Scalable Laws for Stable Network Congestion Control

Fernando Paganini
UCLA Electrical Engineering
IPAM Workshop, March 2002.

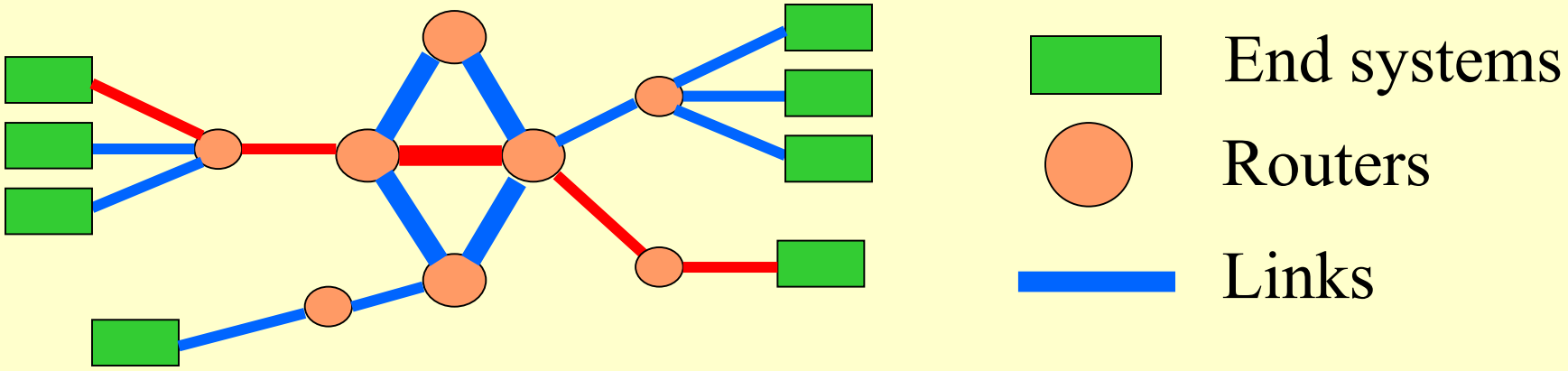
Collaborators:

- Steven Low, John Doyle, Sanjeewa Athuraliya, Jiantao Wang (Caltech).
- Zhikui Wang, Sachin Adlakha (UCLA).

Outline

1. Introduction. Congestion control, models based on prices.
2. Control objectives and linearized design. Local stability theorem.
3. Global, nonlinear implementation. Alternatives to improve fairness.
4. Packet level implementation in ns-2. Results.
5. Conclusions.

Congestion Control Problem



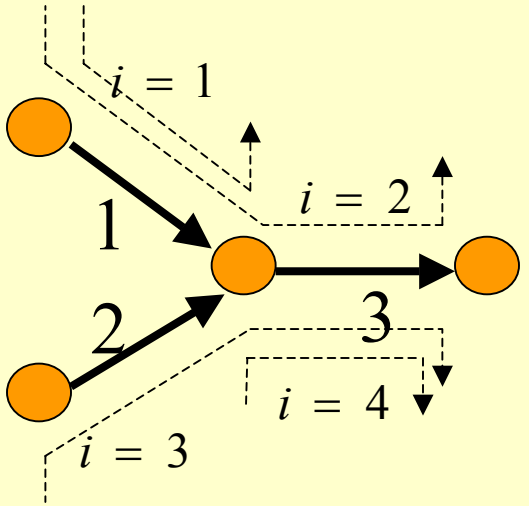
- Regulate transmission rates of end-to-end connections so that they take advantage of the available bandwidth, but avoid exceeding it (congestion).
- Motivation:
  - An interesting, large-scale feedback control problem.
  - Deficiencies of current TCP (long queues, oscillations).
- Aim: regulate large “elephant” flows to a stable point that exploits available capacity, but keep queues small so that uncontrolled “mice” can fly through with minimal delay.

Fluid flow modeling

L communication links shared by S source-destination pairs.

Routing matrix:

$$R_{li} = \begin{cases} 1 & \text{if link } l \text{ uses source } i \\ 0 & \text{otherwise} \end{cases}$$



x_i : Rate of $i-$th source

y_l : Total rate of $l-$th link

c_l : Capacity of the $l-$th link

$$Rx = y$$

$$R = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

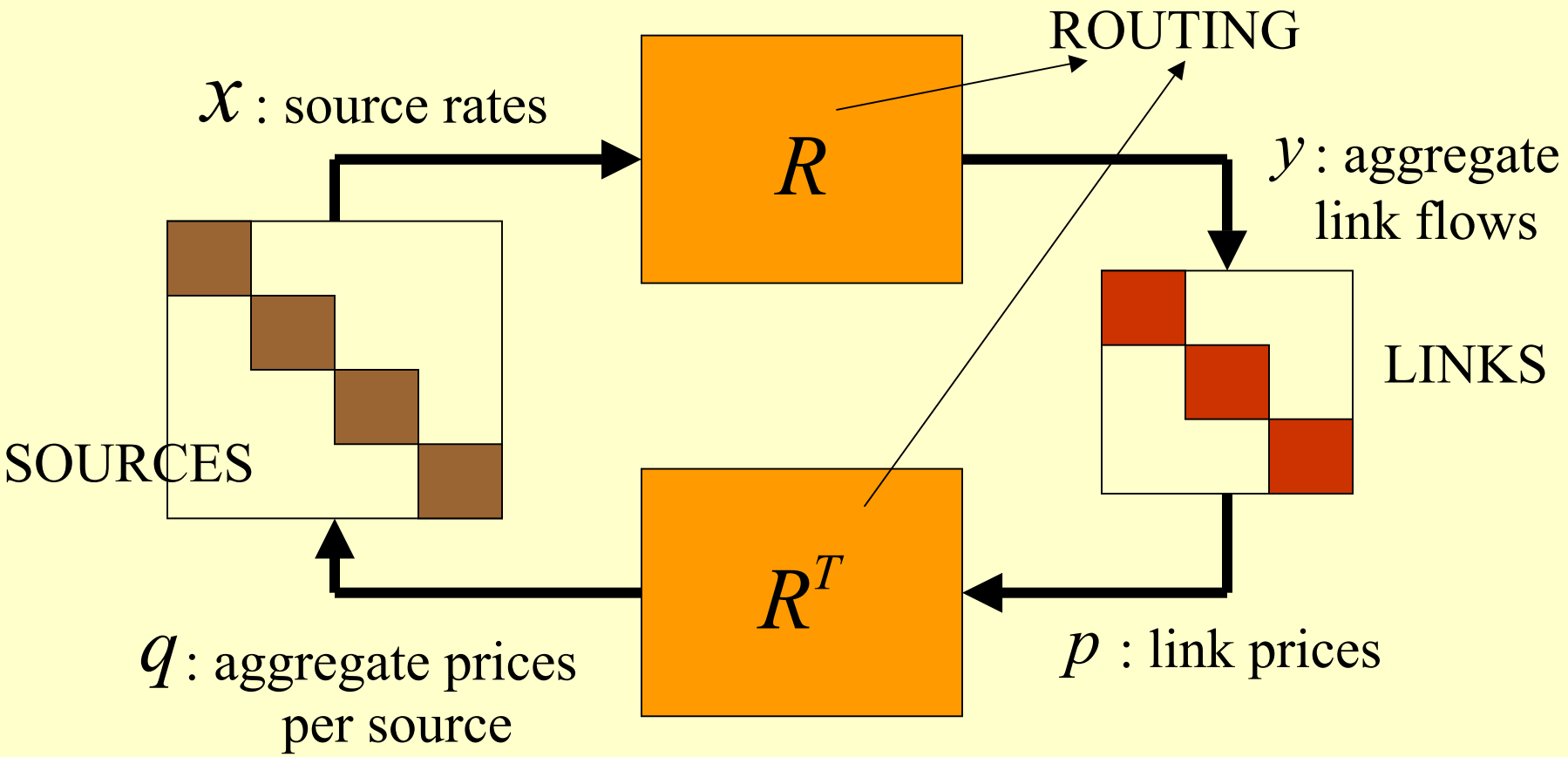
Feedback mechanism:

Each link has a congestion measure or price p_l.

Each source has access to aggregate price q_i of the links in its path.

$$q = R^T p$$

Congestion Control Loop



Decentralized control at links and sources.

Routing assuming fixed, i.e. varying at much slower time-scale.

Optimization interpretation

(Kelly et al, Low et al, Srikant et al.,…)

An equilibrium point x_0, y_0, p_0, q_0 can be interpreted as solving

$$\max_x \sum_i \underbrace{U_i(x_i)}_{\substack{\text{SOURCE}\\ \text{UTILITY}\\ \text{FUNCTION}}}, \quad \text{subject to} \quad \underbrace{Rx \le c}_{\substack{\text{LINK CAPACITY}\\ \text{CONSTRAINTS}}}$$

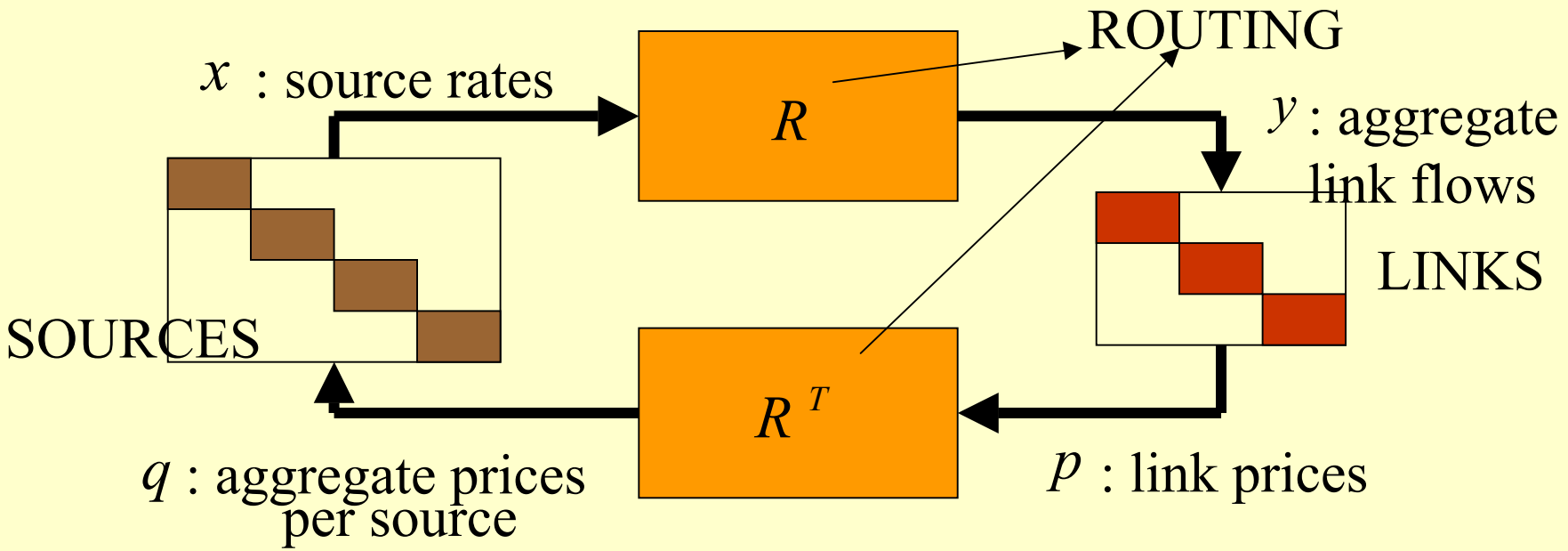
Equilibrium link price signals p_0 provide either a barrier function term, or a Lagrange multiplier for the constraint.

Remark: assuming only:

- Equilibrium rate x_0 is decreasing in the aggregate price q_0.
- For each link, either $y_{0l} = c_l$ or $p_{0l} = 0$.

Then the equilibrium, if achieved, is always the optimum for some utility function, p_0 being the Lagrange multipliers.

"Primal", "dual", and the end-to-end principle.



Usual convention (Kelly, Maulloo, Tan '98):
primal = dynamics at sources, dual = dynamics at links.

It may appear that primal is closer to current TCP, and the end-to-end principle. However:

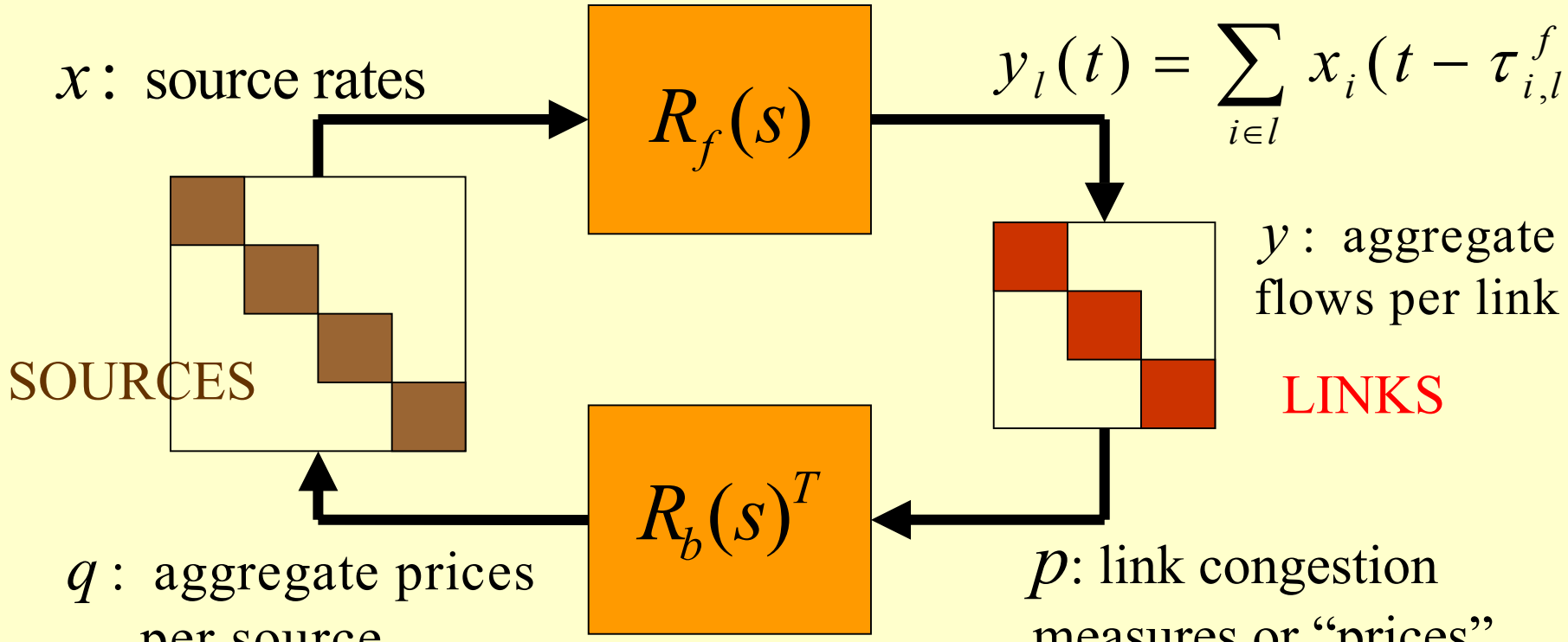
- Current TCP has dynamics in both places.
- End-to-end principle is about complexity, not dynamics.

Dynamics and the role of delay

- Without delay, nothing would stop us from adapting the sources' rates arbitrarily fast.
- In the presence of delay, there is a stability problem: e.g., controlling temperature of your shower.
- Special case of general principle in feedback systems: what limits the performance (e.g. speed of response) are characteristics of the open loop (bandwidth, delay).
- In this case, the only impediment is delay. In particular, this sets the time-scale of our response.

Congestion control loop with delays

Routing/
Delay matrix: $\left[R_f(s)\right]_{li} = \begin{cases} e^{-\tau_{i,l}^f s} & \text{if source } i \text{ uses link } l. \\ 0 & \text{otherwise} \end{cases}$



$$q_i(t) = \sum_{i \in l} p_l(t - \tau_{i,l}^b)$$

RTT: $\tau_i = \tau_{i,l}^f + \tau_{i,l}^b$

Control objectives and design

1. Track available capacity, yet almost empty queues.
2. Stability in the presence of large variations in delay.
3. Dynamic performance: respond as quickly as possible.

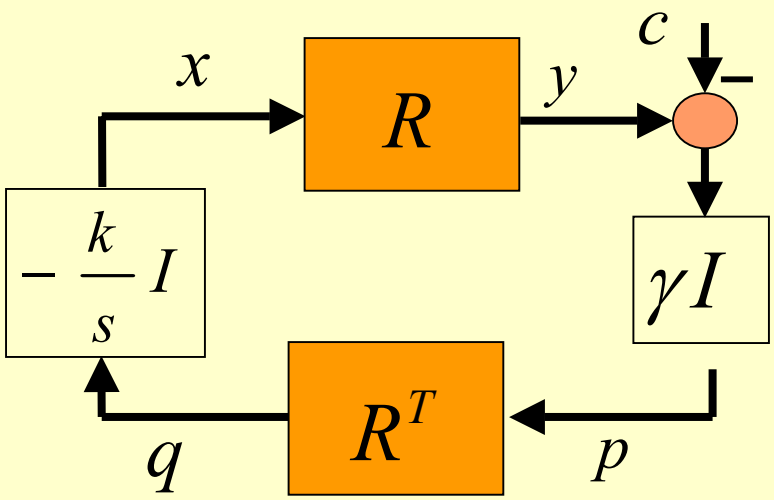
- **Difficulties for control synthesis:**
  - Large-scale, coupled dynamics but **decentralized** information at links and sources. Decentralized control design is hard.
  - Not just global variables, but the **plant** (routing, capacities,…) changes in a way unknown to sources/links. Must be robust.
  - Delay can vary widely. However, sources can adapt to it.
  - To top it off, solution must be simple.
- **Our approach:**
  - Local linear design with classical heuristics.
  - Validated analytically by a local multivariable stability proof.
  - Global nonlinear laws built from the linearization.
  - Performance verified empirically.

Matching capacity through integral control

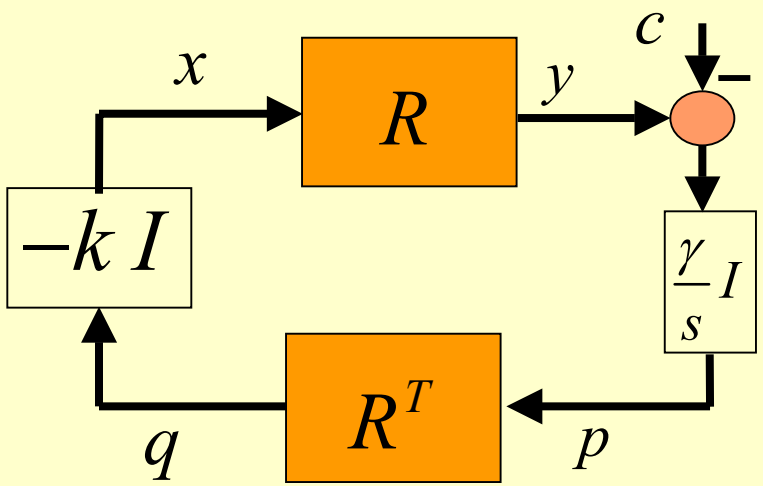
- Tracking of capacity requires integral action in the loop.

Where should we put integrators? A first look (ignore delay):

1. At the sources.



2. At the links.



$$\dot{x} = -kR^T\gamma\left(Rx - c\right)$$

$$\dot{p} = \gamma(R(-k)R^T p - c$$

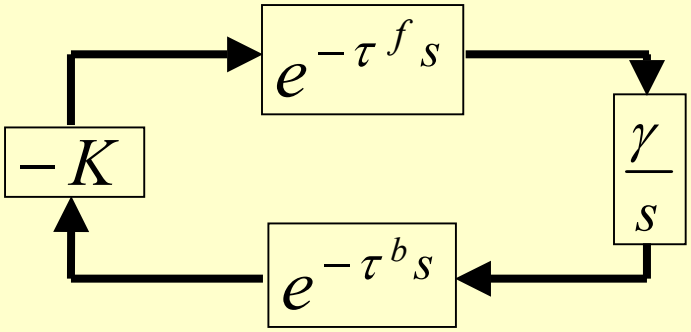
Modes at $-\gamma k\, eig(R^T R)$

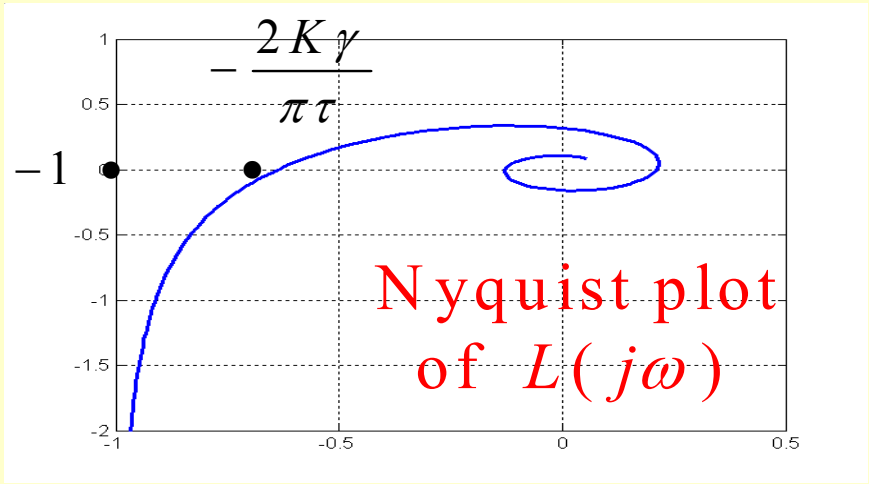
Modes at $-\gamma k\, eig(RR^T)$

Now $R = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ has more rows than columns. So the first case has many additional modes at 0 $\Rightarrow$ unstable.

Compensation for delay

- Include integrators at the links, $\dot{p}_l = \gamma_l(y_l - c_l)$.
- Consider a static source rate control $x_i = f_i(q_i)$, (f_i decreasing). Laws become a special case of those in Low and Lapsley '99. Linearization around equilibrium as $\delta x_i = -K\delta q_i$. For a single link/source, the loop transfer function is $L(s) = K\gamma \dfrac{e^{-\tau s}}{s}$





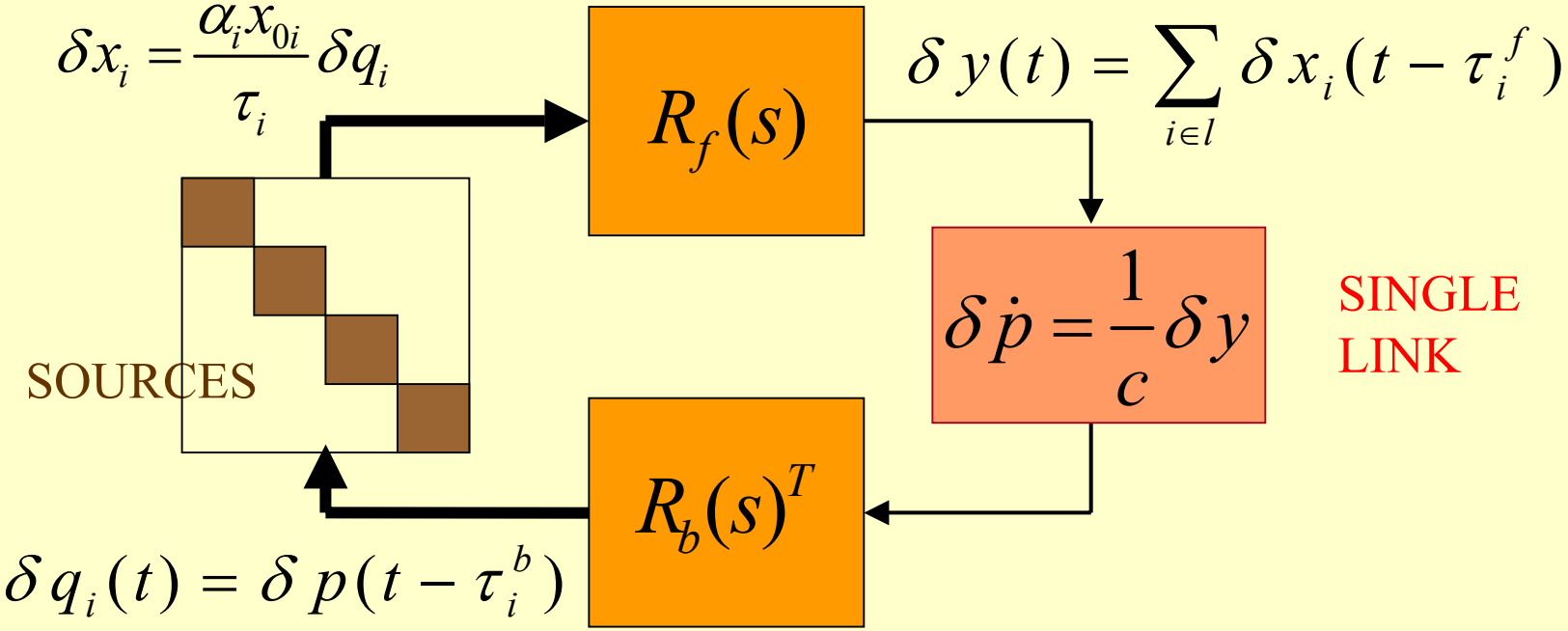
Stable if $K\gamma < \dfrac{\pi}{2\tau}$. Sources know RTT, so pick $K = \dfrac{\alpha}{\tau}$, with $\alpha\gamma < \dfrac{\pi}{2}$.

$\Rightarrow$ Stable for all delays.

Distributed gain compensation

In the multiple source case, we need to bound the overall gain without access to global information.

One solution (δ denotes increments around equilibrium):

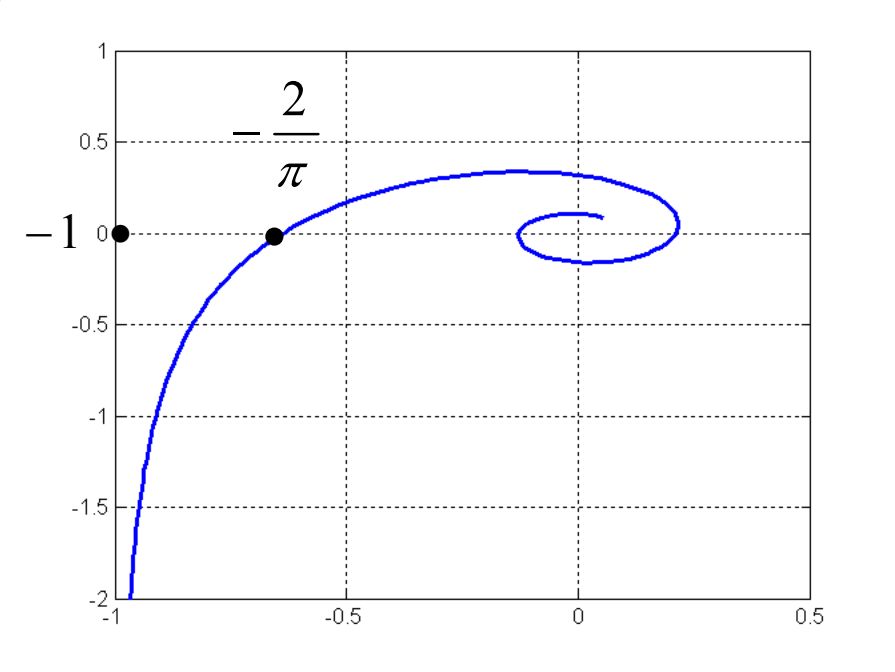


Loop transfer function: $L(s) = \frac{1}{cs} \sum_i \frac{\alpha_i x_{0i}}{\tau_i} \cdot e^{-\tau_i s}$

Nyquist argument for stability

$$L(j\omega) = \sum_i \frac{\alpha_i x_{0i}}{c} \frac{e^{-\tau_i j\omega}}{\tau_i j\omega}$$

Since $\sum_i \frac{x_{0i}}{c} \le 1$, the loop gain is a convex combination of points in the curve $\frac{e^{-j\theta}}{j\theta}$, scaled by $\alpha_i \Rightarrow$ For $\alpha_i < \frac{\pi}{2}$, no encirclements.



Note: if all delays are scaled by some constant, the plot does not change.

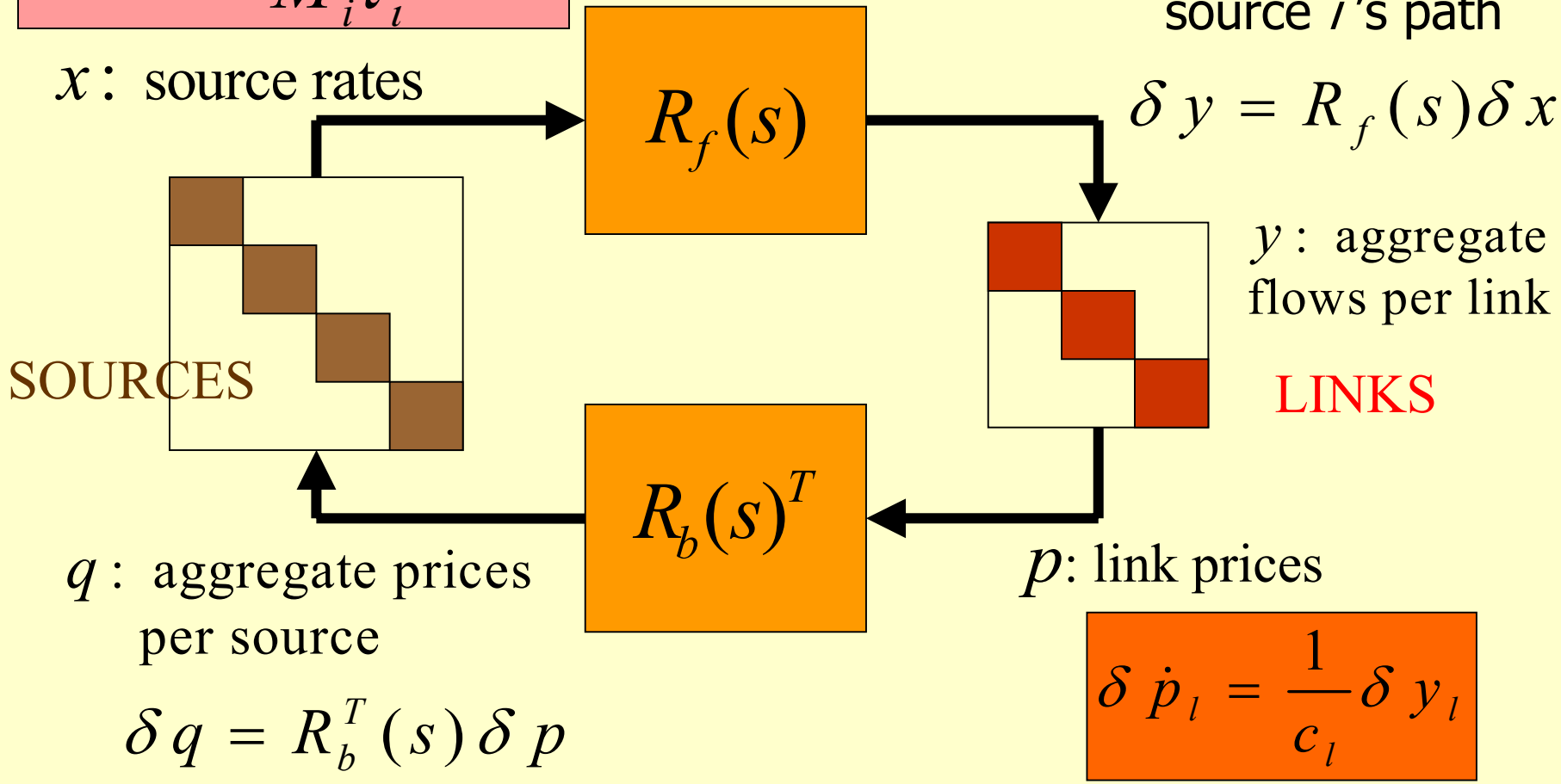
In the time domain, only effect is a change in time-scale of response.

Extension to arbitrary networks

Local analysis around equilibrium. Routing matrices refer here only to bottleneck links.

$$\delta x_i = -\frac{\alpha_i x_{0i}}{M_i \tau_i} \cdot \delta q_i$$

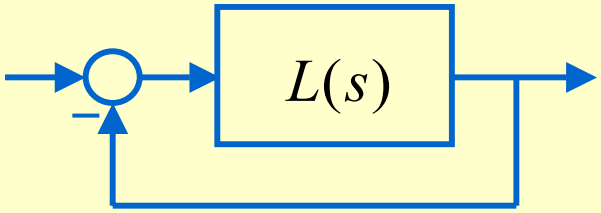
M_i : bound on number of bottlenecks in source *i*'s path



Stability result

Theorem: Assume the matrix $R = R_f(0) = R_b(0)$ (involving only the bottleneck links) is of full row rank, and that $\alpha_i < \frac{\pi}{2}$. Then the feedback system is locally stable for arbitrary delays and capacities.

Steps of the proof:



- Write the loop transfer function

$$L(s) = \underbrace{R_f(s) A X_0 M \mathcal{T} R_b^T(s) \mathcal{C}}_{F(s)} \frac{I_L}{s} = \frac{F(s)}{s}.$$

- $F(s)$ is stable, and $F(0)$ has positive eigenvalues under the rank assumption. This implies stability for small enough α's. Note: integrators at the links!
- A perturbation argument preserves stability as long as $-1 \notin L(j\omega)$. This follows by exploiting $R_b(j\omega) = R_f(j\omega)^* \operatorname{diag}\left(e^{-\tau_1 j\omega}\right)$ (as in Johari -Tan '00), which reduces the eigs$\left(L(j\omega)\right)$ to the same region as in the single link case.

Global, nonlinear implementation

Dynamic Link Control: $\dot{p}_l = \frac{1}{c_l}(y_l - c_l)1_{\{y_l > c_l \text{ or } p_l > 0\}}$

If c_l is the capacity, p_l would be the queueing delay.

But we want to clear the queues!

So replace c_l by a "virtual" capacity $\tilde{c}_l = (1-\varepsilon)c_l$.

Price is now a virtual queueing delay.

Remark: Athuraliya and Low ’00 considered adding another integrator to clear the queue. However, scalable stability for arbitrary delays does not extend to that case.

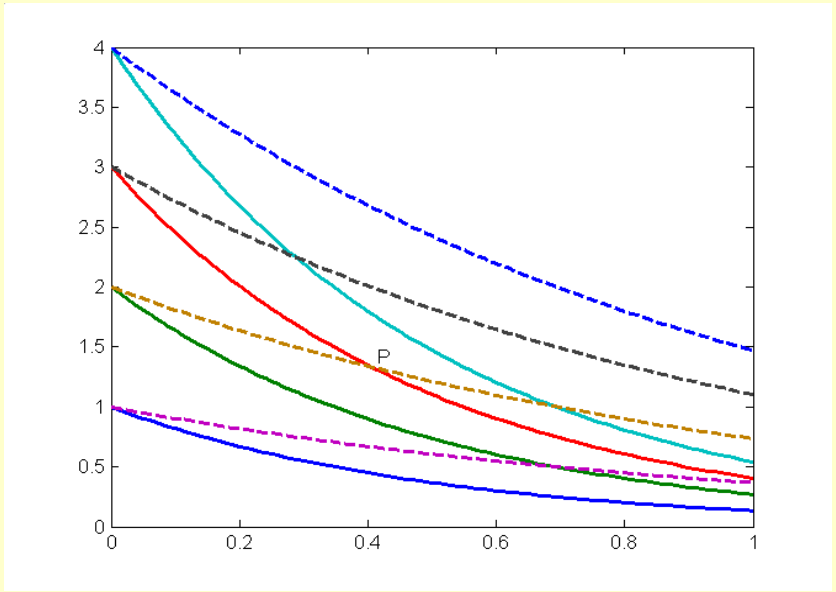
Global, nonlinear implementation

Static control law for sources: linearization requirement is

$$\delta x_i = -\frac{\alpha_i x_{0i}}{M_i \tau_\iota} \cdot \delta q_i \Rightarrow \frac{\partial x_i}{\partial q_i} = -\frac{\alpha_i}{M_i \tau_\iota} x_i(q_i, \tau_i)$$

Assume M_i known, or take a known upper bound. Initially, fix α_i independently of the operating point. Solving the differential equation:

$$x_i = x_{\max,i} e^{-\frac{\alpha_i q_i}{M_i \tau_i}}$$



The utility function would be

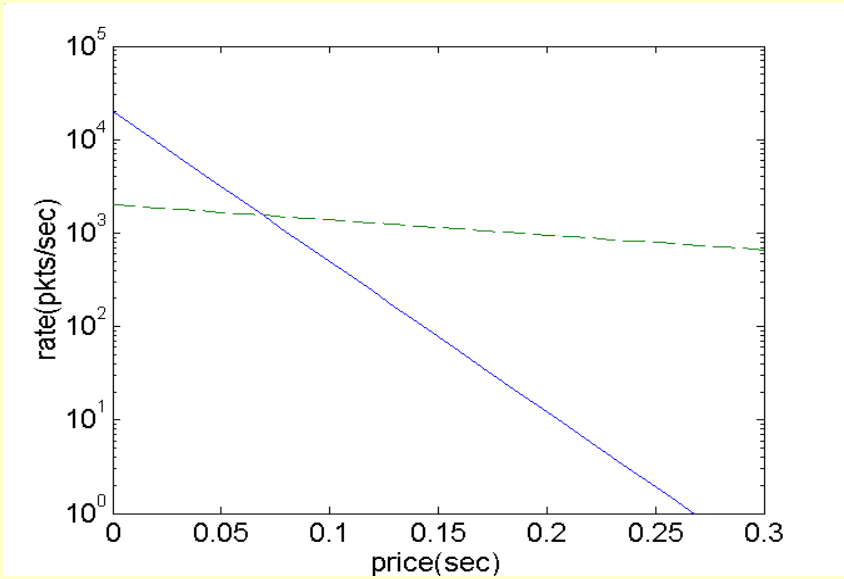
$$U_i(x_i) = \frac{M_i \tau_i}{\alpha_i} x \left[1 - \log\left(\frac{x}{x_{\max,i}} \right) \right]$$

"Elasticity" of demand decreases with delay, number of bottlenecks.

Properties of the nonlinear laws

- Global stability? Validate by
  - Flow simulation of differential equations using Matlab. So far, cases of local stability have been global.
  - Mathematical proof. Tools which combine delay and nonlinearity are very limited! We have partial results for single link, but with further parameter constraints.
- Fairness of equilibrium?

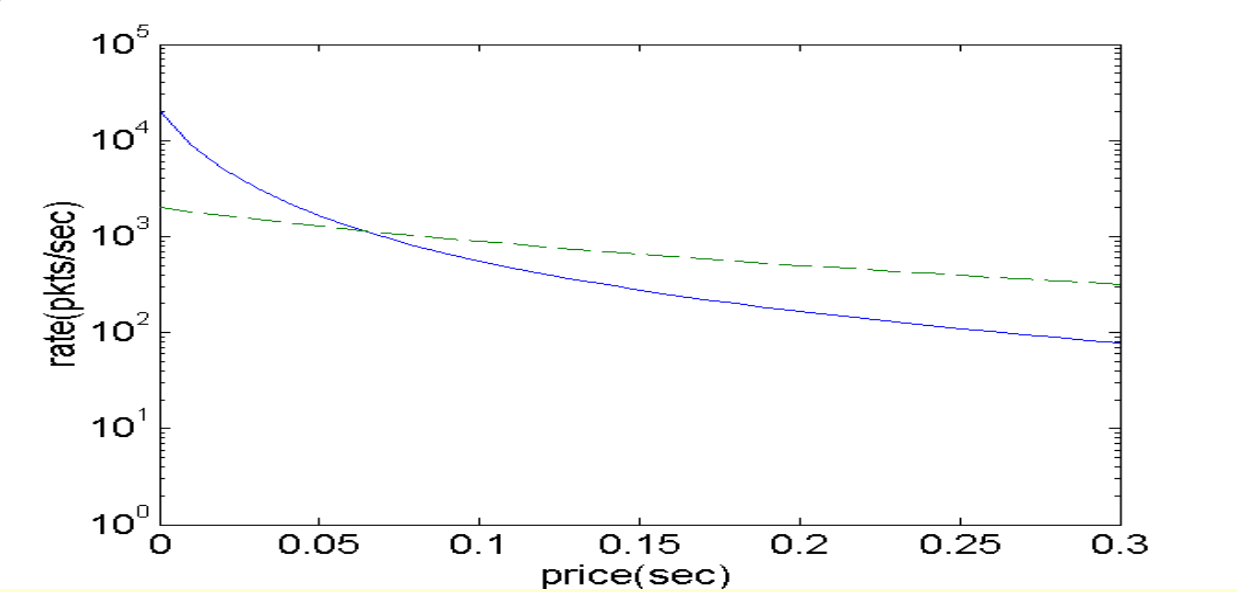
  Difficult with exponential laws, which distinguish too sharply the rates for different delays.



An alternative with fairer allocation

- Back to linearization requirement $\dfrac{\partial x_i}{\partial q_i} = -\dfrac{\alpha_i}{M_i \tau_\iota} x_i(q_i, \tau_i)$
- Alternative where α_i depends on the operating point:

$$x_i = \frac{\varphi(M_i, \tau_i)}{\left(M_i \tau_i + \beta_i q_i\right)^{\gamma_i}}, \qquad \beta_i \gamma_i < \frac{\pi}{2}$$



- More freedom in utility functions, but not arbitrary.

Packet-level implementation

- Links maintain price through a virtual queue counter, incremented on packet arrival, decremented at rate $(1-\varepsilon)c_l$.
- The price can be communicated to sources by "random exponential marking" (REM, Athuraliya and Low '00):

  Link l sets the ECN flag on packet with probability $\mathcal{P}_l = 1 - \Phi^{-p_l}$

  Independence $\Rightarrow$ prob. packet from source i gets marked is

$$1 - \prod_{i \text{ uses } l} \left(1 - \mathcal{P}_l\right) = 1 - \Phi^{-\sum_{i \text{ uses } l} p_l} = 1 - \Phi^{-q_i}$$

- Sources can estimate q_i from packet marking statistics: e.g., counting positive marks on the last N packets. Estimation dynamics adds an extra lag of $\approx \frac{N}{2W}\tau$, where W is the current congestion window. This can be considered in stability analysis.

Packet-level implementation

- The parameter Φ should be "universal". Now for good estimation, the marking probability should not be too close to 0 or 1, which forces prices to vary in some absolute range. Since they are virtual queueing delays, a range of 0.01-1 sec might suffice.
- Window implementation of rate law:

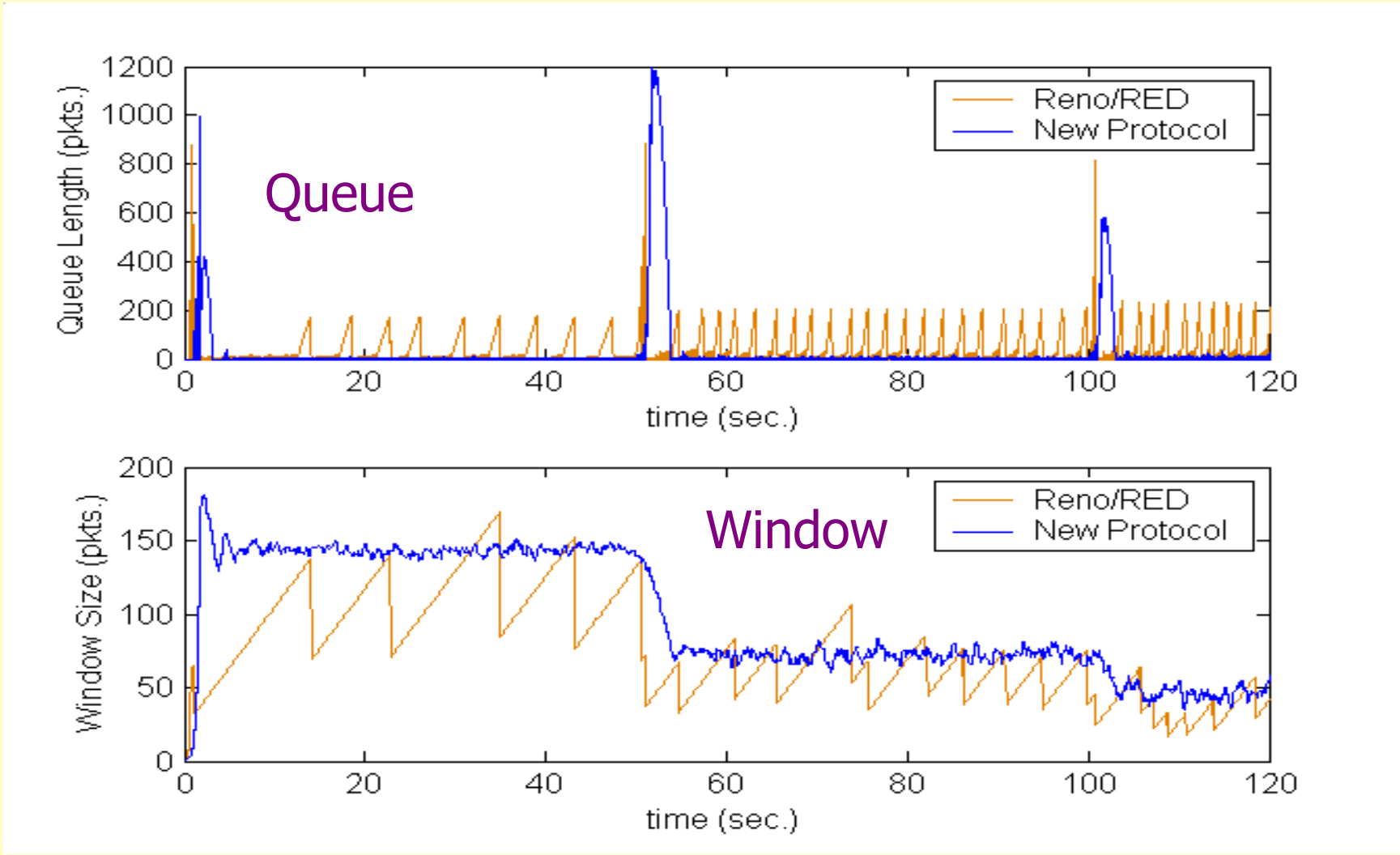
$$W_i = W_{\max}\left(\frac{M_i\tau_i}{M_i\tau_i + \beta_i q_i}\right)^{\gamma_i}, \qquad \beta_i\gamma_i < \frac{\pi}{2}$$

Use minimum observed RTT for the propagation delay τ_i. M_i scaling required in the worst-case, but typically there are enough sources of conservatism so that $M_i = 1$ works.

- ns-2 implementation:
  - Modify REM-module for the links.
  - Modify Vegas module for the sources.

Packet-level simulation in ns-2

60 sources starting in groups of 20, RTT=120ms. 1 link, 25 pkts/ms



Stable, but time-response not slower than existing protocols.

Conclusions

- Classical design heuristics + multivariable analysis lead to a locally stable feedback control under widely varying operating conditions, and within very tight information constraints.
- From local to global: extract nonlinear laws from linearization conditions at every point. This step leaves some degrees of freedom left for addressing equilibrium fairness, etc.
- Pending theory questions:
  - Global stability with nonlinearity and delay. Partial results exist.
  - Equilibrium structure
- Packet implementation based on ECN marking appears to perform well. In particular, fast response, empty queues.

Issues for future studies:

- Parameter settings: some of them must be “universal”.
- Backward compatibility, incremental deployment.

References

http://www.ee.ucla.edu/~paganini

- F. Paganini, J. Doyle and S.H.Low, “Scalable Laws for Stable Network Congestion Control” , Proceedings IEEE Conference on Decision & Control, 2001.
- S. H. Low, F. Paganini, J. Doyle, “Internet Congestion Control: an Analytical Perspective”, IEEE Control Systems Magazine, Feb. 2002.
- F.Paganini, S.H.Low, Z. Wang, S. Athuraliya, J. Doyle, “A new TCP congestion control with empty queues and scalable stability”, submitted to 2002 Sigcomm.
- Z. Wang, F. Paganini “Global Stability with Time Delay in Network Congestion Control”, submitted to IEEE Conference on Decision & Control, 2002.