Transport Equations for Internet Transmission Control

F. Baccelli

INRIA and ENS

ICERM, October 17, 2011

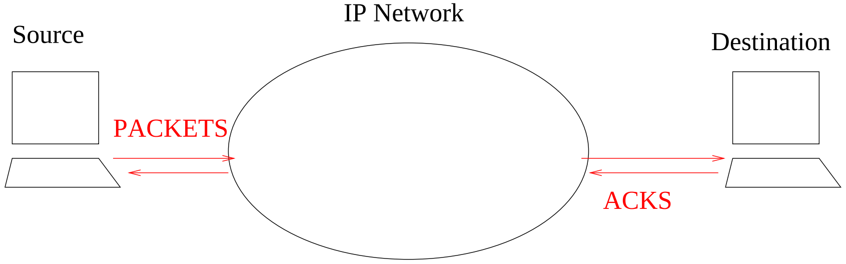
Summary

- Isolated TCP flows
  - Persistent Flows
  - On-Off Flows
- Interaction of Parallel TCP Flows
  - Persistent Flows
  - On-Off Flows
- Interaction of TCP Flows in Series

Persistent Flows

- Dynamics of TCP
- Square root formula
- Markov analysis
- Distributions

TCP Congestion Control



– **Error control** Each packet received by the destination is acknowledged;

– **Congestion control** The number of unacknowledged packets in transit in the network is limited by the source to a maximal value W called the **window**.

If the Round Trip Time (RTT) is R, the **throughput** of the connection is

$$X = \frac{W}{R}$$

Congestion Avoidance Phase of TCP

- **TCP dynamic window size (updated when acks are received)**

$$w_{n+1} = g(w_n, F(n)),$$

$F(n)$: **feedback signal on the state of congestion,**

- **TCP Reno: AIMD**

$$g(w_n, \mathbf{OK}) = w_n + 1 \textbf{ every } w_n \textbf{ acks}, \quad g(w_n, \mathbf{LOSS}) = \left\lfloor \frac{w_n}{2} \right\rfloor$$

- **Scalable TCP: MIMD Kelly 03**

$$g(w_n, \mathbf{OK}) = w_n + a, \quad g(w_n, \mathbf{LOSS}) = \lfloor w_n b \rfloor, \quad 0 < 1 < b.$$

- **TCP Tahoe, HighSpeed TCP Floyd 03, Fast TCP Low 05**

.....

Hybrid Model for TCP RENO Congestion Avoidance

- **AI: the window is increased of 1 unit every W ack:**
  - **In dt, the number of acks that arrive is $X(t)dt$;**
  - **Hence the window increases of $X(t)dt/W(t) = dt/R$.**
- **MD: in case of a loss event, the windows is cut by half.**
- **Differential equation, with $N(t)$ the loss point process:**

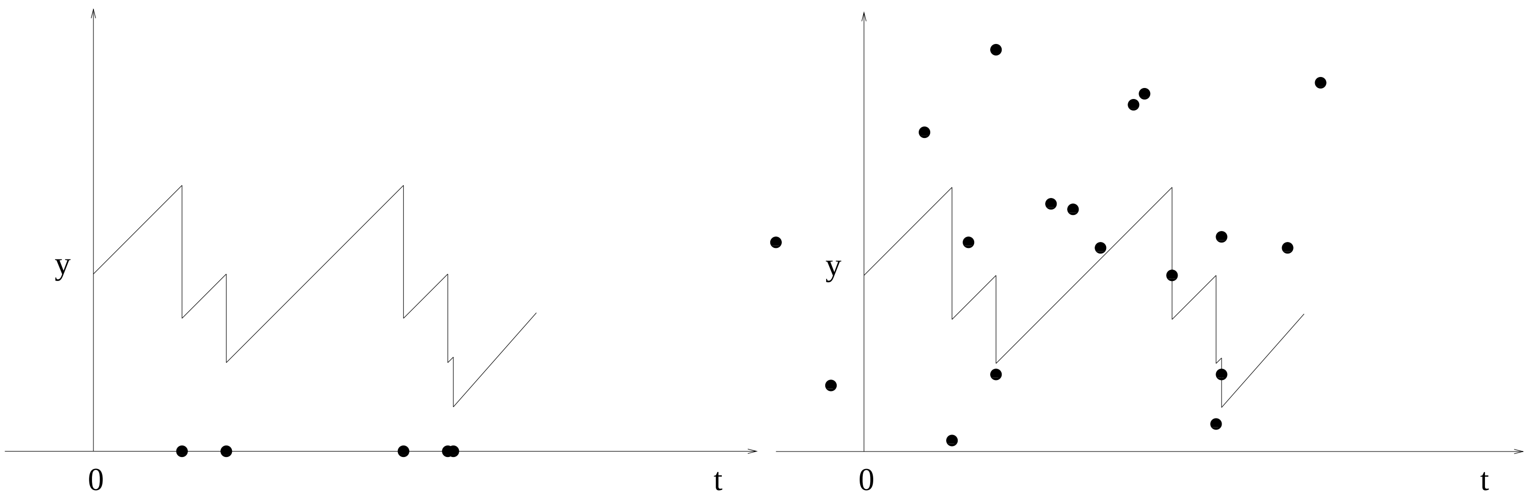
$$dW(t) = \frac{dt}{R} - \frac{W(t-)}{2} N(dt) \qquad dX(t) = \frac{dt}{R^2} - \frac{X(t-)}{2} N(dt)$$

- **First studied by M. Mathis, J. Semke, J. Mahdavi and T. Ott, 97 (square root formula)**

Loss Point Processes

- Losses are modeled by two kinds of point processes $N(t)$:
  1. rate independent (RI) case: homogeneous Poisson point process with intensity λ
  2. rate dependent (RD) case: point process with a stochastic intensity $pX(t)$
- Rationale
  1. RI: losses caused by physical layer events arising on wireless links (fast fading) or DSL links (impulse noise)
  2. RD: PER (packet error rate) due to congestion or transmission errors.

Loss Point Processes (*continued*)



The Square Root Formula in 3 Lines - RI case

- **If there exists a stationary regime with X integrable, the Rate Conservation Principle gives**

$$\frac{1}{R^2} = \frac{\lambda}{2}\mathbb{E}_N^0[X(0-)]$$

with $\mathbb{E}_N^0$ the Palm probability of N;

- **Pasta implies $\mathbb{E}_N^0[X(0-)] = \mathbb{E}[X(0)]$, which gives**

$$\mathbb{E}[X(0)] = \frac{2}{\lambda R^2}$$

- **The packet loss probability p is such that $p\mathbb{E}[X(0)] = \lambda$; Hence**

$$\mathbb{E}[X(0)] = \sqrt{\frac{2}{pR^2}}.$$

RD Case

- If there exists a stationary regime where N has intensity λ and X is integrable, the **Rate Conservation Principle** gives

$$\frac{1}{R^2} = \frac{\lambda}{2}\mathbb{E}^0_N[X(0-)]$$

- From **Papangelou's theorem** $\mathbb{E}^0_N[X(0-)] = \mathbb{E}[X(0)\frac{pX(0)}{\lambda}]$ so that

$$\mathbb{E}[X(0)^2] = \frac{2}{pR^2}$$

- No simple identification of the mean TCP throughput.

Markov Analysis

- $X(t)$ **is a Markov Process**
  - **with continuous time**
  - **with continuous state space**
- **It falls in the Piecewise Deterministic Process framework of Davis.**
- **The embedded chain (at ”discontinuities”) is geometrically ergodic.**

Analytical Results: Distributions - RI Case

- **For all $u > 0$ for all continuity point of $X(t)$,**

$$\dot{X}^u(t) = uX^{u-1}(t)\dot{X}(t) = uX^{u-1}(t)/R^2$$

so that

$$X^u(t) = X^u(0) + \int_0^t \frac{uX^{u-1}(v)}{R^2} dv - \left(1 - \frac{1}{2^u}\right) \int_0^t X^u(v-) N(dv)$$

Thus

$$M(t) = X^u(t) - X^u(0) - \frac{u}{R^2} \int_0^t X^{u-1}(v) dv + \lambda \left(1 - \frac{1}{2^u}\right) \int_0^t X^u(v-) dv$$

is a martingale s.t. $M(0) = 0$ so that whenever moments are finite

$$\frac{\partial}{\partial t} \mathbb{E}[X^u(t)] = \frac{u}{R^2} \mathbb{E}[X^{u-1}(t)] - \lambda \left(1 - \frac{1}{2^u}\right) \mathbb{E}[X^u(t)].$$

Analytical Results: Distributions - RI Case (*continued*)

- **Mellin transforms of the density of X at time t:**

$$\mathbb{E}[X^u(t)] = \int_0^\infty z^u f(t,z)dz = \widehat{f}_t(u+1).$$

- **Functional equation:**

$$\frac{\partial}{\partial t}\widehat{f}_t(u+1) = \frac{u}{R^2}\widehat{f}_t(u) - \lambda\left(1 - \frac{1}{2^u}\right)\widehat{f}_t(u+1)$$

- **PDE**

$$\frac{\partial f(z,\, t)}{\partial t} + \frac{1}{R^2}\frac{\partial f(z,\, t)}{\partial x} + \lambda\left(f(z,\, t) - 2f\left(2z,\, t\right)\right) = 0$$

Analytical Results: Distributions - RI Case (*continued*)

- **Stationary ODE:**

$$\frac{df(z)}{dz} + \xi\left(f(z) - 2f\left(2z\right)\right) = 0$$

with $\xi = \lambda R^2$**.**

- **Stationary functional equation:**

$$u\widehat{f}(u) = \xi\left(1 - \frac{1}{2^u}\right)\widehat{f}(u+1)$$

- $\widehat{f}(u) = g(u)\Gamma(u)\xi^{-u}$. **Then**

$$g(u) = g(u+1)(1-2^{-u}), \quad i.e. \quad g(u) = g(\infty)\prod_{k\geq 0}(1-2^{-u-k}),$$

Analytical Results: Distributions - RI Case (*continued*)

- **Theorem**

The unique stationary distribution solution of this functional equation has for Mellin transform

$$\widehat{f}(u) = \phi\Gamma(u)\xi^{-u}\prod_{k\geq 0}(1-2^{-u-k})$$

with $\xi = \lambda R^2$ **and** $\phi = \xi\left(\prod_{k\geq 1}(1-2^{-k})\right)^{-1}$.
The associated probability density is

$$f(z) = \phi\sum_{n\geq 0} b_n e^{-(\xi 2^n)z}$$

with $b_0 = 1$ **and** $b_n = (-1)^n \prod_{k=1}^{n} \frac{2}{(2^k-1)}$.

Distributions - RD

- **Formal proof of PDE by the same martingale approach:**

$$\frac{\partial f}{\partial t}(z,t) + \frac{1}{R^2}\frac{\partial f}{\partial z}(z,t) = p(4zf(2z,t) - zf(z,t)), \quad z \geq 0,$$

  - **mass leaves the interval** $[z, z+dz]$ **at rate** $pzf(z,t)dz$ **approximately.**
  - **mass enters this interval because of losses among throughputs in the interval** $[2z, 2(z+dz)]$ **at rate** $p2zf(2z,t) \cdot 2dz$

- **Functional equation for the stationary Mellin transform of** f**:**

$$u\widehat{f}(u) = \xi\widehat{f}(u+2)\left(1 - 2^{-u}\right).$$

with $\xi = pR^2$**.**

Distributions - RD (*continued*)

- **Theorem The unique density satisfying the ODE is**

$$f(z) = 2\phi \sum_{n\geq 0} a_n e^{-\left(\frac{\xi}{2}4^n\right)z^2}$$

with

$$\phi = \left(\sqrt{\pi}\left(\frac{2}{\xi}\right)^{\frac{1}{2}} \prod_{k\geq 1}(1-2^{-2k+1})\right)^{-1} \quad \textbf{and} \quad a_n = (-1)^n \prod_{k=1}^{n} \frac{4}{(4^k-1)}.$$

Its Mellin transform is

$$\widehat{f}(u) = \phi\Gamma\left(\frac{u}{2}\right)\left(\frac{2}{\xi}\right)^{\frac{u}{2}} \Pi_\infty(u), \quad \textbf{with} \quad \Pi(u) = \prod_{k=0}^{\infty}\left(1-2^{-u-2k}\right)$$

Its mean is

$$\mathbb{E}[X(0)] = \sqrt{\frac{2}{pR^2}}\sqrt{\frac{1}{\pi}}\frac{\Pi(2)}{\Pi(1)} \sim \frac{1.309}{R\sqrt{p}}.$$

Scalable TCP Distributions - RD

- **SDE: (with $\alpha = \frac{a}{R}$):**

$$dX(t) = \alpha X(t)dt - (1-b)X(t)N(dt)$$

- **PDE:**

$$\frac{\partial f(x,t)}{\partial t} + \alpha x \frac{\partial f(x,t)}{\partial x} = -pxf(x,t) - \alpha f(x,t) + \frac{1}{b^2} pxf\left(\frac{1}{b}x, t\right),$$

- **ODE**

$$\alpha x \frac{df(x)}{dx} = -pxf(x) - \alpha f(x) + \frac{1}{b^2} pxf\left(\frac{1}{b}x\right).$$

Scalable TCP Distributions - RD (*continued*)

- **Mean throughput: Ott 05, Altman 05**

$$\mathbb{E}[X] = -\frac{a}{pR\log b}.$$

- **Mellin of stationary distribution:**

$$\hat{f}(u) = \Psi\Gamma(u-1)\left(\frac{\alpha}{p}\right)^{u-1}\prod_{k\geq 0}\left(1-b^{u+k-1}\right).$$

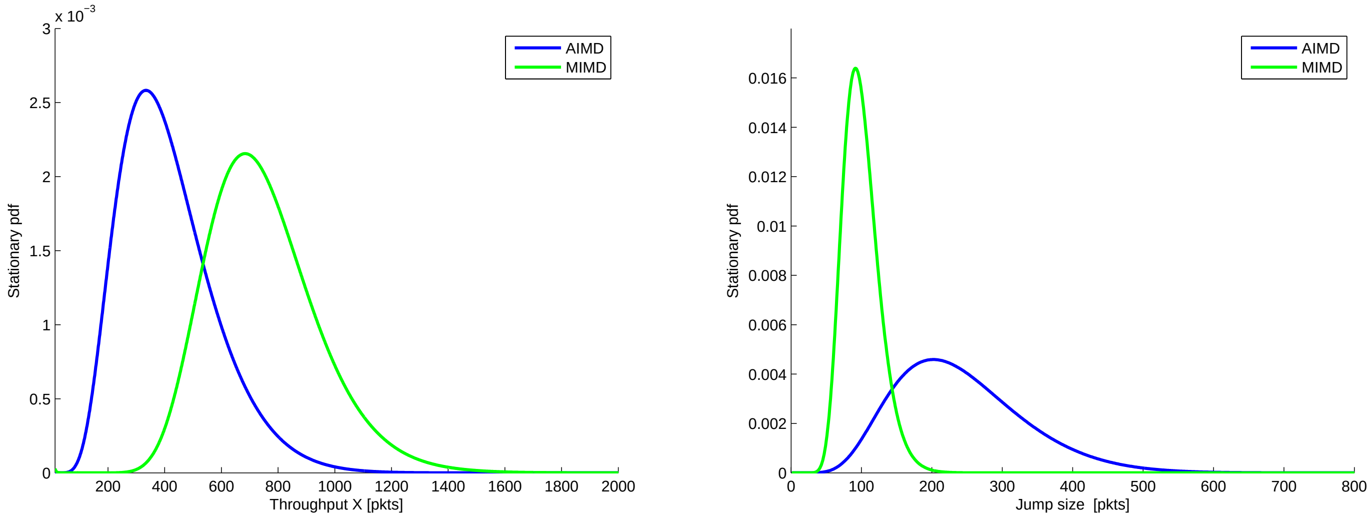
- **Distribution:**

$$f(x) = \Psi\frac{1}{x}\sum_{n\geq 0} c_n e^{-\left(\frac{pR}{a}\left(\frac{1}{b}\right)^n\right)x},$$

where

$$c_n = (-1)^n \prod_{k=1}^{n} \frac{b^{-1}}{(b^{-k}-1)} \quad \textbf{and} \quad \Psi = \frac{1}{\log\frac{1}{b}}\left[\prod_{k\geq 0}\left(1-b^{k+1}\right)\right]^{-1}. \qquad (1)$$

Comparison RENO–Scalable TCP



Left: throughput density; Right: jump size density
Parameters: $p = 0.001$, $R = 100ms$, $a = 0.01$, $b = 0.875$
Jump size: for all bounded measurable $\phi(\cdot)$**:**

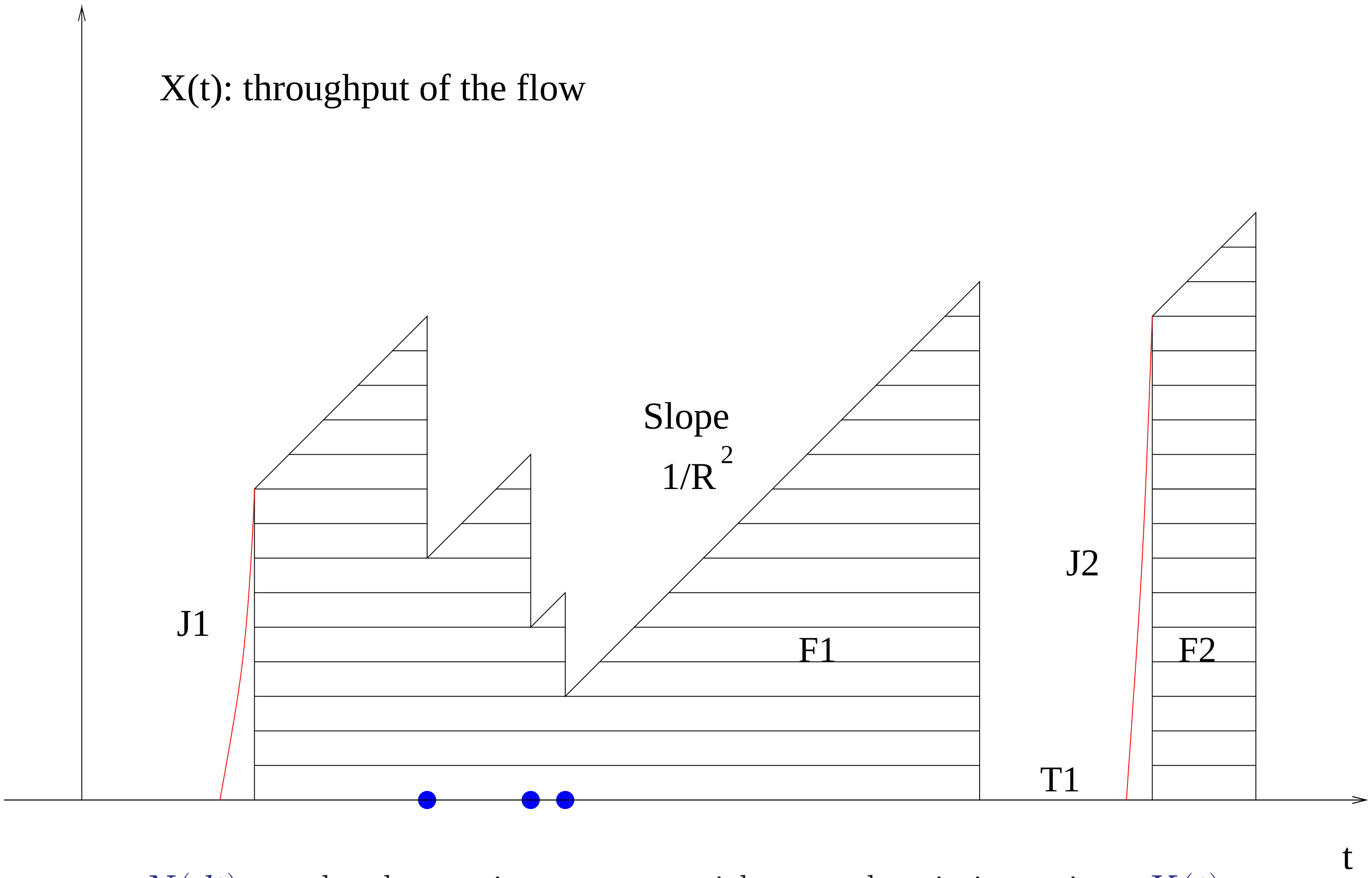
$$\mathbb{E}^0[\phi(X^-)] = \frac{\mathbb{E}[\phi(X)pX]}{p\mathbb{E}[X]} = \frac{1}{\mathbb{E}[X]}\int_0^\infty \phi(x)f(x)x dx,$$

ON-OFF Flows

- **On-Off TCP flow**
  - **PDE for the exponential Model**
  - **Moments**
  - **Distributions**

Dynamics

- RD model with packet loss probability p and RTT R.
- The flow alternates between document downloads and think times, inducing an **ON/OFF flow structure**
  - Document sizes F_i are i.i.d. with mean $1/\mu$
  - Think times T_i are i.i.d. with mean $1/\beta$
- Motivation: HTTP 1.1 where the files successively downloaded by a flow use the same TCP-Reno connection:
  - **Slow Start** jump approximation: jumps J_i are i.i.d.



$N(dt)$: packet loss point process, with a stochastic intensity $pX(t)$.

Exposnential Model

- File sizes are exponential with parameter μ
- Think times are exponential with parameter β
- Slow start jump is a bounded random variable with law H (for example with density h).

- $X(t)$ is a Markov Process
  - with continuous time
  - with continuous state space (with an atom)
- It falls in the Piecewise Deterministic Process framework of Davis.
- The embedded chain (at "discontinuities") is geometrically ergodic.

PDE

- $(f(z,t),\nu(t))$ **distribution of throughput at time t:**
- **PDE**

$$\frac{\partial f}{\partial t}(z,t) + \frac{1}{R^2}\frac{\partial f}{\partial z}(z,t) \;=\; \beta\nu(t)h(z) - \mu z f(z,t) + 4zpf(2z,t) - zpf(z,t)$$

- **Boundary:**

$$\frac{d\nu}{dt}(t) = \int_0^\infty \mu z f(z,t)dz - \beta\nu(t).$$

- **Normalization:**

$$\int_0^\infty f(z,t)dz = 1 - \nu(t).$$

ODE for Stationary Regime

- **ODE: since $\int_0^\infty \mu v f(v) dv = \beta\nu$,**

$$\frac{df(z)}{dz} = \beta\nu R^2 h(z) - \mu R^2 z f(z) + 4zpR^2 f(2z) - zpR^2 f(z)$$
$$= \mu R^2 h(z) \int_0^\infty v f(v) dv - \mu R^2 z f(z) + 4pzR^2 f(2z) - zpR^2 f(z)$$

- **Mellin**

$$u\widehat{f}(u) = -\mu R^2 \widehat{f}(2)\widehat{h}(u+1) + \mu R^2 \widehat{f}(u+2) + pR^2 \widehat{f}(u+2)\left(1 - 2^{-u}\right)$$

Moments

- $\mathbb{E}T$: **mean time to transfer a file**
- $\mathbb{E}X$: **mean stationary throughput**
- **Cycle formula (thanks to the regenerative structure):**

$$\mathbb{E}X = \frac{\frac{1}{\mu}}{\frac{1}{\beta} + \mathbb{E}T}$$

- **Probability that a flow is OFF :** $\nu = \mathbb{E}X \frac{\mu}{\beta}$

No Slow Start

- **Theorem** (From Mellin)
  - The mean time to transfer a file is

$$\mathbb{E}T = \frac{1}{\mu}\sqrt{\frac{\pi}{2}}R\frac{\prod_{l=1}^{\infty}\left(1-\frac{2p}{p+\mu}4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1-\frac{p}{p+\mu}4^{-l}\right)}\sqrt{p+\mu}$$

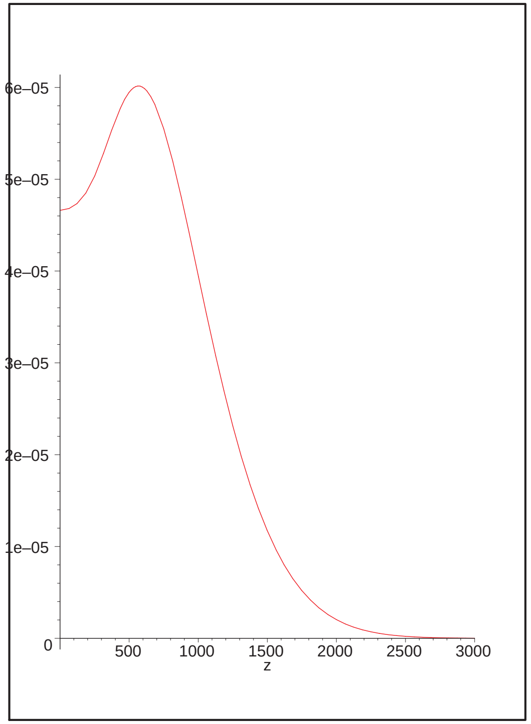
  - The density $f(z)$ of the throughput at $z > 0$ is

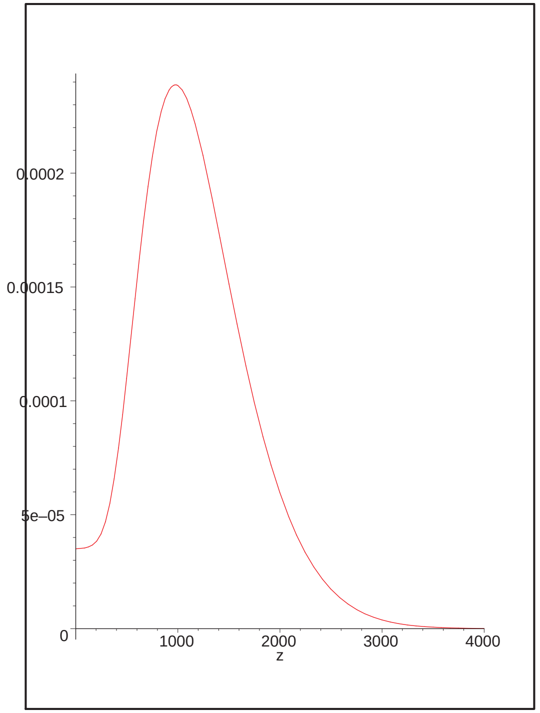
$$f(z) = \frac{\widehat{f}(2)}{\prod_{l=1}^{\infty}(1-\frac{p}{p+\mu}2^{-2l})}(p+\mu)R^2\sum_{n\geq 0} a_n e^{-(\frac{(p+\mu)R^2}{2}4^n)z^2}$$

    with

$$a_n = (-\frac{p}{p+\mu})^n \prod_{l=1}^{n}\frac{4}{4^l-1}.$$

Stationary Density





Left $R = 0.1$ **s.,** $1/\beta = 2$ **s.,** $1/\mu = 100$ **and** $p = 1\%$**.**
Right $R = 0.1$ **s.,** $1/\beta = 2$ **s.,** $1/\mu = 1000$ **and** $p = 1\%$**.**

With Slow Start

- **Theorem (From Mellin)**

$$\mathbb{E}T = \frac{1}{\mu}\sqrt{\frac{\pi}{2}} R \frac{\Pi_\infty(1)}{\Pi_\infty(2)} \sqrt{p+\mu} + \frac{\sqrt{\pi}R^2}{2} \sum_{k=0}^{\infty} \left(\Pi_k(2) \frac{\Pi_\infty(1)}{\Pi_\infty(2)} \frac{\left(\frac{(p+\mu)R^2}{2}\right)^{k+\frac{1}{2}}}{(k+1)!} \widehat{h}(2k+3) - \Pi_k(1) \frac{\left(\frac{(p+\mu)R^2}{2}\right)^{k}}{\Gamma(k+\frac{3}{2})} \widehat{h}(2k+2) \right)$$

with

$$\Pi_k(u) = \prod_{l=0}^{k-1} \left(1 - \frac{p}{p+\mu} 2^{-u-2l}\right).$$

Distribution of Throughput for Slow Start
A More General Class of ODEs

We consider the more general equation

$$\frac{df(z)}{dz} = \delta T(f)A(z) - \mu z^{\gamma-1} f(z) + \beta z^{\gamma-1}(\rho^{\gamma} f(\rho z) - f(z)), \quad z \geq 0,$$

where

$$T(f) = \int_0^\infty z^{\gamma-1} f(z) dz,$$

$\delta \geq \mu$,
A **is a probability density function,**
$\gamma \geq 1$,
$\rho > 1$,
$\beta \geq 0$,
$\mu \geq 0$,
$\mu + \beta > 0$.

RD Special Case

$\gamma = 2$,
$\rho = 2$,
$pR^2 \to \beta$,
$\mu R^2 \to \delta$,
$h(z) \to A(z)$,
$\mu R^2 \to \mu$

- **One gets back the initial RD ODE**

$$\frac{df(z)}{dz} = \mu R^2 \int_0^\infty v f(v) dv h(z) - \mu R^2 z f(z) + 4pzR^2 f(2z) - zpR^2 f(z)$$

RI Special Case

$\gamma = 1$,
$\rho \to \theta$,
$\lambda R^2 \to \beta$,
$\mu R^2 \to \delta$,
$h(z) \to A(z)$,
$\mu R^2 \to \mu$

- **One gets the RI ODE**

$$\frac{df(z)}{dz} = \mu R^2 h(z) - \mu R^2 f(z) + \lambda R^2(\theta f(\theta z) - f(z)), \quad z \geq 0,$$

which represents the AIMD on-off dynamics when

- **losses occur according to a Poisson point process of intensity λ, leading to a division of the throughput by θ;**
- **file lifetime (on-time) is exponentially distributed with parameter μ.**

General ODE

- **Theorem** **Assume that** $\gamma \geq 1$**,** $\rho > 1$**,** $\mu \geq \delta \geq 0$**,** $\beta \geq 0$**,** $\mu + \beta > 0$**. Let** $\theta = \rho^{\gamma}$ **and let** A **be a density such that**

$$\int_0^{\infty} A(z) e^{\frac{(\mu+\beta)}{\gamma} z^{\gamma}} dz < \infty.$$

Then the unique density solution to the ODE is the function

$$f(z) = \frac{1}{C\gamma} \sum_{n \geq 0} \left(\frac{\beta}{\mu + \beta} \right)^n b_n d_n(z) e^{-\left(\frac{\beta+\mu}{\gamma}\right) \theta^n z^{\gamma}}$$

General ODE (*continued*)

with

$$b_n = (-1)^n \prod_{k=1}^{n} \frac{\theta}{(\theta^k - 1)}$$

and

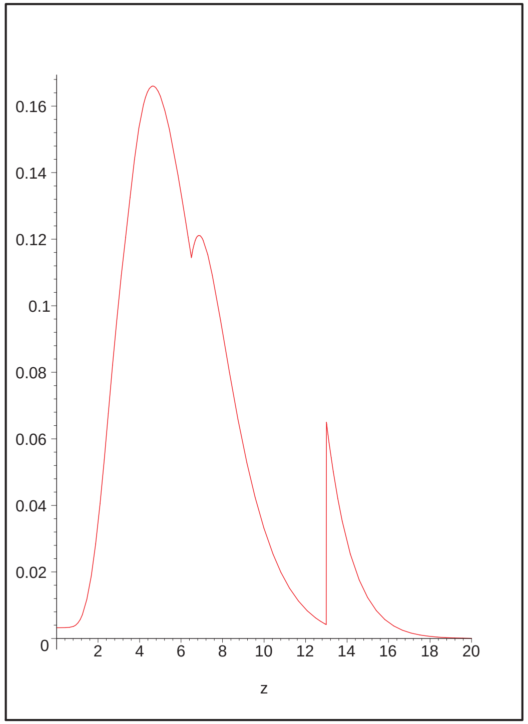
$$d_n(z) = \sum_{m \geq 0} c_m \left(\delta \int_0^{z\theta^{\frac{(n+m)}{\gamma}}} A(x) e^{\frac{\mu+\beta}{\gamma}\theta^{-m}x^{\gamma}} dx + (\mu - \delta) \right)$$

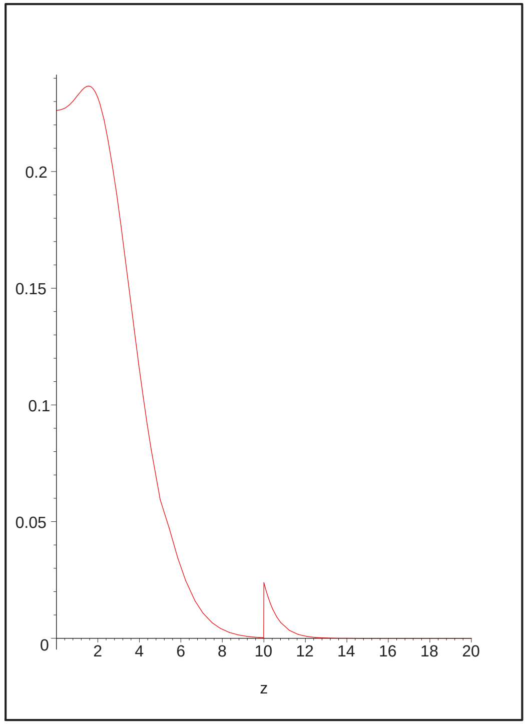
and

$$c_m = \left(\frac{\beta}{\mu + \beta} \right)^m \prod_{i=1}^{m} \frac{1}{1 - \theta^{-i}}$$

C**, the constant which normalizes** f**, is known in closed form.**

Examples of Stationary Densities with Slow Start





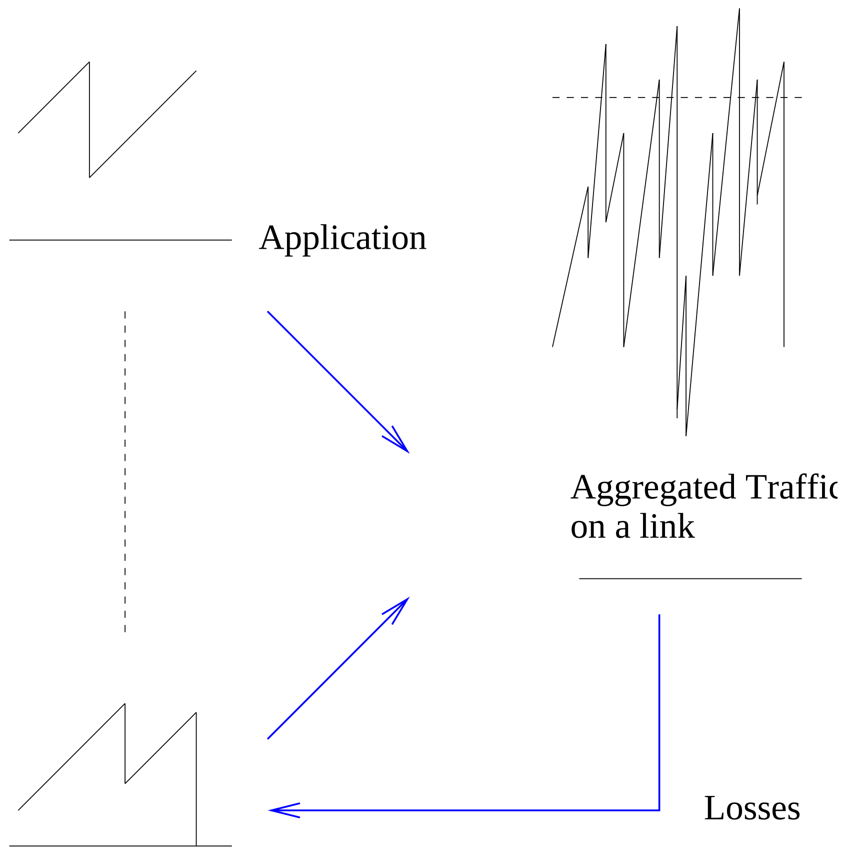
Left: $R = 1$ s., $1/\mu = 100$, $p = 5/100$ and $h = \delta_\eta$ with $\eta = 13$.
Right $R = 1$ s., $1/\mu = 5$ and $p = 5\%$ and $\eta = 13$.

Summary

- Isolated TCP flows
  - Persistent Flows
  - On-Off Flows
- Interaction of Parallel TCP Flows
  - Persistent Flows
  - On-Off Flows
- Interaction of TCP Flows in Series

Parallel Flow Competition

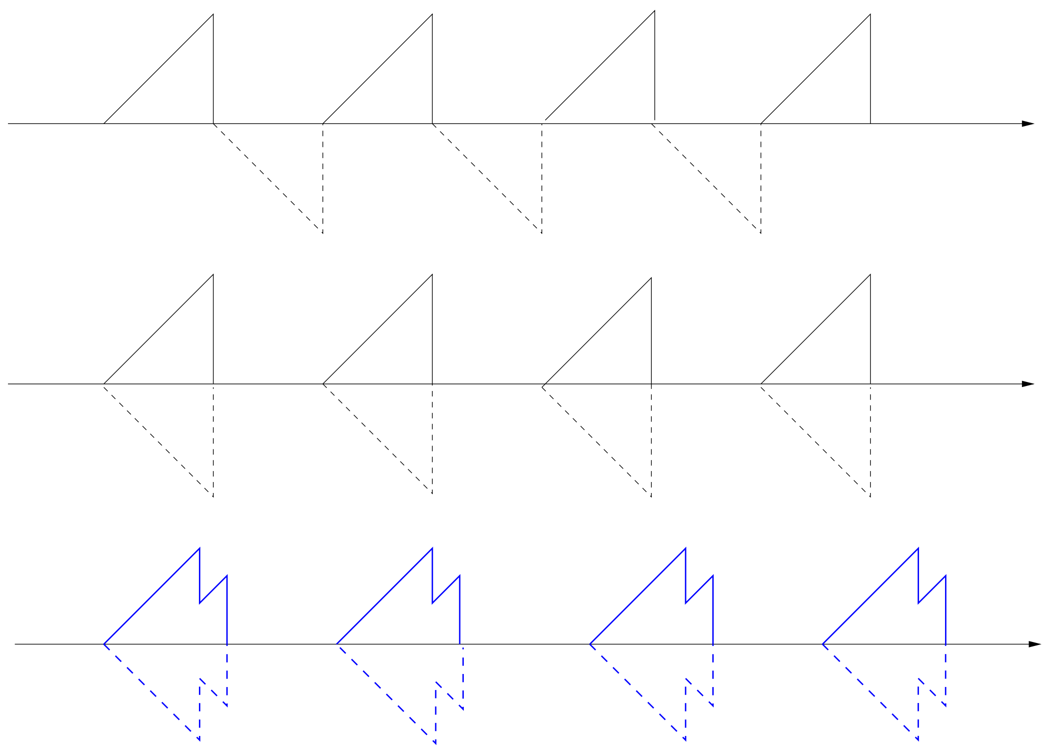
- The instantaneous throughput of a flow depends on losses, which result from the competition with other flows sharing the same links.
- Mean field



Parallel Flow Competition (*continued*)

- **AQM**: a loss Poisson point process of an appropriate intensity (either RI or RD) is used against each individual flow
  - **Mean field, Persistent case**: **F.B. & Mc Donald 02 Mc Donald & Reynier 06** in the AQM setting. Ongoing work of **Graham & Robert**.
- **TD**: when the sum rate raeches the capacity constraint, a proportion p of the flows experiences instantaneous losses
  - **Mean field, Persistent case**: **F.B. & Anantharam 11**

In Phase and Out of Phase AIMD, On-Off Flows in Parallel under TD



Interaction of On-Off TCP Flows on a TD Link

– N homogeneous HTTP users share a link of capacity NC

– Each HTTP user alternates between document downloads and think times, inducing an **ON/OFF flow structure**

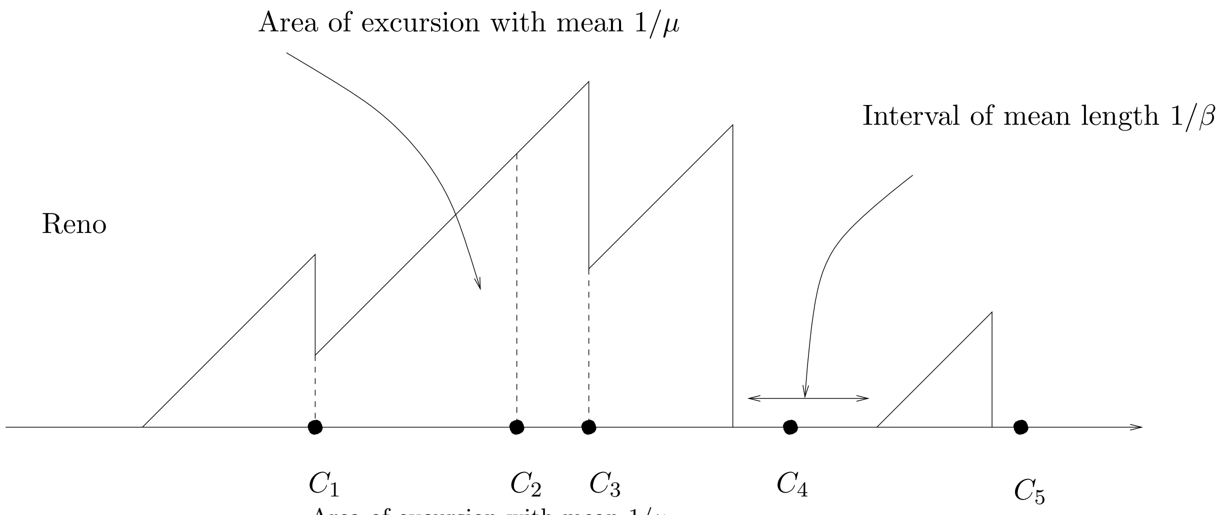
– Document sizes are i.i.d. with mean $1/\mu$

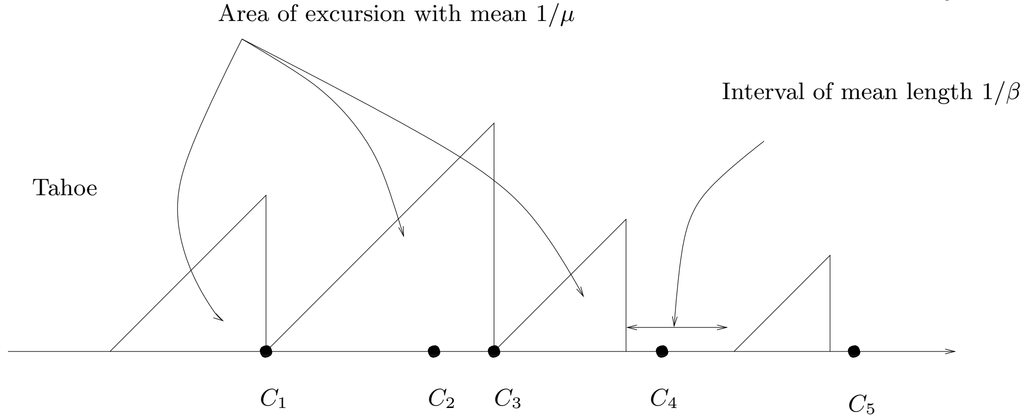
– Think times are i.i.d. $1/\beta$

– All connections have the same RTT R

– **Congestion** takes place as soon as the sum of the rates is equal to or exceeds NC

– Congestions result in an instantaneous halving of rate for a proportion p of the flows (**synchronization rate**).





Sample paths of the rate $X_n(t)$ of flow n

Mean Field Limit

- We let the population parameter N go to ∞
- We analyze

– the limiting aggregated rate

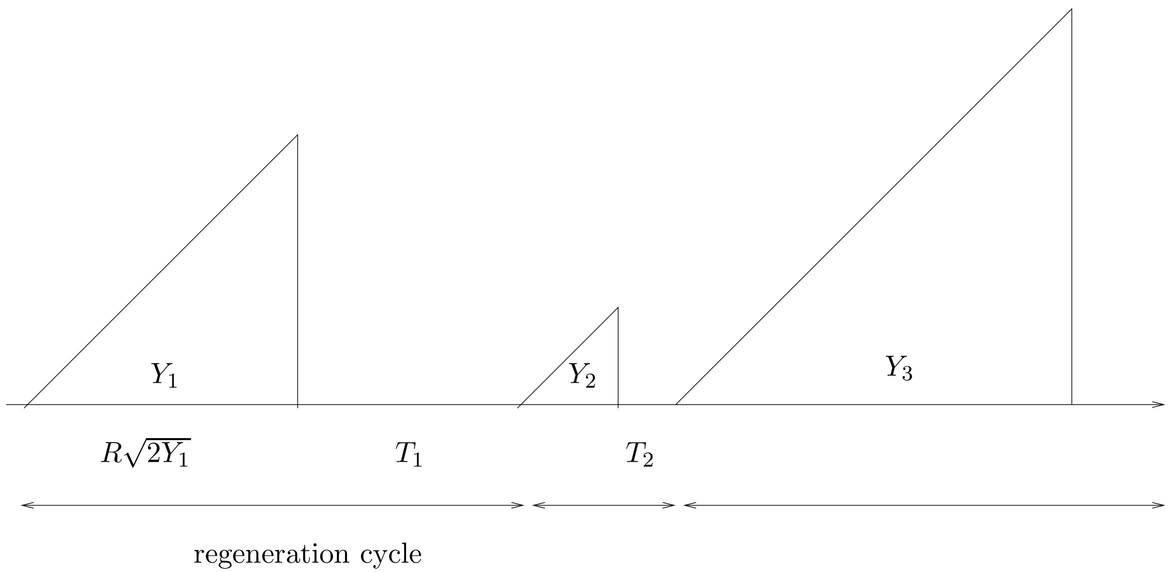
$$\alpha(t) = \lim_N \frac{1}{N} \sum X_n^N(t)$$

– the limiting distribution of rates

$$s(dz, t) = \lim_N \frac{1}{N} \sum \delta_{X_n^N(t) \in dz}$$

The Free Regime Mean Field Limit ($C = \infty$)

$\{F_i\}$ i.i.d. sequence of file sizes for tagged flow
$\{T_i\}$ i.i.d. sequence of think times for tagged flow
The rate $X(t)$ of the tagged flow is a regenerative process.



The Free Regime Mean Field Limit ($C = \infty$) (*continued*)

- **The mean field aggregated rate**

$$\alpha(t) = \lim_N \frac{1}{N} \sum X_n(t)$$

exists and is deterministic as well as

– $s(z, t)$ **the proportion of flows active, with rate** z **at time** t

– $\nu(t)$ **the proportion of flows inactive at time** t

The Free Regime Mean Field Limit ($C = \infty$) (*continued*)

- **Example of analytical characterization in the exponential F and T case: PDE for the congestion-less mean field density:**

$$\begin{aligned} \frac{\partial}{\partial t} s(z,t) + \frac{1}{R^2}\frac{\partial}{\partial z} s(z,t) &= -\mu z s(z,t) \\ \frac{d}{dt}\nu(t) &= -\beta\nu(t) + \mu \int_0^\infty z s(z,t) dz \end{aligned}$$

with

$$s(0,t)/R^2 = \beta\nu(t) \qquad \textbf{and} \qquad \int_0^\infty s(z,t)dz = 1 - \nu(t).$$

The Free Regime Mean Field Limit ($C = \infty$) (*continued*)

- **Transient distribution via Laplace transform in time**

$$s(z,t)=s(z-\frac{t}{R^2},0)\ e^{-\mu\left(tz-\frac{t^2}{2R^2}\right)}+e^{-\mu R^2\frac{z^2}{2}}R^2\beta\left(1-\int_0^\infty s(x,t-zR^2)dx\right)$$

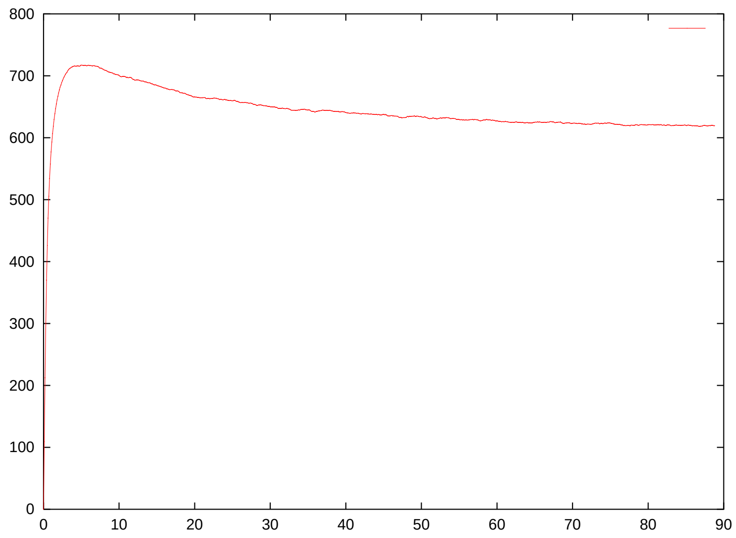
The Free Regime Mean Field Limit ($C = \infty$) (*continued*)

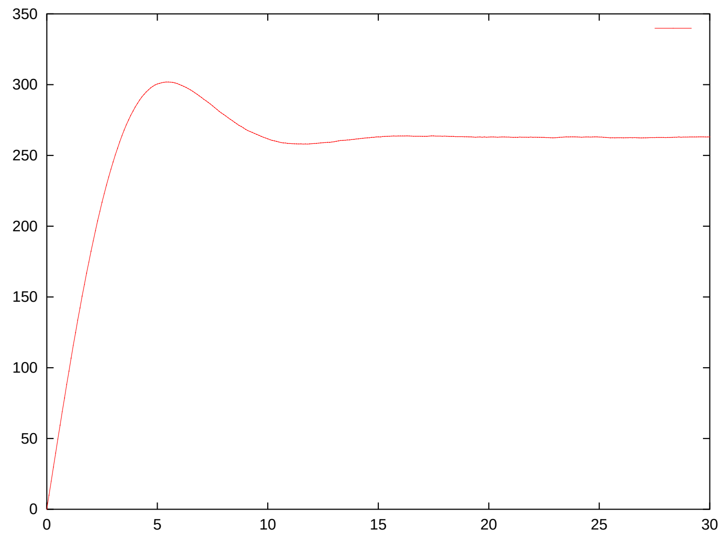
- **Stationary regime (from previous analysis)**

$$\begin{aligned}
\nu(\infty) &= \frac{\frac{1}{\beta}}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}} \\
s(z,\infty) &= \frac{R^2 e^{-R^2\mu z^2/2}}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}} \\
\alpha(\infty) &= \frac{1}{\mu}\frac{1}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}} = \rho.
\end{aligned}$$

- ρ**: load per user in the steady state congestion-less regime.**

Examples of Aggregated Rates





All flows are initially active and with 0 rate;
Mean file size 2000; mean think time 2 sec.

Left: lognormal F with st. dev. 4 $\times$ the mean and R=30 ms.
Right: exponential with R=100 ms.

The Finite Capacity Mean Field Limit ($C < \infty$)

- For all finite C, there exists a deterministic mean-field limit with a sequence of intercongestion times τ_1, τ_2, ... (finite or not).
- Proof based on mean field
- If one of the τ_i's is infinite, the stationary mean field limit for C is an interaction-less regime (similar to the free regime);
- If all τ_i's are finite, the stationary mean field limit for C is an interaction regime; of special interest: periodic interaction regimes with $\tau_i = \tau$.

Necessary Condition for a Periodic Congestion Interaction Mean Field Limit Regime

- **Necessary condition** for the existence of a periodic interaction mean-field regime with intercongestion time $\tau < \infty$: τ should solve the **Rate Conservation Principle** equation:

$$\frac{\mathbb{P}(X(0) > 0)}{R^2} = \frac{pC}{2\tau} + \lambda_\delta \mathbb{E}_0^\delta[X(0^-)] \quad \textbf{[Reno]}$$

- In the exponential case, all terms in this fixed point equation are computable thanks to the **regenerative structure**.

- **Regeneration** when tagged flow is inactive at a congestion.

Congestion Regime: the Invariant Measure Equation

- $s_0(z)$ **the proportion of flows active and with rate** z **at a congestion epoch**
- ν_0 **the proportion of flows inactive at a congestion epoch**

■ **Invariant measure equation associated with** τ**:**

$$s_0(z) = (1-p)S_0(z,\tau) + pS_0(2z,\tau) \quad \textbf{[Reno]},$$

where $S_0(z,t)$ **is the solution of the congestion-less PDE with the initial condition** s_0**.**

Congestion Regime: the Invariant Measure Equation (*continued*)

- **The existence of a "good" solution to the invariant measure equation, i.e. of a probability measure $(\nu_0, s_0(z))$**
  - **solution of the invariant measure equation for τ**
  - **such that**

$$\alpha_0(\tau) = C \quad \textbf{and} \quad \alpha_0(t) < C \quad \textbf{for all} \quad t < \tau$$

is necessary and sufficient for the existence of a congestion periodic regime of period τ

- **The time average mean rate of a flow and the time average rate distribution of a flow can be expressed from this (cycle formula).**

Multiplicity of Stationary Mean Field Regimes

- If $\rho > C$, the congestion-less regime is impossible.
- **Main Finding** (proved in the Tahoe case, numerical evidence in the Reno case):
  1. The condition $\rho < C$ is not sufficient for having an interaction-less mean-field regime only
  2. There exist values of C such that depending on the initial condition, one enters either in an interaction-less or in an interaction stationary regime.

HTTP Turbulence

- We call **turbulent regime** the periodic congestion regime when $\rho < C$.
- **Rationale**:
  - for an appropriate phasing of the flow (e.g. stationary), there would be no congestion
  - for other initial conditions, in phasing and synchronization jointly lead to the perpetuation of a periodic congestion regime

Turbulence: Scenario 1 – Numerical Evidence

- **Exponential model,** $1/\mu = 2000$ **Pkts,** $1/\beta = 2$**,** $p = 0.8$ **and** $R = 0.1s$**. The load factor** ρ **is then around** 263 **Pkts/sec. We take** $C = 270$ **Pkts/sec.**
- **When the initial condition is the stationary law of the interactionless regenerative rate process, no congestions occur at all since** $\rho < C$**.**
- **When the initial condition is with all sources initially active and with 0 rate, periodic congestion regime with** $\tau \sim 3.7s$**.**
- **Backed by the following numerical evidence:**

– τ **is one of the two solutions for the RCP**

– **the invariant measure equation has a "good" solution for** τ

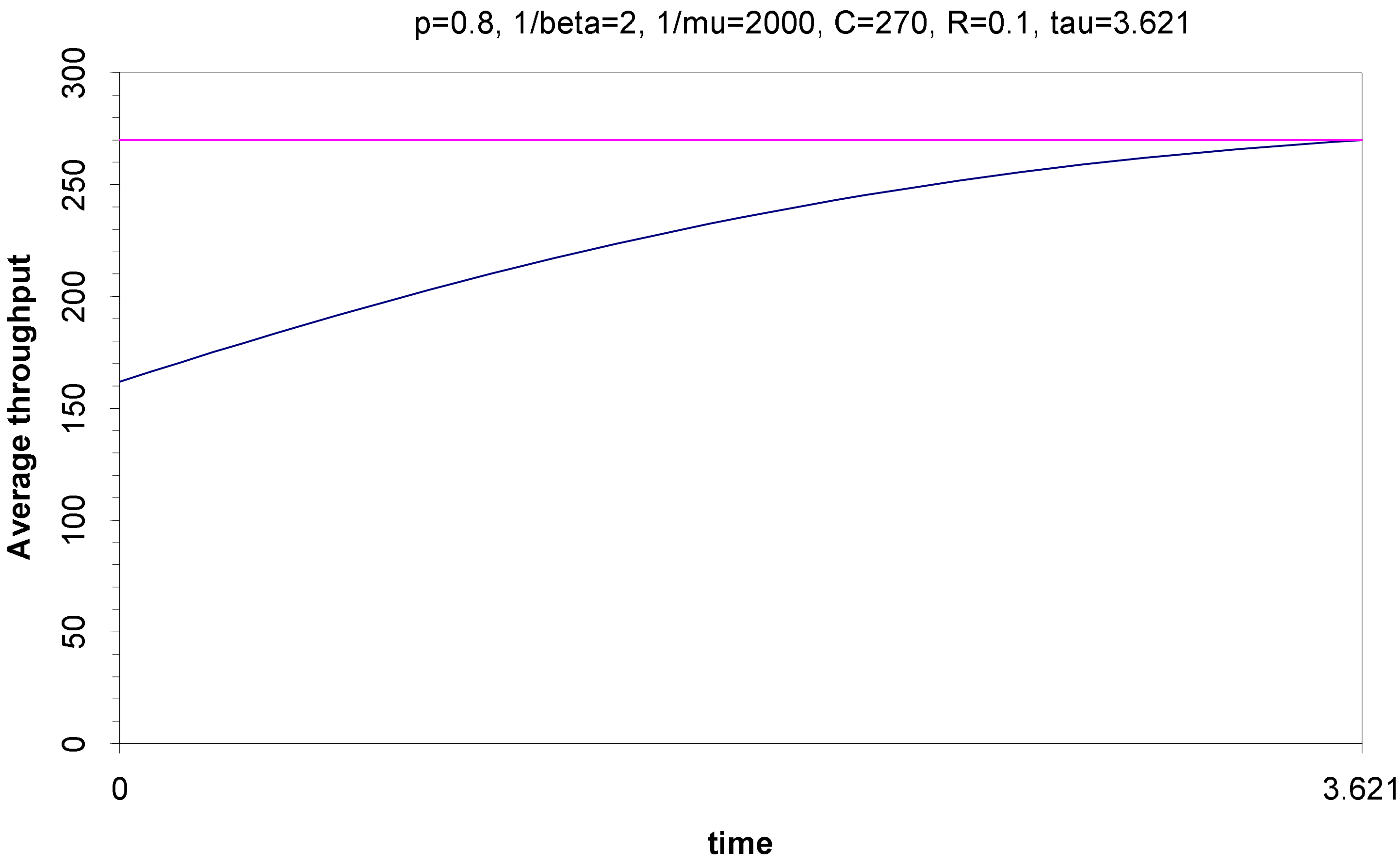


Figure 1: Evolution of the congestion-less aggregated rate with the time

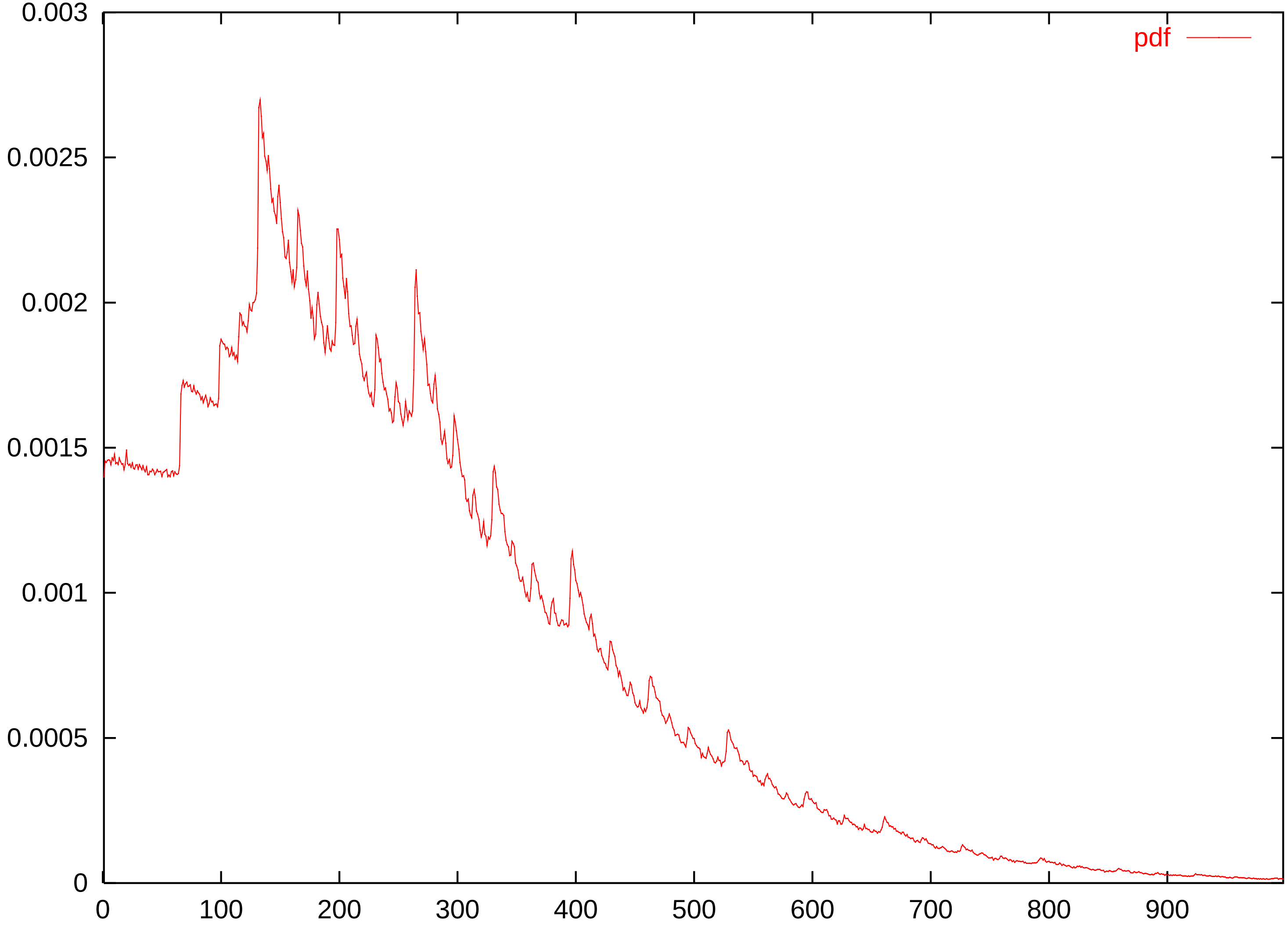


Figure 2: Invariant rate pdf

Turbulence: Scenario 1 – Simulation Evidence

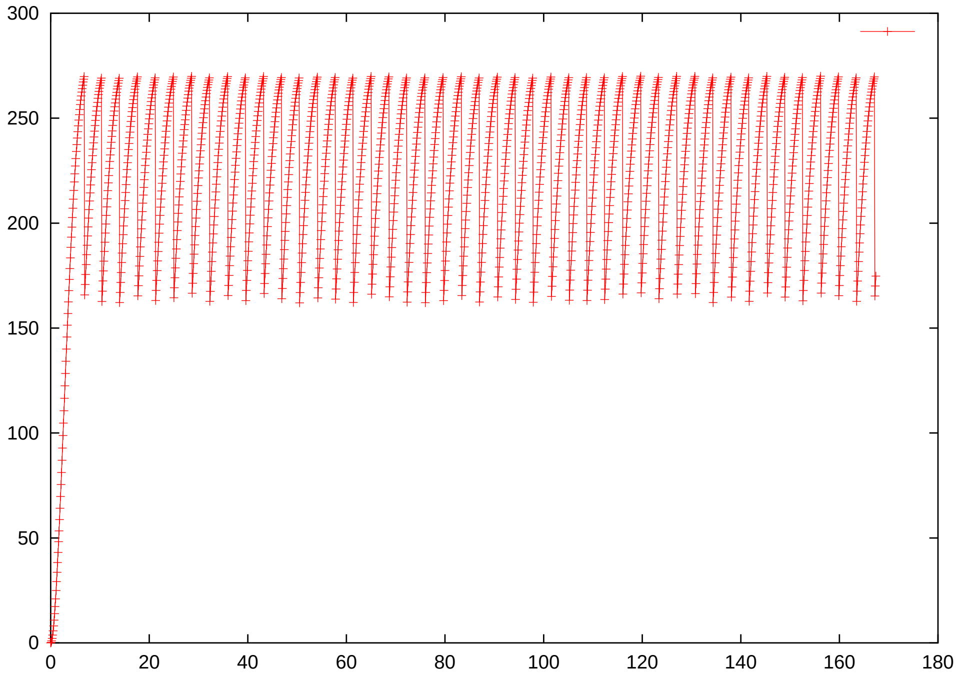
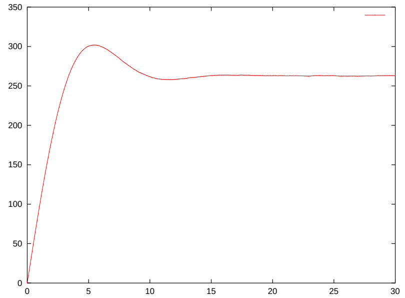


Figure 3: Evolution of aggregated rate when all flows are initially active and with null rate for $C = 270$ Pkts/sec.

Turbulence: – Mathematical proof (Tahoe Case)



– M: **maximum of** $\alpha(t)$ **over all** $t > 0$;

– m: **minimum of** $\alpha(t)$ **over all** $t > \tau$;

– γ: **minimum of** $1 - \nu(t)$ **over all** $t > 0$

for the initial condition with all flows active and with 0 rate.

Let

$$C_T = p\gamma M + (1 - p\gamma)m.$$

Lemma If $C_T > \rho$, then the Tahoe version of the model has turbulence for all C in the interval $\rho \leq C \leq C_T$ for this initial condition.

- **No proof for Reno at this stage.**

Turbulence: Scenario 2 – Simulation

- **Lognormal file size and off-periods; file size has mean value 2000 Pkt and standard deviation 8669 Pkts, and the off-period has a mean value of 2 sec and a standard deviation of 8.7 sec**
- **TCP Reno,** $R = 0.03$ **s.,** $p = 0.8$**;**
- **We observe the same phenomenon concerning** α **as in the exponential case, with a first maximum at 717 Pkts/s, significantly larger than the horizontal asymptote at** $\rho = 620$ **Pkts/s.**
- **The turbulence region goes from** $C = 620$ **to** $C = 680$ **Pkts/s.**

Refinements

- These phenomena are also present when taking into account
  - Slow start (extension of the PDE approach)
  - Maximal window

Bistability of the Finite Population Model – Simulation

- The fact that the mean field limit has two stationary regimes for some values of the parameters translates into the existence of two stable regimes for any finite stochastic system with the same mean parameters, with rare oscillations from one stable regime to the other.
- Ongoing analysis with M. Lelarge & D. McDonald of the rarity of the transitions using Kiffer's discrete version of Wentzell-Freidlin's theory.

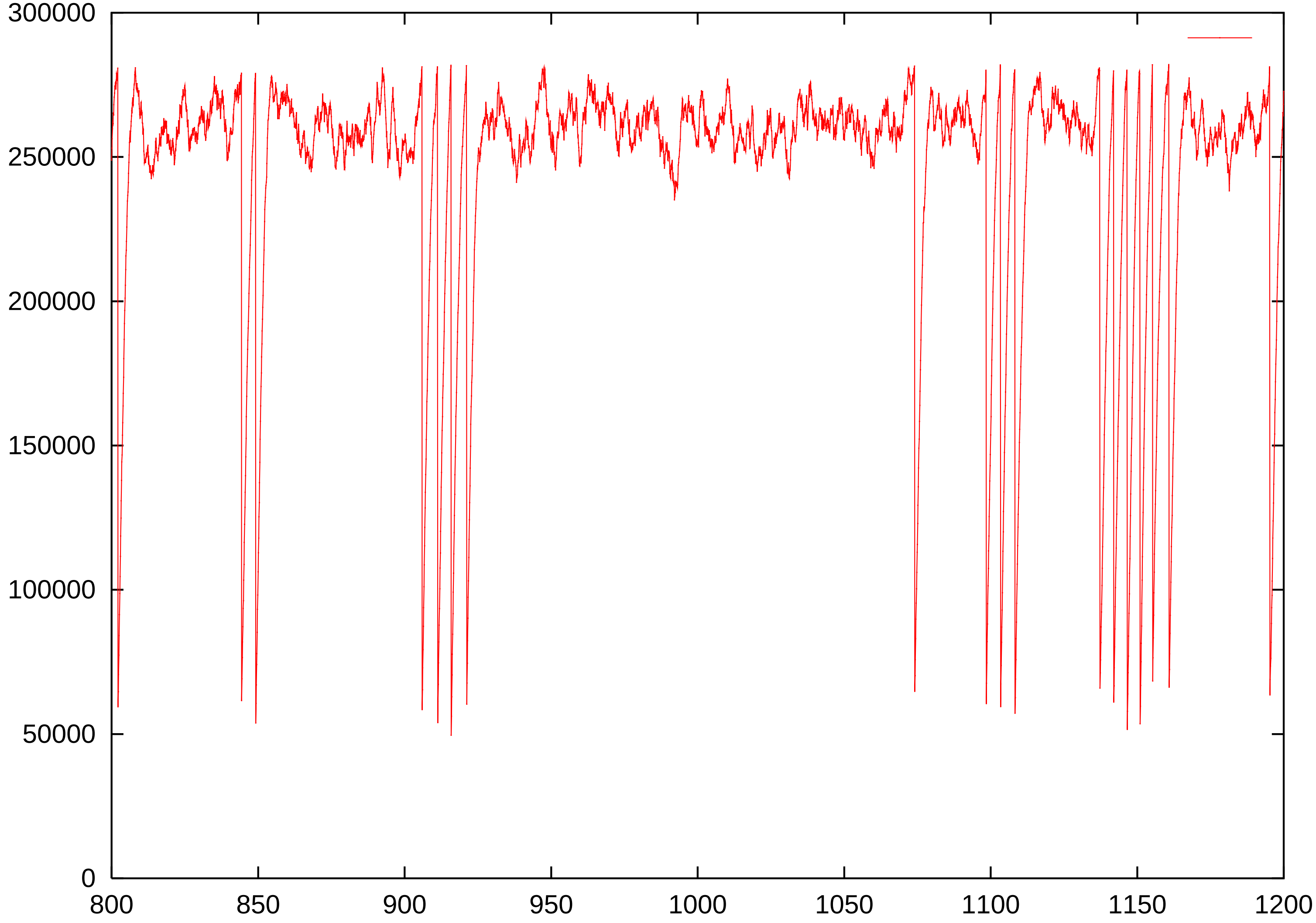


Figure 4: Bi-stability: 1000 Tahoe flows with $C = 282$.

Summary

- Isolated TCP flows
  - Persistent Flows
  - On-Off Flows
- Interaction of Parallel TCP Flows
  - Persistent Flows
  - On-Off Flows
- Interaction of TCP Flows in Series

Split TCP

– **Dynamics**

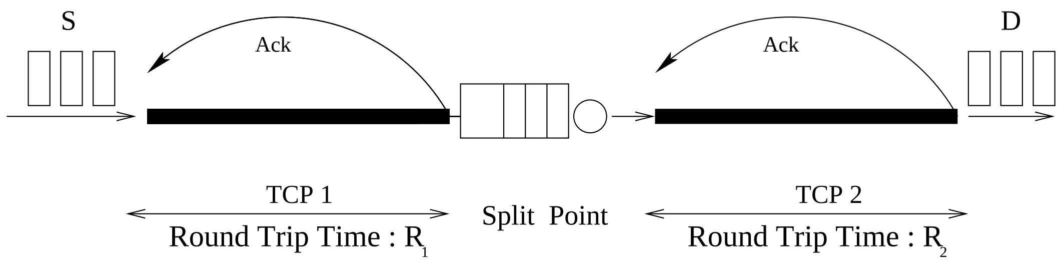
– **Stability**

– **Tails**

– **PDEs**

Split TCP Dynamics

- The split of a multihop TCP connection consists in replacing a plain end-to-end TCP connection by a series of TCP connections.



- Used in overlay networks; dominant in wireless networks (separation of the wireless and the wired parts);
- Used either with infinite buffer or with backpressure.

Notation

- $X(t)$, $Y(t)$: **the throughputs of** $TCP1$, $TCP2$ **time** t
- $M(dt)$, $N(dt)$: **the loss point process on** $TCP1$, $TCP2$
- λ, μ: **the loss point process intensity on** $TCP1$, $TCP2$ **in the RI case**
- p, q: **the packet error rate on** $TCP1$, $TCP2$ **in the RD case**
- R_1, R_2: **the local Round Trip Times**
- $Q(t)$: **the proxy buffer content at time** t
- B: **the proxy buffer size**

Phases

- Phase 1 or the free phase: the buffer is neither empty nor full, and $X(t)$ and $Y(t)$ evolve independently;
- Phase 2 or the starvation phase: the buffer is empty and Y is limited by the input throughput X
- Phase 3 or the backpressure phase: the buffer has reached it storage capacity B and X is forced by the backpressure algorithm to slow down to the rate Y at which the buffer is drained off.

No phase 3 if $B = \infty$.

Dynamics - Phase 1

- **In the free phase:**

$$\textbf{on } \{0 < Q(t) < B\} \quad \begin{cases} dX(t) = \alpha dt - \frac{X(t)}{2} M(dt) \\ dY(t) = \beta dt - \frac{Y(t)}{2} N(dt). \\ dQ(t) = X(t) - Y(t) \end{cases}$$

  where $\alpha = 1/R_1^2$, $\beta = 1/R_2^2$.

- **Rationale: the RENO AIMD rule + fluid dynamics for the queue.**

Dynamics - Phase 2

- **Potential rate of TCP2:** $Y(t) = W_2(t)/R_2$
- **In phase 2, the buffer is empty, which requires** $X(t) \leq Y(t)$**:**

$$\textbf{on} \quad \{Q(t) = 0\} \quad \begin{cases} dX(t) = \alpha dt - \frac{X(t)}{2} M(dt) \\ dY(t) = \beta \frac{X(t)}{Y(t)} dt - \frac{Y(t)}{2} N(dt). \end{cases}$$

- **Rationale for a diff. increase of** $Y(t)$ **proportional to** $\frac{X(t)}{Y(t)} \leq 1$**:**
  - **when the buffer is empty, since** $X(t) < Y(t)$**, the rate at which packets are injected in** $TCP2$ **and hence the rate at which** $TCP2$ **acks arrive is** $X(t)$**.**
  - **the window of** $TCP2$**,** W_2**, increases of** $X(t)dt/W_2(t) = dt\frac{X(t)}{R_2 Y(t)}$ **in the interval** $(t, t + dt)$
  - **the potential rate of** $TCP2$ **thus increases of** $\beta dt \frac{X(t)}{Y(t)}$ **during this interval.**

Dynamics - Phase 3

- **In the backpressure phase, which lasts until the buffer is saturated (this requires that $X(t) \geq Y(t)$):**

$$\textbf{on } \{Q(t) = B\} \quad \begin{cases} dX(t) = \alpha \frac{Y(t)}{X(t)} dt - \frac{X(t)}{2} M(dt) \\ dY(t) = \beta dt - \frac{Y(t)}{2} N(dt). \end{cases}$$

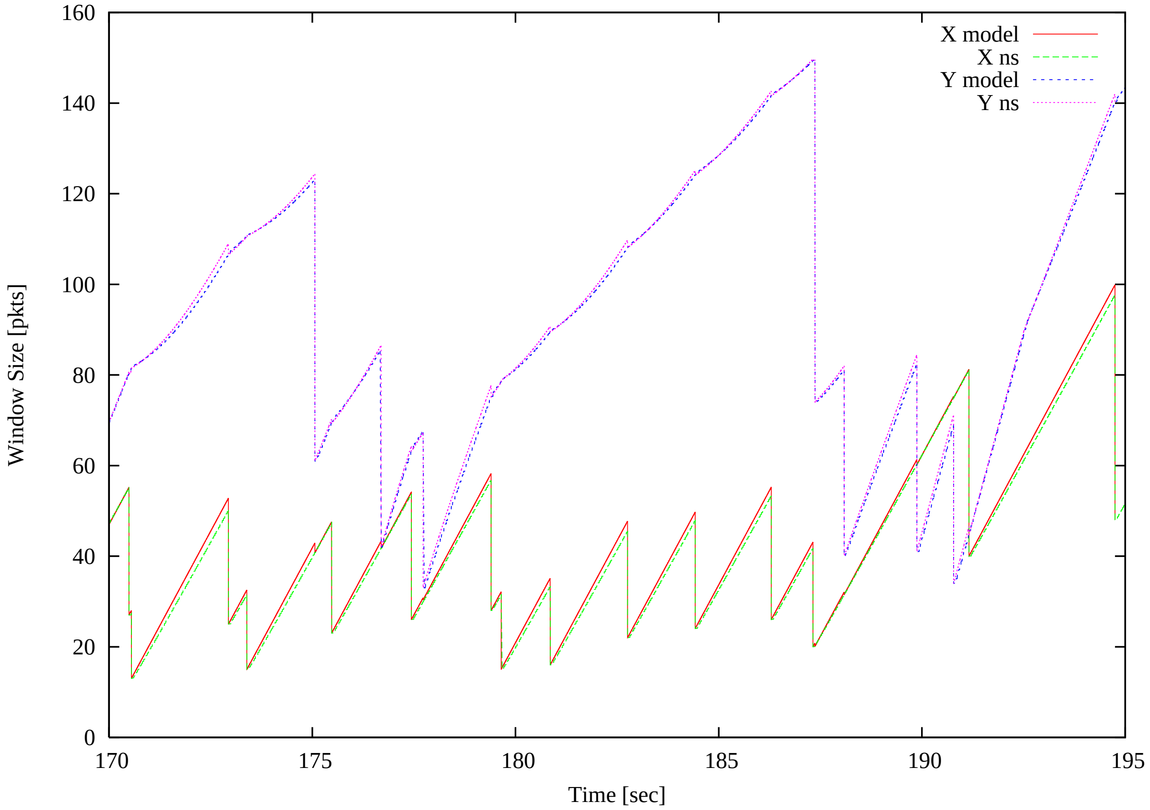
- **Rationale: acks of $TCP1$ now come back at a rate of $Y(t)$. Hence the congestion window, $W_1(t)$ of $TCP1$ grows at the rate $Y(t)/W_1(t)$.**

Global Queue Dynamics

$$Q(t) = \max\left(\sup_{0\le u\le t}\int_u^t (X(v)-Y(v))dv,\ Q(0)+\int_0^t (X(u)-Y(u))du\right).$$

- **Fluid queue with input $X(.)$ and drain $Y(.)$.**

Global Queue Dynamics (*continued*)



Infinite Buffer Case

Congestion windows: comparison between the model and *ns2*

Observations on Dynamics

- **The triple $(X(t), Y(t), Q(t))$ forms a Markov process on $\mathbb{R}_+^3$.**
- **Example of interaction with $B = \infty$**

– $X(.)$ **evolves freely.**

– $Y(.)$ **is slowed down by X whenever phase 2 is visited**

– **This slow down in turn affects the building of $Q(.)$....**

Stability - RI

- **Theorem** **If $\rho < 1$, where $\rho = \frac{\alpha\mu}{\beta\lambda}$, then the RI system is stable. If $\rho > 1$, then it is unstable.**
- **Proof based on a dynamical system backward construction leveraging monotonicity.**

Monotonicity in the RI, $B = \infty$ Case

- Denote by Y^f some fictitious process which evolves according to the dynamics of phase 1 only, built from the same realization of N.
  1. If $Y^f(t), \widehat{Y}^f(t)$ are built on the same realization of N, but depart from different initial conditions,
  $$Y^f(0) \leq \widehat{Y}^f(0) \quad \textbf{implies} \quad Y^f(t) \leq \widehat{Y}^f(t) \quad \forall t \geq 0.$$
  2. Let $Y_v^f(t)$, $t \geq v$ be the process starting from 0 at time v:
  $$Y_{v_1}^f(t) \geq Y_{v_2}^f(t), \quad \forall v_1 < v_2 \leq t.$$
  3. If $Y^f(0) = Y(0)$, then
  $$Y(t) \leq Y^f(t), \quad \forall t \geq 0.$$

Backward Construction

- $Q_t(0)$: **queue size at time** 0 **when departing from the following condition at time** $t < 0$:
  - **Queue size:** $Q(t) = 0$,
  - **TCP1: the stationary rate** $\widetilde{X}(t)$ **of** $TCP1$ **at** t **in isolation,**
  - **TCP2: the stationary rate** $\widetilde{Y}^f(t)$ **of** $TCP2$ **at** t **in isolation.**

$$Q_t(s) = \sup_{t \leq u \leq s} \int_u^s (\widetilde{X}(v) - Y_t(v))dv, \forall s \geq t,$$

with $Y_t(v)$ **the rate of TCP2 in at time** v **in Split TCP under the above assumptions.**

- **Stability: Does** $Q_t(0)$ **have an a.s. finite limsup when** t **tends to** $-\infty$**?**

Lower Bound

- From monotonicity property 3,

$$Q_t(0) \geq L_t = \sup_{t \leq u \leq 0} \int_u^0 (\widetilde{X}(v) - \widetilde{Y}^f(v))dv,$$

  with $\widetilde{Y}^f(.)$ the stationary free process for TCP2.

- This is a fluid input and fluid drain queue with
  input $(\widetilde{X}(.))$
  drain $(\widetilde{Y}^f(.))$

- The stochastic process $(\widetilde{X}(t), \widetilde{Y}^f(t))$ forms a stationary and **geometrically ergodic Harris chain**.

Upper Bound

- $\tau(t)$: **the beginning of the last busy period of** $Q_t(s)$ **before time 0 (0 if** $Q_t(0) = 0$ **and** t **if** $Q_t(s) > 0$ **for all** $t < s \leq 0$**).**

$$Q_t(0) = \int_{\tau(t)}^{0} (\widetilde{X}(v) - Y_t(v))dv \leq \int_{\tau(t)}^{0} (\widetilde{X}(v) - Y^f_{\tau(t)}(v))dv$$
$$\leq U_t = \sup_{t \leq u \leq 0} \int_{u}^{0} (\widetilde{X}(v) - Y^f_u(v))dv$$

- **The first inequality follows from**

1. **the fact that the dynamics on** $(\tau(t), 0)$ **is that of the free phase and**
2. **the monotonicity property 1.**

Proof of Stability Theorem - RI

- **If $\rho > 1$, the lower bound queue is unstable**
- **Assume $\rho < 1$ and $\limsup Q_t(0) = \infty$ with a positive probability. Then $\limsup U_t = \infty$ with a positive probability too. This implies that there exists a sequence t_n tending to $-\infty$ and such that a.s.**

$$\int_{t_n}^{0} (\widetilde{X}(v) - Y_{t_n}^{f}(v))dv \to_{n\to\infty} \infty.$$

Proof of Stability Theorem - RI (*continued*)

- **The pointwise ergodic theorem implies that**

$$\frac{1}{t}\int_{-t}^{0} \widetilde{X}(v)dv = \frac{1}{t}\int_{-t}^{0} \widetilde{X}(0)\circ\theta_v dv \rightarrow_{t\to\infty} \mathbb{E}[\widetilde{X}(0)],$$

- **If we show that the following a.s. limit also holds:**

$$\frac{1}{t}\int_{-t}^{0} Y_{-t}^{f}(v)dv \rightarrow_{t\to\infty} \mathbb{E}[\widetilde{Y}^{f}(0)]$$

this will conclude the proof by contradiction.

Proof of Stability Theorem - RI (*continued*)

- **The function**

$$\phi_t = \int_{-t}^{0} Y^f_{-t}(v)dv$$

is super-additive: $\phi_{t+s} \geq \phi_t \circ \theta_{-s} + \phi_s$

- **Thanks to the sub-additive ergodic theorem, this together with the fact that** ϕ_t **is integrable imply that a.s.**

$$\exists \lim_{t\to\infty} \frac{1}{t} \int_{-t}^{0} Y^f_{-t}(v)dv = K,$$

for some constant K **which may be finite or infinite.**

- **The fact that** K **is finite follows from the bound** $0 < Y^f_{-t}(v) \leq \widetilde{Y}^f(v)$ **and from the pointwise ergodic theorem applied to** $\{\widetilde{Y}^f(v)\}$**.**

Proof of Stability Theorem - RI (*continued*)

- **Since K is finite, the last limit holds both a.s. and in L^1**
- **By the same arguments**

$$K = \lim_t \frac{1}{t} \int_{-t}^{0} Y_{-t}^{f}(v)dv = \lim_t \mathbb{E}\left(\frac{1}{t}\int_0^t Y_0^f(v)dv\right).$$

- **From the fact that $Y_0^f(v)$, $v \geq 0$ is a geometrically ergodic Markov chain,**

$$\exists \lim_{t\to\infty} \frac{1}{t} \int_0^t Y_0^f(v)dv = \mathbb{E}[\widetilde{Y}^f(0)] \quad \textbf{a.s.}$$

Hence $K = \mathbb{E}[Y^{\tilde{f}}(0)]$

Stability - RD

- **Theorem If $\rho < 1$, where $\rho = \frac{\alpha q}{\beta p}$, then the RD system is stable. If $\rho > 1$, then it is unstable.**
- **Uses the coupling based on the 2-d point Poisson point process and the optimization problem:**

What is the infimum over all $y \geq 0$ of the integral

$$\int_u^0 Y^f_{u,y}(v)dv$$

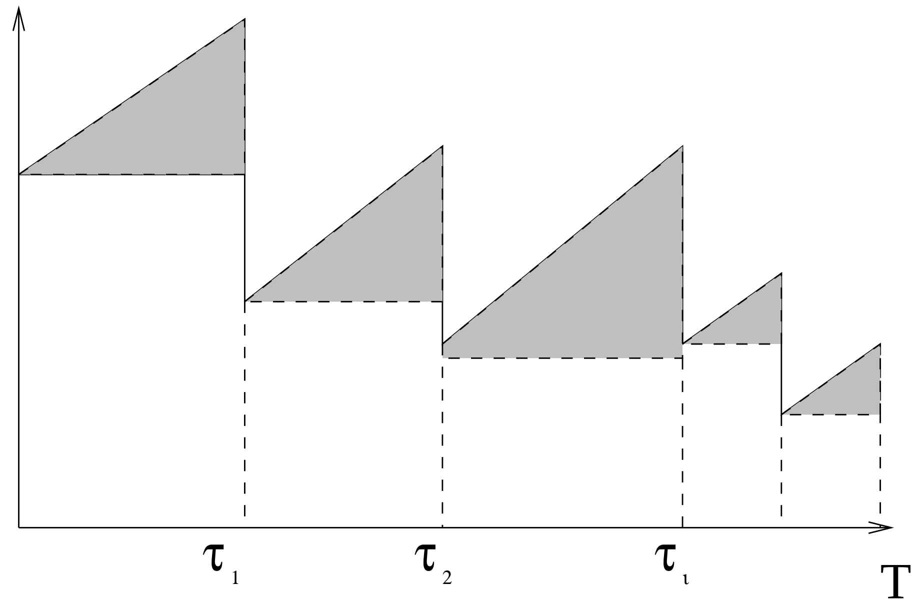
where $Y^f_{u,y}(v)$ is the value of the free process of $TCP2$ at time $v \geq u$ when starting from an initial value of y at time u?

Tails - RI

- **Lemma**
  In the stable RI case, the queue distribution is heavier than a Weibull distribution of shape parameter $k = 0.5$.
- **relies on the lower bound queue L_t**
- **relies on the fact that the fluid input process and the fluid draining process of this queue are jointly stationary and ergodic and have renewal cycles**
- **T the length of the renewal cycle and**

$$\Delta = \int_0^T \widetilde{X}(t) - \widetilde{Y}^f(t)dt = I_x - I_y.$$

Integral of X



Decomposition of the integral of X in a sum of trapezes.

Integral of X (*continued*)

$$\Pr\left(\sum_0^{N_T} Trap_i > q\right) \geq \Pr\left(\sum_0^{N_T} \alpha\frac{\tau_i^2}{2} > q\right),$$

where N_T denotes the number of losses in the cycle.

- **All triangular areas are i.i.d and heavy tailed:**

$$\Pr\left(\alpha\frac{\tau^2}{2} > x\right) = \Pr\left(\tau > \sqrt{\frac{2x}{\alpha}}\right) = e^{-\mu\sqrt{\frac{2x}{\alpha}}},$$

  which is Weibull with shape parameter $k = 0.5$.

- **propagates to Δ (Foss & Zachary 03)**
- **propagates to stationary L_t (Veraverbeke's theorem)**

Functional Equation for Stationary Law - RI - $B = \infty$

- **Phase 1:** $F(u,v,z) = \mathbb{E}(X^u Y^v Q^z 1_{\textbf{phase 1}})$
- **Phase 2:** $G(u,v) = \mathbb{E}(X^u Y^v 1_{\textbf{phase 2}})$
- **Functional equation for Mellin transforms**

$$\begin{aligned} 0 &= \alpha u F(u-1,v,z) + \beta v F(u,v-1,z) \\ &\quad + z(F(u+1,v,z-1) - F(u,v+1,z-1)) \\ &\quad + \left[\lambda\left(\frac{1}{2^u}-1\right) + \mu\left(\frac{1}{2^v}-1\right)\right] F(u,v,z) \\ &\quad + \alpha u G(u-1,v) + \beta v G(u+1,v-2) \\ &\quad + \left[\lambda\left(\frac{1}{2^u}-1\right) + \mu\left(\frac{1}{2^v}-1\right)\right] G(u,v) \end{aligned}$$

- **From PDE obtained by the PDP approach.**

Conclusions

- Generic framework to analyze the fluctuations of TCP troughput.
- Many open problems in the network setting, often approched using mean field analysis.
- Many open PDE problems.

References

– F. B., K.B. Kim and D. McDonald, "Equilibria of a Class of Transport Equations Arising in Congestion Control", Queueing Systems, Volume 55, Number 1, pp. 1–8, 2007.

– F. B. and D. McDonald, "A Stochastic Model for the Throughput of Non-Persistent TCP Flows", Performance Evaluation, Volume 65, Number 6-7, Pages 512-530, 2008.

– F. B., G. Carofiglio and S. Foss, "Proxy Caching in Split TCP: Dynamics, Stability and Tail Asymptotics", CUP Kahn Volume, 2009.

– Ongoing work with V. Anantharam.