Fairness in MIMD Congestion Control Algorithm

E. Altman, K. Avrachenkov, and B. J. Prabhu

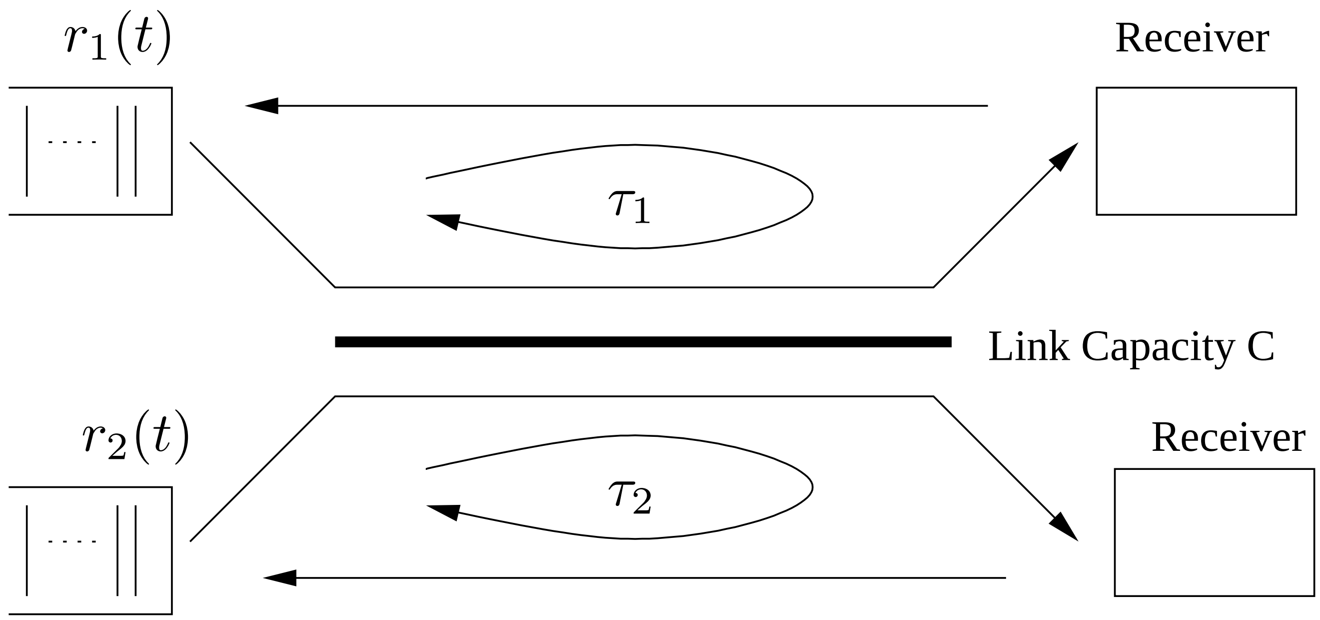
Projet MAESTRO

INRIA Sophia-Antipolis, France.

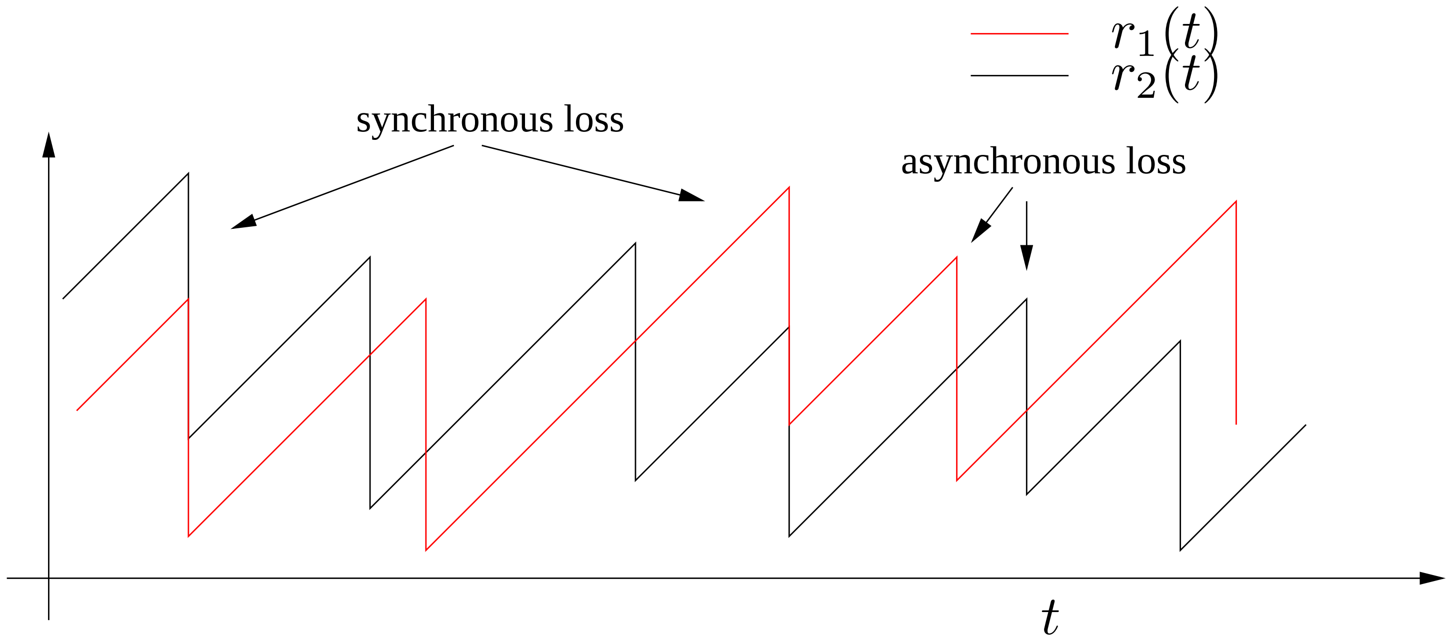
`http://www-sop.inria.fr/maestro`

	Standard TCP(AIMD)	Scalable TCP(MIMD)
no losses	$W_{n+1} \leftarrow W_n + 1$ $\Rightarrow$ linear increase $\frac{dW}{dt} = \frac{1}{\tau}$ $\Rightarrow$ linear growth	$W_{n+1} \leftarrow \alpha \times W_n$ $\Rightarrow$ multiplicative increase $\frac{dW}{dt} = \frac{\log[\alpha]}{\tau} W$ $\Rightarrow$ exponential growth
≥ 1 loss	$W_{n+1} \leftarrow 0.5 \times W_n$ $\Rightarrow$ multiplicative decrease	$W_{n+1} \leftarrow \beta \times W_n$ $\Rightarrow$ multiplicative decrease

- How two sources share a link of capacity C?



- Losses occur when the capacity is reached but could also occur before



- In the absence of control signals (i.e., losses), rate increases as

$$\mathsf{r_i}(t+\delta) = \alpha^{\delta/\tau_i} \cdot \mathsf{r_i}(t), \quad i = 1, 2.$$

- Reaction to control signals

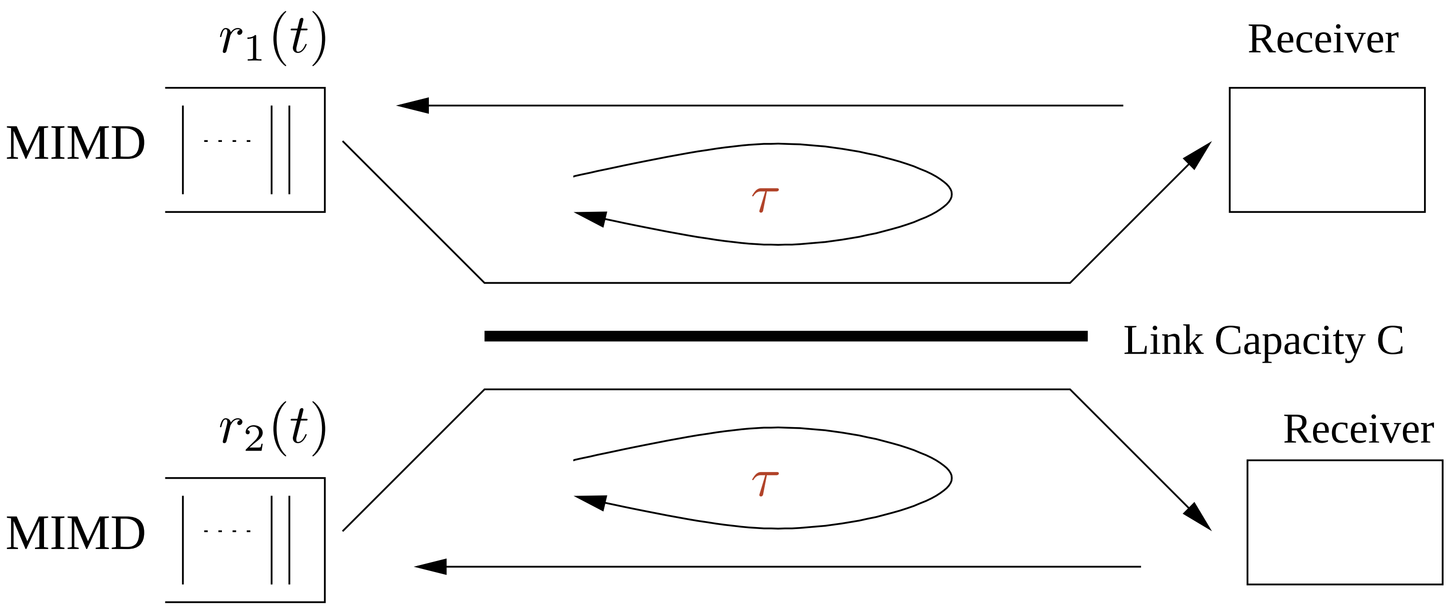
control vector	$\mathsf{r}_1(t_j+)$	$\mathsf{r}_2(t_j+)$
$(0,0)$	$\mathsf{r}_1(t_j)$	$\mathsf{r}_2(t_j)$
$(0,1)$	$\mathsf{r}_1(t_j)$	$\beta \cdot \mathsf{r}_2(t_j)$
$(1,0)$	$\beta \cdot \mathsf{r}_1(t_j)$	$\mathsf{r}_2(t_j)$
$(1,1)$	$\beta \cdot \mathsf{r}_1(t_j)$	$\beta \cdot \mathsf{r}_2(t_j)$

Asynchronous loss (rows $(0,1)$ and $(1,0)$)

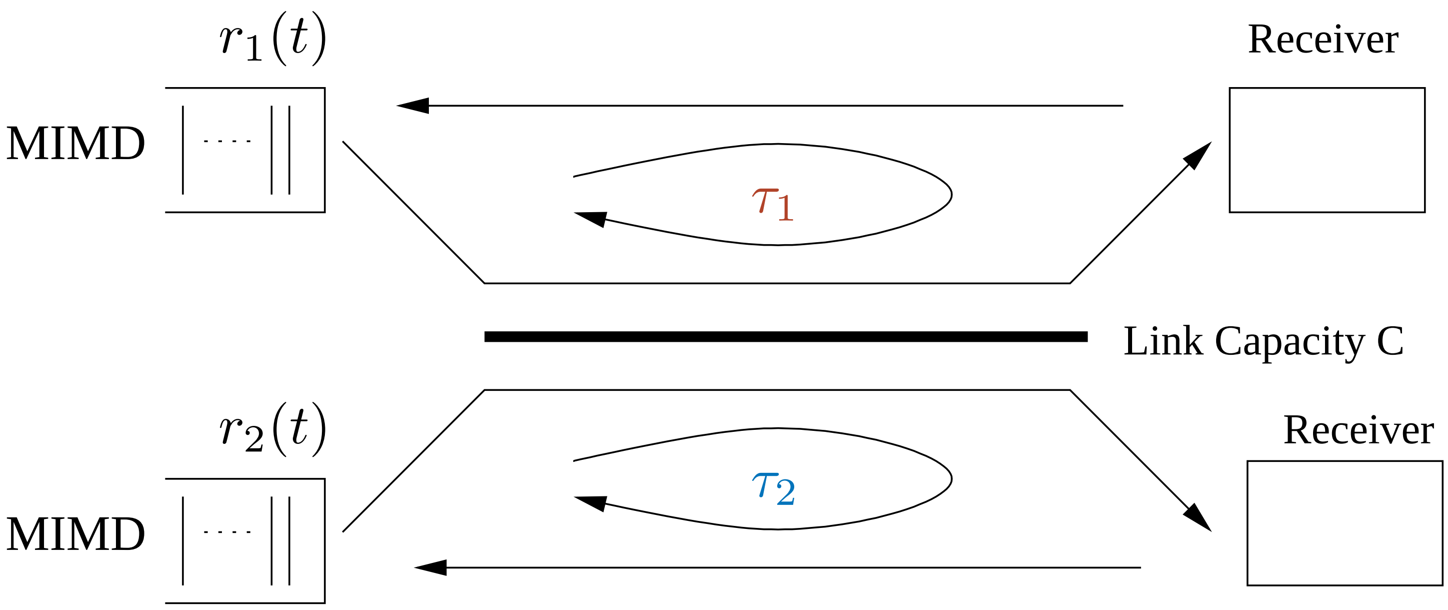
Synchronous loss (row $(1,1)$)

Outline

- Intra protocol (MIMD) fairness: same RTT

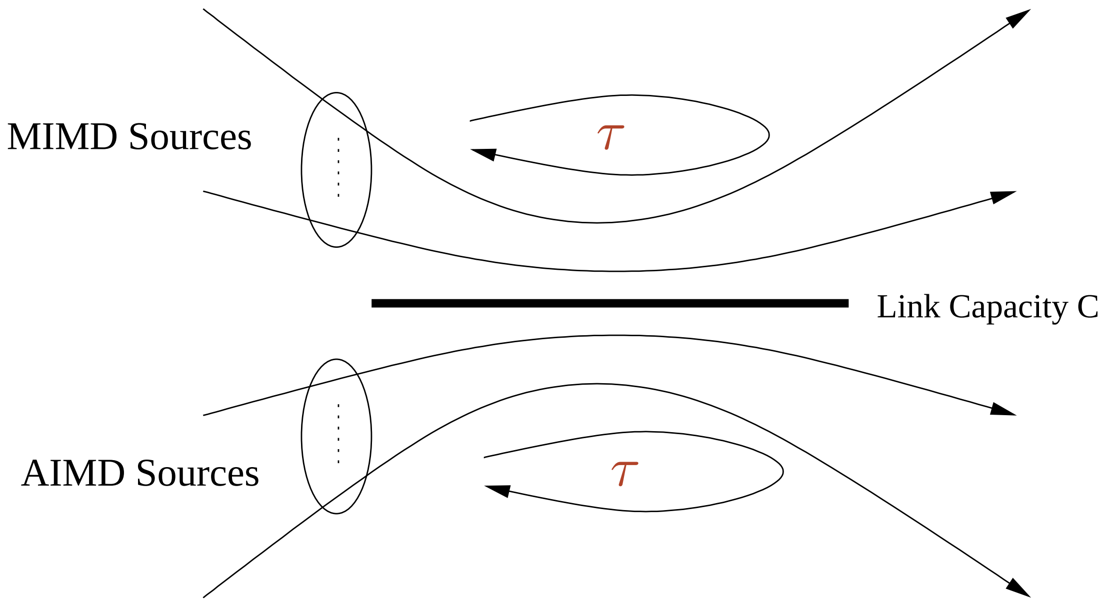


- Intra protocol (MIMD) fairness: different RTTs



Outline

- Inter protocol (MIMD & AIMD) fairness: same RTT



Outline

- Intra protocol (MIMD) fairness: same RTT
- Intra protocol (MIMD) fairness: different RTTs
- Inter protocol (MIMD & AIMD) fairness: same RTT

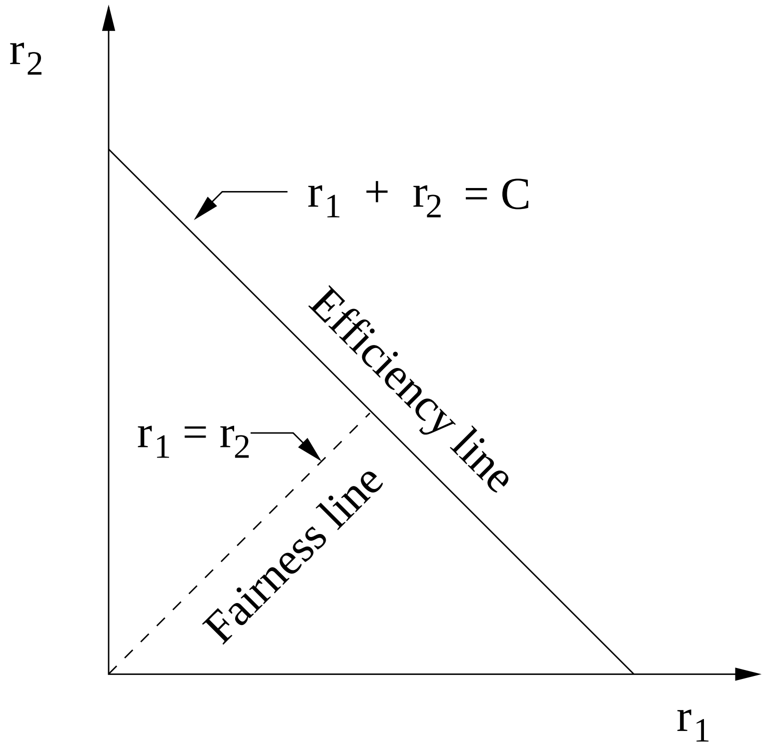
Intra-protocol Fairness: Same RTT

- Chiu and Jain (1989) studied the long term fairness index

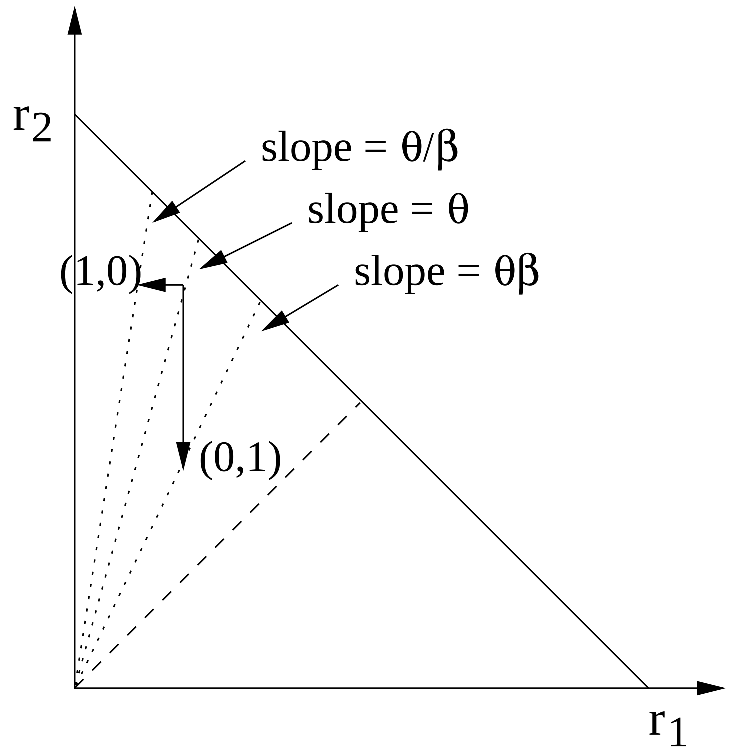
$$F_\infty = \lim_{t\to\infty} \frac{1}{2} \frac{(\mathsf{r}_1(t) + \mathsf{r}_2(t))^2}{\mathsf{r}_1(t)^2 + \mathsf{r}_2(t)^2}.$$

- They showed that, in the presence of only synchronous losses, MIMD algorithm can be extremely unfair, i.e., F_∞ strongly depends on the initial rate allocation vector, $(\mathsf{r}_1(0), \mathsf{r}_2(0))$.
- Gorinsky (2003) argued that MIMD can be fair in the presence of rate dependent (i.e., asynchronous) losses
- In this work, we
  - give a stochastic model for the rate allocation vector.
  - show that rate dependent losses indeed improve fairness by removing the dependence on initial state.

- Let $\theta(t) = \frac{r_2(t)}{r_1(t)}$ be the instantaneous throughput ratio process.



(a) Rate vector, r



(b) Evolution of $\theta(t)$

- Let θ_n denote $\theta(t)$ embedded before at the n^{th} control instant.
- For a given θ_0 there exists a $\lambda \in (\beta^{-1/2}, \beta^{1/2}]$ such that $\theta_0 = \lambda\beta^j$ for some $j \in \mathbb{Z}$.

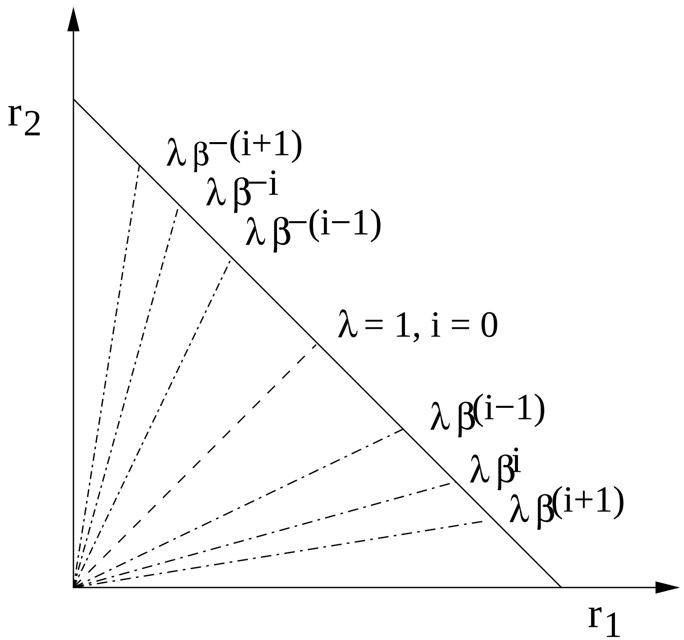


Figure 1: State space, $\mathbb{S}$, of θ.

State space of θ_n is

$$\mathbb{S} = \{\lambda\beta^i, \forall i \in \mathbb{Z}\}$$

- The process, $\{s_n = i\} \Rightarrow \{\theta_n = \lambda\beta^i\}$. $\{s_n, n \geq 0\}$ is modelled as a discrete-time Markov chain with state space $\mathbb{Z}$.
- Probability of asynchronous loss is ϵ.

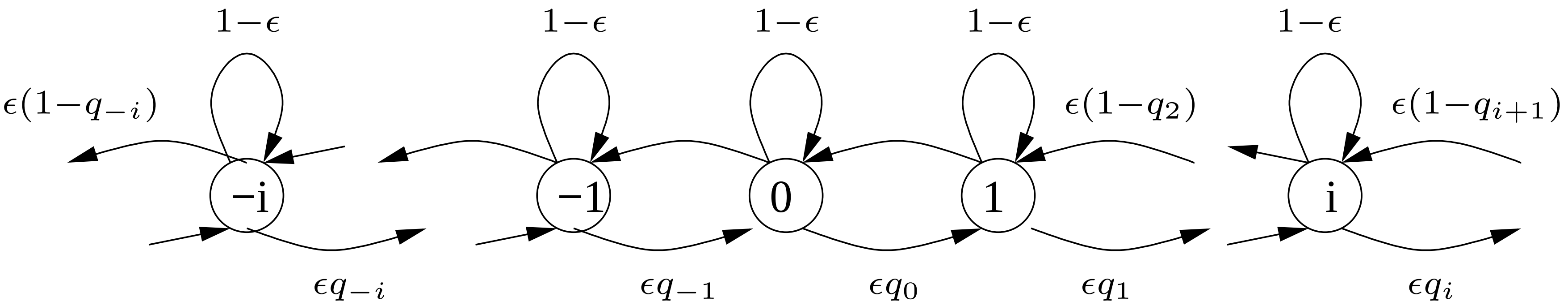


Figure 2: Markov chain of s_n.

- Transition probabilities can be state independent or state dependent.

- F_∞ defined as

$$\lim_{n\to\infty} F_n = \lim_{n\to\infty} \frac{1}{2}\frac{(r_1(n)+r_2(n))^2}{r_1(n)^2+r_2(n)^2}$$

$$F_\infty = \frac{1}{2} + \sum_{i=-\infty}^{\infty} P(s_\infty = i)\frac{\beta^i}{1+\beta^{2i}}.$$

- Other characterstics : Mean first passage time to "Fairness line", rate of convergence.

- Each source gets a decrease signal with equal probability.

$$s_{n+1}|(s_n = i) = \begin{cases} i-1 & \text{w.p.} \quad \frac{\epsilon}{2} \\ i+1 & \text{w.p.} \quad \frac{\epsilon}{2}, \\ i & \text{w.p.} \quad 1-\epsilon \end{cases} \qquad \forall i$$

- s_n is null-recurrent.
- Two accumulation points $(C, 0)$ and $(0, C)$.
- Asynchronous rate independent losses are insufficient to improve fairness.

- Source with a larger rate has a higher probability to get a decrease signal.
- Probability of getting a decrease signal $\propto \frac{r_i}{(r_1+r_2)}$

$$s_{n+1}|(s_n = i) = \begin{cases} i+1 & \text{w.p.} \quad \epsilon\frac{\beta^i}{1+\beta^i} \\ i-1 & \text{w.p.} \quad \epsilon\frac{1}{1+\beta^i} \\ i & \text{w.p.} \quad 1-\epsilon \end{cases}.$$

- s_n is positive recurrent for all $\beta \in [0, 1)$.
- Asynchronous rate dependent losses improve fairness by removing dependence on initial state.

Observations

- Steady state probability $p_i \overset{\text{def}}{=} P(s = i)$ is independent of ϵ.
- $p_i \propto \beta^{i(i-1)/2}(1 + \beta^i) \Rightarrow$ tail decreases rapidly.
- Fairness index

β	0.95	0.875	0.75	0.6	0.5	0.1
F_∞	0.987	0.97	0.942	0.91	0.88	0.777

- Mean first passage time is proportional to ϵ^{-1}, and can be computed using recursion.

Outline

- Intra protocol (MIMD) fairness: same RTT
- Intra protocol (MIMD) fairness: different RTTs
- Inter protocol (MIMD & AIMD) fairness: same RTT

Intra-protocol Fairness: different RTTs

- Xu *et al* (2004) showed that the MIMD algorithm is RTT unfair, i.e., the session with the shortest RTT grabs all the capacity.
- We provide a sufficient condition to reduce the RTT unfairness.

- Increase algorithm

$$\mathsf{r_i}(t+\delta) = \alpha^{\delta/\tau_i} \cdot \mathsf{r_i}(t), \quad i = 1, 2.$$

- Let $\mathsf{z}(t) = \log\left[\frac{\mathsf{r}_2(t)}{\mathsf{r}_1(t)}\right]$.
- In the absence of asynchronous losses,

$$\mathsf{z}(t+\delta) = \mathsf{z}(t) + \gamma\delta,$$

where $\gamma = \log[\alpha]\left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)$.

- Drift towards $+\infty$ (resp. $-\infty$) if $\tau_2 < \tau_1$ (resp. $\tau_1 < \tau_2$).

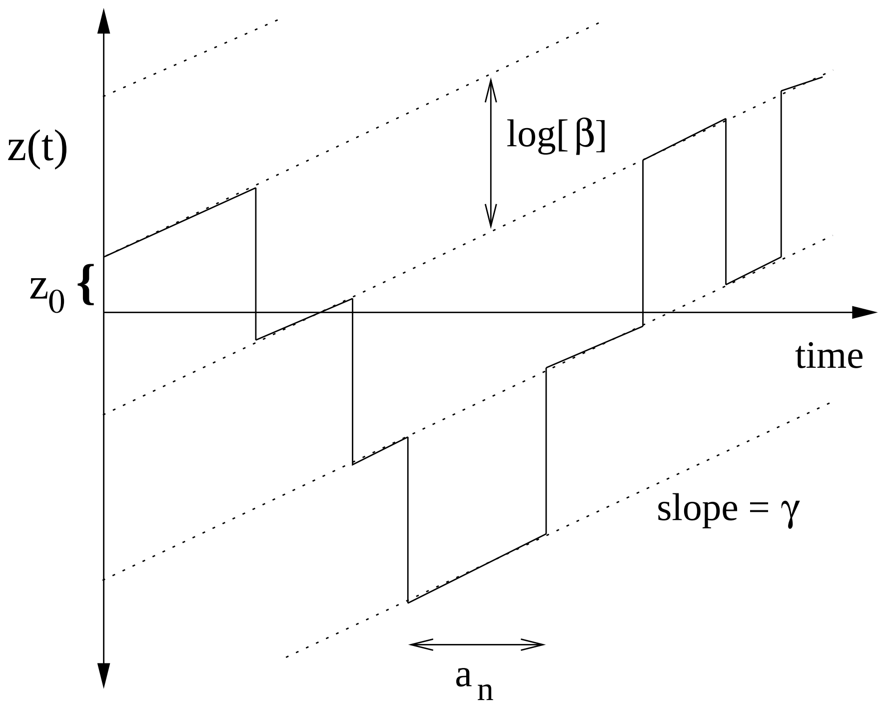


Figure 3: Evolution in time of $z(t)$. $\tau_2 < \tau_1$.

- Let z_n denote $z(t)$ embedded just before the n^{th} control signal.
- Let a_n be the interarrival time of the control signals. a_n is i.i.d. with mean λ^{-1}.

$$z_{n+1} = z_n + \gamma a_n + c_n,$$

where c_n is defined as

$$c_n = \begin{cases} -\log[\beta] & \text{w.p. } \frac{1}{e^{z_n}+1} \\ +\log[\beta] & \text{w.p. } \frac{e^{z_n}}{e^{z_n}+1} \end{cases} .$$

- z_n is positive recurrent if

$$\lambda > \frac{\gamma}{-\log[\beta]}.$$

Simulation Results

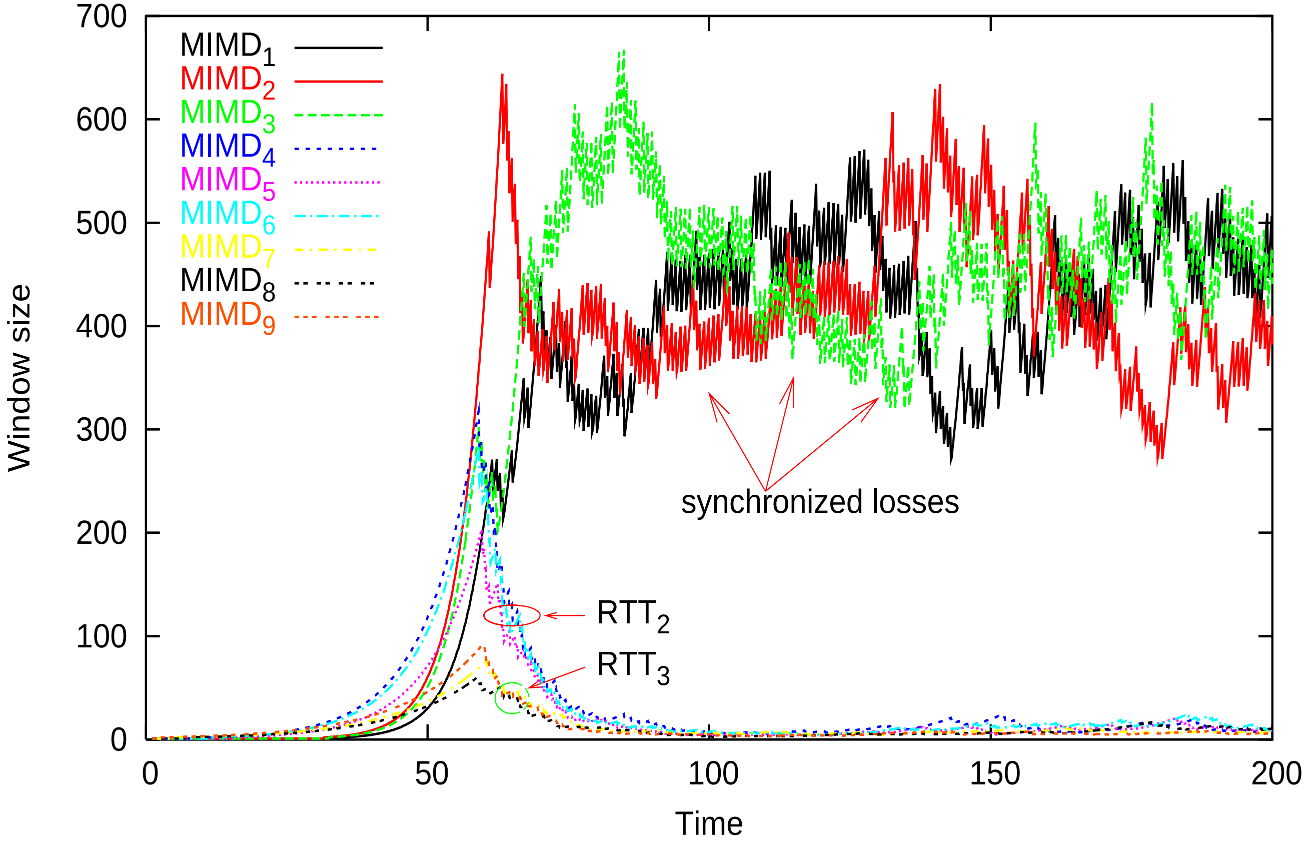


Figure 4: $\epsilon = 0$.

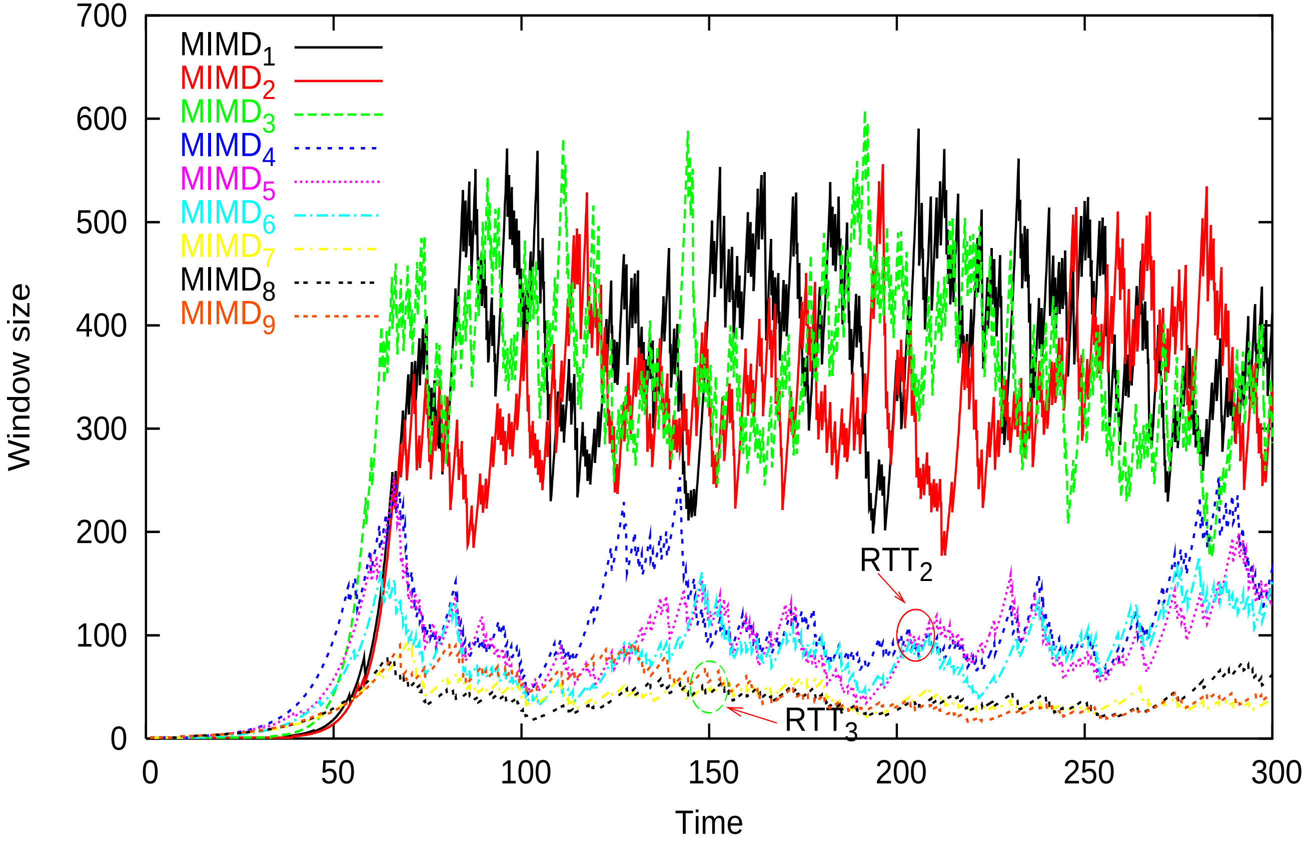


Figure 5: $\epsilon = 0.00015$.

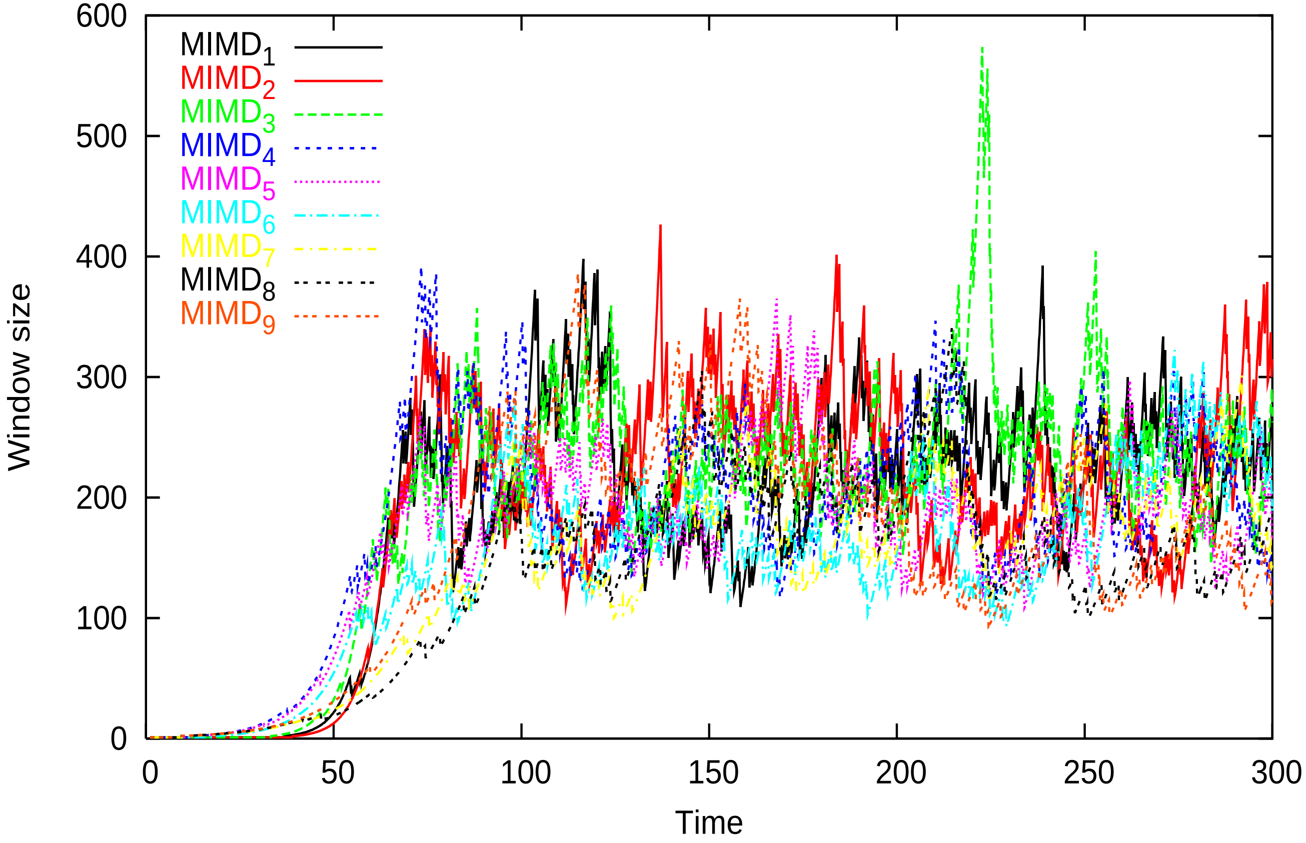


Figure 6: $\epsilon = 0.0003$.

Table 1: Throughput for each RTT class and overall efficiency

ϵ	η_1(Mbps)	η_2(Mbps)	η_3(Mbps)	$\frac{\eta_1+\eta_2+\eta_3}{C}$
0	178	2.8	1	0.91
0.00015	148	25.5	7.14	0.905
0.0003	101	48	28.4	0.89

Outline

- Intra protocol (MIMD) fairness: same RTT
- Intra protocol (MIMD) fairness: different RTTs
- Inter protocol (MIMD & AIMD) fairness: same RTT

- Capacity sharing between MIMD and AIMD sources.
- N_m : number of MIMD sessions. N_a : number of AIMD sessions.
- Same RTT τ.
- Losses are synchronous.

- Rate evolution:

$$r_i(t+\delta) = \alpha_m^{\delta/\tau} r_i(t) \qquad i = 1, .., N_m$$
$$r_i(t+\delta) = r_i(t) + \frac{\alpha_a}{\tau^2}\delta \qquad i = 1, .., N_a$$

If a loss ocuurs at time t

$$r_i(t+) = \beta_m r_i(t) \qquad i = 1, .., N_m$$
$$r_i(t+) = \beta_a r_i(t) \qquad i = 1, .., N_a$$

- Losses occur when capacity is reached.

- Let $t_n, n \geq 0$ denote the n^{th} loss instant.
- $\Delta_n = t_{n+1} - t_n$ is the inter-loss time.

$$r_i(n+1) = \begin{cases} \beta_m r_i(n) \alpha_m^{\Delta_n/\tau} & i = 1,..,N_m \\ \beta_a \left(r_i(n) + \frac{\alpha_a}{\tau^2} \Delta_n \right) & i = 1,..,N_a \end{cases}$$

$$\sum_{i=1}^{N_m} \frac{r_i(n)}{\beta_a} + \sum_{i=1}^{N_a} \frac{r_i(n)}{\beta_m} = C.$$

- Assumption : rate of each session has an equilibrium behaviour, i.e., inter loss time is constant.

$$\lim_{n \to \infty} \Delta_n = \Delta, \quad \lim_{n \to \infty} r_i(n) = \psi_i.$$

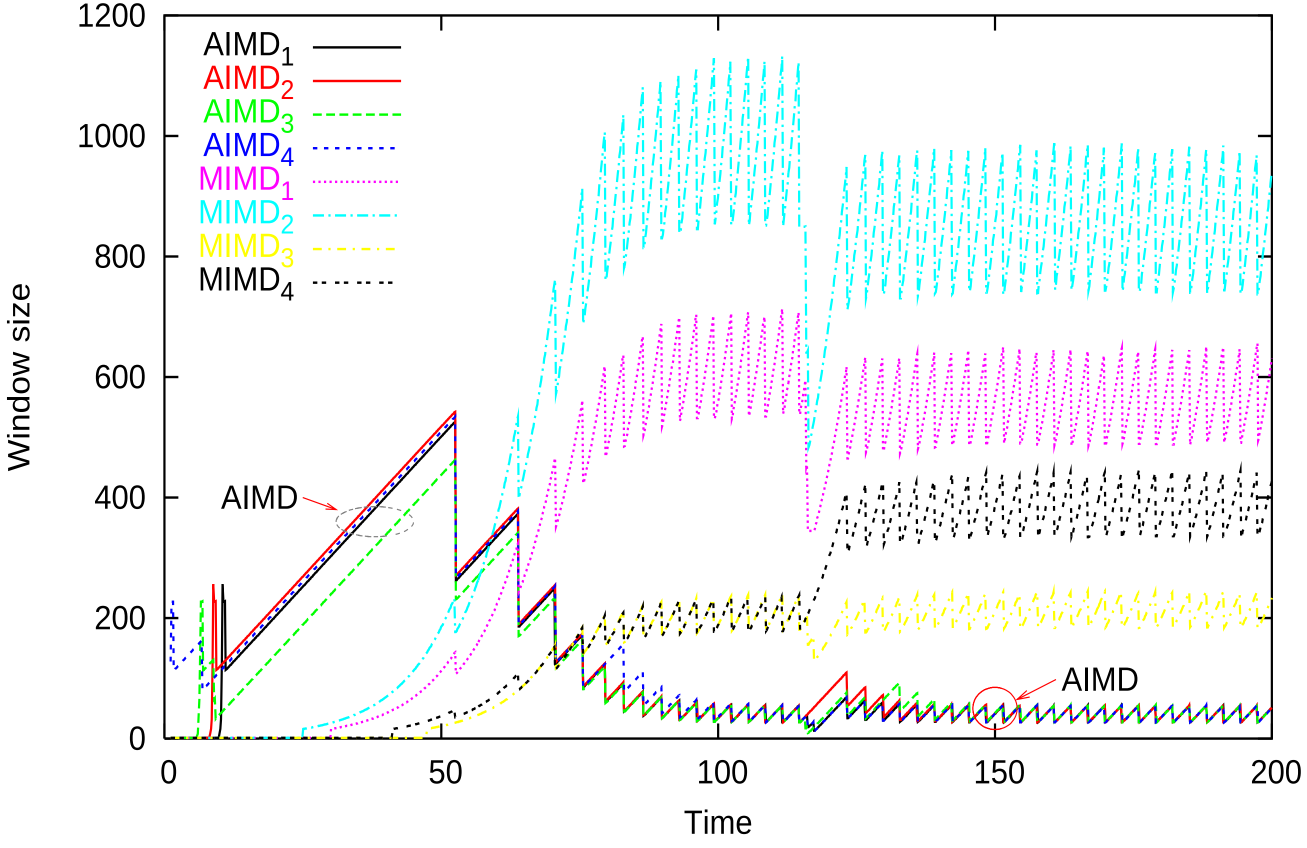


Figure 7: Rates of four AIMD and four MIMD sources sharing a link.

- Inter-loss time depends only on the increase and decrease parameters of the MIMD algorithm. $\Delta = -\tau \frac{\log[\beta_m]}{\log[\alpha_m]}$.
- Rate of an AIMD session, $\psi = \alpha_a \frac{\beta_a}{1-\beta_a} \frac{\Delta}{\tau^2}$,
  - is independent of C.
  - depends only on the increase and decrease parameters of the algorithms.
- Total rate of MIMD sessions: they utilize the rest of the capacity.

- As capacity increases, MIMD user gets a larger share of the bandwidth.
- There exists a Λ_l below which AIMD user gets a better throughput compared to a MIMD user.
- AIMD user can improve its throughput by opening multiple connections whereas the same is not true for the MIMD user.

Thank You.