Scheduling Algorithms for 5G Networks with Mid-haul Capacity Constraints

Abhishek Sinha*

Matthew Andrews†, Prasanth Ananth†

*IIT Madras, †Nokia Bell Labs, Murray Hill, NJ

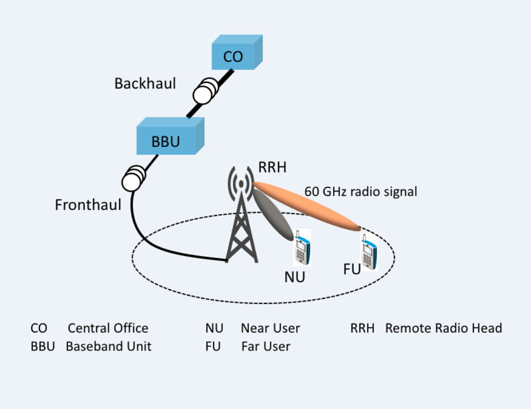
WiOpt 2019, Avignon, France

May 24, 2019

NOKIA Bell Labs

Introduction

We consider a scheduling problem for efficiently integrating a 5G backhaul with the front-haul (RUs).

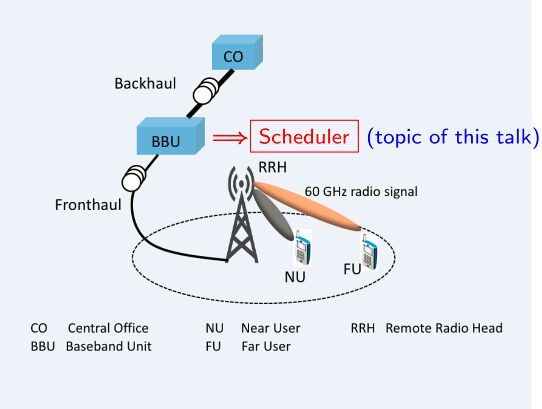


A vRAN architecture

Figure courtesy: IEEE future networks

Introduction

We consider a scheduling problem for efficiently integrating a 5G backhaul with the front-haul (RUs).



A vRAN architecture

Figure courtesy: IEEE future networks

Technical Background

- vRAN Architecture: UE scheduling and part of baseband processing are done at Central Units (CU) located at the edge cloud
- This *split-processing* architecture reduces computational overhead on the remote units (RRH)

Technical Background

- vRAN Architecture: UE scheduling and part of baseband processing are done at Central Units (CU) located at the edge cloud
- This *split-processing* architecture reduces computational overhead on the remote units (RRH)
- The scheduled data is transported
  1. First, from the CU to RUs via a Passive Optical Network PON
  2. Then, the data is *immediately* transmitted over the air at the *same slot*
- In particular, no queueing takes places at RUs, which improves the latency.

The Scheduling Problem with PON capacity constraint

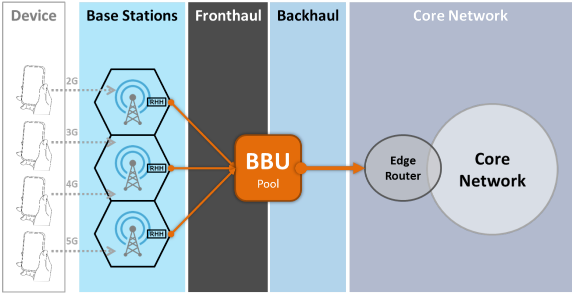


Figure courtesy: Medium Technology

The Scheduling Problem with PON capacity constraint

Time-Varying
UE channels

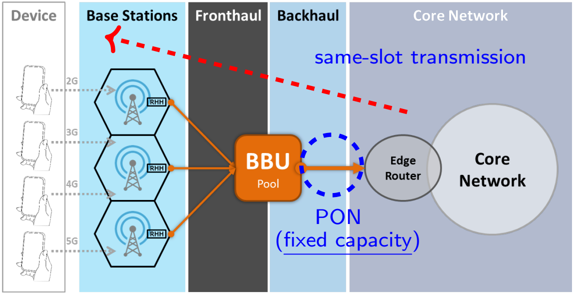
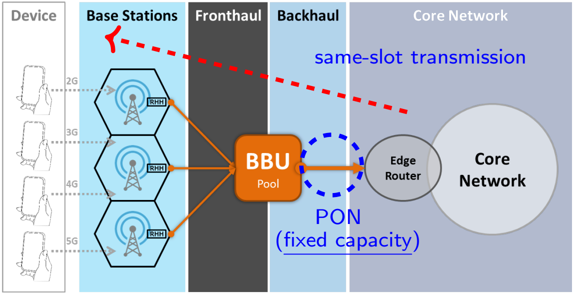


Figure courtesy: Medium Technology

The Scheduling Problem with PON capacity constraint

Time-Varying UE channels



Problem: How to efficiently schedule data to the UEs with time-varying wireless channels through a fixed-capacity PON?

Figure courtesy: Medium Technology

Our results

Limitation of the State-of-the-art

- We show that the well-known Proportional Fair scheduler is **not optimal** in this architecture

Our results

Limitation of the State-of-the-art

- We show that the well-known Proportional Fair scheduler is **not optimal** in this architecture

Our contributions - `Single Cell`

- Polynomial-time LP-based algorithm with a guaranteed 2-approximation
- Pseudo polynomial-time Optimal scheduling using Dynamic Programming

Our contributions- `Multi Cell`

- A Matroid-based greedy 2-approximation algorithm

Our results

Limitation of the State-of-the-art

- We show that the well-known Proportional Fair scheduler is **not optimal** in this architecture

Our contributions - `Single Cell`

- Polynomial-time LP-based algorithm with a guaranteed 2-approximation
- Pseudo polynomial-time Optimal scheduling using Dynamic Programming

Our contributions- `Multi Cell`

- A Matroid-based greedy 2-approximation algorithm

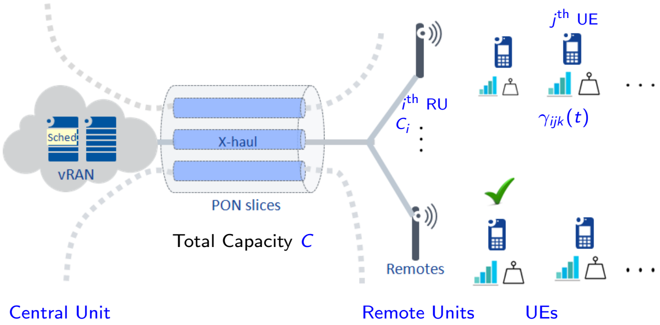
Main Challenge

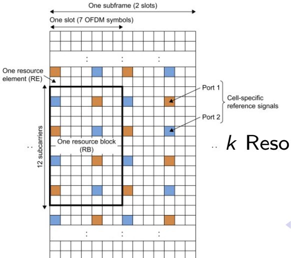
- **Scalable solution** to a hard combinatorial packing problem.

Disruption

- Simulation shows that the proposed algorithm achieves $> 2X$ gain over the PF scheduler.

System Model





k Resource Blocks (RB) per RU

Service Constraints and Long-term objective

1. Each RU can transmit over k RBs
2. A RU can allocate a RB to at most one UE per slot.
3. The channel rates differ across the RUs and RBs. The maximum air-interface rate for the j^{th} UE of the i^{th} UE for the k^{th} RB at time t is $\gamma_{ijk}(t)$.
4. The aggregate service rate allocated to all users at a given time-slot is limited by the PON capacity C.

Service Constraints and Long-term objective

1. Each RU can transmit over k RBs
2. A RU can allocate a RB to at most one UE per slot.
3. The channel rates differ across the RUs and RBs. The maximum air-interface rate for the j^{th} UE of the i^{th} UE for the k^{th} RB at time t is $\gamma_{ijk}(t)$.
4. The aggregate service rate allocated to all users at a given time-slot is limited by the PON capacity C.

Long-Term objective: Design a scheduling policy to maximize sum-log utility of the users:

$$\max \sum_{ij} \log(\bar{r}_{ij})$$

where, $\bar{r}_{ij} = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_{ij}(t)$ are the long-term rates.

Slot-by-Slot Optimization

- Using the gradient-based scheduling algorithm by *Stolyar (2005)*, the long-term objective reduces to the following slot-by-slot optimization problem.

Slot-by-Slot Optimization

- Using the gradient-based scheduling algorithm by *Stolyar (2005)*, the long-term objective reduces to the following slot-by-slot optimization problem.

Decision Variables:

- Let the binary variable $x_{ijk}(t) \in \{0,1\}$ denote whether the k^{th} RB is allocated to the j^{th} UE of the i^{th} RU.
- Let the non-negative real variable $y_{ijk}(t)$ denote the corresponding allocated rate.

The exponentially-weighted average rate $R_{ij}(t)$ is computed for the UE (i,j) as follows

$$R_{ij}(t+1) = (1-\beta)R_{ij}(t) + \beta \underbrace{\sum_k y_{ijk}(t)}_{current\ \ rate},$$

for some fixed small parameter $\beta > 0$.

Mixed Integer Linear Program (MILP) formulation

Problem: Single Shot

$$\max_{\mathbf{x}(t),\mathbf{y}(t)} \sum_{i,j} \frac{\sum_k y_{ijk}(t)}{R_{ij}(t)}.$$

Subject to,

$$\sum_j x_{ijk}(t) \quad \leq \quad 1 \qquad \text{(at most one UE per RB)}$$

Mixed Integer Linear Program (MILP) formulation

Problem: Single Shot

$$\max_{\mathbf{x}(t),\mathbf{y}(t)} \sum_{i,j} \frac{\sum_k y_{ijk}(t)}{R_{ij}(t)}.$$

Subject to,

$$\sum_j x_{ijk}(t) \leq 1 \qquad \text{(at most one UE per RB)}$$

$$y_{ijk}(t) \leq \gamma_{ijk}(t) x_{ijk}(t) \qquad \text{(instantaneous air-interface rate constraint per RB)}$$

Mixed Integer Linear Program (MILP) formulation

Problem: Single Shot

$$\max_{\mathbf{x}(t),\mathbf{y}(t)} \sum_{i,j} \frac{\sum_k y_{ijk}(t)}{R_{ij}(t)}.$$

Subject to,

$$\sum_j x_{ijk}(t) \leq 1 \quad \text{(at most one UE per RB)}$$

$$y_{ijk}(t) \leq \gamma_{ijk}(t)x_{ijk}(t) \quad \text{(instantaneous air-interface rate constraint per RB)}$$

$$\sum_{i,j,k} y_{ijk}(t) \leq C \quad \text{(PON capacity constraint)}$$

$$\sum_{j,k} y_{ijk}(t) \leq C_i, \quad \text{(RU-specific capacity constraints (for multi cell))}$$

$$\underbrace{x_{ijk}(t)}_{binary} \in \{0,1\}, \quad \underbrace{y_{ijk}(t)}_{continuous} \geq 0.$$

Structural Results - `Single Cell`

For the `Single Cell` problem, there is no RU-specific capacity constraint (*i.e.*, $C_i = \infty$, $\forall i$) and the problem is equivalent to a single RU. Hence, we drop the index i in this section.

Structural Results - Single Cell

For the Single Cell problem, there is no RU-specific capacity constraint (*i.e.*, $C_i = \infty, \forall i$) and the problem is equivalent to a single RU. Hence, we drop the index i in this section.

Definition (Almost Discrete (AD) Allocation)

A feasible rate-allocation vector $(\boldsymbol{x}(t), \boldsymbol{y}(t))$ is called Almost Discrete if $y_{jk}(t) = \gamma_{jk}(t)x_{jk}(t)$ for all but (at most) one RB.

Structural Results - Single Cell

For the Single Cell problem, there is no RU-specific capacity constraint (*i.e.*, $C_i = \infty$, $\forall i$) and the problem is equivalent to a single RU. Hence, we drop the index i in this section.

Definition (Almost Discrete (AD) Allocation)

A feasible rate-allocation vector $(\boldsymbol{x}(t), \boldsymbol{y}(t))$ is called Almost Discrete if $y_{jk}(t) = \gamma_{jk}(t)x_{jk}(t)$ for all but (at most) one RB.

Theorem (Optimality of AD)

There exists an optimal solution to Single Shot *which is* Almost Discrete.

- We present two different proofs of this theorem in the paper.
- The first one is constructive and algorithmic
- The second one utilizes combinatorial properties of a resulting LP.

MILP to LP Relaxation

- The previous theorem proves that the constraint $y_{jk}(t) \leq \gamma_{jk}(t)x_{jk}(t)$ is tight in *almost all* RBs.
- Hence, it is natural to consider the following LP relaxation by $y_{jk}(t) \leftarrow \gamma_{jk}x_{jk}(t)$:

Problem: RLP

$$\max_{\boldsymbol{x}(t)} \sum_{jk} x_{jk}(t)\frac{\gamma_{jk}(t)}{R_j(t)}$$

Subject to,

$$\begin{aligned} \sum_{j} x_{jk}(t) &\leq 1, \quad \forall k. \\ \sum_{jk} \gamma_{jk}x_{jk} &\leq C, \\ \boldsymbol{x} &\geq \boldsymbol{0}. \end{aligned}$$

MILP to LP Relaxation

- The previous theorem proves that the constraint $y_{jk}(t) \leq \gamma_{jk}(t)x_{jk}(t)$ is tight in *almost all* RBs.
- Hence, it is natural to consider the following LP relaxation by $y_{jk}(t) \leftarrow \gamma_{jk}x_{jk}(t)$:

Problem: `RLP`

$$\max_{\boldsymbol{x}(t)} \sum_{jk} x_{jk}(t)\frac{\gamma_{jk}(t)}{R_j(t)}$$

Subject to,

$$\begin{aligned} \sum_{j} x_{jk}(t) &\leq 1, \quad \forall k. \\ \sum_{jk} \gamma_{jk}x_{jk} &\leq C, \\ \boldsymbol{x} &\geq \boldsymbol{0}. \end{aligned}$$

☞ Clearly, the solution to `RLP` will be a good approximation to `Single Shot` if `RLP` also has the AD property (*i.e.*, mostly 0-1 solutions).

Solution Structure of RLP

Theorem (RLP has the AD property)

An optimal solution to RLP allocates every RB to at most one UE, excepting, at most one RB, which is shared between two UEs.

The proof of this theorem crucially utilizes the properties of the Basic Feasible Solutions.

Solution Structure of RLP

Theorem (RLP has the AD property)

An optimal solution to RLP allocates every RB to at most one UE, excepting, at most one RB, which is shared between two UEs.

The proof of this theorem crucially utilizes the properties of the Basic Feasible Solutions.

The above theorem suggests the following policy which we prove to be 2-optimal.

Algorithm 2 LP-based 2-Approximation Algorithm for SINGLE SHOT

1: Find the maximum possible objective value obtainable by using a *single* RB, i.e.,

$$F_{\max} = \max_{j,k} \frac{1}{R_j} \min\{\gamma_{jk}, C\}.$$

2: Solve the Linear Program RLP. Let I be the objective value obtained by the standalone RBs (*i.e.*, for which $x_{jk} = 1$ for some j) in its optimal solution.

3: Choose the solution corresponding to the maximum of I and $F_{\max}$.

☞ In the paper, we also design a pseudo-polynomial time Optimal algorithm for SINGLE SHOT using DP.

Structural results - Multi-Cell

In the Multi-Cell case, the RU-specific capacity constraints C_i are active.

Let $\mathcal{I}$ be the set of all feasible RB assignments, E be the ground set.

Structural results - Multi-Cell

In the Multi-Cell case, the RU-specific capacity constraints C_i are active.

Let $\mathcal{I}$ be the set of all feasible RB assignments, E be the ground set.

Lemma

The system $(E, \mathcal{I})$ *is a* PARTITION MATROID.

Structural results - Multi-Cell

In the Multi-Cell case, the RU-specific capacity constraints C_i are active.

Let $\mathcal{I}$ be the set of all feasible RB assignments, E be the ground set.

Lemma

The system $(E, \mathcal{I})$ *is a* PARTITION MATROID.

Let $f : \mathcal{I} \to \mathbb{R}_+$ be the optimal objective function for a given RB assignment. Note that $f(\cdot)$ can be evaluated efficiently by solving an LP.

Lemma

The set function $f(\cdot)$ *is submodular.*

By the well-known *Fisher-Nemhauser-Wolsey* (1978) paper, the above two properties readily shows that a greedy algorithm is within a factor of 2 of the optimal.

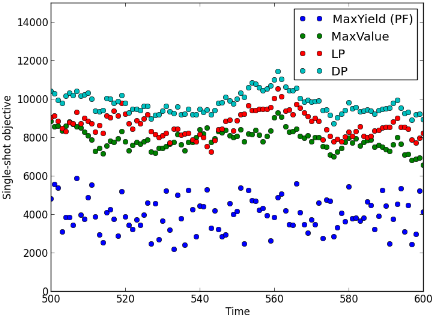
2-approximation Algorithm for SINGLE SHOT

Algorithm 3 Greedy Algorithm for SINGLE SHOT (`Multi-Cell`)

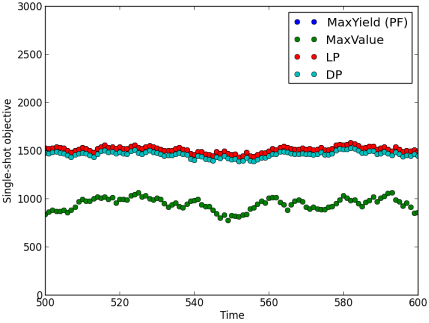
```
S ← φ
while 1 do
    Find a feasible augmentation S̄ ∈ 𝓘 of S that maximizes f(S̄) subject to the
    constraint |S̄ \ S| = 1.
    if f(S̄) = f(S) then
        break
    else
        S ← S̄
    end if
end while
```

Simulation Results-I

SETUP: 1 km^2 area, 1000 users distributed according to PPP, 100 cells, 20 MHz BW.

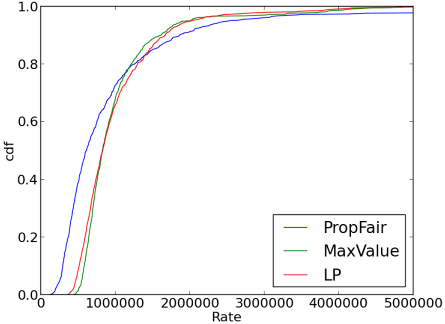


Single-Shot objective $(C = 1 \text{ Gbps})$

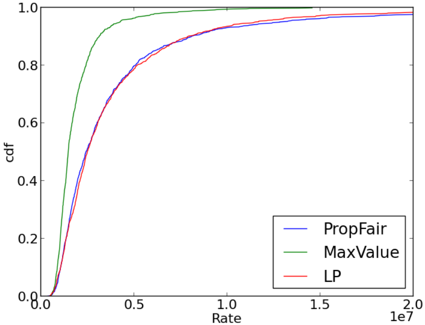


Single-Shot objective $(C = 10^3 \text{ Gbps})$

Simulation Results-II



Long-term UE rate distribution ($C = 1$ Gbps)



Long-term UE rate distribution ($C = 10^3$ Gbps)

Conclusion and Future Works

- We considered the problem of downlink vRAN scheduling with mid-haul constraints.
- We have proposed an LP-based (2-approx.), a DP based (pseudo-poly, optimal) algorithms for `single cell`
- We have also proposed a matroid-based 2-approx. algorithm for `multi-cell`
- Our model assumed that there is no inter-cell interference (due to CoMP). We will be extending our methodologies when this assumption does not hold.
- In future, we are looking forward in implementing these algorithms in our 5G-test bed