Capacity Provisioning for Schedulers with Tiny Buffers

Yashar Ghiassi
yashar@comm.utoronto.ca

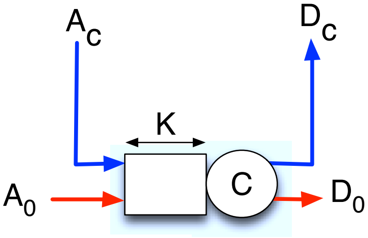
Department of ECE
University of Toronto

Joint work with:
J. Liebeherr

April 18, 2013

University of Waterloo

Resource Provisioning for Link Schedulers



- Input traffic: Through flows A_0 and cross flows A_c
- Output traffic: Through flows D_0 and cross flows D_c
- Link capacity C, buffer size K
- Backlog $b_0(t)$ and delay $d_0(t)$ of the through flows at time t

Size C and K such that:

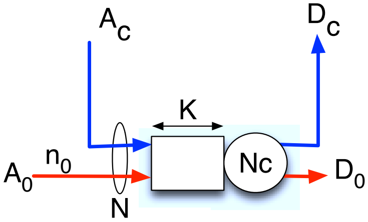
$P\{b_0(t) > K\} \leq \varepsilon^*$ and/or $P\{d_0(t) > \bar{d}\} \leq \varepsilon^*$, where $\bar{d}$ is the delay bound

Towards Small Buffers

There are arguments in favour of small buffers:

- Small buffers enable fast memory technologies (e.g., SRAM).
  (Enachescu et al.' 05)
- Small buffers might even mitigate traffic burstiness.
  (Likhanov and Mazumdar' 98), (Mao and Panwar' 01)
- In case of many sources, adding small buffers satisfies loss probability.
  (Mao and Panwar' 01)

Asymptotic Observations



Define

- c: per-flow capacity
- $\bar{a}$: per-flow average rate $:= \lim_{t \to \infty} \frac{1}{N} \frac{A_0(t)+A_c(t)}{t}$

Given: a loss probability constraint (using large deviation techniques)

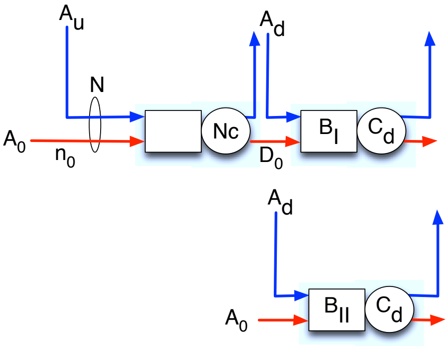
For *any* work-conserving scheduling $\lim_{N \to \infty} c \to \bar{a}$. *(Eun and Shroff' 05)*

The results hold for small buffers (i.e., O(1))$\Rightarrow$ network decomposition

Network Decomposition in an Asymptotic Regime

- Convergence of D_0 to A_0:
  (Wischik' 99), (Ying et al.' 94)

- Convergence of B_{I} to B_{II}:
  (Eun and Shroff' 05), (Ciucu and Hohlfled' 09), (Ciucu and Liebeherr' 09)



Does Link Scheduling Matter if N is Finite?

Some existing non-asymptotic results for schedulers:

- $D_0 \to A_0$ for FIFO scheduling even when N is few hundreds under some statistical independence assumptions. *(Ciucu and Liebeherr' 09)*
- A non-asymptotic capacity size is computed for a given per-flow delay bound constraint in a FIFO scheduler. It scales by $c = O(\frac{1}{N})$. *(Ciucu and Hohlfled' 09)*

Open question:

How does link scheduling impact capacity requirement and decomposition for finite N?

Contributions

We show that for finite N, the choice of link scheduling has a big impact on

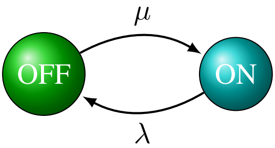
- Buffer overflow probability
- Capacity provisioning
- Viability of network decomposition

In particular

$c - \bar{a}$ ranges from $O\left(\sqrt{\frac{\log N}{N}}\right)$ to $O(\frac{1}{N})$ depending on the scheduling algorithm.

Traffic Source (MMOO)

Markov-modulated On-Off (MMOO) source:



- P Kbps in ON state, idle in OFF state
- Average time to return to the same state: $T^* = \frac{\lambda+\mu}{\lambda\mu}$
- The larger the T^*, the more bursty the traffic

Exponentially Bounded Burstiness

Exponentially Bounded Burstiness (EBB) sources (Yaron, Sidi'93)

An arrival process A is EBB with parameters (M, ρ, α) if for any $s \leq t$

$$P(A(s,t) > \rho(t-s) + \sigma) \leq Me^{-\alpha\sigma} := \varepsilon(\sigma)\ .$$

We write it by $A \sim (M, \rho, \alpha)$.

Suppose: A is the aggregate of n iid MMOO flows with parameters λ, μ, and P.

Then, $A \sim (1, nr(\alpha), \alpha)$ for any $\alpha \geq 0$, with

$$r(\alpha) = \frac{1}{2\alpha}(P\alpha - \lambda - \mu + \sqrt{(P\alpha - \mu + \lambda)^2 + 4\mu\lambda}\ .$$

We use this flexibility (a family of EBB characterizations) to get new insights.

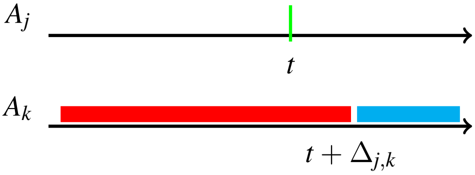
Δ-Schedulers

A scheduler whose operation is entirely determined by a matrix of constants $(\Delta_{j,k})_{j,k \in \mathcal{N}}$.

- The followings are Δ-schedulers:
  - FIFO: $\Delta_{j,k} = 0$
  - SP, BMux: $\Delta_{j,k} = \begin{cases} -\infty & \text{if flow j has higher priority} \\ +\infty & \text{if flow k has higher priority} \end{cases}$
  - EDF: $\Delta_{j,k} = d_j^* - d_k^*$
- GPS is not a Δ-scheduler.

Δ-Schedulers

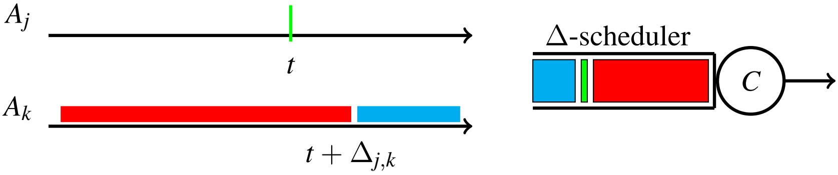
A scheduler whose operation is entirely determined by a matrix of constants $(\Delta_{j,k})_{j,k \in \mathcal{N}}$.



- The followings are Δ-schedulers:
  - FIFO: $\Delta_{j,k} = 0$
  - SP, BMux: $\Delta_{j,k} = \begin{cases} -\infty & \text{if flow j has higher priority} \\ +\infty & \text{if flow k has higher priority} \end{cases}$
  - EDF: $\Delta_{j,k} = d_j^* - d_k^*$
- GPS is not a Δ-scheduler.

Δ-Schedulers

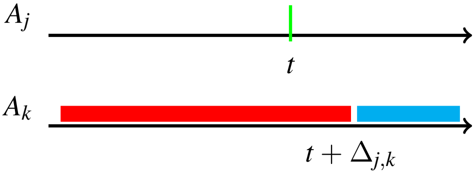
A scheduler whose operation is entirely determined by a matrix of constants $(\Delta_{j,k})_{j,k\in\mathcal{N}}$.

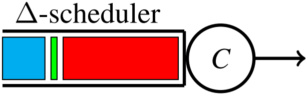


- The followings are Δ-schedulers:
  - FIFO: $\Delta_{j,k} = 0$
  - SP, BMux: $\Delta_{j,k} = \begin{cases} -\infty & \text{if flow j has higher priority} \\ +\infty & \text{if flow k has higher priority} \end{cases}$
  - EDF: $\Delta_{j,k} = d_j^* - d_k^*$
- GPS is not a Δ-scheduler.

Δ-Schedulers

A scheduler whose operation is entirely determined by a matrix of constants $(\Delta_{j,k})_{j,k \in \mathcal{N}}$.





- The followings are Δ-schedulers:
  - FIFO: $\Delta_{j,k} = 0$
  - SP, BMux: $\Delta_{j,k} = \begin{cases} -\infty & \text{if flow j has higher priority} \\ +\infty & \text{if flow k has higher priority} \end{cases}$
  - EDF: $\Delta_{j,k} = d_j^* - d_k^*$
- GPS is not a Δ-scheduler.

A Backlog Bound for EBB flows in Δ-Schedulers

A backlog bound for Δ-schedulers [Ghiassi, Liebeherr, Burchard' 11]

- $A_0 \sim (M_0, \rho_0, \alpha_0)$ and $A_c \sim (M_c, \rho_c, \alpha_c)$.
- $\Delta_{0,c} = \Delta$ and capacity C.

For any $\sigma_0, \sigma_c \geq 0$ and $0 \leq \gamma \leq \frac{C-\rho_c-\rho_0}{2}$

$$\theta^* = \min\left(\frac{\sigma_c}{C-\rho_c-\gamma}, \frac{[\sigma_c + (\rho_c+\gamma)\Delta]_+}{C}\right)$$

$$b(\sigma_0, \sigma_c) = \sigma_0 + (\rho_0 + \gamma)\theta^*$$

$$\varepsilon(\sigma_0, \sigma_c) = M_0 e\left(1 + \frac{\rho_0}{\gamma}\right)e^{-\alpha_0\sigma_0} + M_c e\left(1 + \frac{\rho_c}{\gamma}\right)e^{-\alpha_c\sigma_c} .$$

Then,

$$\Pr\{B_0(t) > b(\sigma_0, \sigma_c)\} \leq \varepsilon(\sigma_0, \sigma_c) .$$

Capacity Sizing of a Δ-scheduler

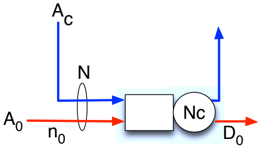
Corollary (Per-flow capacity scaling properties)

The per-flow capacity of a Δ-scheduler with a fixed (arbitrary small) buffer size, a target loss probability, and MMOO input flows satisfies

$$c - \bar{a} = \begin{cases} O\left(\sqrt{\frac{\log N}{N}}\right) & \Delta \geq 0 \\ O\left(\frac{1}{N}\right) & \Delta < 0 \end{cases}$$

$\lim_{N\to\infty} c \to \bar{a}$ for all work-conserving schedulers.
The speed of convergence is highly affected by the scheduling algorithm.

Network Decomposition ($D_0 \to A_0$)



Output EBB characterization

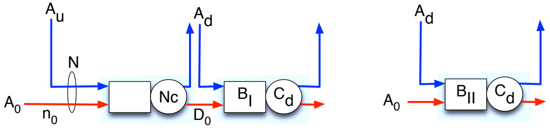
Given: $A_0 \sim (1, \rho_0, \alpha_0)$ and A_c are MMOO input flows to a Δ-scheduler. Then, $D_0 \sim (M_0^{out}, \rho_0, \alpha_0^{out})$, with

$$\alpha_0^{out} = \alpha_0 - O(\frac{1}{N}); \qquad M_0^{out} = \begin{cases} L(N)N^{\frac{1}{N}} & \Delta \geq 0 \\ L(N)\left(Ne^{-N\beta}\right)^{\frac{1}{N}} & \Delta < 0 \end{cases}.$$

where $\lim_{N\to\infty} L(N) = 1$.

- $D_0 \to A_0$ as $N \to \infty$ for any work-conserving schedulers.
- The speed of convergence is substantially affected by the schedulers.

Network Decomposition ($B_I \rightarrow B_{II}$)



Theorem (a.s. convergence of B_{I} to B_{II})

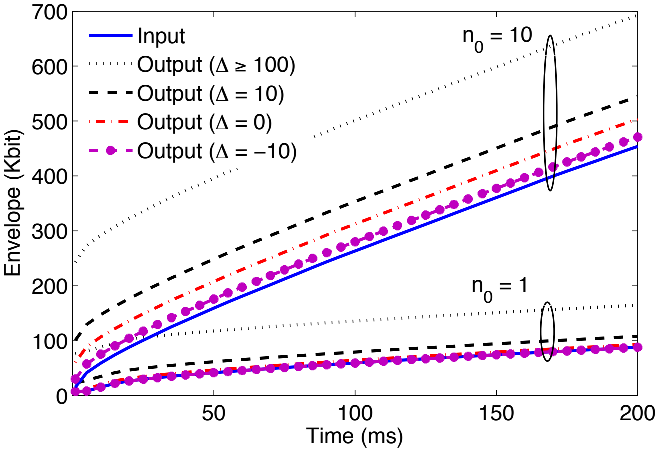
For MMOO traffic sources and Δ-schedulers, there exists a constant $\alpha > 0$ and a non-negative function L such that for any $\sigma \geq 0$

$$\Pr\{|B_I(t) - B_{II}(t)| > \sigma\} = \begin{cases} O(N^2)e^{-N\alpha\sigma} & \Delta \geq 0 \\ O(N^2 e^{-N\beta})e^{-N\alpha\sigma} & \Delta < 0 \end{cases}$$

$\lim_{N\rightarrow\infty} B_I \rightarrow B_{II}$ for all work-conserving schedulers.

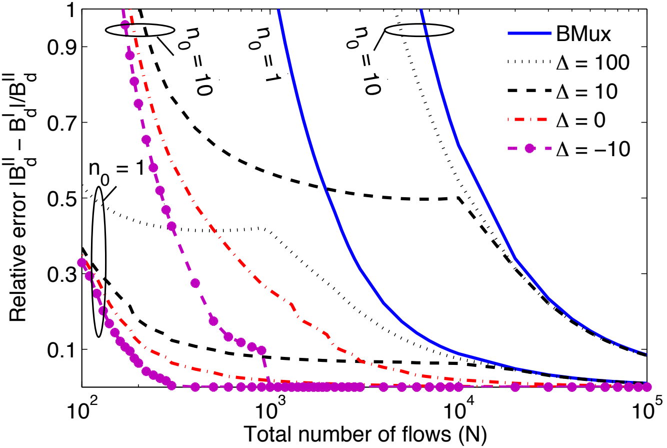
The speed of convergence is highly affected by the scheduling algorithm.

Example 1: Network Decomposition ($D_0 \to A_0$)



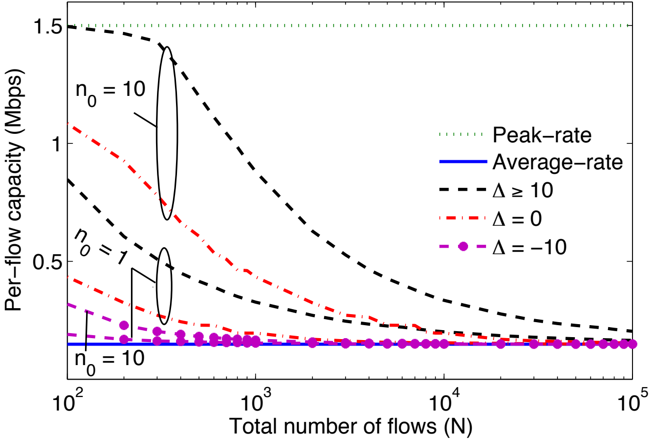
- $n_0 = 1, 10$, $C = 100$ Mbps, $U = 90\%$, and $\varepsilon^* = 10^{-6}$
- MMOO iid flows each with $P = 1.5$ Kbits and $T^* = 10$ ms

Example 2: Network Decomposition ($B_{II} \to B_I$)



- $n_0 = 1, 10$, $C = 100$ Mbps, $U = 90\%$, and $\varepsilon^* = 10^{-6}$
- MMOO iid flows each with $P = 1.5$ Kbits and $T^* = 10$ ms

Example 3: Capacity Provisioning



- $n_0 = 1, 10$, $b_0 = 1.5$ Kbits, $U = 90\%$, and $\varepsilon^* = 10^{-6}$
- MMOO iid flows each with $P = 1.5$ Kbits and $T^* = 10$ ms

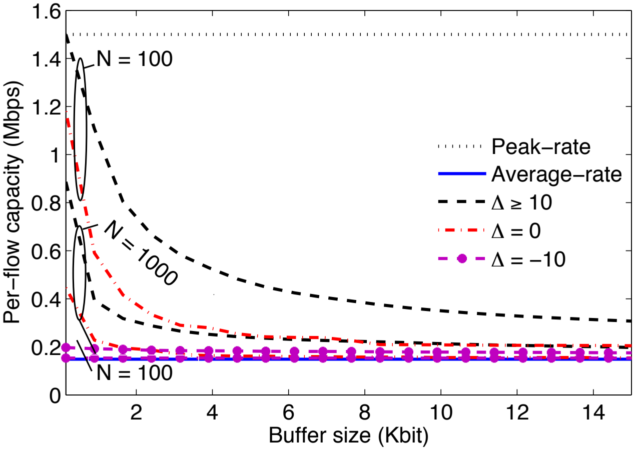
Conclusions

- $c - \bar{a}$ ranges from $O\left(\sqrt{\frac{\log N}{N}}\right)$ to $O(\frac{1}{N})$ depending on the scheduling algorithm.
- Capacity provisioning is highly affected by the scheduling algorithm.
- Network decomposition is valid for some schedulers even for moderate values of N (e.g., few hundreds).

Thank You

Questions?

Example 4: Capacity Provisioning



- $n_0 = 1$, $U = 90\%$, and $\varepsilon^* = 10^{-6}$
- MMOO iid flows each with $P = 1.5$ Kbits and $T^* = 10$ ms