Universität Stuttgart Alcatel·Lucent Bell Labs

Institut für Kommunikationsnetze und Rechnersysteme

Self-Organizing QoS Optimization by Context-Aware Resource Allocation

Problem & Motivation

- Modern Smartphones have diverse traffic requirements, from various applications: multimedia, data, voice, ...
  - ➨ Traffic is heterogeneous and bursty
  - ➨ Heavy load peaks can degrade the user's experience
- Bottleneck in mobile cellular networks: Radio access link
- **Observation:** Plenty of traffic can wait

Software updates, browser background tabs,...

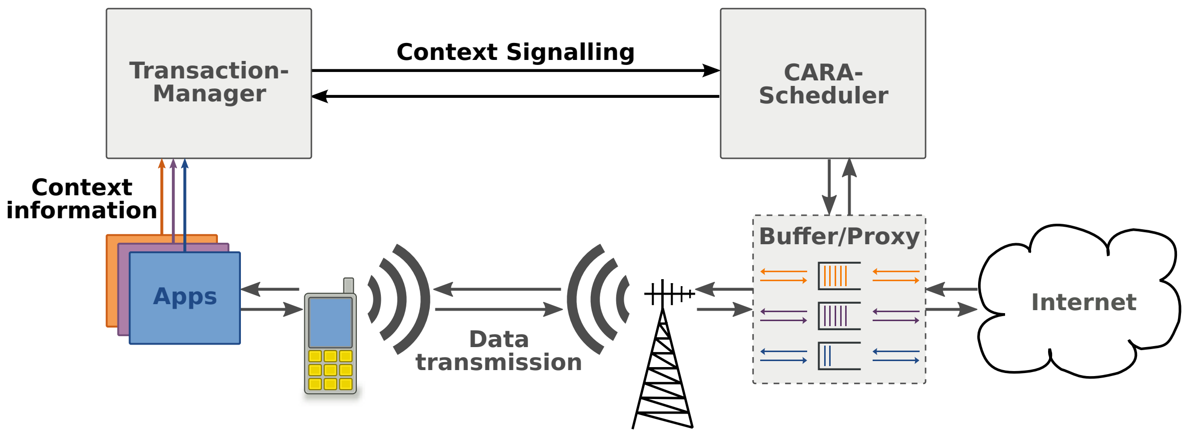
Approach

- Exploit more information about the user's context at the scheduler
- Example for context information:
  „Which part of the user's traffic can wait?"

  ➨ Shifting transmissions in time can **improve real-time services** and **increase multi-user diversity**
- Here: CARA architecture and theoretical framework to access and profit from context information

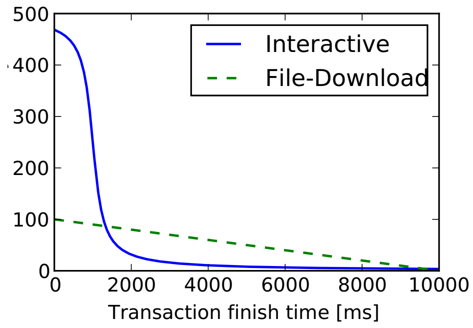
CARA Architecture

- Proposed traffic representation:
  Transactions reflecting all traffic between a user interaction and its observable result
- For each transaction: Obtain context information
  - Derived from local application knowledge
  - Signaled to the base station
- Advantage: Allows to plan scheduling ahead
- Utility functions for each transaction derived from local context information



Context Features & Time-Variant Utility Functions

- **Context Features**
  - Allow to collect and to aggregate local context information
  - **Examples:** User focus, speed and environment, device orientation, process activation, user preferences, ...
- **Time-Variant Utility Functions**
  - Express individual delay requirements of a transaction
  - Describe user experience w.r.t. transaction finish times
  - Have higher Utility when transaction finishes earlier
  - Express different delay classes by different shapes



Optimization With Ideal Knowledge

- Assume ideal channel and traffic knowledge
- Determine the optimal scheduling solution for a predefined time span
- Formulation as Utility optimization problem:

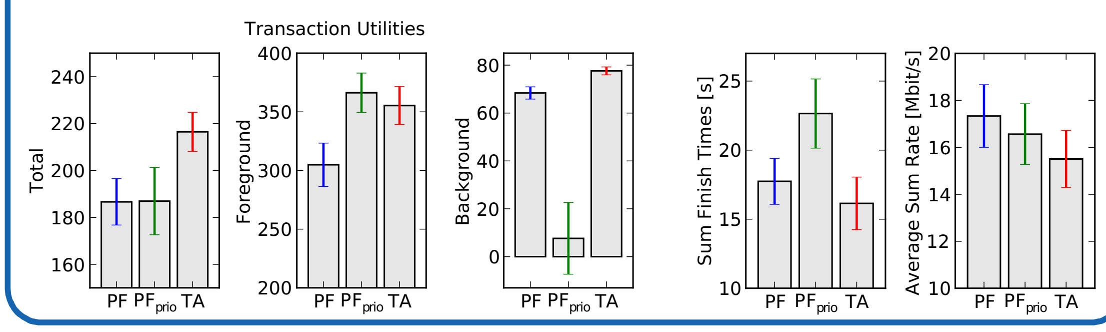
$$\text{maximize} \quad U_{total} = \sum_t \sum_T U_T(t) f_{T,t}$$

- With the constraints:

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

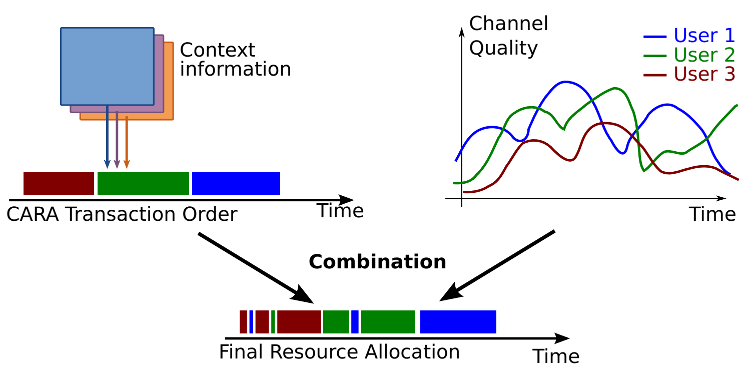
Simulation Results

- **Scenario:** *5* foreground and *5* background transactions starting at *t = 0 s*, *10 s* simulation time, fast fading only, *10 MHz* bandwidth
- **Comparison:** Transaction Aware (TA) and Proportional Fair (PF) scheduling without and with static priorization
- **Results**
  - Overall Utility for foreground ***and*** background traffic increased
  - Static priorization cannot improve total Utility
  - Average finish times improved
  - Average sum rate slightly decreased



Scheduling Heuristic

- **Objective:** Implement a context-aware scheduler
- **Is based on:** Determining beneficial transaction order by using context information
- **Involves:** Prediction of channel and traffic states
- **Exploits:** Multi-user diversity by preferring a transaction with good channel quality
- **Is flexible:** Scaling factor to trade off CARA-sequence and channel-awareness



Magnus Proebster (magnus.proebster@ikr.uni-stuttgart.de), Matthias Kaschub, and Stefan Valentin