Context-Aware Resource Allocation for Cellular Networks

Magnus Proebster, Matthias Kaschub, Thomas Werthmann, Stefan Valentin
magnus.proebster@ikr.uni-stuttgart.de
stefan.valentin@alcatel-lucent.com
13.03.2012

Universität Stuttgart
Institute of Communication Networks
and Computer Engineering (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

Alcatel·Lucent Bell Labs

Motivation

Modern Smartphones: diverse traffic requirements from various applications

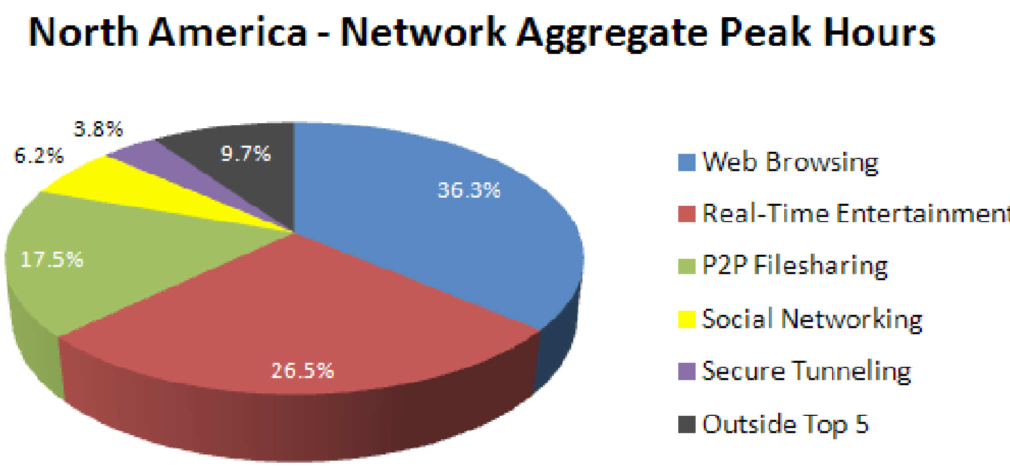
→ Traffic is heterogeneous and bursty

→ Load peaks can degrade the user's experience

Assumption: Bottleneck in mobile cellular networks → Radio access link

Observation: Plenty of traffic can wait

- App downloads / updates
- Browser background tabs
- File downloads



Sandvine, *"2010 Mobile Internet Phenomena Report"*

Exploit information about user's *context* at the scheduler

"Which part of the user's traffic can wait?"

➨Serve more users in the cell without reducing user satisfaction!

Overview

Signaling & Transaction Framework

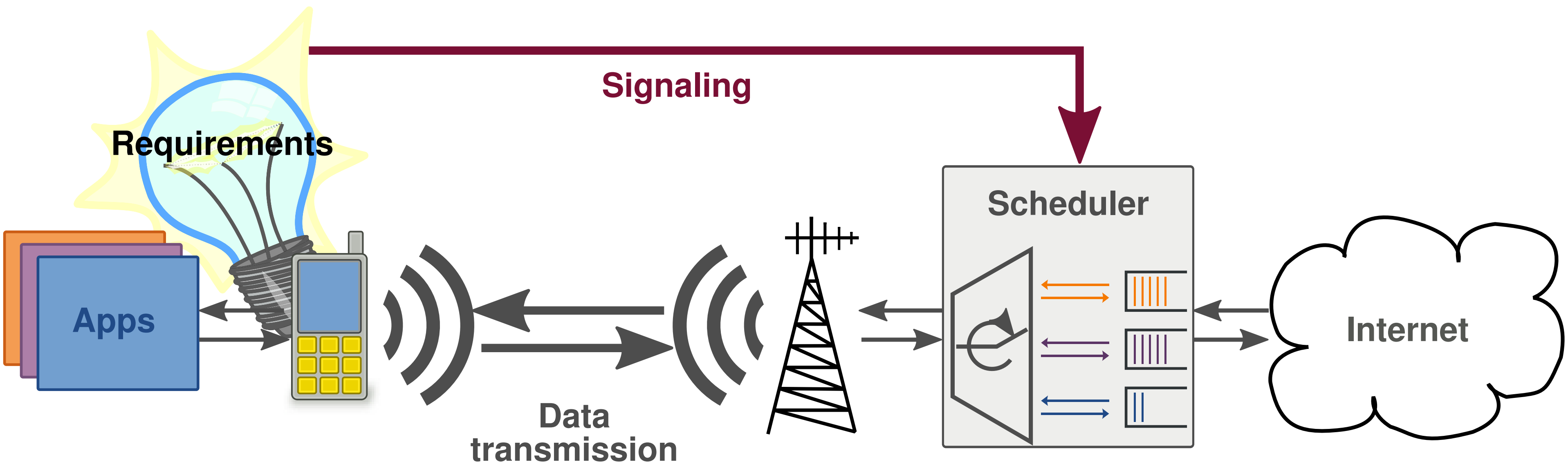
Optimal & Heuristic Scheduler

Simulation Results

Conclusions

Approach: Context signaling

- Signal information about application's environment
- More than just application awareness: Information on users' OS & environment
- Examples
  application foreground / background state, screen-saver, location, orientation, ...

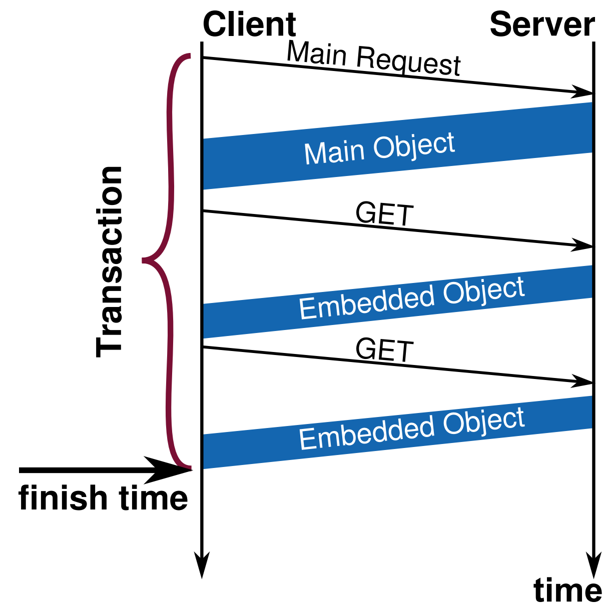


➨**Make the scheduler in the base station aware of context information**

Approach: Transactions

Definition

- Transactions reflect all traffic between a **user's request** and its **observable result**
- Mapping of MAC / IP packets to application layer objects
- Transaction can be segmentation / aggregation of transport connections (e.g. TCP)
- QoS requirements referring to transactions



Example Transaction (HTTP)

Advantages

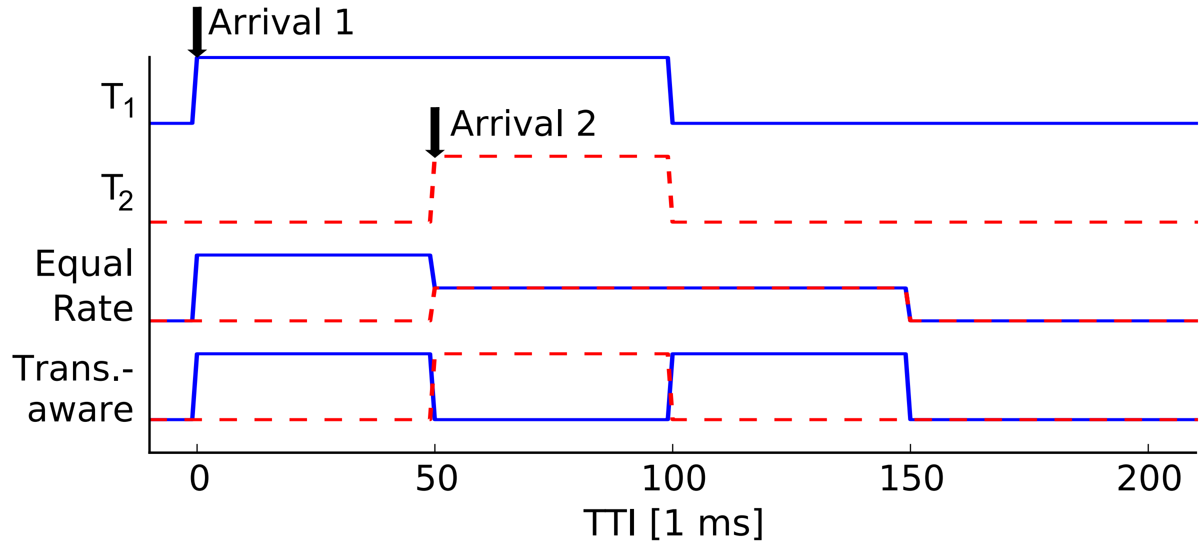
- Scheduler considers user-visible objects relevant for user experience
- Allows to shift complete application layer objects in time
- Allows to reduce interleaving between transactions which is bad for transaction finish-times

➨ **Focus on user-visible latency & Quality of Experience**

Illustration Example

Scenario

- Arrival of two transactions
- Constant bandwidth per user



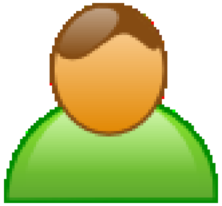
Comparison

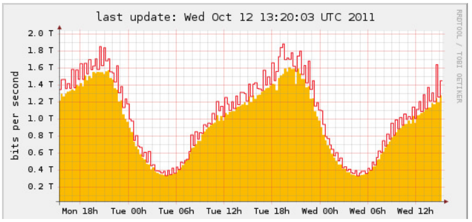
- Fair scheduler **without** context knowledge
  - Distributes bandwidth equally
- CARA scheduler **with** context knowledge
  - Additional degree of freedom
  - Total finish time improved

➨ **"Put the latency where it doesn't hurt!"**

Context Information Sources

- **User**
  - Explicit feedback
  - Preferences, configuration
- **Applications**
  - Type of application, transaction, priority, ...
  - Activity (foreground tab?)
  - Size of transaction (often estimation)
- **Platform**
  - Event source (click, timer)
  - Parallel or interactive activity
  - Reasonable defaults for application values
- **Device / operating system**
  - Screensaver, device orientation, proximity sensor
  - Foreground / background
- **Network**
  - Current and future network load

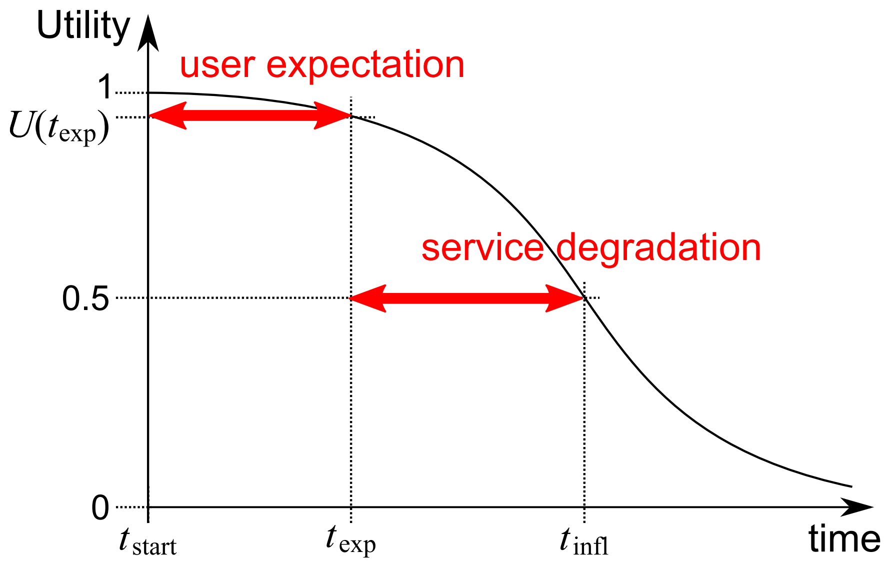


Utility Functions

How to make use of context information?

- Formalize user requirements
  → Utility functions
- Express individual latency requirements of a transaction
- Describe user experience in terms of transaction finish times
- Parameters are derived from context information (e.g., user focus)



Examples

- Foreground: Web pages
- Background: File Downloads

➨ **Here: U(t) instead of U(r)**

➨ **Units: Transactions instead of single packets**

Utility Optimization Problem

Context-aware resource allocation as a constrained maximization problem

Objective Function

$$\text{maximize} \quad U_{total} = \sum_t \sum_T U_T(t) f_{T,t}$$

Constraints

$$\begin{aligned}
\forall T:\ \sum_t f_{T,t} &= 1 \\
\forall T,t:\ f_{T,t} &\leq \frac{1}{B_T}\left(\sum_{t_1=1}^{t} r_{T,t_1}\gamma_{T,t_1}\right) \\
\forall t:\ R &\geq \sum_T r_{T,t} \\
\forall T:\ B_T &= \sum_t r_{T,t}\gamma_T(t) \\
\forall t < t_{0T}:\ r_{T,t} &= 0
\end{aligned}$$

- Assumes ideal channel and traffic knowledge
- Determines the optimal scheduling solution for a predefined time span
- Size of the solution space (decision variables): $\mathrm{O}(\mathrm{n_T} \cdot \mathrm{n_{TTI}})$

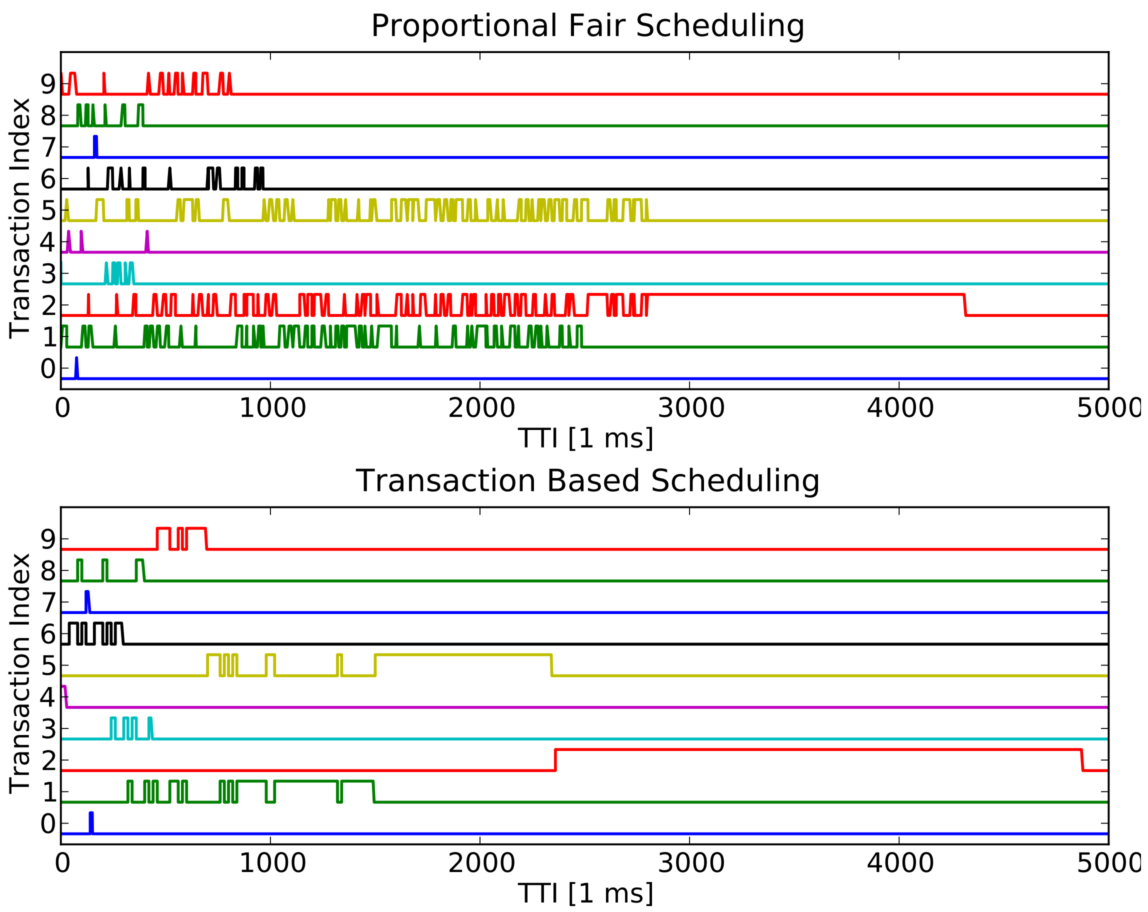
Resource Allocation Comparison

Proportional Fair

- Focus on channel conditions
- Interleaved resource allocation

→ Good for capacity

→ Not so good for finish times

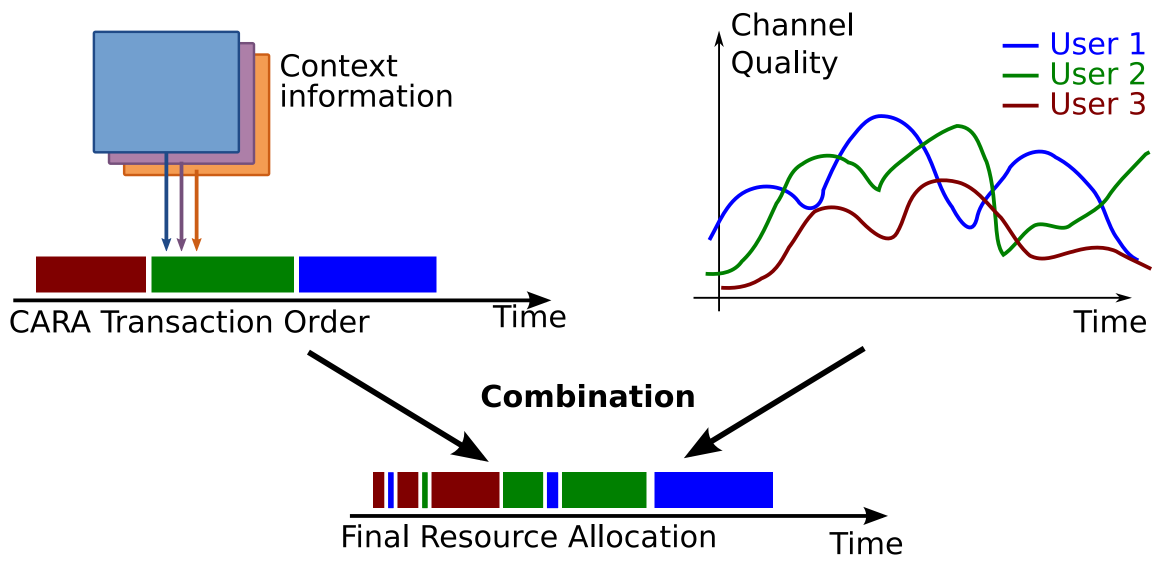


Transaction-based

- Focus on finish times
- Channel-aware
- Context-aware (Utility-max)

→ Reduced finish times, higher Utility

2-Step CARA Scheduling Heuristic



Step 1: Sequence Planning

- Optimize utility by planning a sequence of active transactions
- Account for average & predicted channel situations

→ **Reduces interleaving, increases overall utility**

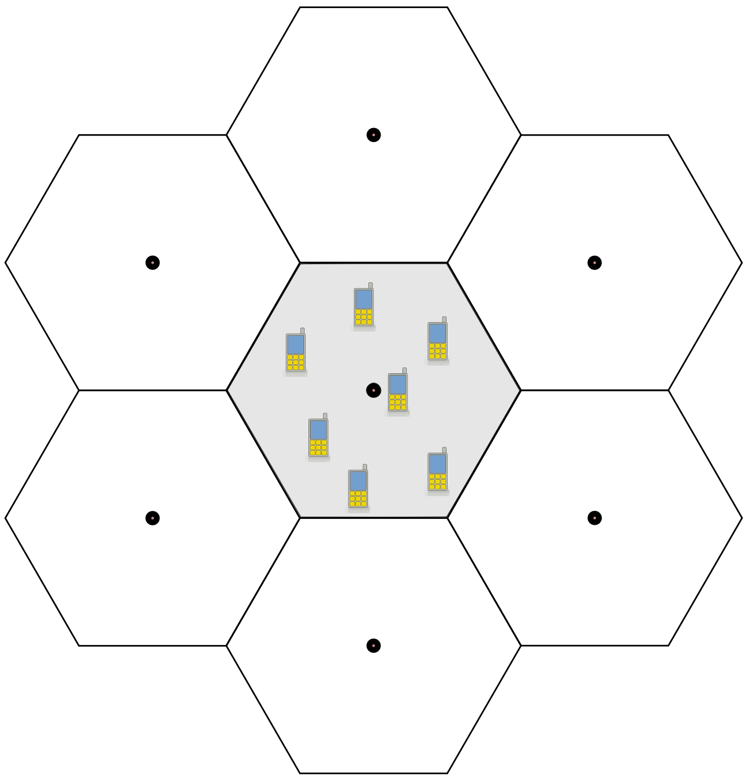
Step 2: Accounting for Short-Term Channel Fluctuations

- Channel fluctuations can be faster than transaction granularity
- Deviate from scheduling sequence in dependence of instantaneous channel condition

→ **Increases total cell throughput**

Simulation Scenario

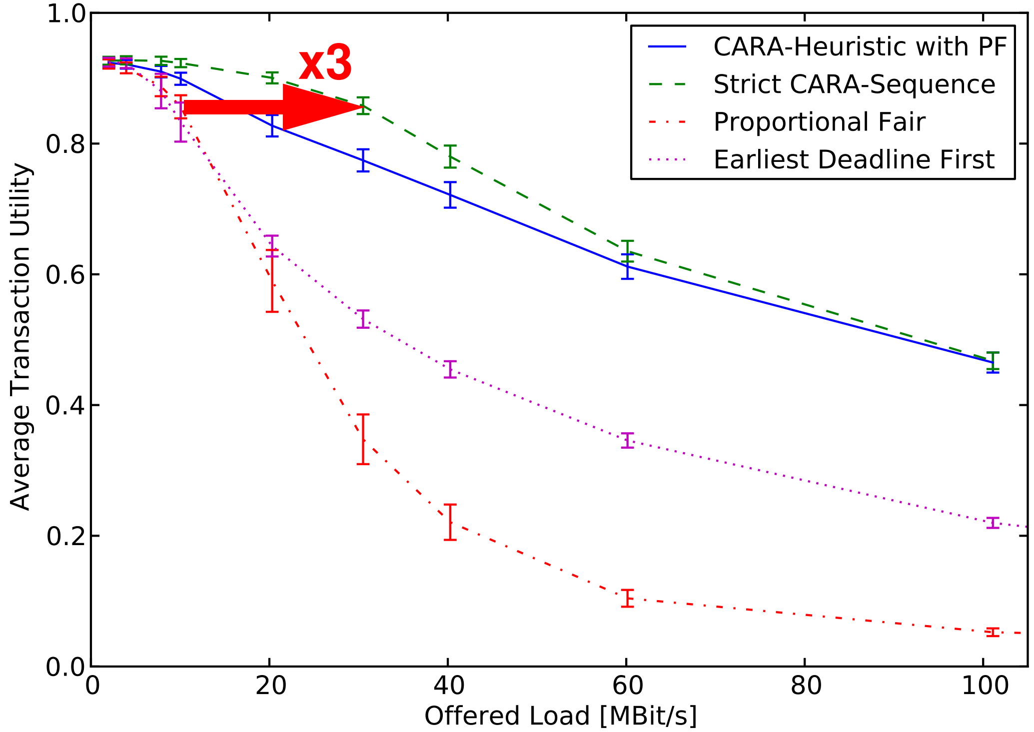
- Cellular network, 7 sites (center cell evaluated)
- OFDMA System (e.g., 3GPP LTE)
  2 GHz band, 10 MHz bandwidth
- Channel model includes
  - Path loss
  - Shadowing
  - Fast fading (Rayleigh, Veh. A, 10 km/h)
- Shannon capacity, SINR clipped at 20 dB
- Transmission time interval: 1 ms
- 20 user equipments
- Multiple transactions per UE possible
- Simulation duration: 10 s
- Traffic Mix
  - Foreground: HTTP
  - Background: FTP



Utility Improvements with CARA Scheduling

Utility vs offered load

- Utility range 1 (good) .. 0 (bad)
- Load variations by changing IAT
- System fully occupied at ~20Mbit/s

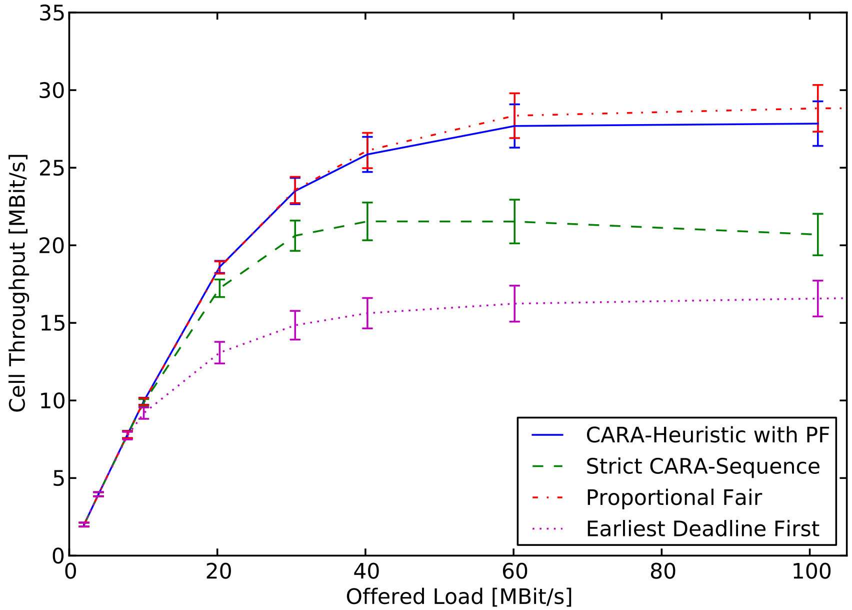


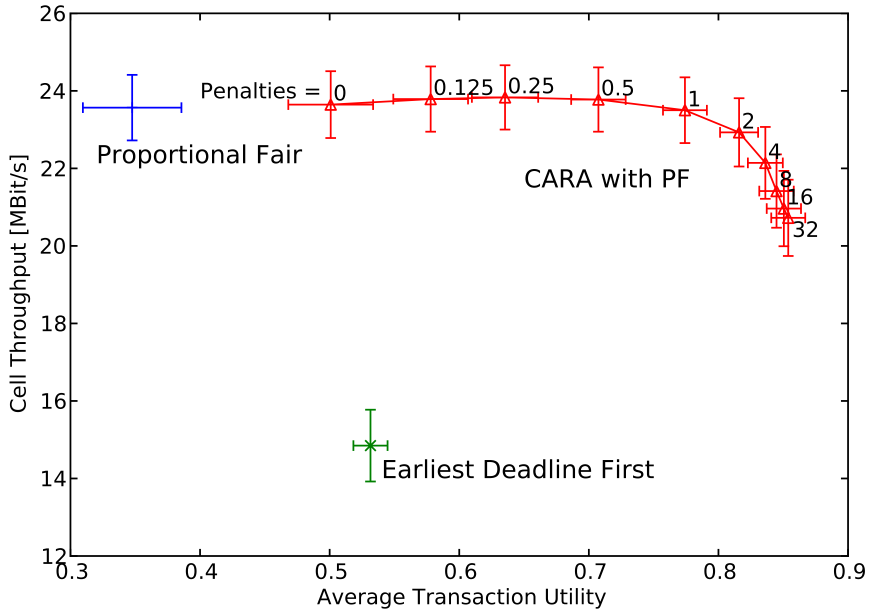
Observations

- Utility of PF strongly decreases for higher load
- EDF considers utility but is not channel-aware
- CARA heuristics maintain good utility performance even at high loads
- Strict sequence (no short-term channel-adaptation) offers best utility

➨ **Serve more users without decreasing user experience!**

Cell Throughput vs Utility





- PF achieves good cell throughput but bad utility
- EDF not channel-aware → worst throughput results
- Combination of CARA and Proportional Fair delivers best throughput AND utility
- Trade-off in CARA-PF combination with penalty factor

➨ **Strong utility increase without throughput degradation possible**

Conclusions

Exploit context information for scheduling by new transaction-based framework

From packet-level to transaction-level $\rightarrow$ Plan resource allocation into the future

Quality of experience expressed by utility functions
$\rightarrow$ Improve the user observable result

Practical scheduling heuristic

Combines the advantage of opportunistic and context-aware scheduling

Increases the number of users at the same utility level

Outlook

Improved scheduling heuristic, publication in preparation

[1] M.Proebster, M.Kaschub, and S.Valentin *"Context-Aware Resource Allocation to Improve the Quality of Service of Heterogeneous Traffic"*, IEEE ICC 2011, Kyoto

[2] M.Proebster, M.Kaschub, T.Werthmann, and S.Valentin *"Context-Aware Resource Allocation for Cellular Wireless Networks"*, submitted to EURASIP JWCN