Delay-Optimal Scheduling in Bandwidth-Sharing Networks

Maaike Verloop
CWI

Rudesindo Núñez-Queija
CWI & Eindhoven University of Technology

Sem Borst
Bell Labs, CWI & Eindhoven University of Technology

CISS 2006
Princeton University, March 22-24, 2006

Background & motivation

Bandwidth-sharing networks

- Network consists of several resources (links) indexed by set $\mathcal{L}$

- Resources shared by several classes of users indexed by set $\mathcal{K}$

- Class-k users require simultaneous capacity from subset of resources $\mathcal{L}_k \subseteq \mathcal{L}$

Background & motivation (cont'd)

Find rate allocation $(r_k)_{k\in\mathcal{K}}$ that solves utility maximization problem (Kelly et al.):

$$\begin{aligned} \text{max} \quad & \sum_{k\in\mathcal{K}} N_k U_k(r_k) \\ \text{sub} \quad & \sum_{k\in\mathcal{K}_l} N_k r_k \le C_l \qquad l \in \mathcal{L} \\ & r_k \ge 0 \qquad k \in \mathcal{K} \end{aligned}$$

with

- $U_k(\cdot)$: **concave utility function**
- N_k: **number of class-k users**
- C_l: **capacity of resource l**
- $\mathcal{K}_l$: **classes that require capacity from resource l**

Background & motivation (cont'd)

α-fair rate allocation policies (Mo & Walrand):

$$U(r) = \frac{r^{1-\alpha}}{1-\alpha}$$

- $\alpha \downarrow 0$: maximum throughput
- '$\alpha = 1$': proportional fairness
- $\alpha = 2$: 'TCP'
- $\alpha \to \infty$: max-min fairness

Background & motivation (cont'd)

Dynamic setting (Bonald, Massoulié, Roberts)

- **Class-k users arrive as Poisson process of rate λ_k and have random service requirements with mean β_k**
- **Traffic load $\rho_k := \lambda_k \beta_k$**

(Minimization of) total transfer delay might matter more to users than (maximization of) instantaneous rate utility

Background & motivation (cont'd)

Finite expected transfer delay is first requirement

Stability condition: $\sum_{k \in \mathcal{K}_l} \rho_k < C_l$ **for all** $l \in \mathcal{L}$

- **Evidently necessary**

- **Sufficient in case of α-fair policies (Bonald & Massoulié, De Veciana, Lee & Konstantopoulos)**

How close to optimal are α-fair policies in terms of transfer delay and user-perceived throughput?

Background & motivation (cont'd)

Objectives

- **Identify non-anticipating scheduling discipline that minimizes mean transfer delay**

- **Compare with α-fair policies to assess potential scope for improvement**

Implementation issues need to be considered as well

Model description

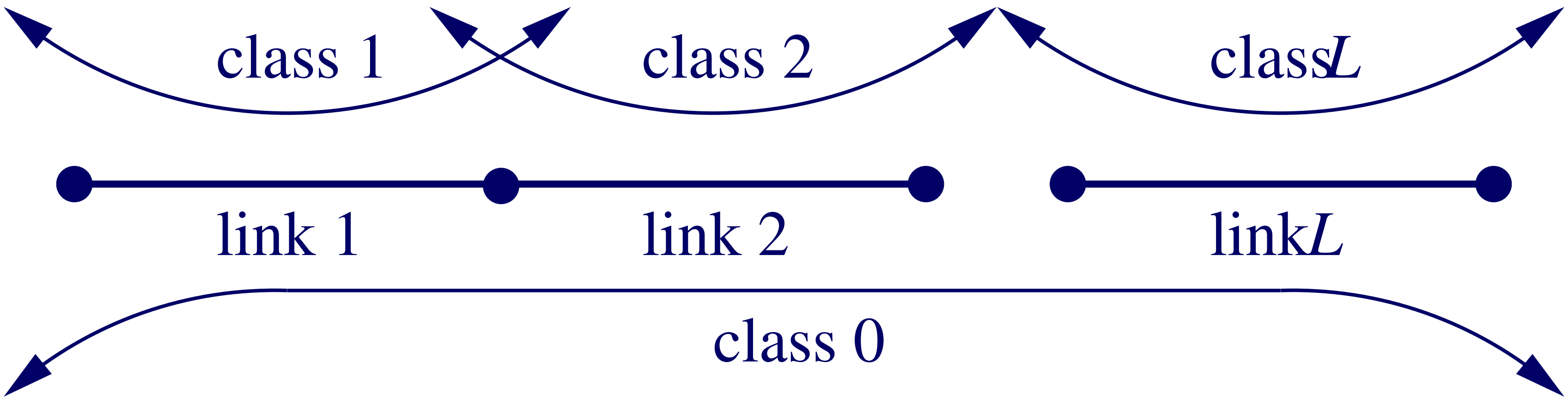
- **Linear network with L links, each of unit capacity**

- **Network is offered traffic from $L + 1$ flow classes**

- **Class-k flows arrive as Poisson process of rate λ_k and have exponentially distributed service requirements with mean $1/\mu_k$**

- **Traffic load $\rho_k := \lambda_k/\mu_k$**

Model description (cont'd)



- **Class-0 flows require capacity from all L links simultaneously**

- **Class-l flows require capacity from link l only, $l = 1, \ldots, L$**

Delay minimization

Suppose that $\mu_0 \le \mu_l$ **for all** $l = 1, \dots, L$

Consider policy π_0 **that gives preemptive priority to classes** $1, \dots, L$ **over class 0**

Denote by $s_k(t)$ **capacity allocated to class** k **at time** t

Policy π_0 **maximizes total instantaneous service completion rate** $\sum_{k=0}^{L} \mu_k s_k(t)$

Delay minimization (cont'd)

Stability condition for class 0:

$$\rho_0 < \mathbb{P}\{N_1 = 0, N_2 = 0, \ldots, N_L = 0\} = \prod_{l=1}^{L} \mathbb{P}\{N_l = 0\} = \prod_{l=1}^{L} (1 - \rho_l)$$

Above criterion more stringent than $\rho_0 + \rho_l < 1$ **for all** $l = 1, \ldots, L$

Policy π_0 **could unnecessarily cause instability, and hence cannot be optimal**

Reflects that policy π_0 **fails to be work-conserving**

Shortest Remaining Processing Time first (SRPT) and Least Attained Service first (LAS) strategies involve similar instability issues

Delay minimization (cont'd)

Two policies will play key role

- **Policy π^* gives preemptive priority to class 0 over classes $1, \ldots, L$**
- **Policy π^{**} serves all classes $1, \ldots, L$ simultaneously whenever possible, and gives preemptive priority to class 0 otherwise**

By construction, both policies utilize full capacity whenever possible: $s_0(t) + s_l(t) = 1$ whenever $N_0(t) + N_l(t) > 0$

Thus, both policies maximize work depletion rate and minimize workload at each link (work conservation)

Delay minimization (cont'd)

For certain values of μ_k, either policy π^* or π^{**} additionally maximizes total instantaneous service completion rate

In those cases, either policy π^* or π^{**} is optimal

In remaining cases, trade-off occurs between maximizing work depletion rate and maximizing service completion rate

Optimal policy then has more complicated structure

Delay minimization (cont'd)

Case I: $\mu_0 \geq \sum_{l=1}^{L} \mu_l$

Policy π^* **gives preemptive priority to class 0 over classes** $1, \ldots, L$

- **minimizes workload at each link**
- **maximizes total instantaneous service completion rate**

Policy π^* **stochastically minimizes total number of users** $\sum_{k=0}^{L} N_k(t)$ **at any point in time**

Delay minimization (cont'd)

Proof sketch:

Work conservation:

$$W_0^*(t) + W_l^*(t) \leq_{\text{st}} W_0^\pi(t) + W_l^\pi(t)$$

for any policy π

Priority to class-0 users:

$$N_0^*(t) \leq_{\text{st}} N_0^\pi(t)$$

for any policy π

Adding inequalities, taking expectations, noting $\mathbb{E}\{W_k(t)\} = \mathbb{E}\{N_k(t)\}/\mu_k$**, and invoking** $\mu_0 \geq \sum_{l=1}^{L} \mu_l$ **yields**

$$\sum_{k=0}^{L} \mathbb{E}\{N_k^*(t)\} \leq \sum_{k=0}^{L} \mathbb{E}\{N_k^\pi(t)\}$$

for any policy π

Delay calculation for policy π^*

$$\mathbb{E}\{N_0^*\} = \frac{\rho_0}{1-\rho_0}$$

$$\begin{aligned}
\mathbb{E}\{N_l^*\} &= \mu_l \mathbb{E}\{W_l^*\} \\
&= \mu_l(\mathbb{E}\{W_0\} + \mathbb{E}\{W_l^*\}) - \mathbb{E}\{W_0^*\}) \\
&= \mu_l \left(\frac{\lambda_0/\mu_0^2 + \lambda_l/\mu_l^2}{1-\rho_0-\rho_l} - \frac{\lambda_0/\mu_0^2}{1-\rho_0} \right) \\
&= \frac{\rho_l}{1-\rho_0} \frac{1-\rho_0(1-\mu_l/\mu_0)}{1-\rho_0-\rho_l}
\end{aligned}$$

$$\mathbb{E}\{N^*\} = \sum_{k=0}^{L} \mathbb{E}\{N_k^*\} = \frac{\rho_0}{1-\rho_0} + \sum_{l=1}^{L} \frac{\rho_l}{1-\rho_0} \frac{1-\rho_0(1-\mu_l/\mu_0)}{1-\rho_0-\rho_l}$$

Comparison with proportional fair allocation

$$\mathbb{E}\{N_l^{PF}\} = \frac{\rho_l}{1 - \rho_0 - \rho_l}$$

$$\mathbb{E}\{N_0^{PF}\} = \frac{\rho_0}{1 - \rho_0}\left(1 + \sum_{l=1}^{L} \frac{\rho_l}{1 - \rho_0 - \rho_l}\right)$$

$$\mathbb{E}\{N^{PF}\} = \sum_{k=0}^{L} \mathbb{E}\{N_k^{PF}\} = \frac{\rho_0}{1 - \rho_0} + \sum_{l=1}^{L} \frac{\rho_l}{1 - \rho_0 - \rho_l}$$

$$\Delta = \mathbb{E}\{N^{PF}\} - \mathbb{E}\{N^*\} = \frac{\rho_0}{1 - \rho_0} \sum_{l=1}^{L} \frac{\rho_l}{1 - \rho_0 - \rho_l}(1 - \frac{\mu_l}{\mu_0})$$

Note that $\Delta \uparrow \frac{\rho_0}{1-\rho_0} \sum_{l=1}^{L} \frac{\rho_l}{1-\rho_0-\rho_l} = \mathbb{E}\{N_0^{PF}\} - \mathbb{E}\{N_0^*\}$ **as** $\mu_0 \to \infty$

Delay minimization (cont'd)

Case II: $\sum_{l=1}^{L} \mu_l - \mu_{l^*} \leq \mu_0 \leq \sum_{l=1}^{L} \mu_l$, **with** $l^* := \arg\min_{l=1,\ldots,L} \mu_l$

Policy π^{**} **serves all classes** $1, \ldots, L$ **simultaneously whenever possible, and gives preemptive priority to class 0 otherwise**

- **minimizes workload at each link**

- **maximizes total instantaneous service completion rate**

Policy π^{**} **stochastically minimizes mean total number of users** $\sum_{k=0}^{L} \mathbb{E}\{N_k(t)\}$ **at any point in time**

Delay minimization (cont'd)

Proof sketch:

Work conservation:

$$W_0^{**}(t) + W_l^{**}(t) \leq_{\text{st}} W_0^{\pi}(t) + W_l^{\pi}(t)$$

for any policy π

In addition, at every point in time, there are some $i \neq j$ **such that**

$$W_0^{**}(t) + W_i^{**}(t) + W_j^{**}(t) \leq_{\text{st}} W_0^{\pi}(t) + W_i^{\pi}(t) + W_j^{\pi}(t)$$

for any policy π

Adding inequalities, taking expectations, noting that $\mathbb{E}\{W_k(t)\} = \mathbb{E}\{N_k(t)\}/\mu_k$, **and invoking** $\sum_{l=1}^{L} \mu_l - \mu_{l^*} \leq \mu_0 \leq \sum_{l=1}^{L} \mu_l$ **yields**

$$\sum_{k=0}^{L} \mathbb{E}\{N_k^{**}(t)\} \leq \sum_{k=0}^{L} \mathbb{E}\{N_k^{\pi}(t)\}$$

for any policy π

Delay minimization (cont'd)

Remaining cases: $\mu_0 \leq \sum_{l=1}^{L} \mu_l - \mu_{l^*}$

Suppose that $N_{l^*}(t) = 0$ **and** $N_l(t) > 0$ **for all** $l \neq l^*$

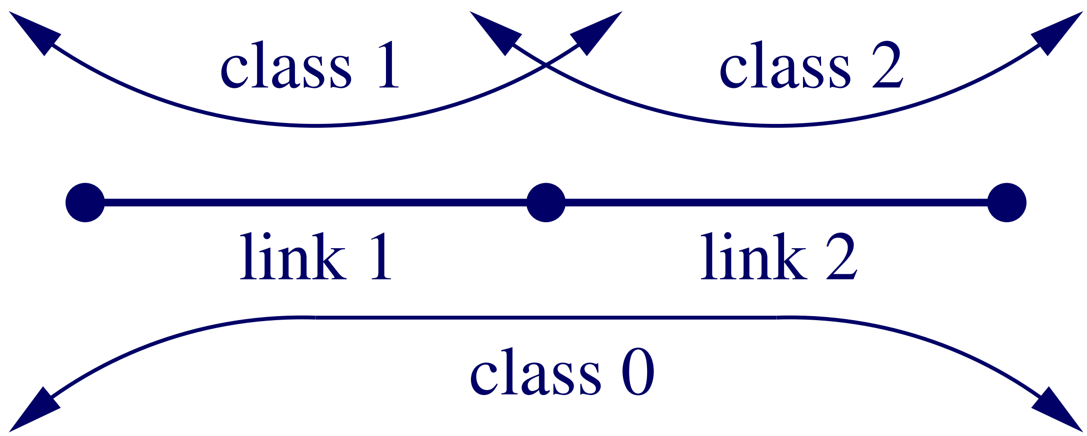
Trade-off arises

- **Serving class 0 maximizes total work depletion rate**
- **Serving all other classes (except** l^***) maximizes total instantaneous service completion rate**

Optimal policy has complicated structure

Delay optimization (cont'd)

Consider network with $L = 2$ links



- $\mu_0 \geq \mu_1 + \mu_2$: **policy π^* is optimal**
- $\mu_1, \mu_2 \leq \mu_0 \leq \mu_1 + \mu_2$: **policy π^{**} is optimal (with $\mu_0 = \mu_1 = \mu_2$ as important special case)**
- $\mu_0 \leq \mu_1$: **serve classes 1 and 2 when possible; when $N_2(t) = 0$, serve class 0 or 1 (switching curve)**

Switching curve

Switching curve may be approximated using fluid analysis

$$\min \quad \int_{t=0}^{\infty} (x_0(t) + x_1(t) + x_2(t))\mathrm{d}t$$

$$\text{sub} \quad \frac{\mathrm{d}x_k(t)}{\mathrm{d}t} = \lambda_k - \mu_k s_k(t) \qquad k = 0,1,2$$

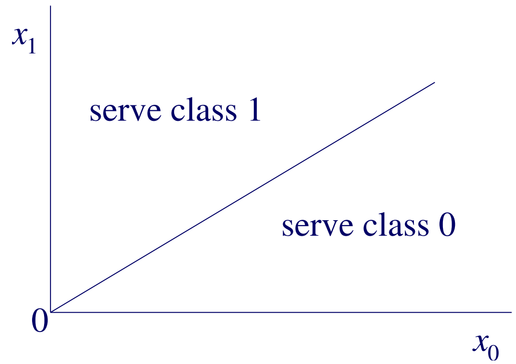
$$s_0(t) + s_l(t) \leq 1 \qquad l = 1,2$$

$$x_k(t) \geq 0 \qquad k = 0,1,2$$

$$x_k(0) = y_k \qquad k = 0,1,2$$

Switching curve (cont'd)

Assume $\mu_0 \le \mu_1$, $\rho_1 \le \rho_2$



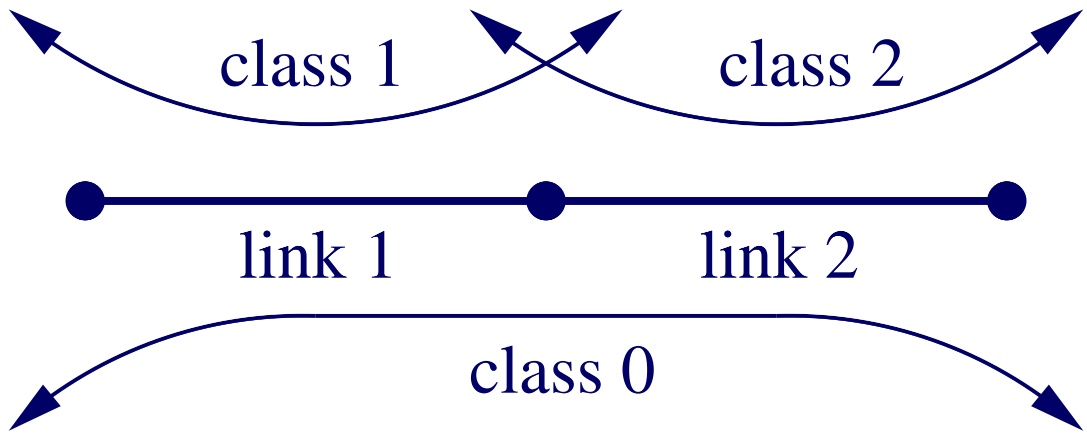
When $x_2(t) = 0$ optimal actions characterized by linear switching curve

- $\mu_0 \le \mu_1, \mu_2$: $x_1 = \frac{\rho_2 - \rho_1}{1 - \rho_0 - \rho_2} x_0$

- $\mu_2 \le \mu_0 \le \mu_1$: $x_1 = \frac{\mu_1}{\mu_1 + \mu_2 - \mu_0} \frac{\mu_2}{\mu_0} \frac{\rho_2 - \rho_1}{1 - \rho_0 - \rho_2} x_0$

Numerical experiments

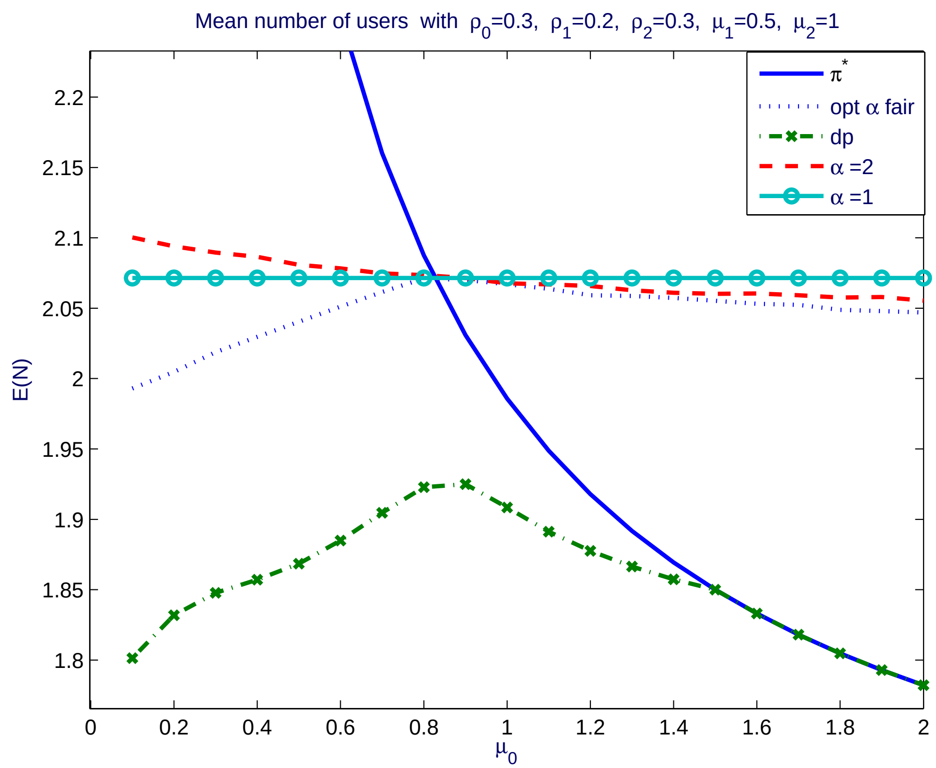
Consider network with $L = 2$ links



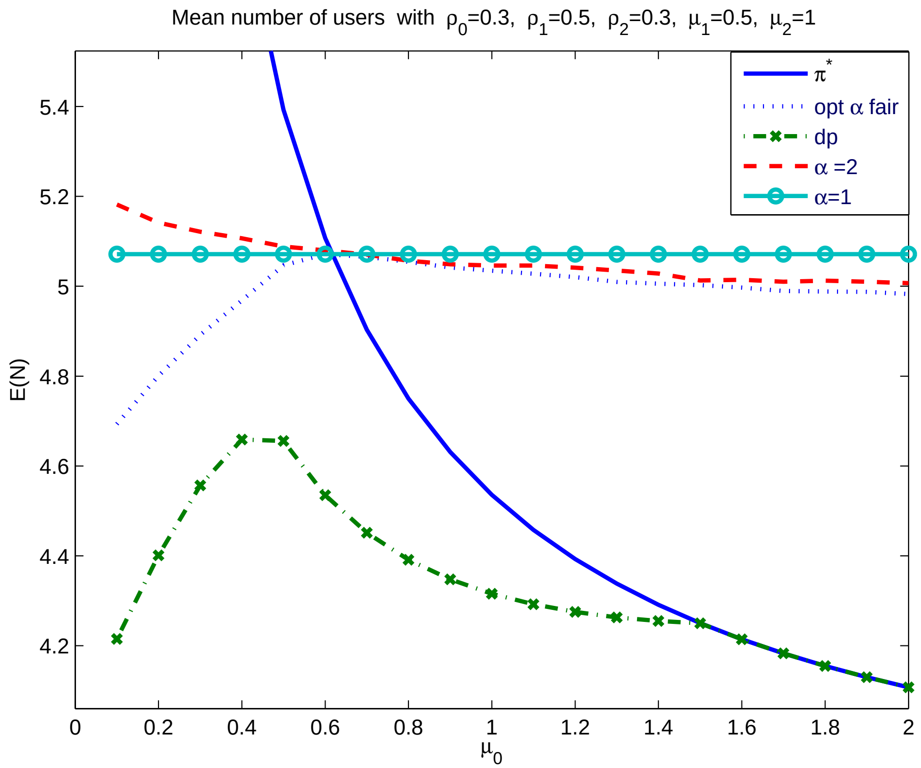
Comparison of

- Optimal policy (using dynamic programming)
- Heuristic policy (based on fluid analysis)
- Optimal α-fair policy

Numerical experiments (cont'd)



Numerical experiments (cont'd)



Conclusions

Optimal policy achieves modest performance improvement over α-fair policies

Performance of α-fair policies fairly insensitive to value of α, provided α is not too small

Heuristic policy based on fluid analysis provides near-optimal performance

Disclaimers

Only considered scheduling across classes with different routes

Standard size-based scheduling across classes with different routes can cause instability

Size-based scheduling within routes may however still produce gains

Preliminary results indicate that these gains can be substantial and arbitrarily large in heavy-traffic conditions