Deterministic architectures for low latency in 5G and beyond systems

Maël Guiraud

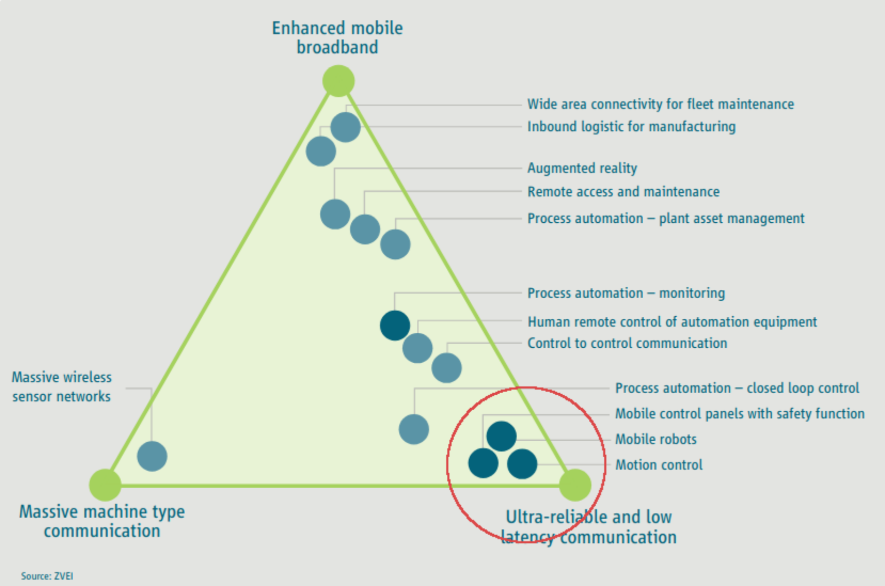
Nokia Bell Labs France - DAVID, Université de Versailles Saint Quentin

October 14, 2020

NOKIA Bell Labs

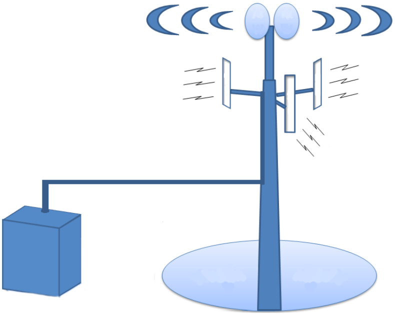
david
données et algorithmes
pour une ville intelligente et durable

Context



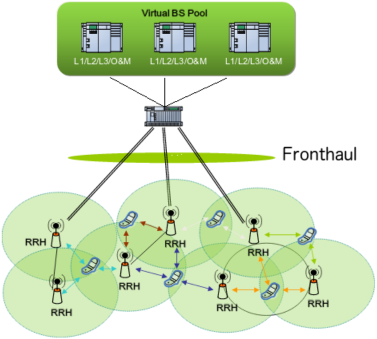
Enhanced mobile broadband
Wide area connectivity for fleet maintenance
Inbound logistic for manufacturing
Augmented reality
Remote access and maintenance
Process automation – plant asset management
Process automation – monitoring
Human remote control of automation equipment
Control to control communication
Massive wireless sensor networks
Process automation – closed loop control
Mobile control panels with safety function
Mobile robots
Motion control
Massive machine type communication
Ultra-reliable and low latency communication
Source: ZVEI

Use case: Cloud-RAN

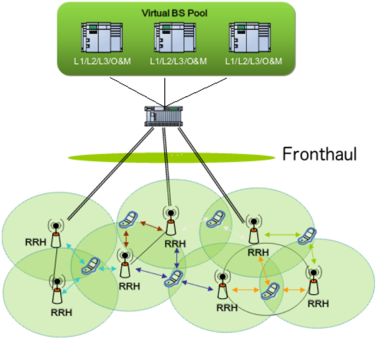


RU=RRH, Distributed/Centralized Unit=BBU

Use case: Cloud-RAN



Use case: Cloud-RAN



C-RAN aims to mutualize the computation ressources.

Since the processing time is constrained by protocol, the lower the transmission delay, the more important the sharing (ie. the coverage area of a data center).

The minimum latency allows to increase the mutualization.

Problematic

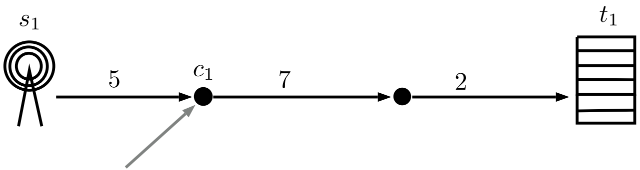
Specificity of Cloud-RAN:

- Minimize latency: increase the cover area.
- Periodic traffic

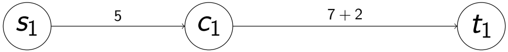
Current approaches:

- Dedicated network $\rightarrow$ Too expensive
- Statistical multiplexing (TSN/Qbv) $\rightarrow$ bounded latency only

Model

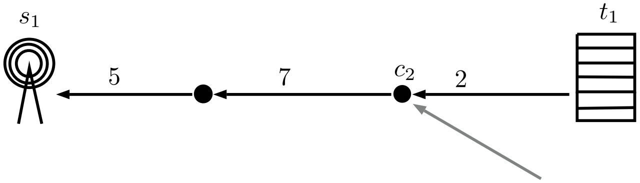


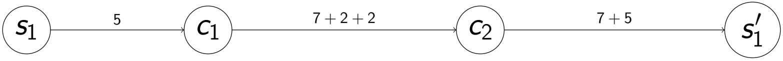
- Network $\rightarrow$ Weighted DAG
- Physical Delay of a link $\rightarrow$ Weight of the arc
- Only the contention points are represented in the graph



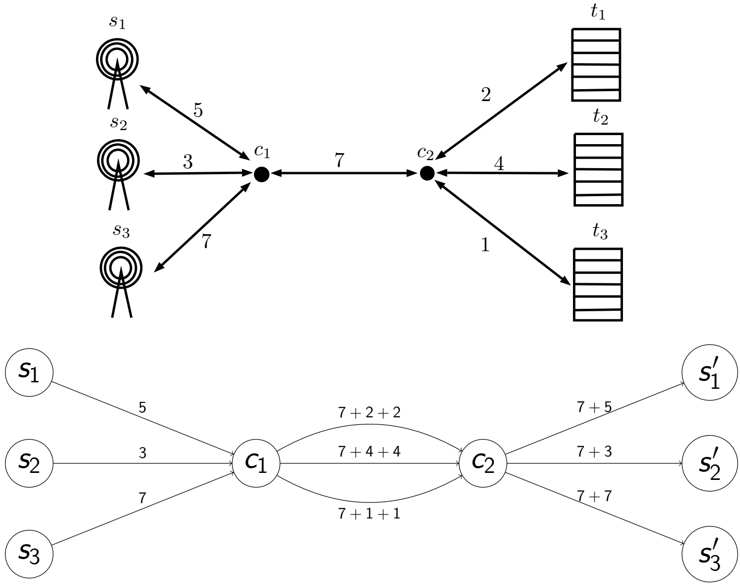
Model

Both ways: from RRH to BBU (forward) then from BBU to RRH (backward)





Model

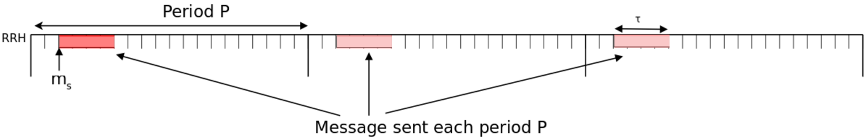

s_1
s_2
s_3
c_1
c_2
t_1
t_2
t_3
5
3
7
7
2
4
1
s_1
s_2
s_3
c_1
c_2
s'_1
s'_2
s'_3
$7+2+2$
$7+4+4$
$7+1+1$
$7+5$
$7+3$
$7+7$

The communication process

The time is discretized.
Two fixed parameters determined by the use case.

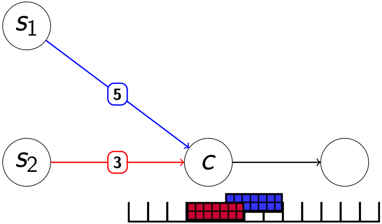
- The period P
- The size of a message τ

Every P units of time, a message of size τ is emitted from each RRH.



The process is periodic : the message is emitted in each period at the same time, called offset.

Collisions

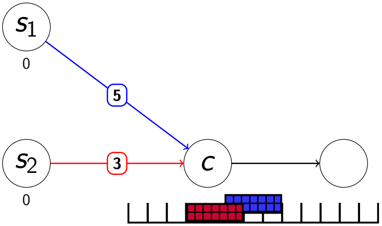


There is a collision between two routes when their messages go through the first vertex of a common arc at the same time.

Periodicity must be taken into consideration

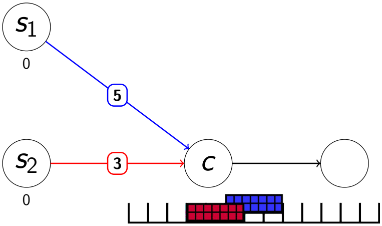
Assignment

$P = 13, \tau = 3$

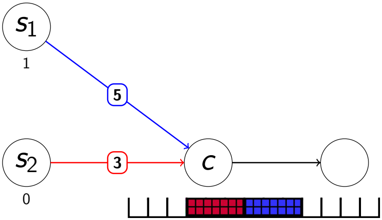


Assignment

$P = 13, \tau = 3$



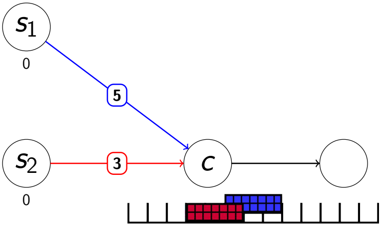
$\rightarrow$



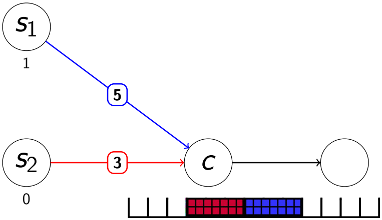
Choosing the offset such that there are no collisions.

Assignment

$P = 13, \tau = 3$



$\rightarrow$

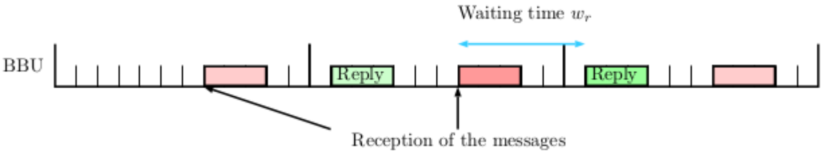


Choosing the offset such that there are no collisions.

An assignment is a choice of offsets for each route without collisions.

Full process

In each BBU, one can choose the waiting time before sending back the answer.



Problem:
Given a routed network, a period and a message size, find an assignment such that there is no collisions.

Two measures to optimize.

PALL:

- The sources may send their message at different dates in the period.
- Local latency constraint.
- Minimizing the process time on the longest route.
- Use cases: URLLC/Indutry 4.0

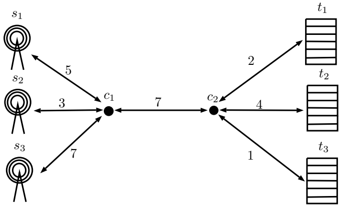
SPALL:

- The sources send their message at the same date: **Synchronized**.
- Global time constraint.
- Minimize the time between the emission of the first message and the reception of the last message.
- use cases: C-RAN

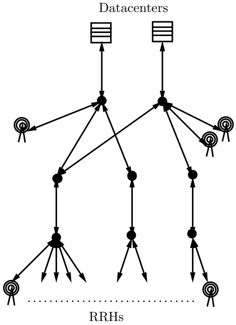
The optimal solutions for PALL and SPALL are not the same.

Different topologies

The **conflict depth** of a route is defined as the number of contention point on the route.
The conflict depth of a routed network is the maximal conflict depth of the routes in the network.



Conflict depth 2.



Conflict depth 3 and more.

NP-Hardness

PALL:

- NP-hard on conflict depth ≥ 3: reduction from k-coloring problem
- NP-hard on conflict depth 2 if the shared arc is not bidirectional.

SPALL:

- NP-hard on conflict depth ≥ 3
- NP-hard on conflict depth 2: reduction from two processors scheduling problem.

Algorithmic solutions

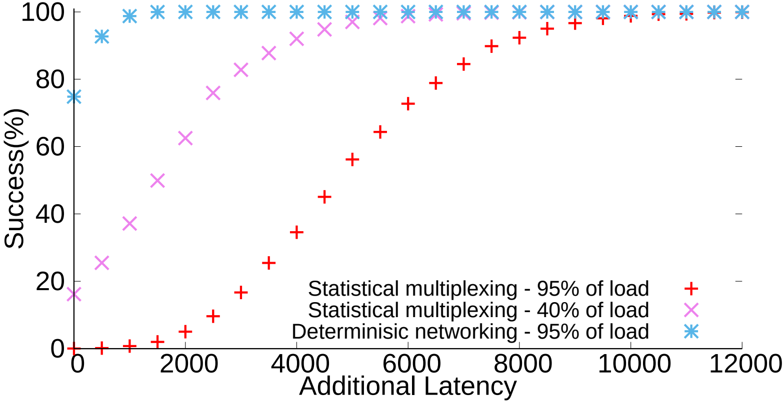
Conflict depth 2:

- Algorithms with theoretical guarantees for moderate load.
- FPT algorithm based on single processor scheduling. FPT: exponential in n, the number of routes but linear in P and τ

Conflict depth ≥ 3:

- Local search algorithms (hill climbing, simulated anealing)
- Optimized Branch and bound: computes realistic instances.

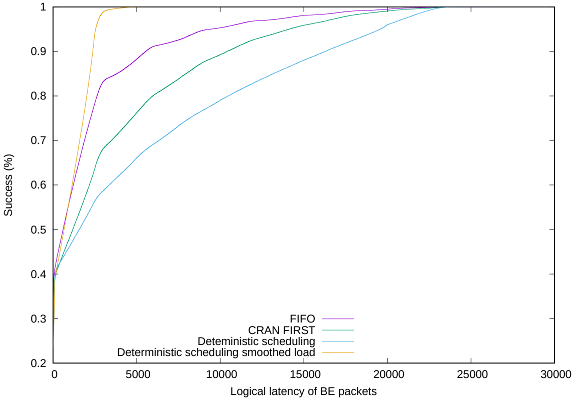
Deterministic vs Statistic



Conflict depth 2 - Deterministic = optimal solution

Period : 20.000 tics

BE Latency optimization



Conflict depth 2 - Deterministic = optimal solution
Period : 33.000 tics - CRAN Load : 60% - BE Load $\simeq$ 20%

Conclusion

- Minimal latency allows higher area coverage: better OPEX
- Deterministic traffic:
  - Zero contention
  - Transparent optical transmission (i.e. without opto-electronic convertion)
  - Energy efficiency
- Better overall traffic on loaded networks

Conclusion

- Minimal latency allows higher area coverage: better OPEX
- Deterministic traffic:
  - Zero contention
  - Transparent optical transmission (i.e. without opto-electronic convertion)
  - Energy efficiency
- Better overall traffic on loaded networks

Future work

- Performance evaluation for conflict depth ≥ 3
- Writing the PhD manuscript

Conclusion

- Minimal latency allows higher area coverage: better OPEX
- Deterministic traffic:
  - Zero contention
  - Transparent optical transmission (i.e. without opto-electronic convertion)
  - Energy efficiency
- Better overall traffic on loaded networks

Future work

- Performance evaluation for conflict depth ≥ 3
- Writing the PhD manuscript

Thank you for your attention.