Saving Resources on Wireless Uplinks: Models of Queue-aware Scheduling

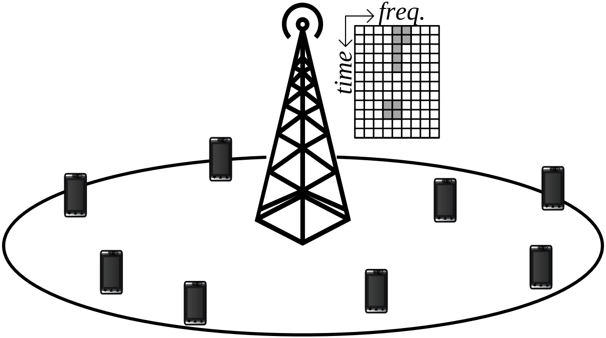
TECHNISCHE UNIVERSITÄT DARMSTADT

Amr Rizk
TU Darmstadt

- joint work with Markus Fidler

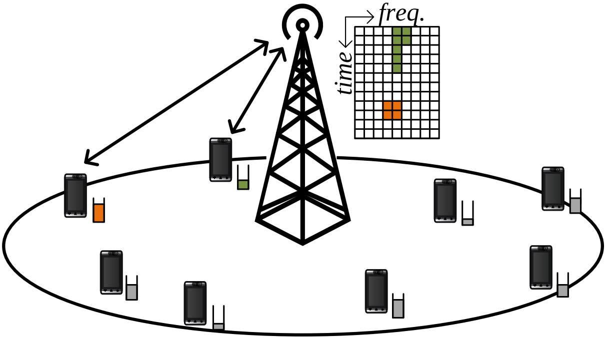
Cellular Uplink Scheduling

TECHNISCHE UNIVERSITÄT DARMSTADT



Cellular Uplink Scheduling

TECHNISCHE UNIVERSITÄT DARMSTADT



Adaptive Resource Allocation in Cellular Uplink Direction

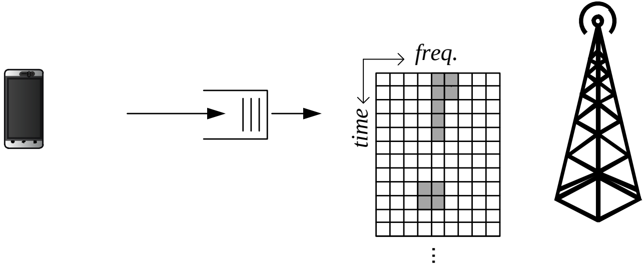
TECHNISCHE UNIVERSITÄT DARMSTADT

Input metrics (LTE)

- Buffer status reports (BSR)
- Channel quality indicators (CQI)

Goal

- Statistical QoS guarantee



Adaptive Resource Allocation in Cellular Uplink Direction

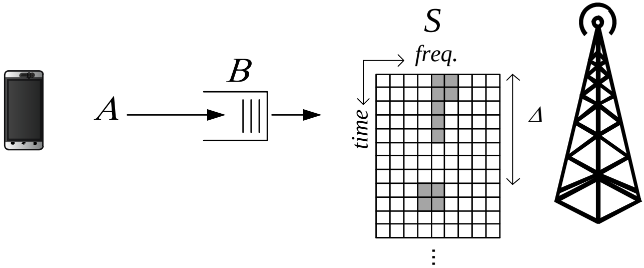
TECHNISCHE UNIVERSITÄT DARMSTADT

Input metrics (LTE)

- Buffer status reports (BSR)
- Channel quality indicators (CQI)

Goal

- Statistical QoS guarantee



- Arrival traffic A
- Buffer filling B

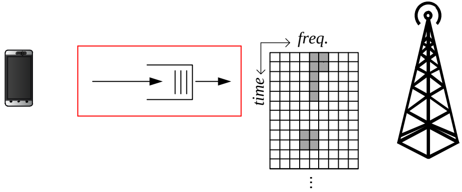
- Service S
- Scheduling epoch Δ

Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

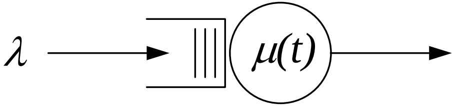
- Exact analysis for Poisson traffic
- Analytical framework for general arrival and service processes

Exact Analysis for Poisson Traffic



Exact Analysis for Poisson Traffic

TECHNISCHE UNIVERSITÄT DARMSTADT

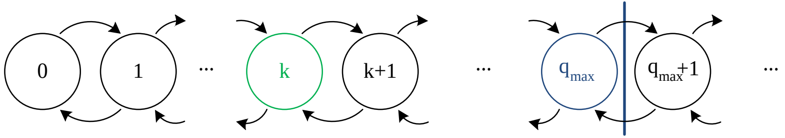


Model

- Single M/M/1 queue: fixed λ, variable $\mu(t)$
- Given λ and the queue length at epoch start
- Epoch based resource allocation

$\Rightarrow$ Find $\mu(t)$ that provides a probabilistic bound on the queue length at the end of the epoch

Exact Analysis for Poisson Traffic



1. Queue initially in state k
2. Fix $\mu(t)$ during Δ such that
3. Probability that the queue at time Δ is longer than $q_{\max}$ is less than ε

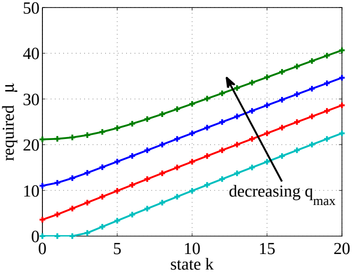
- Based on the transient behavior of the $M/M/1$ queue [Kleinrock].

Model parameters: λ, q_{max} , ε , Δ

Exact Analysis for Poisson Traffic

TECHNISCHE UNIVERSITÄT DARMSTADT

- Required μ for various parameters



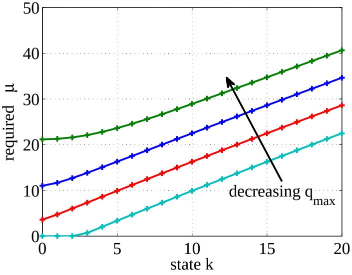
(a) $q_{\max} \in \{5, 10, 15, 20\}$

Parameters: $\lambda = 10, \varepsilon = 10^{-2}, \Delta = 1$ and 10^4 epochs for (b).

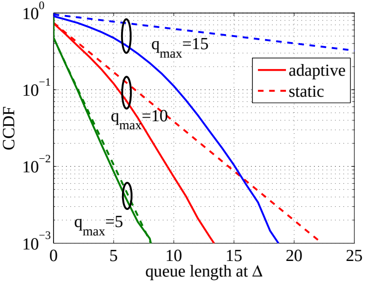
Exact Analysis for Poisson Traffic

TECHNISCHE UNIVERSITÄT DARMSTADT

- Required μ for various parameters
- Improvement w.r.t. the static system with equivalent $\bar{\mu}$



(a) $q_{\max} \in \{5, 10, 15, 20\}$



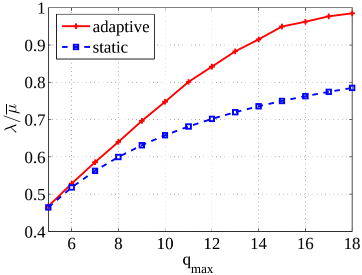
(b) Static system parameterized with $\bar{\mu}$.

Parameters: $\lambda = 10, \varepsilon = 10^{-2}, \Delta = 1$ and 10^4 epochs for (b).

Exact Analysis for Poisson Traffic

TECHNISCHE UNIVERSITÄT DARMSTADT

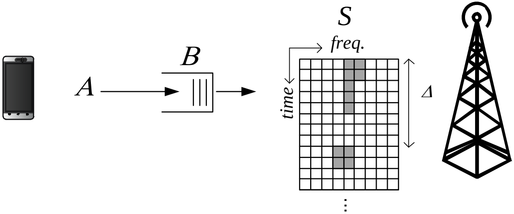
Utilization comparison:



- Key relation of $\lambda\Delta$ to $q_{\max}$ for a given ε
- Initial queue length is less helpful if the unknown traffic amount in during the epoch, i.e., $\lambda\Delta$, predominates $q_{\max}$
  $\rightarrow$ Operation of queue-aware scheduling is non-trivial
- Resource savings in the adaptive case $\rightarrow$ Proof of concept

Beyond the Poisson Model

Generalization w.r.t. service and arrival traffic models:

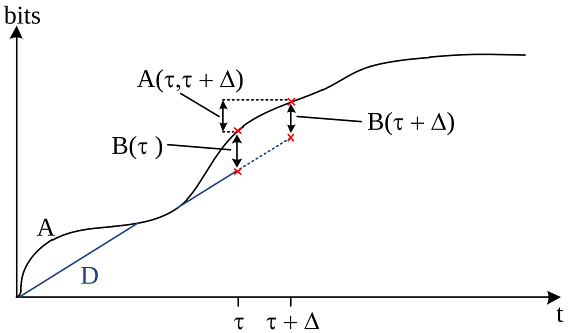


Framework: Stochastic Network Calculus

- Cumulative arrivals $A(\tau)$ resp. departures $D(\tau)$ up to time τ
- Backlog at τ: $B(\tau) = A(\tau) - D(\tau)$
- Service in $(\tau, t]$ as random process $S(\tau, t)$
- Assume strict service resp. adaptive service curve [Burchard et. al'06]

Queueing Model

TECHNISCHE UNIVERSITÄT DARMSTADT



- Evaluation requires a lower bound on the service process

$$\mathsf{P}\left[S(u,t) \geq \mathcal{S}(t-u),\ \forall u \in [\tau, t]\right] \geq 1 - \varepsilon_s$$

- To derive a lower bound on the departures $D(\tau + \Delta)$

Wireless Channel Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Basic block fading model for a wireless transmission [Fidler, Al-Zubaidy]

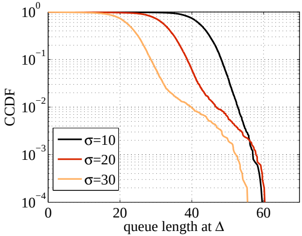
- Time slotted model with iid increments $c_i = \beta \ln(1 + \gamma_i)$
- Rayleigh fading channel: γ_i is exp distributed with parameter η
- Lower bounding function for the service process

$$\mathcal{S}(t) = \frac{1}{\theta}\Big(\ln(\varepsilon_p) - t\big[\eta + \theta\beta\ln(\eta) + \ln\left(\Gamma(1-\theta\beta, \eta)\right)\big]\Big)$$

with $\theta > 0$, incompl. Gamma fct. Γ and a violation probability $\varepsilon_s = (t - \tau)\varepsilon_p$.

Derivation using Boole's inequality, Chernoff's bound and the Laplace transform of the increments.

Infrequent Adaptation



- Bound on backlog at the end of epoch $\mathrm{P}\left[B(\tau+\Delta) \leq b_{\max}\right] \geq 1-\varepsilon_s$
- Requirements on $\mathcal{S}(t)$: Allocate resources β during epoch Δ such that the follwoing holds

$$\mathcal{S}(\Delta) \geq B(\tau) + A(\tau, \tau+\Delta) - b_{\max}, \quad \text{and} \tag{1}$$

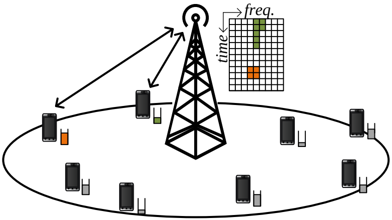
$$\mathcal{S}(\tau+\Delta-u) \geq A(u, \tau+\Delta) - b_{\max}, \quad \forall u \in [\tau, \tau+\Delta] \tag{2}$$

Multi-user Scheduling with Infrequent Adaptation

TECHNISCHE UNIVERSITÄT DARMSTADT

Scenario:

- M homogeneous, statistically independent MS channels
- Base station decides on amount of resource blocks β_j for MS $j \in [1, M]$ based on the infrequent adaptation technique
- Overall amount of resource blocks β_s in epoch Δ



Multi-user Scheduling with Infrequent Adaptation

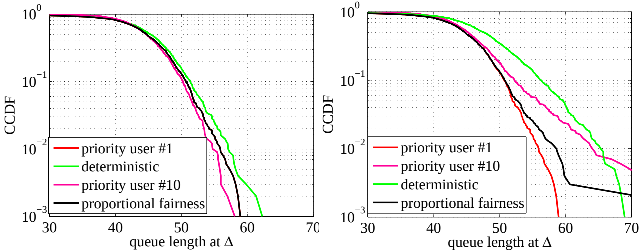
TECHNISCHE UNIVERSITÄT DARMSTADT

Scenario:

- M homogeneous, statistically independent MS
- Base station decides on amount of resource blocks β_j for MS $j \in [1, M]$ based on the infrequent adaptation technique
- Overall amount of resource blocks β_s in epoch Δ
- Three basic resource allocation algorithms:
  1. "deterministic fair": jth MS receives $\hat{\beta}_j = \min\{\beta_j, \beta_s/M\}$
  2. "priority": MS in class j receives $\hat{\beta}_j = \min\{\beta_j, \beta_s - \sum_{k=1}^{j-1} \hat{\beta}_k\}$
  3. "proportional fair emulation": Priority scheduler with priorities reordered every epoch Δ according to a score $S_j(\tau, \tau + \Delta)/(D_j(\tau)/\tau)$ similar to [Kelly, et al. '98].

Multi-user Scheduling with Infrequent Adaptation: Simulation

TECHNISCHE UNIVERSITÄT DARMSTADT



(b) 90% utilization (c) 95% utilization

- Adaptive system retains statistical backlog bound depending on scheduling algorithm
- Notable difference only at very high utilization

Parameters: $M = 10$, $\lambda = 0.65, \varepsilon_s = 10^{-2}, \Delta = 100$ slots, $b_{\max} = 65$, SNR $1/\eta = 3$dB

Key Takeaway Points

TECHNISCHE UNIVERSITÄT DARMSTADT

- Poisson case: Analytical results to quantify best-case resource savings.
- Model reveals important relation of $\lambda\Delta$ to $q_{\max}$.
- Analytical framework identifies two regimes, one where adaptive scheduling is effective and one where it is not.
- A mathematical treatment of queue-aware scheduling that is applicable to a broad class of arrival and service processes.

Related Models

1. Optimization of queueing service policies
2. Optimization of power and rate allocation in cellular systems

Difference to 1:

- online, epoch-based technique for general arrival and service processes

Difference to 2:

- dynamic programming to minimize a cost function of weighted power and rate consumption
- sample path as input