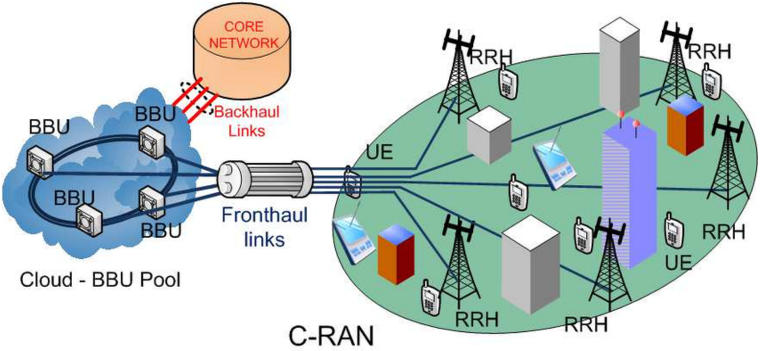
Real Traffic-Aware Scheduling of Computing Resources in Cloud-RAN

Hatem Ibn-Kheder[1,3], Sahar Hoteit[1], Patrick Brown, Véronique Vèque[1], Ruby Krishnaswamy[2], William Diego[2] and Makhlouf Hadji[3]

[1]Laboratoire des Signaux et Systèmes, Université Paris Saclay-CNRS-CentraleSupélec, France
[2]Orange Labs, Chatillon, France
[3]Institut de Recherche Technologique SystemX, France

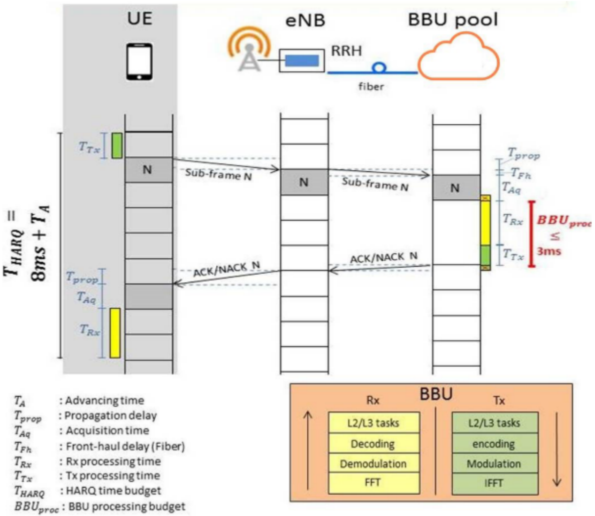
December 15, 2020

Cloud-RAN Architecture



Source: V. N. Ha, L. B. Le and N. ào, "Coordinated Multipoint Transmission Design for Cloud-RANs With Limited Fronthaul Capacity Constraints," in IEEE Transactions on Vehicular Technology, vol. 65, no. 9, pp. 7432-7447, Sept. 2016.

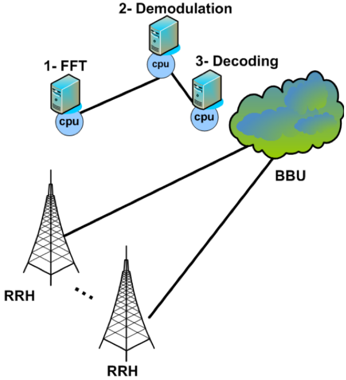
HARQ process in Cloud-RAN architectures



Source: V. Q. Rodriguez and F. Guillemin, "Towards the deployment of a fully centralized Cloud-RAN architecture," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, 2017, pp. 1055-1060.

BBU modeling

- BBUs: a set of CPUs executing virtual BBU sub-functions
- Main sub-functions:
  - IFFT/FFT
  - Encoding/Decoding
  - Modulation/Demodulation
- BBUs allocate resources to remote radio heads (RRH)



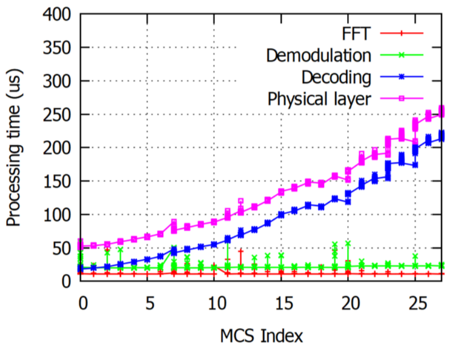
Our Contributions

Scheduling Radio Remote Heads (RRHs) subframes on the **computing resources** (i.e., a set of CPU cores in the shared BBU pool).

- Propose two optimal scheduling algorithms to place RRHs' subframes on the CPU cores of the shared BBU pool.
- Propose three heuristics that give close results to those obtained from optimization problems solutions.

Processing time vs MCS index

- Using OpenAirInterface (OAI) simulator, **[Kheder et al.]** evalute the average processing time of the BBU sub-functions as a function of the Modulation Coding Scheme (MCS) index.



Source: H. Kheder, S. Hoteit, P. Brown, R. Krishnaswamy, W. Diego and V. Vèque, "Processing Time Evaluation and Prediction in Cloud-RAN", In Proc. of ICC, Shanghai, May 2019.

Optimal Scheduling Algorithms Formulation

Optimal Scheduling Algorithm for maximizing the number of subframes (OSA1)

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i \in \mathcal{N}} \sum_{c \in \mathcal{C}} x_i^c \quad && \textbf{Maximize the nb of decoded subframes} \\
\text{subject to} \quad & x_i^c \in \{0,1\},\ \forall i \in \mathcal{N}, c \in \mathcal{C} && \textbf{Boolean decision variable} \\
& \sum_{c \in \mathcal{C}} x_i^c \leq 1,\ \forall i \in \mathcal{N} && \textbf{Single Core Assignment} \\
& \sum_{i \in \mathcal{N}} x_i^c t_i \leq d,\ \forall c \in \mathcal{C} && \textbf{Subframe Deadline}
\end{aligned}
$$

where

$$
x_i^c = \begin{cases} 1, & \text{if the subframe } i \in \mathcal{N} \text{ is assigned} \\ & \text{ to the CPU core } c \in \mathcal{C} \\ 0 & \text{Otherwise} \end{cases}
$$

Optimal Scheduling Algorithm for maximizing the throughput (OSA2)

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i \in \mathcal{N}} \sum_{c \in \mathcal{C}} x_i^c b_i && \textbf{Maximize the throughput} \\
\text{subject to} \quad & x_i^c \in \{0, 1\},\ \forall i \in \mathcal{N}, c \in \mathcal{C} && \textbf{Boolean decision variable} \\
& \sum_{c \in \mathcal{C}} x_i^c \leq 1,\ \forall i \in \mathcal{N} && \textbf{Single Core Assignment} \\
& \sum_{i \in \mathcal{N}} x_i^c t_i \leq d,\ \forall c \in \mathcal{C} && \textbf{Subframe Deadline}
\end{aligned}
$$

where

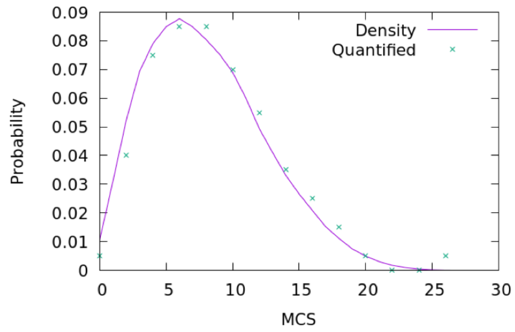
b_i is the ratio between the subframe's number of bytes and the transmission time window

Heuristics

- **Round Robin (RR)**: Scheduling the incoming subframes from the RRHs to the CPU cores of the BBU pool in a round robin fashion.
- **Shortest Time First (STF)**: Sorting the subframes in an increasing order according to their processing time requirement; then applying round robin.
- **Highest Throughput First (HTF)**: Sorting the subframes according to their throughput in a decreasing order; then scheduling using round robin.

Real Traffic Model

We derive the probability density function of a subframe to carry traffic of a given MCS index, based on real measurements collected in **[Trinh et al.]**.

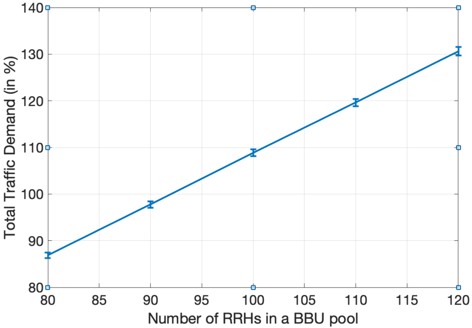


Source: H. D. Trinh, N. Bui, J. Widmer, L. Giupponi, and P. Dini, "Analysis and modeling of mobile traffic using real traces" in IEEE PIMRC, 2017.

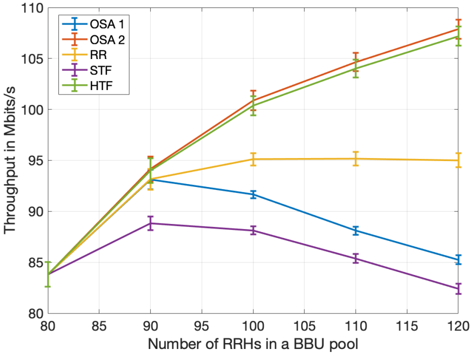
Performance Evaluation

Simulation Parameters

Parameters	**Value**
CPU Cores	4
RRHs per BBU pool	80 to 120
Subframe's Deadline	2 ms
TTI	1 ms



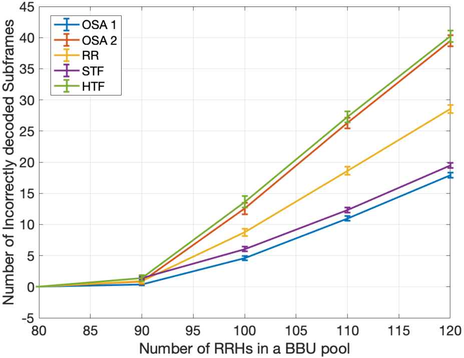
Offered throughput



Observations

- OSA_2 outperforms all others.
- HTF policy is very efficient; very close to OSA_2.
- STF shows the worst performance.

Number of incorrectly decoded subframes



Observations

- OSA_1 has the best performance
- STF is the one that is closer to OSA_1.
- HTF shows the worst performance.

Conclusion

- We propose real-traffic based **scheduling algorithms for computing resources** in Cloud-RAN.
- We compare them in different network scenarios, different load densities and performance metrics.
- We **bring recommendations to mobile network operators** on the best scheduling algorithm that should be adopted to increase network performance according to their needs.

Thanks for your attention !!!