Prediction-based Optimization in Mobile Networks

Joerg Widmer

Nicola Bui, Foivos Michelinakis, Guido Fioravantti

Developing the Science of Networks

Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

Multimedia in mobile networks

- Main source of mobile traffic [1]
- Expected to continue its growth (>70%) [1]
- Predictable requests and multimedia bitrate [2]
- Adaptable video quality [3]
- Buffering

[1] “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update,” 2014.
[2] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, “A peek into the future: predicting the evolution of popularity in user generated content,” in ACM WSDM 2013.
[3] R. Pantos and W. May, “HTTP live streaming,” IETF Draft, June, 2010.

Objectives

- Optimizing resource allocation for multimedia streaming will improve both network efficiency and users quality of experience (less streaming interruptions, higher average bitrate)

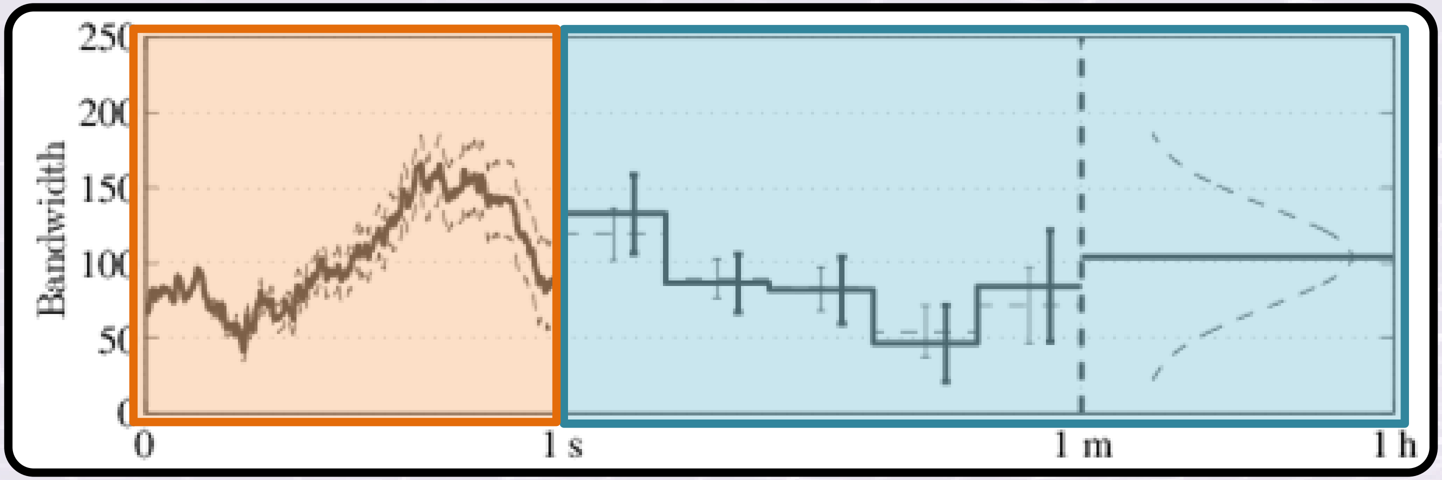
- Develop a general prediction model [1] accounting for different time scales and reliability
- Single user optimization [2] dealing with uncertainties
- Multi user optimization [3] accounting for variable quality

[1] Nicola Bui, Foivos Michelinakis, Joerg Widmer “A Model for Throughput Prediction for Mobile Users”, in European Wireless 2014
[2] Nicola Bui, Joerg Widmer “Mobile Network Resource Optimization under Imperfect Prediction”, in WoWMoM 2015
[3] Nicola Bui, Stefan Valentin, Joerg Widmer “Anticipatory Quality-Resource Allocation for Multi-User Mobile Video Streaming”, in CNTCV workshop 2015

Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

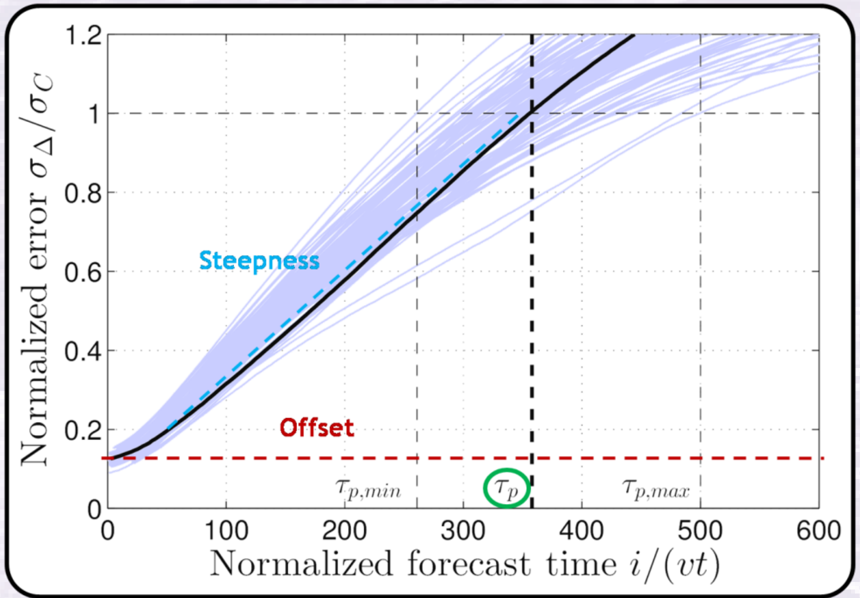
Model



- Idea
  - Short term: approximate the exact variation of the throughput time series as a sum of random variables using ARMA filters
  - Medium-long term: approximate the statistic distribution of the throughput accounting for uncertainties (user position, cell congestion, fading, etc.)

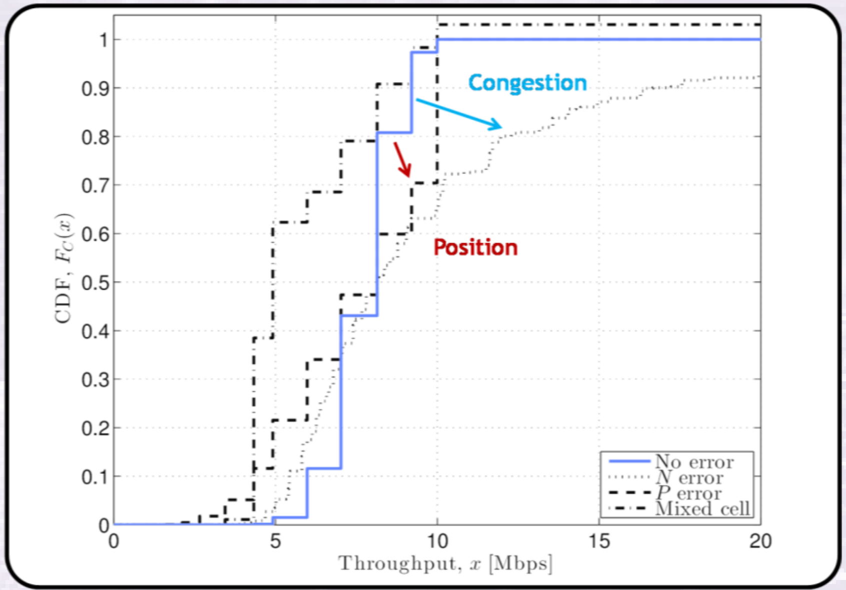
Short-term prediction

- ARMA filters to model achievable rates (throughput) fluctuations varying the sequence dynamics
- After a normalization over dynamics the prediction error shows a characteristic shape
- Two main components:
  - Offset dependent on the intrinsic noise
  - Steepness dependent on the dynamics speed
- Definition of reliability time



Medium-long term prediction

- The objective is to characterize the impact of imperfect knowledge on the achievable rate distribution
- Congestion -> the number of active users in a cell is not easily obtained
- Position -> in a given time frame the user position (both actual and measured) may vary



Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

System model: inputs

- Users' achievable rates are predicted:

$$R = \{r_{i,j} \in [0, r_M],\ i \in \mathcal{N}, j \in \mathcal{T}\}$$

- Can be assigned to users

$$A = \{a_{i,j} \in [0, 1],\ i \in \mathcal{N}, j \in \mathcal{T}\}$$

- The can buffer data

$$b_{i,j+1} = b_{i,j} + a_{i,j} r_{i,j} - \underline{d_{i,j}} - \underline{u_{i,j}}$$

- To stream the content at and bitrate between Minimum and Maximum

Optimization problem

- Optimization priorities
  1. Minimize total interruption time
  2. Minimize total missing extra quality
- K >> 1 enforce priorities

$$\underset{A}{\text{minimize}} \quad \sum_{k \in \mathcal{N}} \sum_{m \in \mathcal{T}} \left(K l_{k,m} / d_{k,m} + e_{k,m} / u_{k,m} \right)$$

$$\text{subject to:} \quad a_{i,j} \geq 0; \quad \sum_{k \in \mathcal{N}} a_{k,j} \leq 1$$

$$\forall i \in \mathcal{N}; j \in \mathcal{T}$$

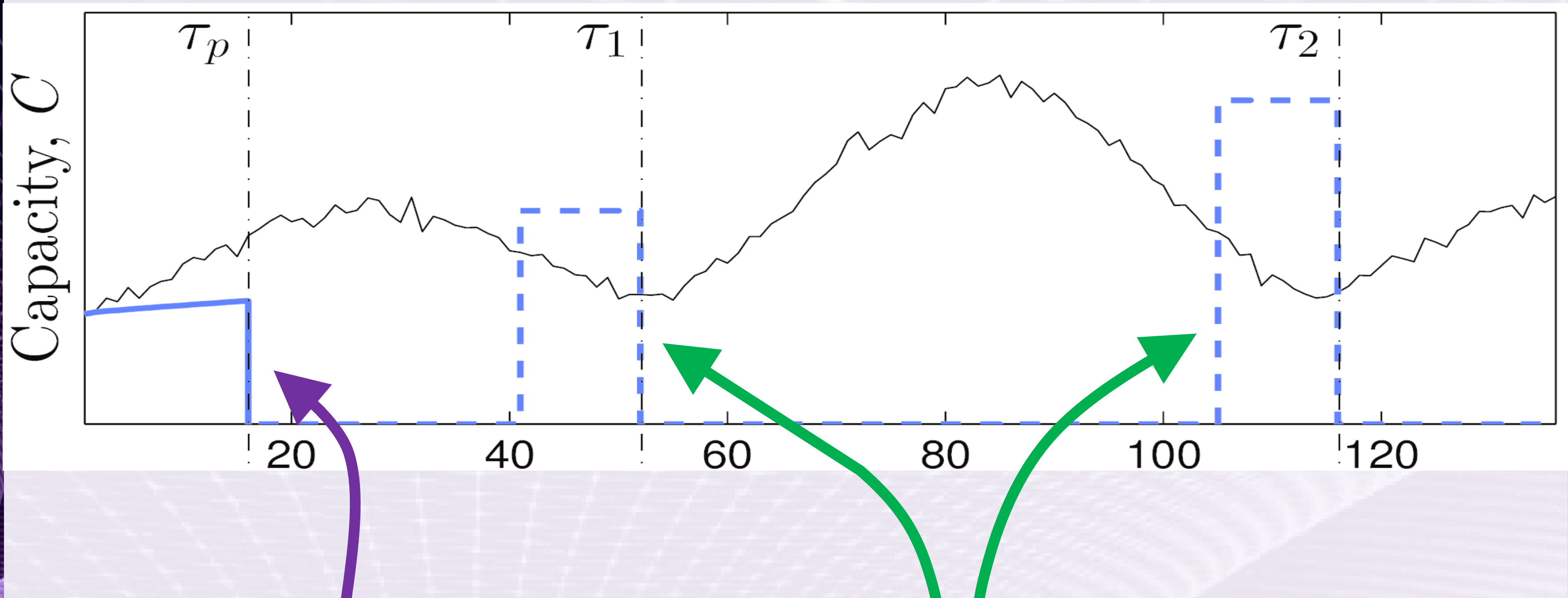
Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

Optimization with imperfect prediction

- Using predictors and statistic to create a predicted achievable rate sequence
- Solve the optimization problem
- Adopt the solution for a few time slot
- Update the prediction and repeat

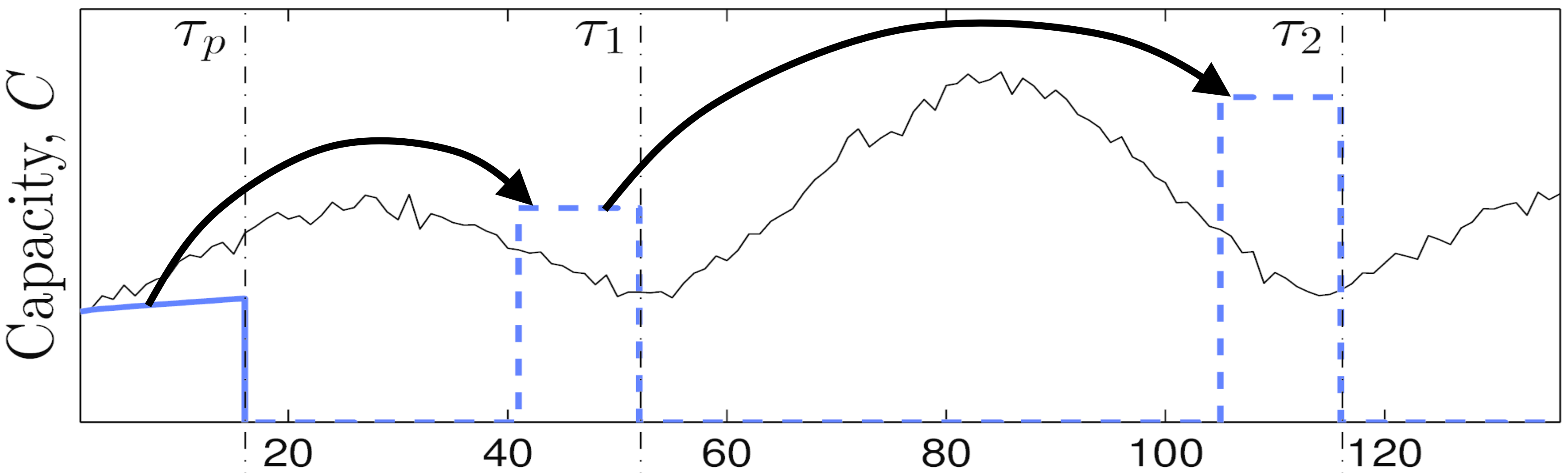
Combined prediction example



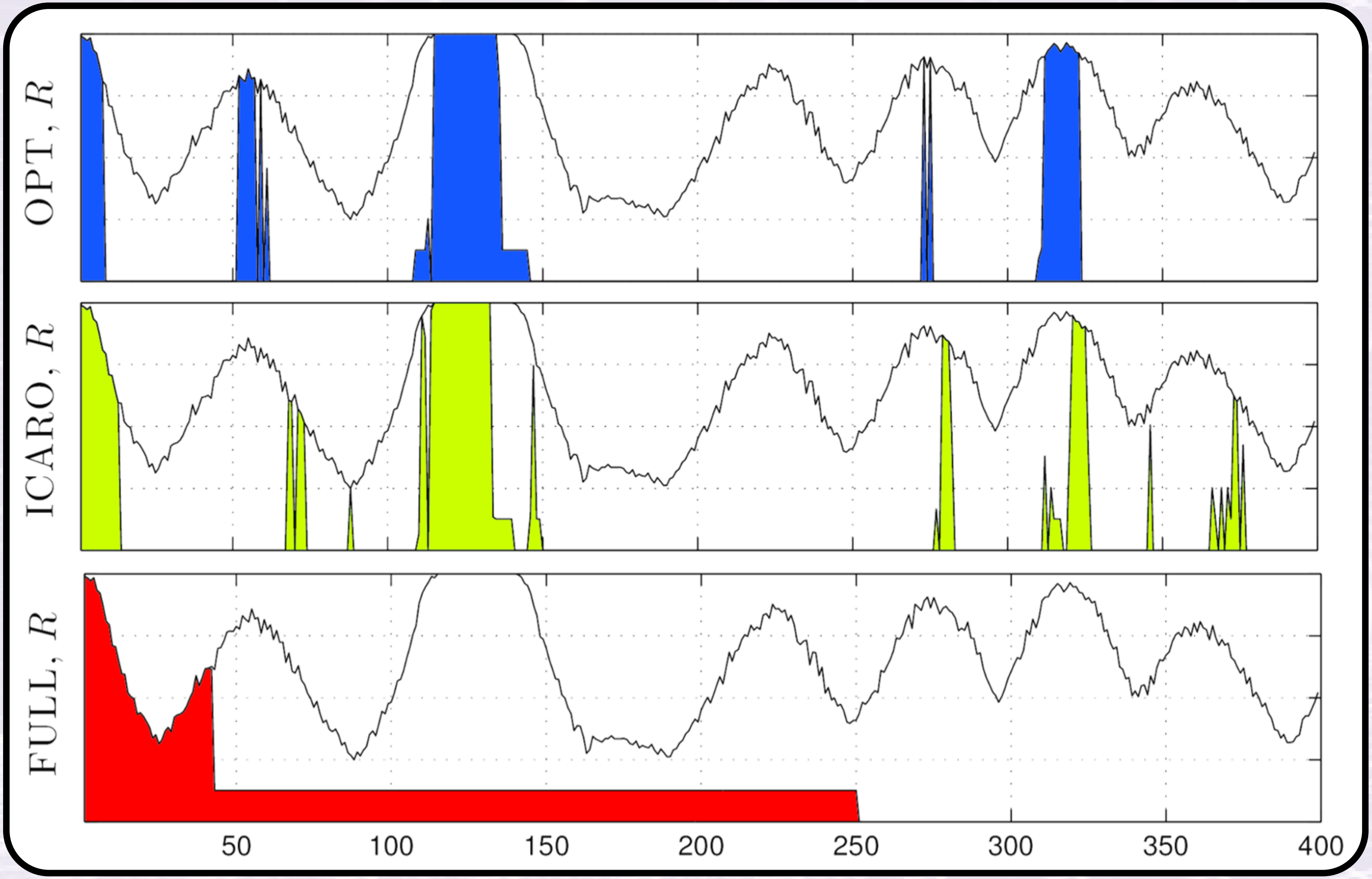
- Short term prediction
- Worst case scenario from the medium-long term models

ICARO: concept

- Imperfect Capacity prediction-Aware Resource Optimization
- Idea: to iteratively decide how much of the current rate to use according to the Split and Sort algorithm applied to the combined prediction sequence
- The current slot is not taken if the buffer allows to wait for cheaper slots

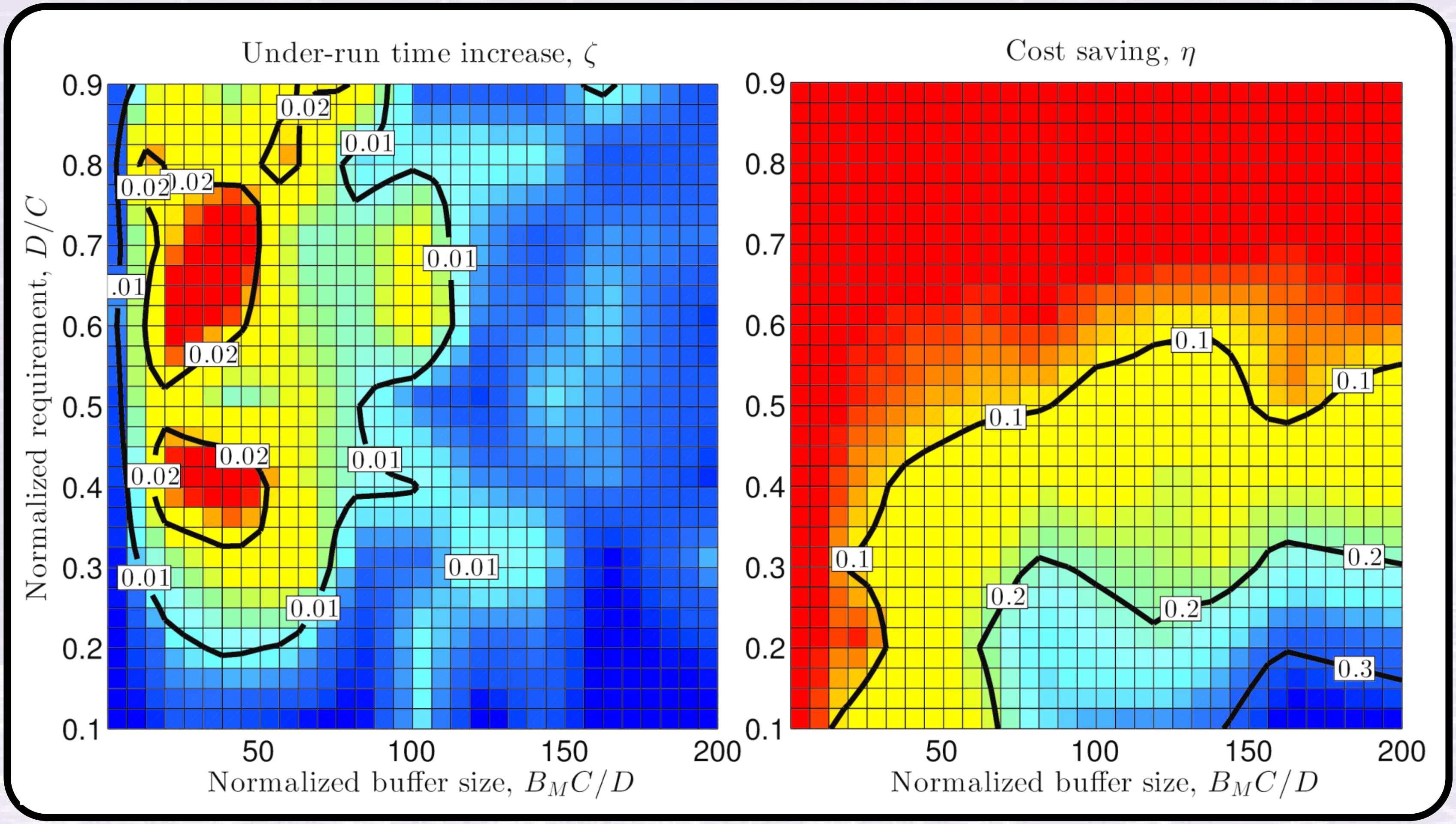


ICARO: result example



ICARO: overall results

- Almost optimal interruption time
- Up to 30% resource savings



Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

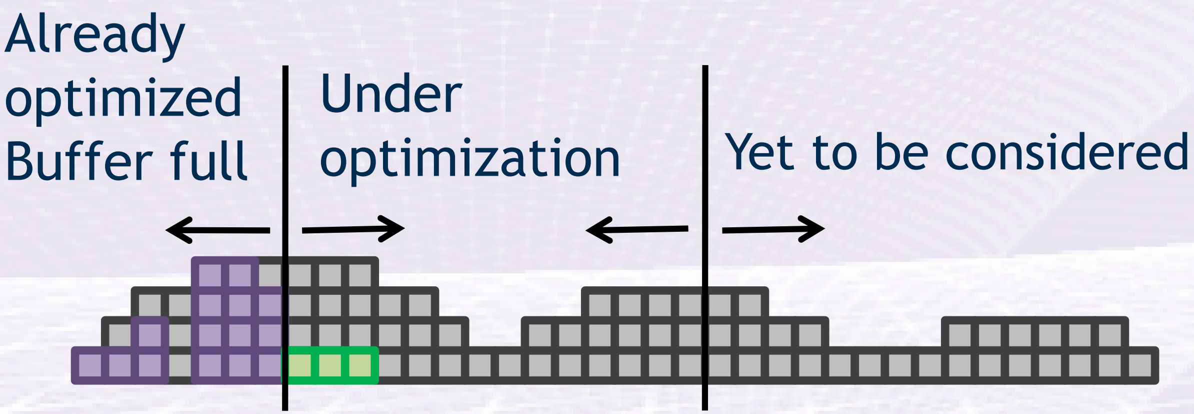
Multiple users optimization

- Heuristic algorithm to solve the continuous streaming and quality maximization problem
- Problem extension to support guaranteed quality of service
- Problem approximation to allow for a fast computation of the solution

Algorithm

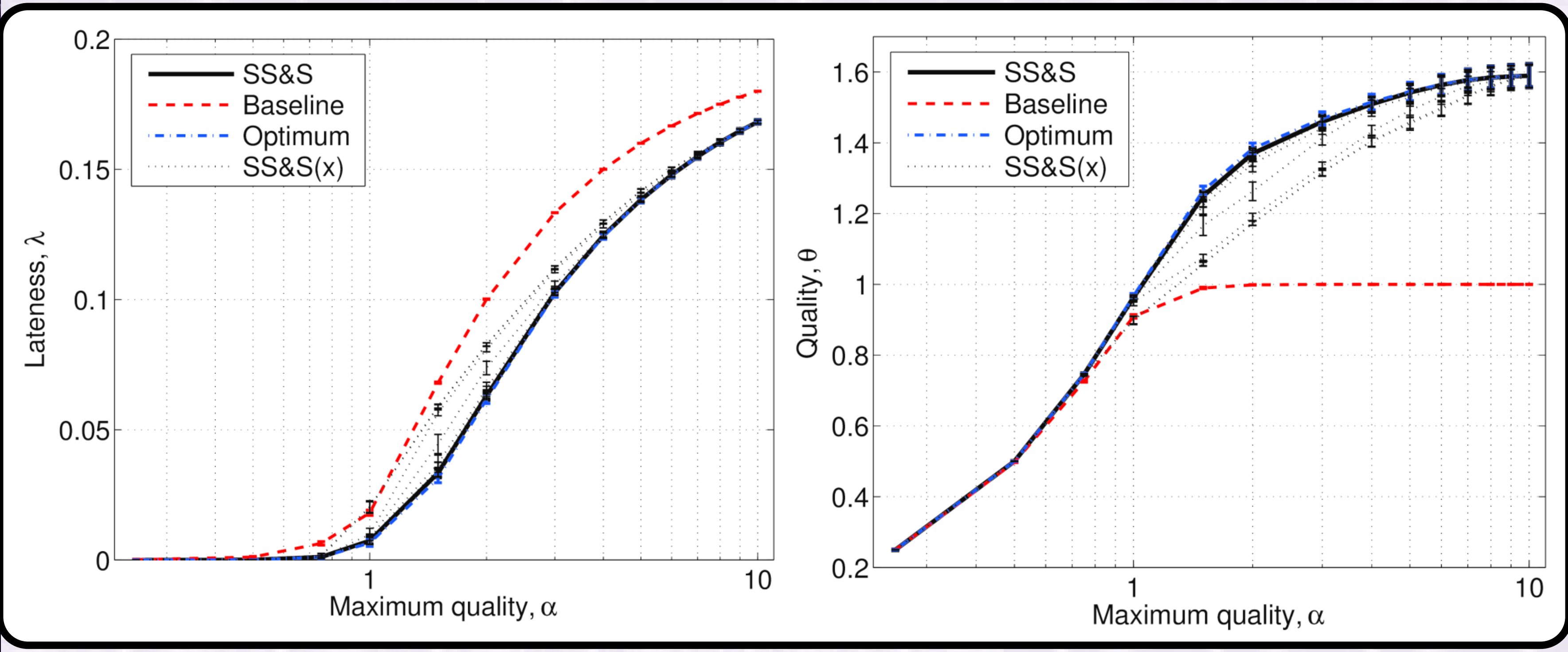
- The Split, Sort & Swap algorithm iterates the following steps:

1. Split & Sort: greedily optimize a region until no more resource can be allocated due to either complete assignment of the best slot and users or buffer violation.
2. Swap: within each of the split windows resources are swapped to improve the objective function.



Results: multi user improvements

- SS&S is achieving quasi optimal performance both in terms of saved interruption time (lateness on the left) and maximum quality (right)
- A higher number of iterations is required when $1<\alpha<3$: in this region the greedy approach is less likely to obtain an optimal allocation



Supporting QoS

- MILP formulation

$$\underset{A,B,L,E,S}{\text{maximize}} \quad \sum_{k \in \mathcal{N}} \left(K(\lambda_k + \underline{s_k}) + \theta_k \right)$$

$$\text{subject to:} \quad a_{i,j} \geq 0; \quad \sum_{k \in \mathcal{N}} a_{k,j} \leq 1$$

$$\lambda_i \geq \underline{\lambda_i^*}\,\underline{s_i}; \quad \theta_i \geq \underline{\theta_i^*}\,\underline{s_i}$$

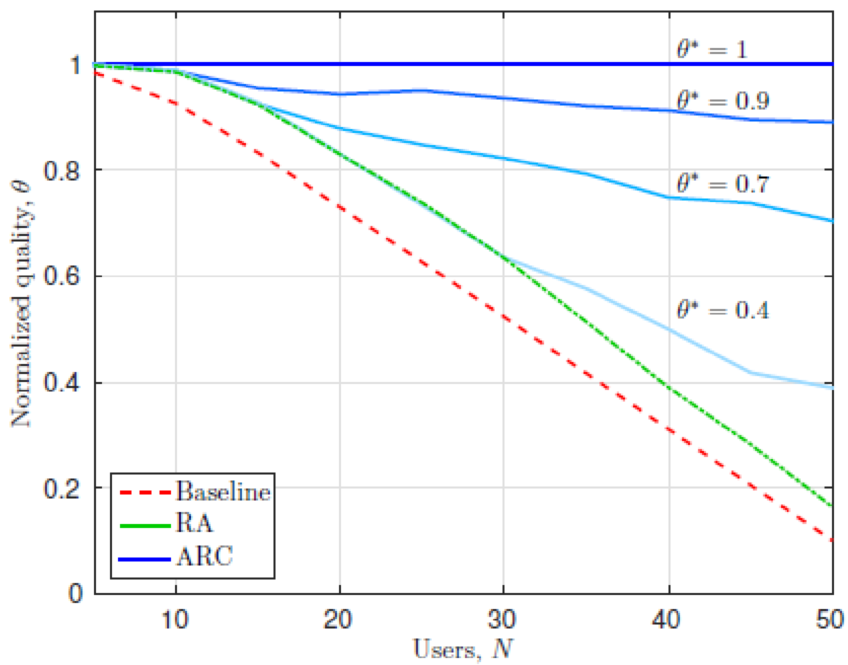
- New integer binary variable for Scheduled users
- QoS constraints for scheduled users

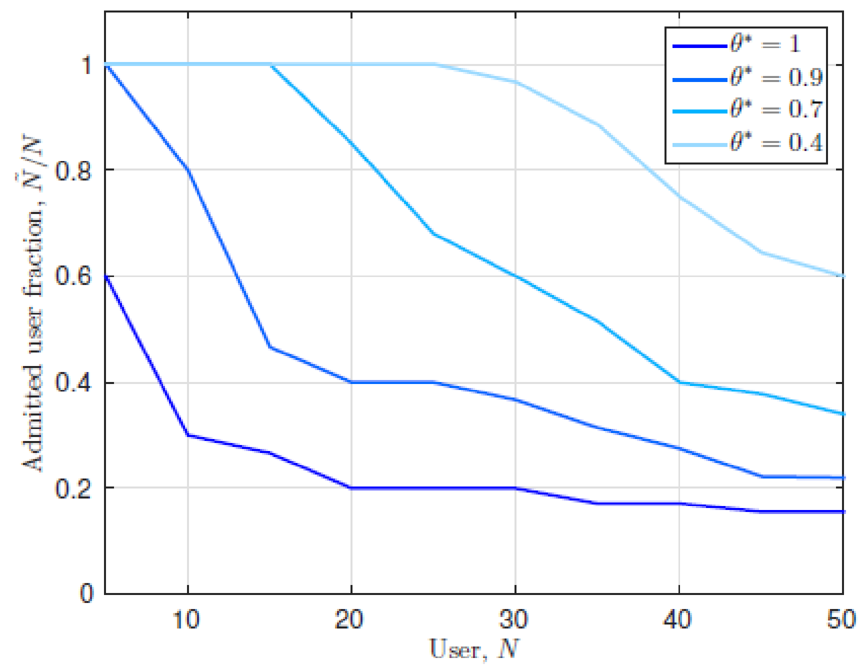
Approximation

- We rank users according to the problem objective function
- We compute the largest set of admitted users
  - Running a dichotomic search over the set size
  - For each size checking whether the problem is feasible with our SS&S algorithm or an LP solver
  - The last iteration gives also the solution

Results: guaranteed QoS

- The approximated problem is effective in guaranteeing QoS
- Quality (left) is consequently traded with the number of admitted users (right)





Outline

- Introduction
- Prediction model
- Optimization framework
- Single user optimization with uncertainties
- Multi-user optimization
- Conclusions

Conclusions

- Prediction model
  - Short term based on filters
  - Medium/long term statistic model
- Single user optimization
  - Accounting for uncertainties
  - Quasi optimal interruption time
  - 30% resource saving
- Multi-user optimization
  - Accounting for variable quality
  - Guaranteed QoS

Thanks

Prediction-based Optimization in Mobile Networks

Nicola Bui

Advisor: Joerg Widmer

Collaborators: Foivos Michelinakis, Guido Fioravantti, Stefan Valentin, Ilaria Malanchini

Questions?