
2) DBA latency
3) Queuing time at ONUs
4) Queuing and processing time at MEC servers



E2E Latency analysis

- Propagation latency, out of control (removed from modeling)
- DBA latency, real-time (on demand) bandwidth calculation latency in DBA can be eliminated by either advance bandwidth assignment or burst scheduling. (removed from modeling)
- Queuing latency at ONU, determined by allocated priority and bandwidth to ONU for upstream remission. (representative of MFH latency)
- Queuing and processing latency at MEC, determined by allocated priority and the amount of processing resource at MEC server (representative of MEC latency)

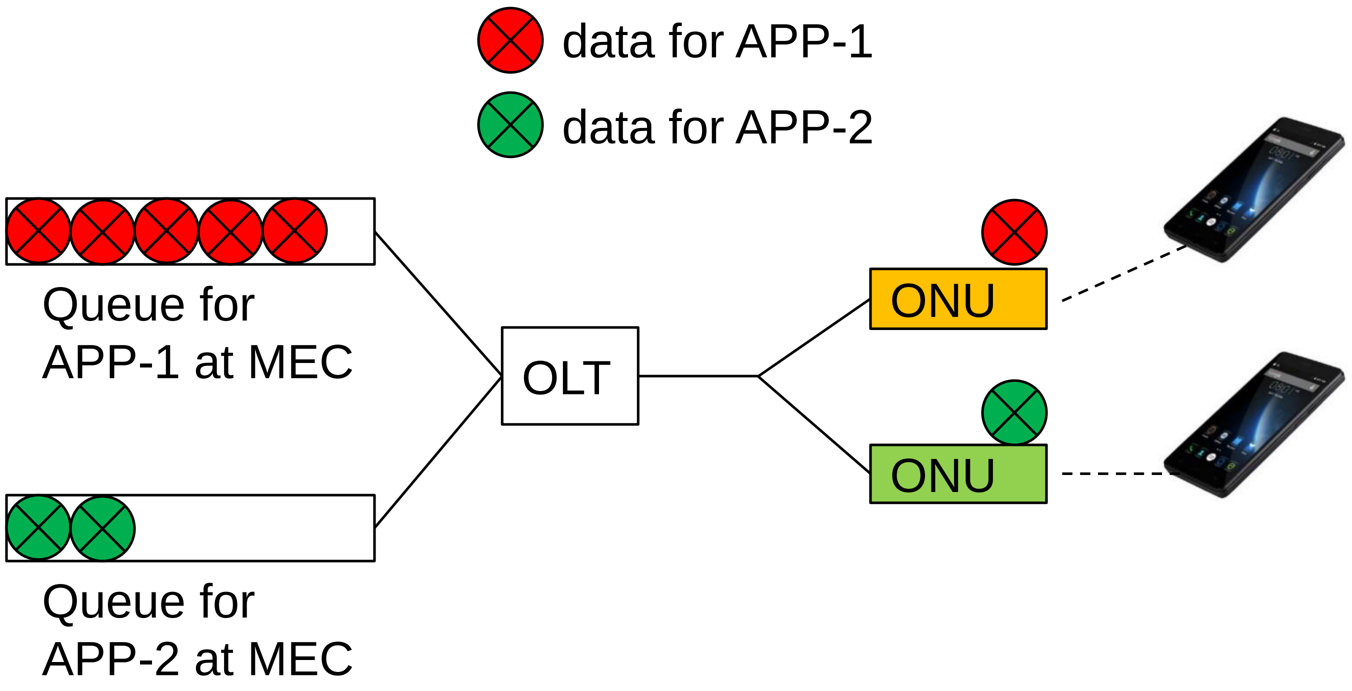


Coordinating MEC with PON

- Motivation for coordination

- Which data go first?

- Case 1: data for APP-1 go first, and data for APP-2 experiences queuing latency in ONU, but data for APP-1 still need to wait at MEC. (invalid optimization)

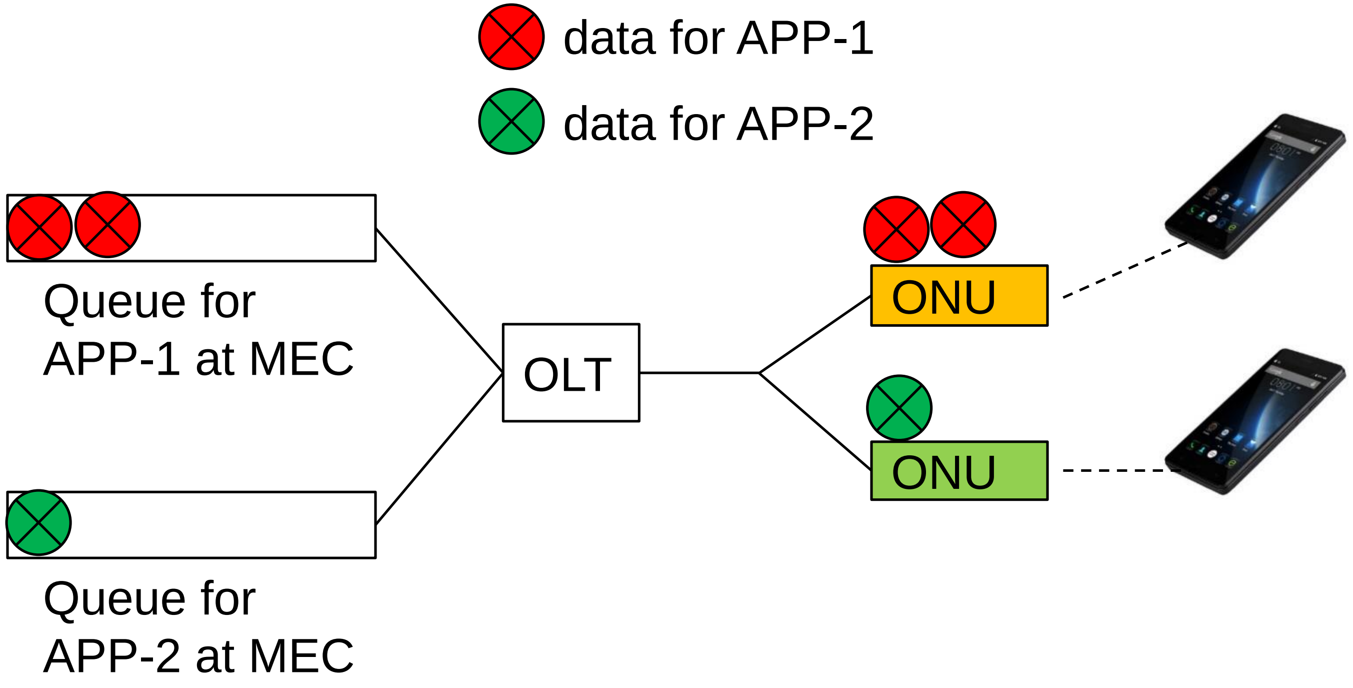


Coordinating MEC with PON

- Motivation for coordination

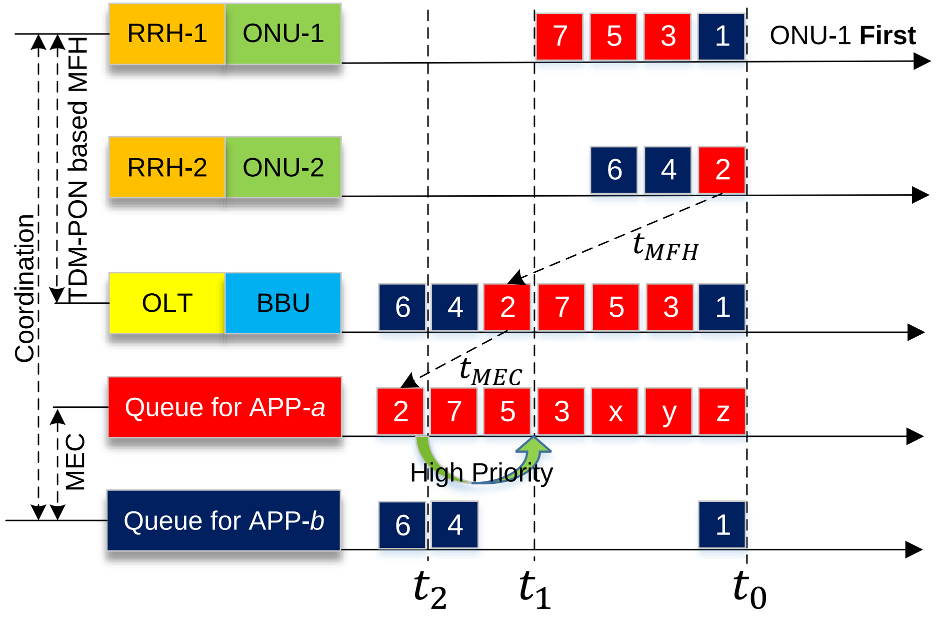
- Process which data with higher priority?

- Case 2: data for APP-2 goes first, and data for APP-1 experiences longer latency at ONU. Processing data for APP-2 with higher priority? (unnecessary optimization)



Use cases for coordinating MEC with PON

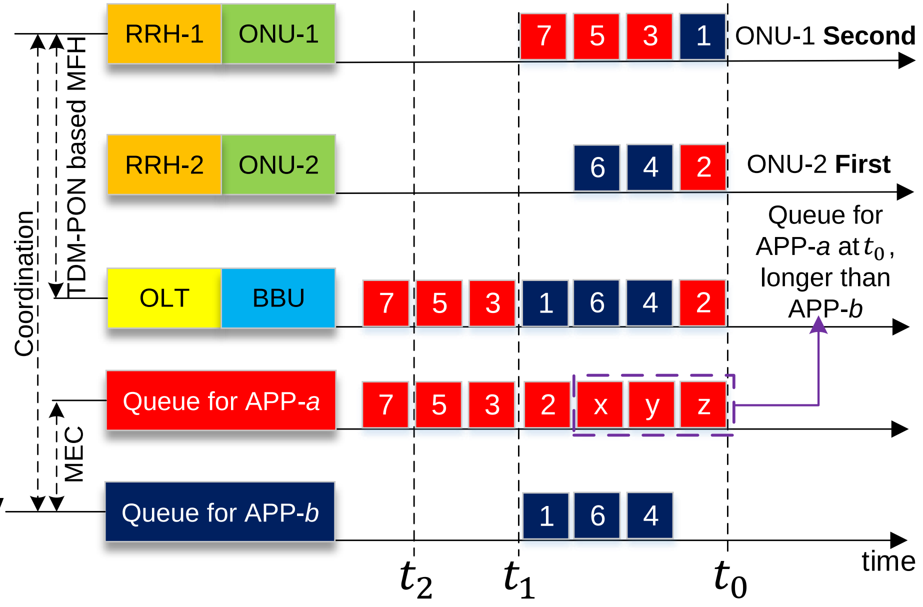
- MFH Latency driven MEC Scheduling
- t_p^{MFH} might be higher or lower than an expected MFH latency according to the specific condition it experienced, but is an important parameter for MEC scheduling, as it decides the tolerated time left for MEC processing, which is $T_a - t_p^{MFH}$.
- If t_p^{MFH} is higher than expected, we can compensate it by processing packet p with higher priority and more CPU time slices.
- Else if t_p^{MFH} is lower than expected, we can take its advantage and process other urgent tasks with higher priority and more CPU time slices, as long as t_p^{MEC} is less than $T_a - t_p^{MFH}$



Use cases for coordinating MEC with PON

- MEC Scheduling aware MFH Bandwidth Allocation

- Queuing status of current pending tasks at a VM (inside MEC servers) affects t_p^{MEC} of each upcoming packet p for the same application, and it in return can be an important indicator for future MFH bandwidth allocation.
- When the queue for an application a is longer (in terms of waiting time) than that of another application b, it is meaningless to transmit the packets for application a through MFH with higher priority and bandwidth, as they need to wait anyway, at either ONUs or MEC servers.
- On the other hand, packets for the application with a shorter queue can be transmitted first in this case.



Use cases for coordinating MEC with PON

- Difficulties for MEC Scheduling aware MFH Bandwidth Allocation
- 1) In practice, it is not easy to know which ONU has upcoming packets for which application, as ONU is not aware of applications.
- Solution for 1): Fortunately, there are existing works on application-aware PON system, and they may be helpful for addressing this problem. In addition, thus information can be predicted at MEC side using data analyzing (e.g., Machine Leaning).
- 2) Another difficulty for this use case is that we cannot manage packets per application in TDM-PON based MFH.
- Solution for 2): Nevertheless, this case can be compromised as “the ONU, whose buffer has more packets for the application with a shorter queue, can be transmitted first” at least.

Summery

- MEC servers can be placed at BBU side in C-RAN
- MEC can help enhance C-RAN latency by both processing user traffic directly and assisting MFH
- Located near each other, MEC and MFH can coordinate in two directions for E2E latency optimization.
  - 1) MFH Latency driven MEC Scheduling
  - 2) MEC scheduling aware MFH Bandwidth Allocation

Thank you!

Wei Wang