Convex Optimization and Congestion Control

Fabian Wirth

University of Würzburg

23.07. – 03.08.2012

Outline

• Part I:	Convexity and Convex Functions	Lectures 1, 2, 3
• Part II:	Convex Optimization	Lectures 4 and 5
• Part III:	Numerical Methods	Lectures 6 and 7
• Part IV:	Congestion Control	Lecture 8
• Part V:	Utility Based Congestion Control	Lecture 9
• Part VI:	Miscellaneous Problems in Networks	Lecture 10 (we shall see)

Part IV: Congestion Control

- IV.1: Basics of TCP
- IV.2: Dynamics of Deterministic AIMD
- IV.3: Utility Based Congestion Control

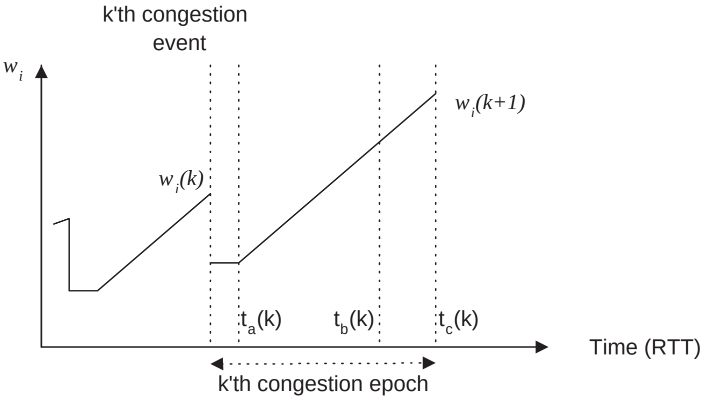
Outline

1. Modeling of TCP flows
   - A linear model I
   - A linear model II
   - Model Validation
2. Analysis of TCP
3. Utility Based Congestion Control

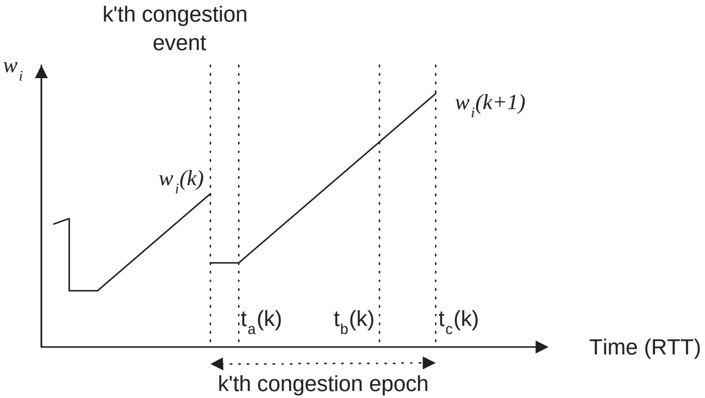
Assumptions I

- Congestion occurs in a single bottleneck.
- Congestion is noticed one RTT after it happens.
- Buffer size of bottleneck is small, i.e. RTT can be approximated by a constant.
- RTT is the same for all sources.
- The network is synchronized.

Many sources, one bottleneck

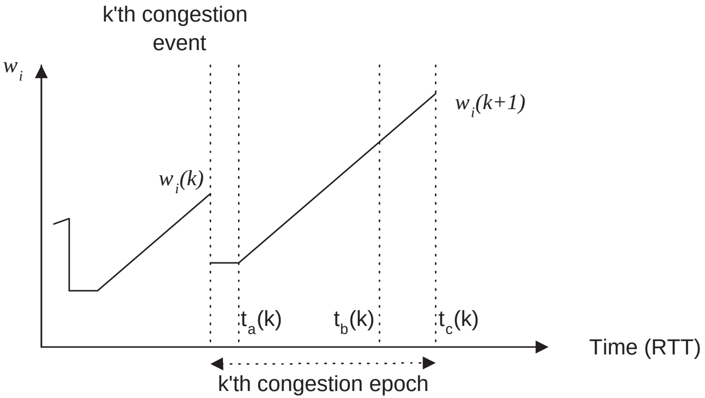


Many sources, one bottleneck



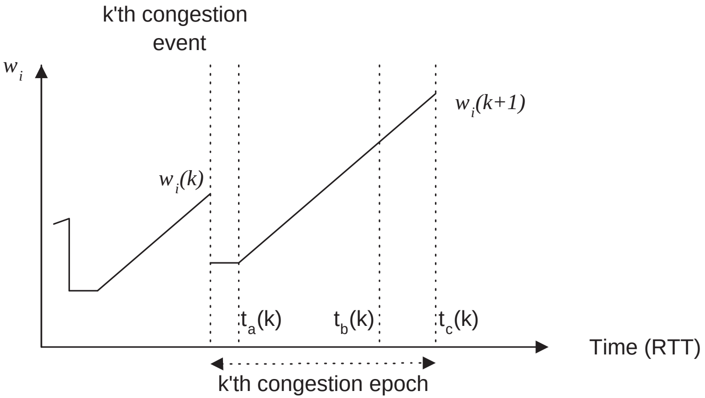
$$w_i(k+1) = \beta_i w_i(k) + \alpha_i(t_c(k) - t_a(k))$$

Many sources, one bottleneck



$$w_i(k+1) = \beta_i w_i(k) + \alpha_i(t_c(k) - t_a(k))$$
$$\sum \alpha_i(t_c(k) - t_a(k)) = P - \sum \beta_i w_i(k) + \sum \alpha_i$$

Many sources, one bottleneck



$$w_i(k+1) = \beta_i w_i(k) + \alpha_i(t_c(k) - t_a(k))$$
$$\sum \alpha_i(t_c(k) - t_a(k)) = P - \sum \beta_i w_i(k) + \sum \alpha_i$$
$$\sum w_i(k) = P + \sum \alpha_i$$

A linear model

A little manipulation shows that

$$w(k+1) = Aw(k)\,, \quad k \geq 1$$

where A is given by

$$\begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0 \dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1 & \cdots & 1-\beta_n \end{bmatrix}.$$

A linear model

A little manipulation shows that

$$w(k+1) = Aw(k)\,, \quad k \geq 1$$

where A is given by

$$\begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1 & \cdots & 1-\beta_n \end{bmatrix}.$$

Here α_i is the additive increase parameter of the i-th source

A linear model

A little manipulation shows that

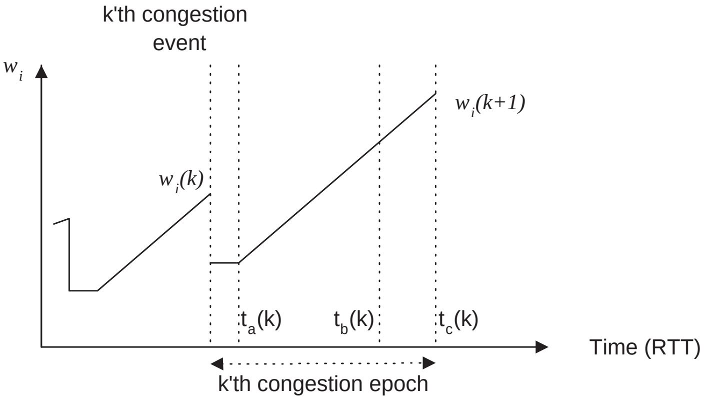
$$w(k+1) = Aw(k)\,, \quad k \geq 1$$

where A is given by

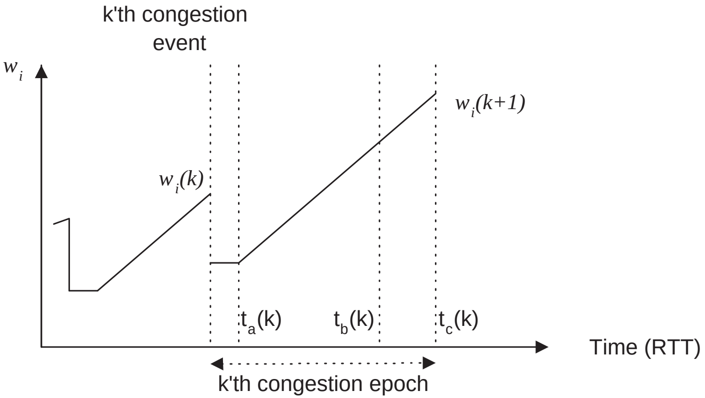
$$\begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1 & \cdots & 1-\beta_n \end{bmatrix}.$$

Here α_i is the additive increase parameter of the i-th source and β_i is the multiplicative decrease parameter.

Digression: Different RTT's

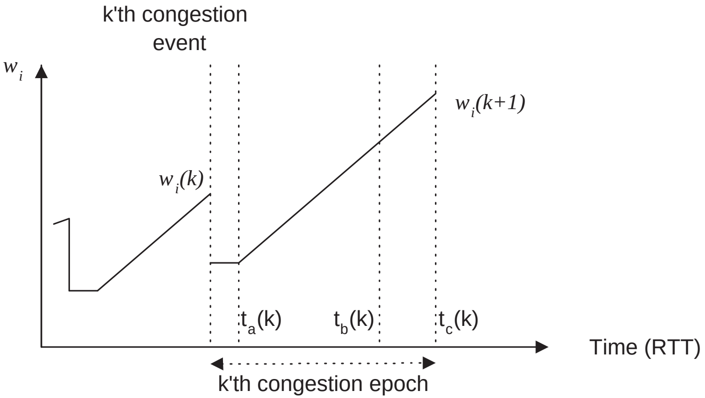


Digression: Different RTT's



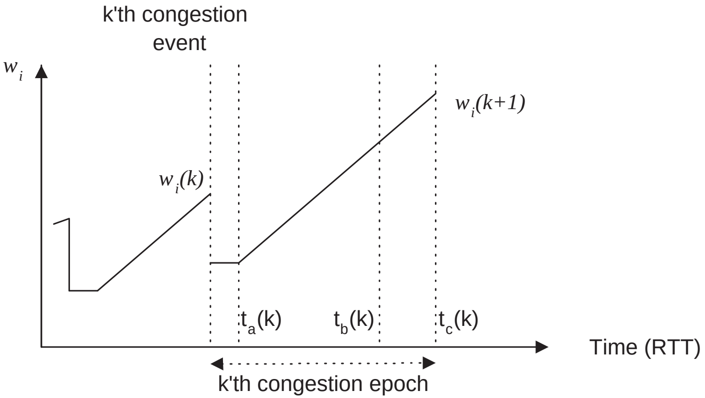
$$w_i(k+1) = \beta_i w_i(k) + \frac{\alpha_i}{\mathrm{RTT}_i}(t_c(k) - t_a(k))$$

Digression: Different RTT's



$$w_i(k+1) = \beta_i w_i(k) + \frac{\alpha_i}{\mathrm{RTT}_i}(t_c(k) - t_a(k))$$
$$\sum \frac{\alpha_i}{\mathrm{RTT}_i}(t_c(k) - t_a(k)) = P - \sum \beta_i w_i(k) + \sum \frac{\alpha_i}{\mathrm{RTT}_i}$$

Digression: Different RTT's



$$w_i(k+1) = \beta_i w_i(k) + \frac{\alpha_i}{\mathrm{RTT}_i}(t_c(k) - t_a(k))$$
$$\sum \frac{\alpha_i}{\mathrm{RTT}_i}(t_c(k) - t_a(k)) = P - \sum \beta_i w_i(k) + \sum \frac{\alpha_i}{\mathrm{RTT}_i}$$
$$\sum w_i(k) = P + \sum \frac{\alpha_i}{\mathrm{RTT}_i}$$

Digression: Different RTT's

A little manipulation leads to a linear equation

$$w(k+1) = \tilde{A} w(k)\,.$$

The similarity transformation

$$\hat{w} := \begin{bmatrix} \mathrm{RTT}_1^{-1} & 0 & \dots & 0 \\ 0 & \mathrm{RTT}_2^{-1} & \dots & 0 \\ & & \ddots & \\ 0 & \dots & & \mathrm{RTT}_n^{-1} \end{bmatrix} w$$

leads to an evolution equation for $\hat{w}$ of the same form as in the case for equal RTT's.

A linear model

$$w(k+1) = Aw(k)\,, \quad k \geq 1$$

where A is given by

$$\begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1 & \cdots & 1-\beta_n \end{bmatrix}.$$

Here α_i is the additive increase parameter of the i-th source and β_i is the multiplicative decrease parameter.

The Synchronized Case

Assumption: Every source experiences all congestions

$$A = \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1 & \cdots & 1-\beta_n \end{bmatrix}.$$

The matrix A is positive and column stochastic. Thus by the Perron-Frobenius Theorem the evolution of A^k is very well understood.

The Synchronized Case

$$A = \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & & \vdots \\ \vdots & & \ddots & \\ 0 \dots & 0 & & \beta_n \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1 - \beta_1 & \cdots & 1 - \beta_n \end{bmatrix}.$$

Theorem (Berman, Shorten, Leith)

1. A has an eigenvalue one with eigenvector

$$v = \begin{bmatrix} \frac{\alpha_1}{1-\beta_1} & \cdots & \frac{\alpha_n}{1-\beta_n} \end{bmatrix}.$$

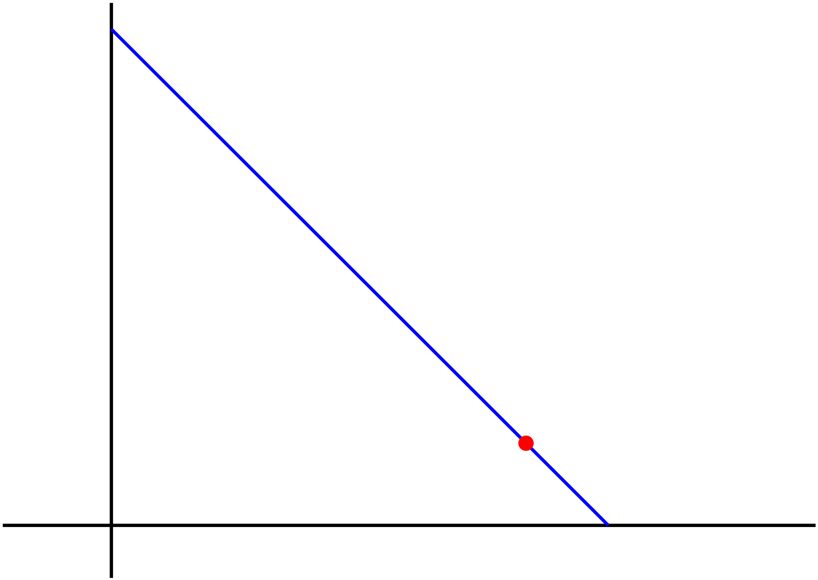
2. For all initial conditions w_0 we have

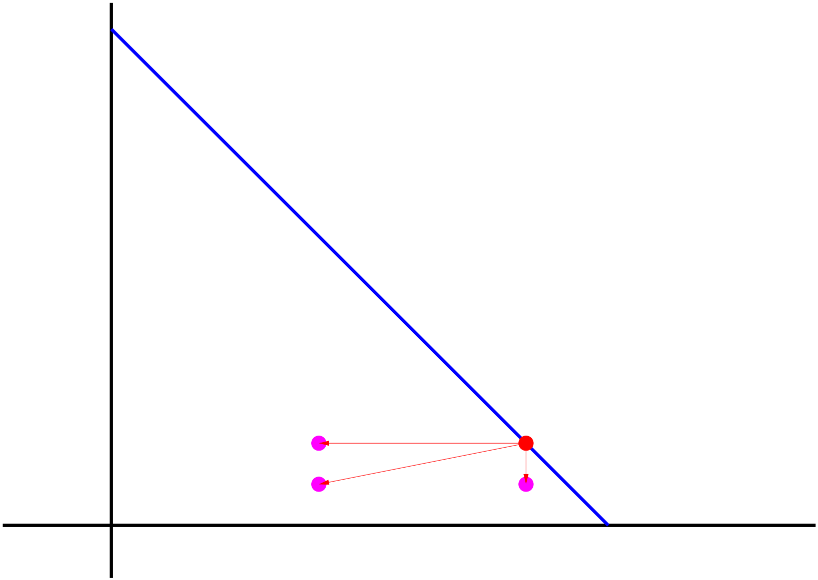
$$A^k w_0 \to \theta v.$$

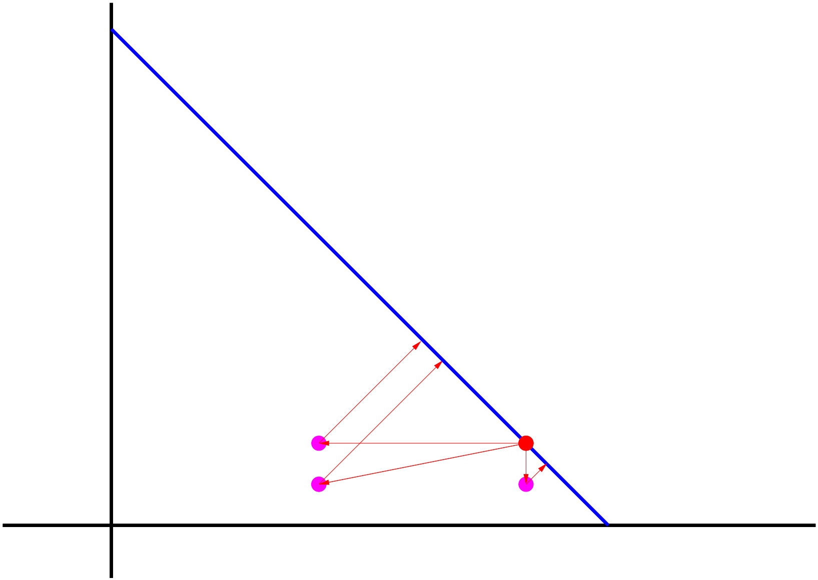
3. the rate of convergence is exponential, bounds for this rate can be given in terms of the β_i.

Assumptions II

- Congestion occurs in a single bottleneck.
- Congestion is noticed one RTT after it happens.
- Buffer size of bottleneck is small, i.e. RTT can be approximated by a constant.
- RTT is the same for all sources.
- ~~The network is synchronized.~~







The unsynchronized case

$$w(k+1) = A(k)w(k)\,, \quad k \geq 1$$

where $A(k)$ is given by

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1(k) & \cdots & 1-\beta_n(k) \end{bmatrix}.$$

The unsynchronized case

$$w(k+1) = A(k)w(k)\,, \quad k \geq 1$$

where $A(k)$ is given by

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1(k) & \cdots & 1-\beta_n(k) \end{bmatrix}.$$

Here α_i is the additive increase parameter of the i-th source

The unsynchronized case

$$w(k+1) = A(k)w(k)\,, \quad k \geq 1$$

where $A(k)$ is given by

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1(k) & \cdots & 1-\beta_n(k) \end{bmatrix}.$$

Here α_i is the additive increase parameter of the i-th source and $\beta_i(k)$ is equal to the multiplicative decrease parameter or 1, depending on whether the $i-th$ source experiences congestion at the k-th congestion event or not.

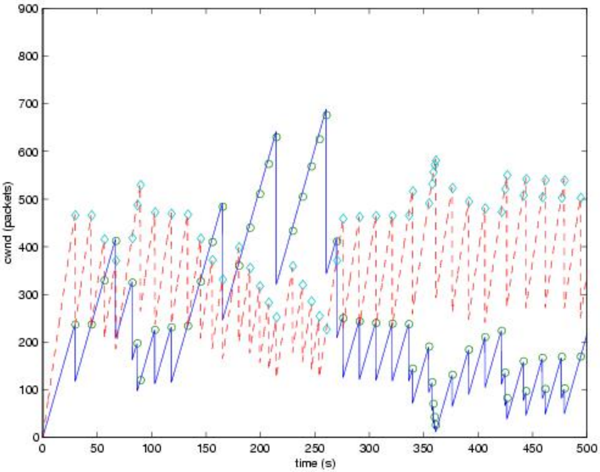
The unsynchronized case

In the analysis of the dynamics of TCP flows we are led to the consideration of a linear inclusion of the form

$$w(k+1) \in \{Aw(k) \mid A \in \mathcal{M}\},$$

where $\mathcal{M} \subset \mathbb{R}^{n \times n}$ is the set of 2^{n-1} matrices obtained by setting $\beta_i, i = 1, \dots, n$ either to 1 or to a constant in $(0,1)$.
(Note that the identity is omitted.)

Model Validation I



Outline

1 Modeling of TCP flows

2 Analysis of TCP
- Analysis
- A fairness result
- Model Validation

3 Utility Based Congestion Control

Application to the TCP model

The matrices modeling TCP flows were of the form

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0 \dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1 - \beta_1(k) & \cdots & 1 - \beta_n(k) \end{bmatrix}.$$

Application to the TCP model

The matrices modeling TCP flows were of the form

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0 \dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1 - \beta_1(k) & \cdots & 1 - \beta_n(k) \end{bmatrix}.$$

The matrices are all column stochastic.

Application to the TCP model

The matrices modeling TCP flows were of the form

$$\begin{bmatrix} \beta_1(k) & 0 & \dots & 0 \\ 0 & \beta_2(k) & & \vdots \\ \vdots & & \ddots & \\ 0\dots & 0 & & \beta_n(k) \end{bmatrix} + \frac{1}{\sum_i \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \begin{bmatrix} 1-\beta_1(k) & \cdots & 1-\beta_n(k) \end{bmatrix}.$$

The matrices are all column stochastic.
The set $\mathcal{M}$ is irreducible on the invariant subspace

$$S := \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\perp}.$$

Weak ergodicity

Theorem Let $\{A(k)\}_{k\in\mathbb{N}} \subset \mathcal{M}^{\mathbb{N}}$ be a sequence with the property that at most one source does not see a drop infinitely often, then

$$\lim_{k\to\infty} A(k)A(k-1)\dots A(0)_{|S} = 0\,.$$

Weak ergodicity

Theorem Let $\{A(k)\}_{k\in\mathbb{N}} \subset \mathcal{M}^{\mathbb{N}}$ be a sequence with the property that at most one source does not see a drop infinitely often, then

$$\lim_{k\to\infty} A(k)A(k-1)\dots A(0)_{|S} = 0\,.$$

In other words, the sequence $\{A(k)\}_{k\in\mathbb{N}}$ is weakly ergodic.

Weak ergodicity

Theorem Let $\{A(k)\}_{k\in\mathbb{N}} \subset \mathcal{M}^{\mathbb{N}}$ be a sequence with the property that at most one source does not see a drop infinitely often, then

$$\lim_{k\to\infty} A(k)A(k-1)\dots A(0)_{|S} = 0\,.$$

In other words, the sequence $\{A(k)\}_{k\in\mathbb{N}}$ is weakly ergodic.
This result extends results of Leizarowitz (LAA, 2000) in this special case.

Weak ergodicity

Theorem Let $\{A(k)\}_{k \in \mathbb{N}} \subset \mathcal{M}^{\mathbb{N}}$ be a sequence with the property that at most one source does not see a drop infinitely often, then

$$\lim_{k \to \infty} A(k)A(k-1)\dots A(0)_{|S} = 0\,.$$

In other words, the sequence $\{A(k)\}_{k \in \mathbb{N}}$ is weakly ergodic.
This result extends results of Leizarowitz (LAA, 2000) in this special case.
Proof builds on properties of extremal norms

A fairness result

Assume now that the matrices $A(k)$ are i.i.d. random variables.
These random variables induce for the i-th source a probability

$$\lambda_i := P(\text{ source } i \text{ experiences congestion at the } k\text{-th congestion event}).$$

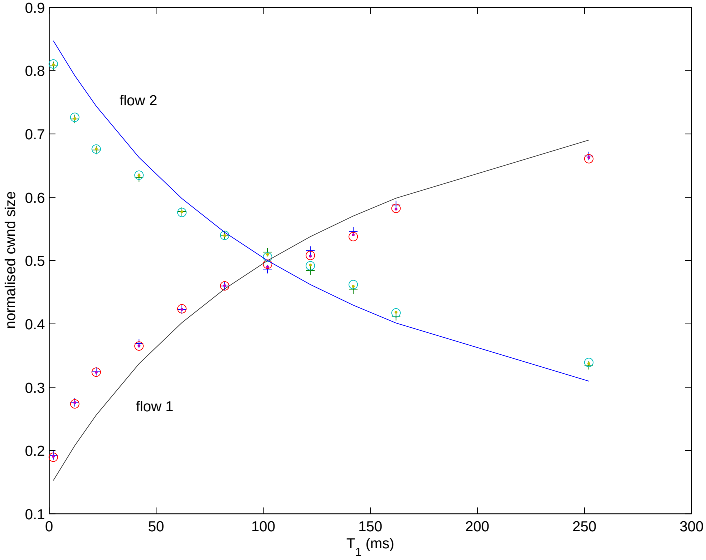
Theorem
If for each $i = 1, \ldots, n$ we have

$$\lambda_i > 0\,,$$

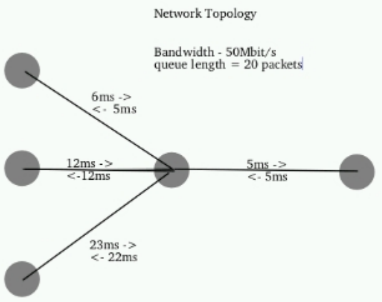
then, almost surely, (for the right γ)

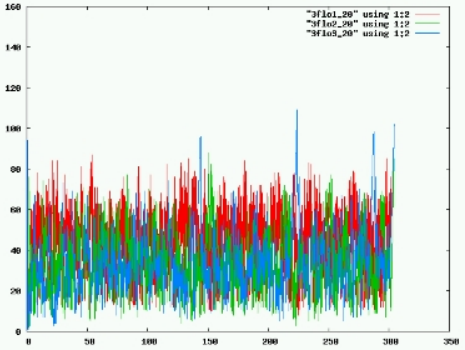
$$\lim_{k\to\infty} \frac{1}{k+1} \sum_{l=0}^{k} w(l) = \gamma \left[\begin{matrix} \frac{\alpha_1}{\lambda_1(1-\beta_1)} & \cdots & \frac{\alpha_n}{\lambda_n(1-\beta_n)} \end{matrix} \right]^T .$$

Model Validation II, NS Simulation



Model Validation III, Data from a Real Network





Model Validation III, Data from a Real Network

Queue		rtt=21ms	rtt=34ms	rtt=55ms
20	Measurement	0.4055	0.3014	0.2931
	Theorem 3.3	0.4054	0.3061	0.2886
	% difference	0.0247	1.5594	1.5353
40	Measurement	0.4122	0.2849	0.3029
	Theorem 3.3	0.4121	0.2915	0.2964
	% difference	0.0243	2.3166	2.1459
60	Measurement	0.4024	0.3093	0.2882
	Theorem 3.3	0.4204	0.3087	0.2709
	% difference	4.4732	0.1940	6.0028

Outline

User and Resources

Setup

Consider a situtation in which users $i = 1, \ldots, n$ use a network or resources $\mathcal{R}$.

User and Used Resources

We identify user i with his subset of used resources, i.e.

$$i \subset \mathcal{R}\,.$$

Capacity Constraint

Each resource is limited, so we need

$$\sum_{i: r \in i} x_i \leq C_r\,.$$

Fairness

Definition

An allocation vector $x = (x_1, \ldots, x_n)$ is called *min-max fair*, if

(i) it satisfies the constraint conditions,

(ii) if y is another vector satisfying the constraints and for some i we have $y_i > x_i$, then there is an index j such that

$$y_j < x_j \leq x_i \, .$$

Bottlenecks

Definition

Given an allocation vector $x = (x_1, \dots, x_n)$ a resource r is called a *bottleneck* for user i, if

(i) full use of the resource, i.e.

$$\sum_{j, r \in j} x_j = C_r \,,$$

(ii) $x_i \geq x_j$ for all users j with $r \in j$.

Characterization of Min-Max Fairness

Lemma

An allocation vector $x = (x_1, \ldots, x_n)$ is min-max fair if and only if every user has a bottleneck resource.

Utility Functions

Utility

We assume from now on that the utility obtained by user i of having x_i bandwidth/rate is given by the utility function U_i.

Assumptions

(i) U_i are strictly concave and increasing, $i = 1, \ldots .n$.

(ii) $U_i(x_i) \to -\infty$ as $x_i \to 0$

Optimization Problem

Definition

$$\text{maximize} \quad \sum_{i=1}^{n} U_i(x_i) \tag{1}$$

$$\text{subject to} \quad \sum_{i:r\in i} x_i \leq C_r \tag{2}$$

$$x_i \geq 0\,, \quad i = \{1, \dots, n\}\,.$$

Barrier Functions

Definition

For every $r \in \mathcal{R}$ let $f_r : [0, \infty) \to [0, \infty)$ be a continuous, non-decreasing function. We assume that

$$\int_0^y f_r(s)ds \to \infty\,, \qquad \text{for } y \to \infty\,. \tag{3}$$

Use these integrals as a general form of barrier function.

$$V(x) = \sum_{i=1}^{n} U_i(x_i) - \sum_{r \in \mathcal{R}} \int_0^{\sum_{i:r\in i} x_i} f_r(s)ds \tag{4}$$

Lemma

Assume that for each i the utility function U_i is continuously differentiable, nondecreasing and strictly concave. The function V is strictly concave.

Convex Optimization Problem

$$\text{maximize} \quad V(x)$$

$$x_i \geq 0\,, \quad i = \{1, \ldots, n\}\,.$$

Characterization of the Optimum

$$U_i'(x_i^*) - \sum_{r:r\in i} f_r \left(\sum_{j:r\in j} x_j^* \right) = 0\,, \quad i = 1, \ldots, n\,.$$

Prices

Barrier Functions as Prices

Price for router r

$$p_r = f_r\Big(\sum_{j:r\in j} x_j\Big)\,. \tag{5}$$

Total price per user

$$q_i = \sum_{r\in i} p_r\,.$$

Optimality Condition

$$U_i'(x_i^*) - q_i(x^*) = 0\,.$$