Prescriptive Analytics for MEC Orchestration

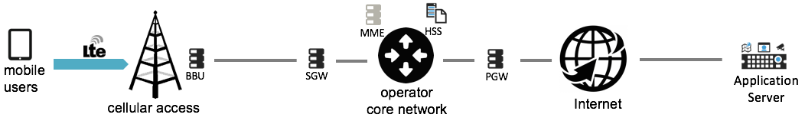
Alberto Ceselli [1]
joint work with Marco Fiore [2], Marco Premoli [1], Stefano Secci [4]

[1]Università Degli Studi di Milano, Dept. of Computer Science, Crema, Italy
[2]CNR–IEIIT, Torino, Italy
[4]Sorbonne Universitè, CNRS, LIP6, F-75005 Paris, France

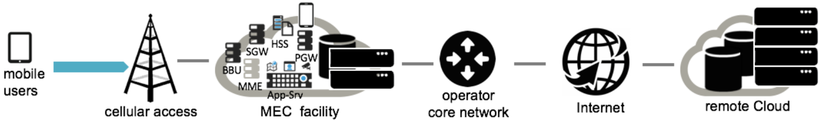
Milan - May 24th 2018

Mobile Access Network Evolution

- Current Trend: softwarization of mobile access networks
- Legacy 4G access network:



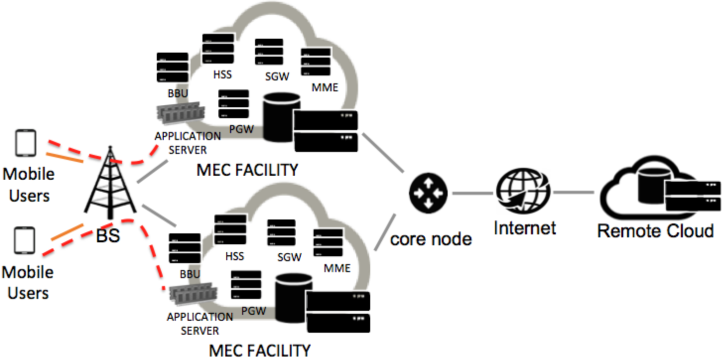
- Virtualizable nodes in MEC facilities (extreme scenario):



Mobile Access Network Evolution

Mgmt. of virtualized nodes running at MEC facilities:

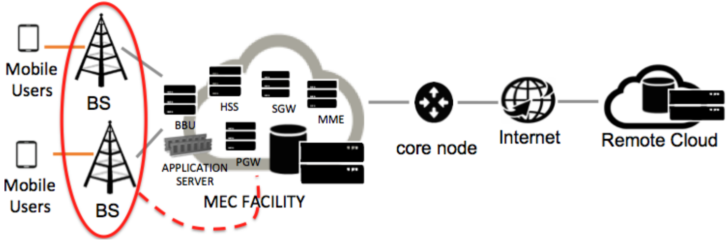
1. user-to-VM associations (application servers)



Mobile Access Network Evolution

Mgmt. of virtualized nodes running at MEC facilities:

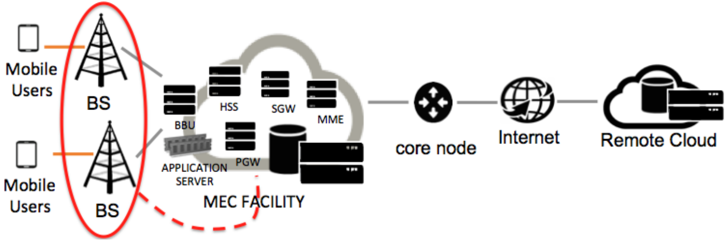
2. BS-to-MEC facility association (network functions)



Mobile Access Network Evolution

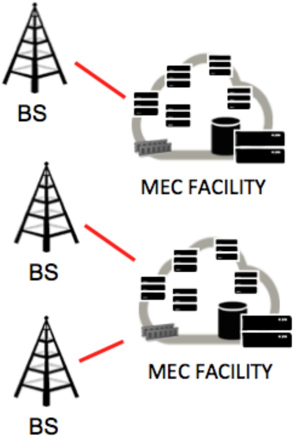
Mgmt. of virtualized nodes running at MEC facilities:

2 BS-to-MEC facility association (network functions)



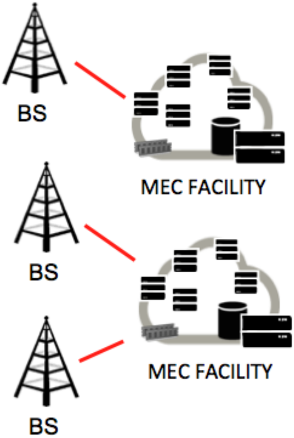
- Operators Current Focus!

MEC Orchestration



Focus of this work on BS-to-MEC facility associations:

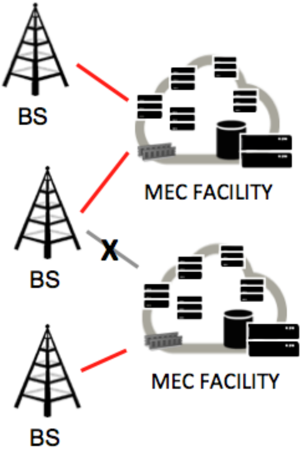
MEC Orchestration



Focus of this work on BS-to-MEC facility associations:

- changing mobile access demand

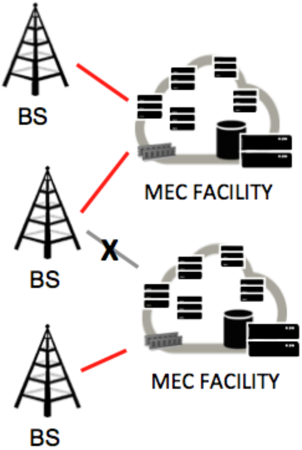
MEC Orchestration



Focus of this work on BS-to-MEC facility associations:

- changing mobile access demand
- orchestrator manages system

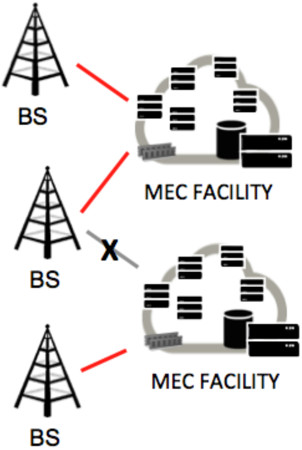
MEC Orchestration



Focus of this work on BS-to-MEC facility associations:

- changing mobile access demand
- orchestrator manages system
- orchestration actions:
  - have a cost: migration or switching cost
  - not continuous: implicit time-discretization

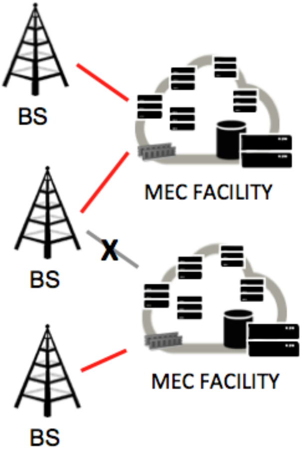
MEC Orchestration



Problem:

- Planning horizon split in time-periods with different traffic demands
- Limited capacity MEC facilities
- Find optimal assignments patterns of BSs to MEC facilities over time
- Minimize costs:
  - MEC facility access latency
  - orchestration costs

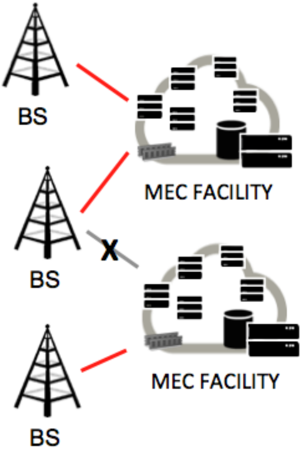
MEC Orchestration



Output:

- Tactical Proactive decisions
- Basis for more complex approaches

MEC Orchestration



Output:

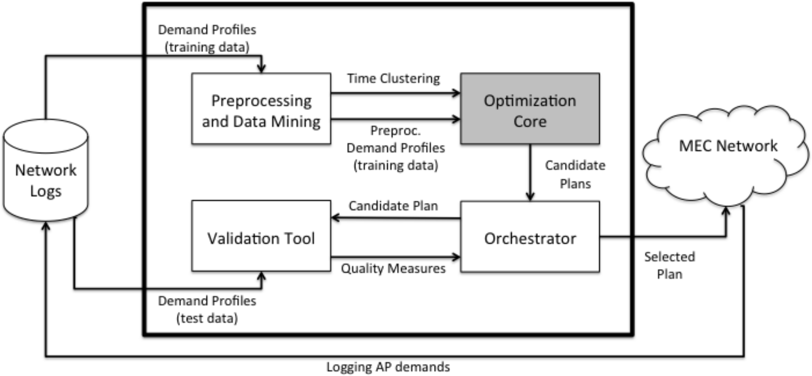
- Tactical Proactive decisions
- Basis for more complex approaches

Issues:

- what demand to use?
- what time-periods to use?

Data-Driven MEC Orchestration Decisions

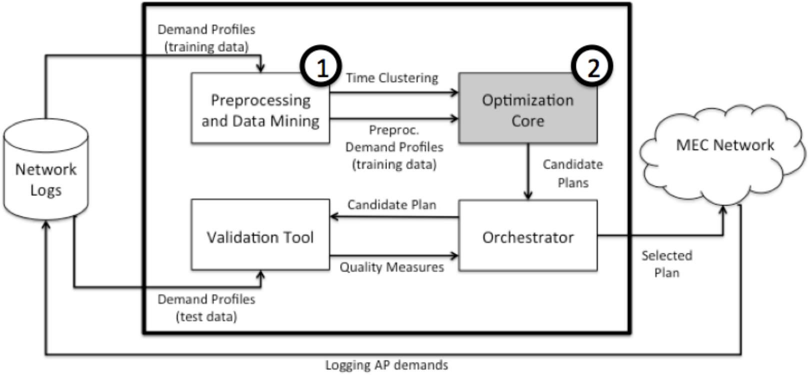
- Integrate the result of data analytics in MEC orchestration decisions



Data-Driven MEC Orchestration Decisions

1 - Time-Periods Aggregation:

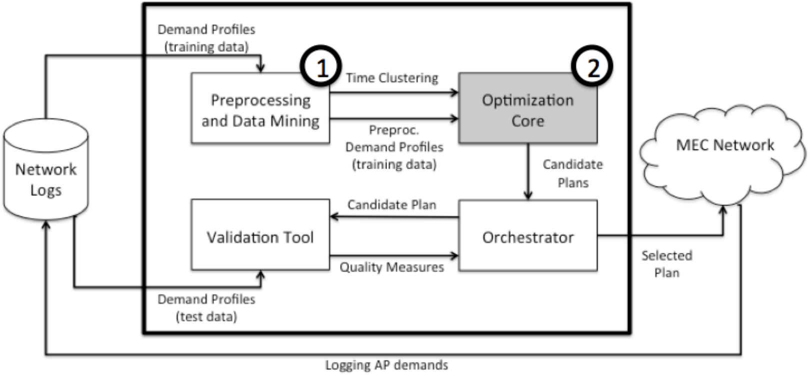
- existing analytics for the spatiotemporal classification of traffic
- extract long-timescale patterns of the mobile traffic demand
- patterns guide the orchestration algorithm



Data-Driven MEC Orchestration Decisions

1 - Time-Periods Aggregation:
  - existing analytics for the spatiotemporal classification of traffic
  - extract long-timescale patterns of the mobile traffic demand
  - patterns guide the orchestration algorithm

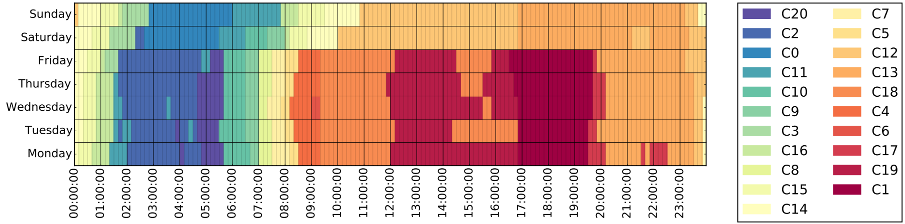
2 - Mathematical Programming based orchestration algorithm.



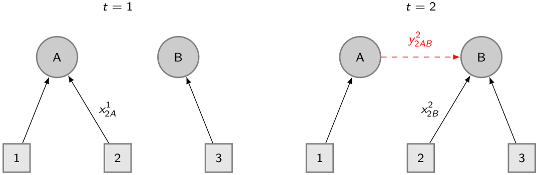
Data Analytics For Time-Periods Aggregation

- agglomerative hierarchical clustering from previous work
- supporting evidence of weekly mobile demand periodicity
- accurate estimation of mobile demand in each aggregated time-period

Example:



Orchestrating Mathematical Programming Model



NP-Hard ILP problem, multi-period extension of GAP:

$$\min\ \alpha \sum_{t \in T} \sum_{i \in A} \sum_{\substack{(j,k) \in \\ K \times K}} d_i^t l_{jk} y_{ijk}^t + \beta \sum_{t \in T} \sum_{i \in A} \sum_{k \in K} d_i^t m_{ik} x_{ik}^t \tag{1}$$

$$\text{s.t.} \sum_{i \in A} d_i^t x_{ik}^t \leq C_k \ , \ \forall t \in T, \forall k \in K \tag{2}$$

$$\sum_{k \in K} x_{ik}^t = 1 \ , \ \forall i \in A, \forall t \in T \tag{3}$$

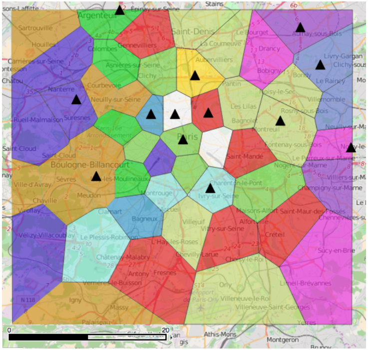
$$x_{ik}^t = \sum_{l \in K} y_{ilk}^t \ , \ \forall i \in A, \forall t \in T \setminus \{1\}, \forall k \in K \tag{4}$$

$$x_{ik}^t = \sum_{l \in K} y_{ikl}^{t+1} \ , \ \forall i \in A, \forall t \in T \setminus \{T\}, \forall k \in K \tag{5}$$

$$\boldsymbol{x} \in \{0,1\}, \boldsymbol{y} \in \{0,1\} \tag{6}$$

Data to Analyze

- collected in the core network of mobile operator Orange in France
- three months in 2016 (12 weeks)
- French metropolitan areas of Lyon (300 BSs) and Paris (1900 BSs)
- specific mobile service, *i.e.*, Facebook
- traffic aggregated in 10-minute time periods, no spatial aggregation

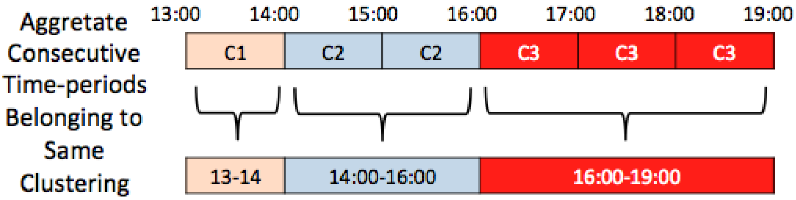


Experiments

- training using average week data on first 4 weeks of dataset

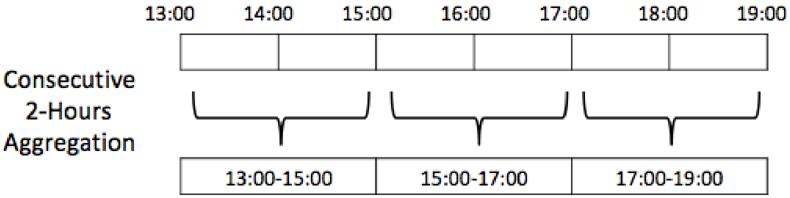
Experiments

- training using average week data on first 4 weeks of dataset
- time-period aggregations comparison:
  1- aggregation resulting from clustering



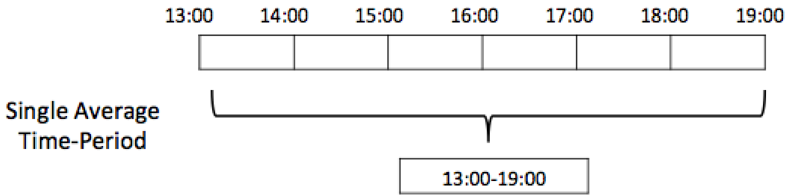
Experiments

- training using average week data on first 4 weeks of dataset
- time-period aggregations comparison:
  1- aggregation resulting from clustering
  2- simple aggregations of consecutive periods



Experiments

- training using average week data on first 4 weeks of dataset
- time-period aggregations comparison:
  1- aggregation resulting from clustering
  2- simple aggregations of consecutive periods
  3- single time-period assignment (benchmark)



Experiments

- training using average week data on first 4 weeks of dataset
- time-period aggregations comparison:
  1- aggregation resulting from clustering
  2- simple aggregations of consecutive periods
  3- single time-period assignment (benchmark)

- Model parameters:
  - different cardinalities of MEC facilities
  - equal weights of access latency costs and orchestration costs

Experiments

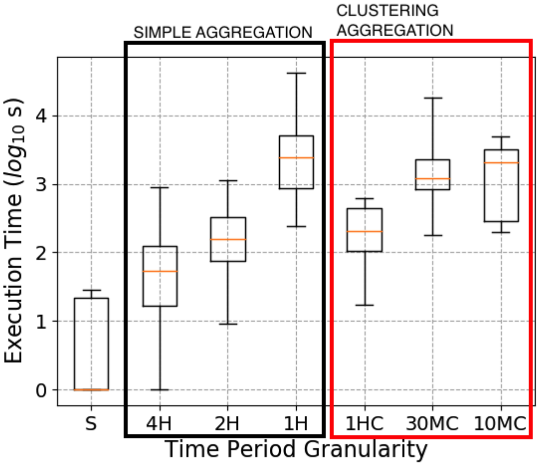
- training using average week data on first 4 weeks of dataset
- time-period aggregations comparison:
  1- aggregation resulting from clustering
  2- simple aggregations of consecutive periods
  3- single time-period assignment (benchmark)

- Model parameters:
  - different cardinalities of MEC facilities
  - equal weights of access latency costs and orchestration costs
- Test with complete 12 weeks dataset:
  - Training Computational Results
  - Cost Component Assessment
  - Capacity Violation

Training Computational Results

Median Execution Time:

- benchmark (S): seconds
- same cardinality of time-periods leads to similar execution time
- few minutes up to 2 hours

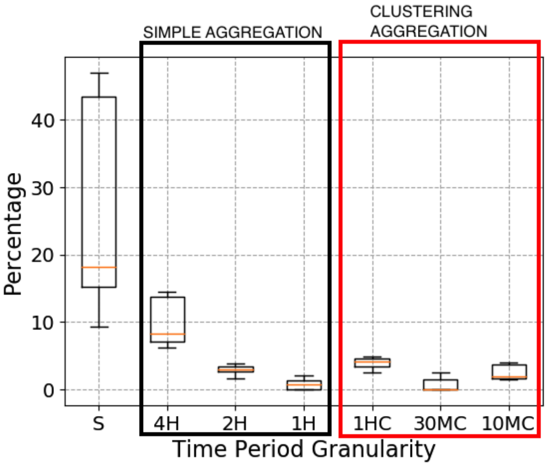


Cost Component Assessment

- Training retrieves one week assignment
- Test the one week assignment against the 12 weeks of data
- Cost Comparisons:
  - MEC access latency (Assignment Costs)
  - Orchestration Cost
  - Total Cost
- Comparison as percentage w.r.t. lowest

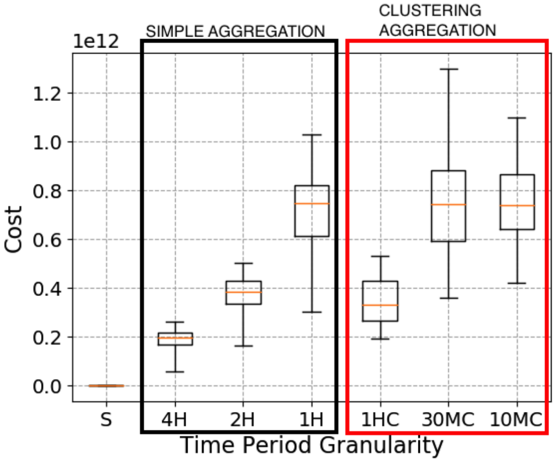
Cost Assessment - 1/3 - MEC access latency

- benchmark (S) always highest (+20%)
- clustering aggregation out-performs simple aggregation with same cardinality of time-periods



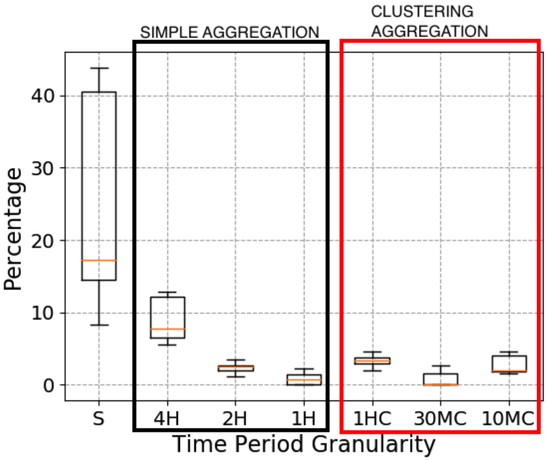
Cost Assessment - 2/3 -Orchestration Costs

- benchmark (S) always 0 (by design)
- Intuitively: lower number of time-periods → lower orchestration costs



Cost Assessment - 3/3 -Total Costs

- Total Costs is mostly consisting of assignment costs



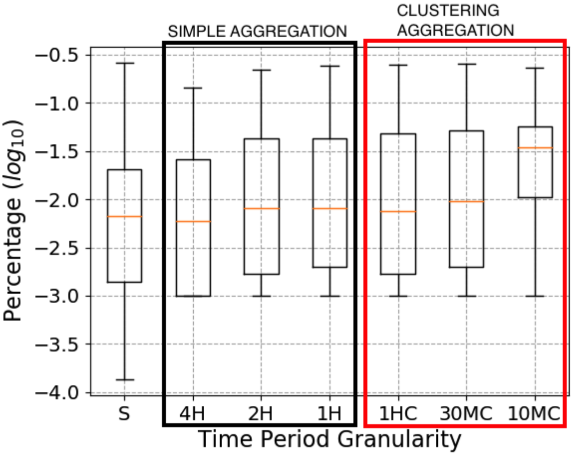
Capacity Violation

Demand variation from training week to test weeks leads to capacity violations:

- 3 indices:
  1- average capacity excess
  2- number of times a capacity is exceeded
  3- average of strictly positive capacity excess

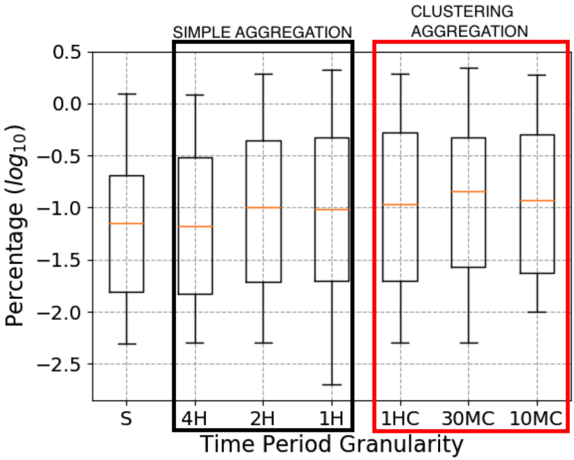
Average Capacity Excess

- always very low
- highest is 0.05% for clustering aggregation of 10-minutes (10MC) scenario



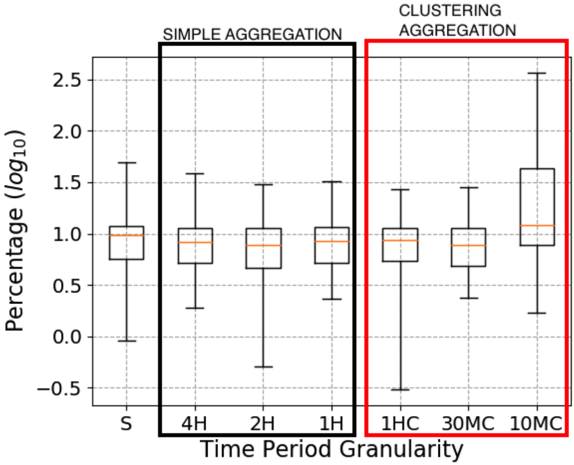
Percentage of time capacity is exceeded

- Always lower than 0.5%
- no clear differences between time aggregations



Average of strictly positive capacity excess

- Median value is constant
- worst-case for clustering of 10-minute (10MC): $> 350\%$



Take away messages

We understood that:

- data-driven approaches have great potential in next generation networking
- ingredient #1 is time-series analysis and spatiotemporal clustering to reduce an *infinite* horizon to a *finite* set of slots
- ingredient #2 is mathematical programming to optimize over (simple) *forecasting* to produce *decision patterns*
- ingredient #3 is easy usage of decision patterns on new data

Further Work:

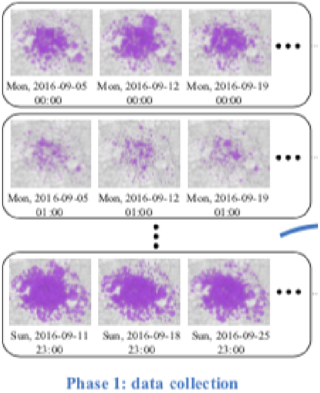
- Refine clustering algorithm
- Improve integration of data analytics and orchestration algorithm

Thanks for your attention.

Data Analytics

A. Furno, D. Naboulsi, R. Stanica, M. Fiore, “Mobile demand profiling for cellular cognitive networking”, IEEE Transactions on Mobile Computing, vol. 16, no. 3, March 2017

1- collecting data

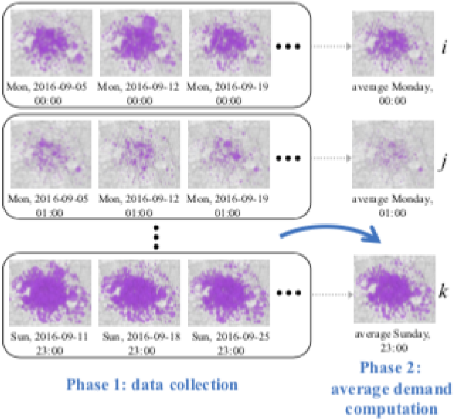


Data Analytics

A. Furno, D. Naboulsi, R. Stanica, M. Fiore, “Mobile demand profiling for cellular cognitive networking”, IEEE Transactions on Mobile Computing, vol. 16, no. 3, March 2017

1- collecting data

2- computation of weekly average demand



Data Analytics

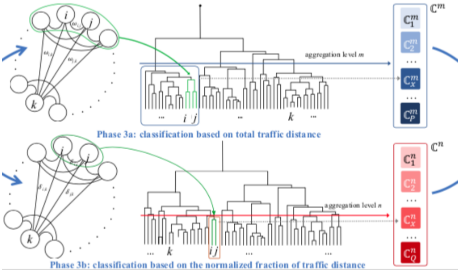
A. Furno, D. Naboulsi, R. Stanica, M. Fiore, “Mobile demand profiling for cellular cognitive networking”, IEEE Transactions on Mobile Computing, vol. 16, no. 3, March 2017

1- collecting data

2- computation of weekly average demand

3- Agglomerative hierarchical clustering with 2 distance metrics:

- difference of total traffic volumes
- difference of the normalized fraction of traffic (spatial distributions of the demand)



Data Analytics

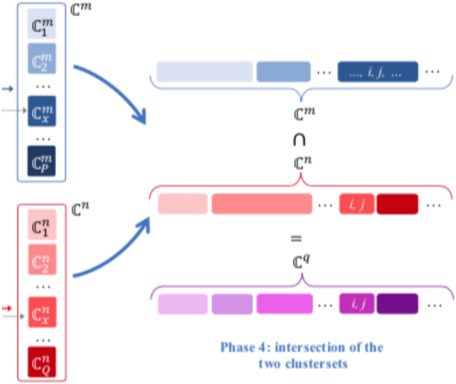
A. Furno, D. Naboulsi, R. Stanica, M. Fiore, “Mobile demand profiling for cellular cognitive networking”, IEEE Transactions on Mobile Computing, vol. 16, no. 3, March 2017

1- collecting data

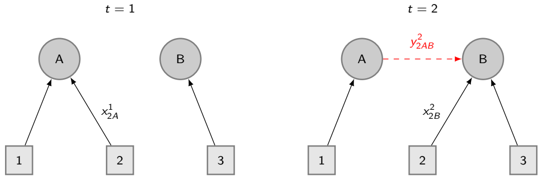
2- computation of weekly average demand

3- Agglomerative hierarchical clustering with 2 distance metrics:
   - difference of total traffic volumes
   - difference of the normalized fraction of traffic (spatial distributions of the demand)

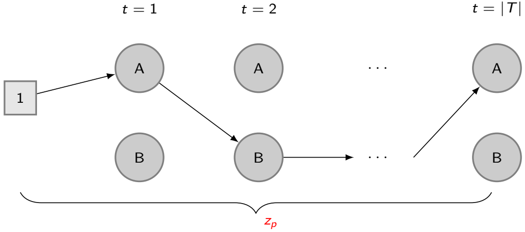
4- final clusters as intersection of clusters of two previous results



Orchestrating Mathematical Programming Model



Following Dantzig-Wolfe Decomposition principle:



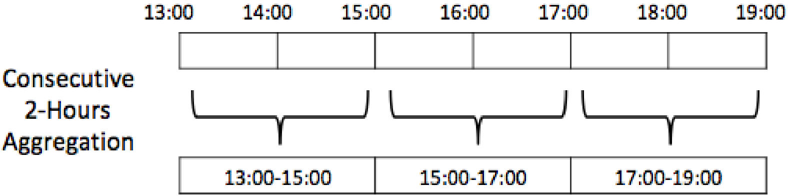
Dynamic variable generation (column generation technique)

Time-Periods Aggregation - 1/3

Simple Aggregation:

- aggregate consecutive 10-min periods to form a period of:
  - 4 hours (label 4H)
  - 2 hours (label 2H)
  - 1 hour (label 1H)

Example:

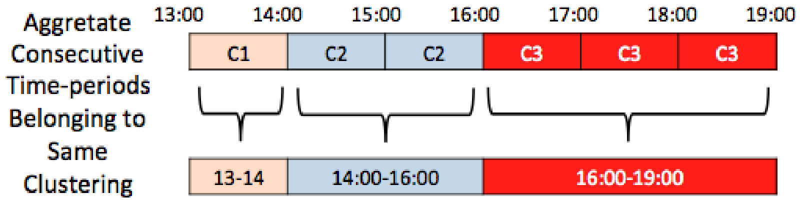


Time-Periods Aggregation - 2/3

Clustering-driven Aggregation:

- aggregate consecutive 10-min periods belonging to the same clustering profile with clusterset of:
  - 1 hour periods (label 1HC)
  - 30 minutes periods (label 30MC)
  - 10 minutes periods (label 10MC)

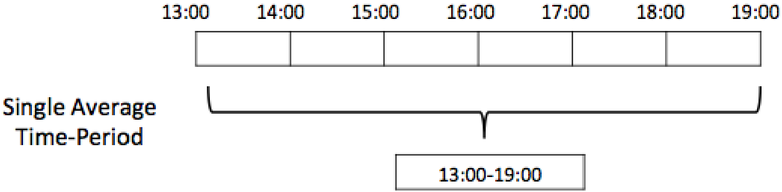
Example:



Time-Periods Aggregation - 3/3 - Benchmark

Baseline Approach (label S):

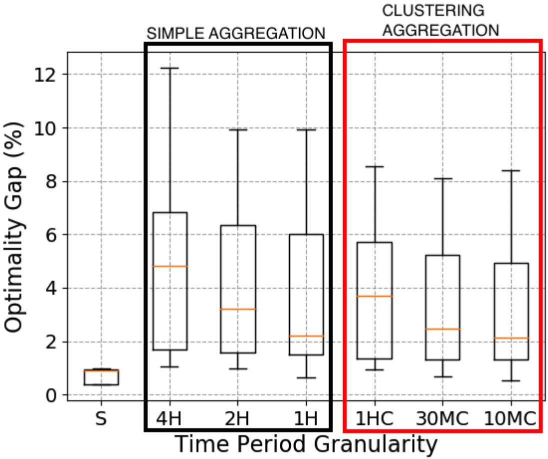
- single time-period
- fixed BS-MEC assignment $\rightarrow$ no switching
- training on average demand of all time-periods
- ILP problem solved with ILP solver of CPLEX until opt.gap $< 1\%$



Optimality Gap

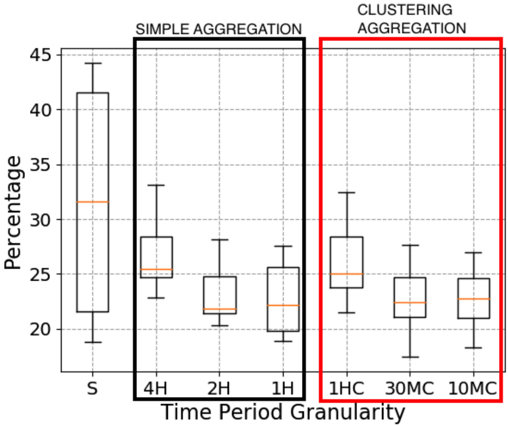
Median Optimality Gap:

- S: $< 1\%$
- 10MC/30MC: 3%
- 1H/1HC: 4%
- 4G: 5%



Non-Nearest Assignment

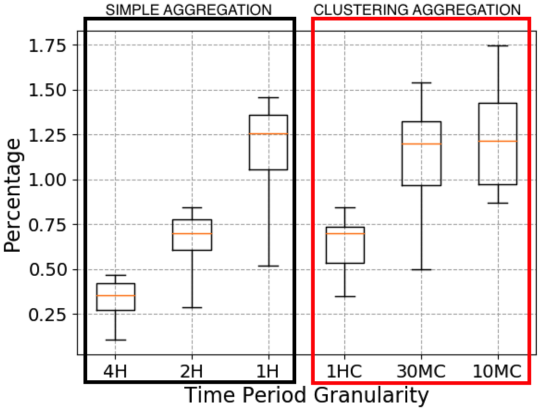
- Rationale behind assignment cost variations
- Assignment to nearest MEC facility is best option
- Percentage of time BS is not assigned to nearest MEC facility
- S highest $> 30\%$
- 4H/1HC: 26%
- all remaining: 22%



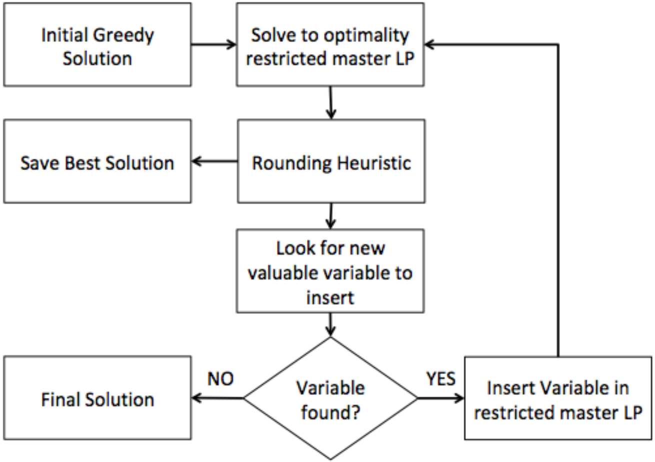
Switching occurrences

Rationale behind switching cost variations:

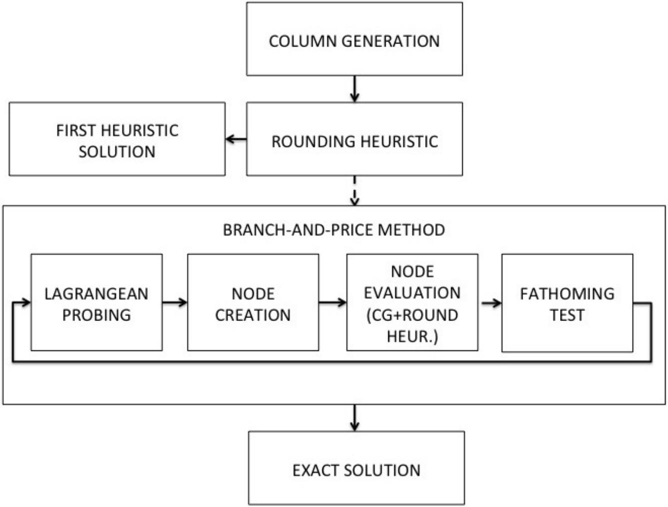
- percentage of number of switching occurring w.r.t. total number of possible switching
- median value is always very low
- highest value is 1.25% of possible switching occurring
- weight of switching cost in total cost is very low



Heuristic Framework



Resolution Framework



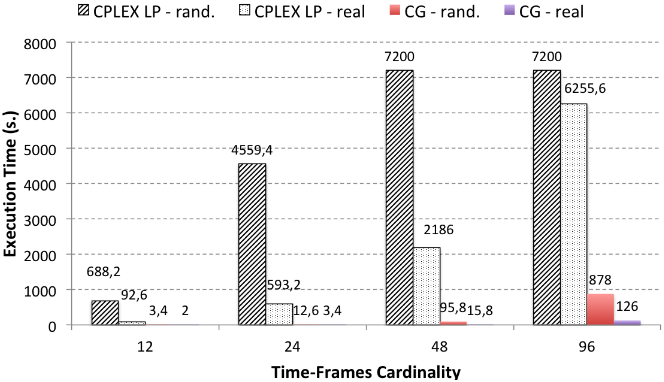
Computational Assessment

Extensive Computational Assessment can be found in:

"Optimized Assignment Patterns in Mobile Edge Cloud Networks", A. Ceselli, M. Fiore, M. Premoli, S. Secci, *Computers and Operation Research*, Elsevier, 2018, In Press

- two synthetic traffic demand datasets:
  - generated uniformly at random
  - single day real traffic demand perturbed uniformly at random
- increasing time-slot cardinality
- comparison of:
  - CG + rounding heur. vs. IBM CPLEX LP + rounding heur.
  - Branch-And-Price vs. IBM CPLEX ILP

Computational Assessment



We found that:

- our CG quicker than CPLEX LP (up to 2 order of magnitude)
- our CG\BaP scale better than CPLEX LP\ILP
- BaP provides good solution quicker than CPLEX ILP