On a Class of Partial Differential Equations Associated with Interacting TCP Flows

F. Baccelli & D. Mc Donald

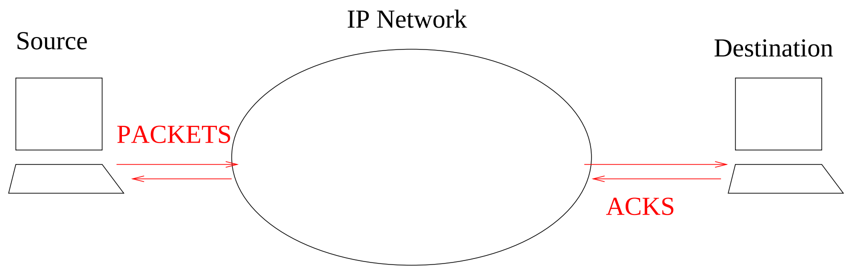
ENS-INRIA

Workshop Mathematical Models of Traffic Flows
Nice, Novembre 2005

Summary

- I. TCP, AQM and TD links
- II. Equations for persistent TCP flows sharing an AQM link (with D. McDonald, K. B. Kim and J. Reynier)
- III. Equations for non persistent TCP flows in an AQM link (with D. McDonald)
- IV. Interaction of on-off TCP flows on a TD link (with A. Chaintreau, D. McDonald and D. De Vleeshchauwer)

I. Congestion and Flow Control



TCP Transmission Control Protocol of the Internet: distributed, end-to-end

Error control Each packet received by the destination is acknowledged;

Flow 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

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

TCP

- TCP dynamic window size

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

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

$$\text{phase} \uparrow : \quad g(w_n, \text{OK}) \geq w_n, \quad \text{phase} \downarrow : \quad g(w_n, \text{LOSS}) \leq w_n.$$

- Reno: congestion avoidance: AIMD

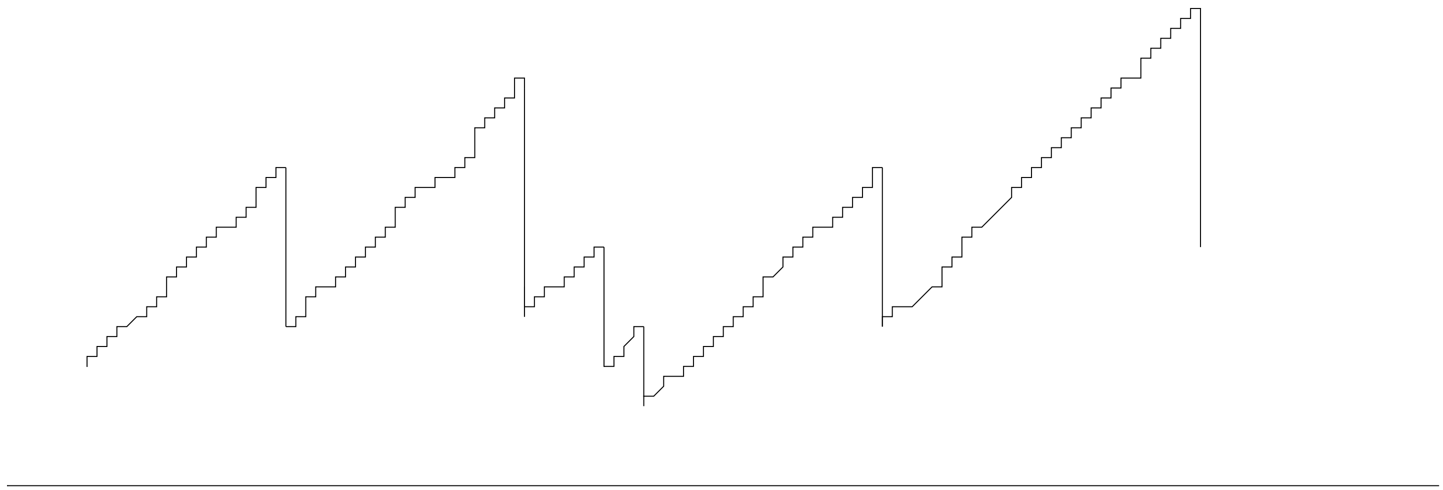
$$g(w_n, \text{OK}) = w_n + \frac{1}{w_n}, \quad g(w_n, \text{LOSS}) = \left\lfloor \frac{w_n}{2} \right\rfloor$$

- Tahoe:

$$g(w_n, \text{OK}) = w_n + \frac{1}{w_n}, \quad g(w_n, \text{LOSS}) = 1$$

TCP (*continued*)

- Typical Reno window trajectory



Types of Links

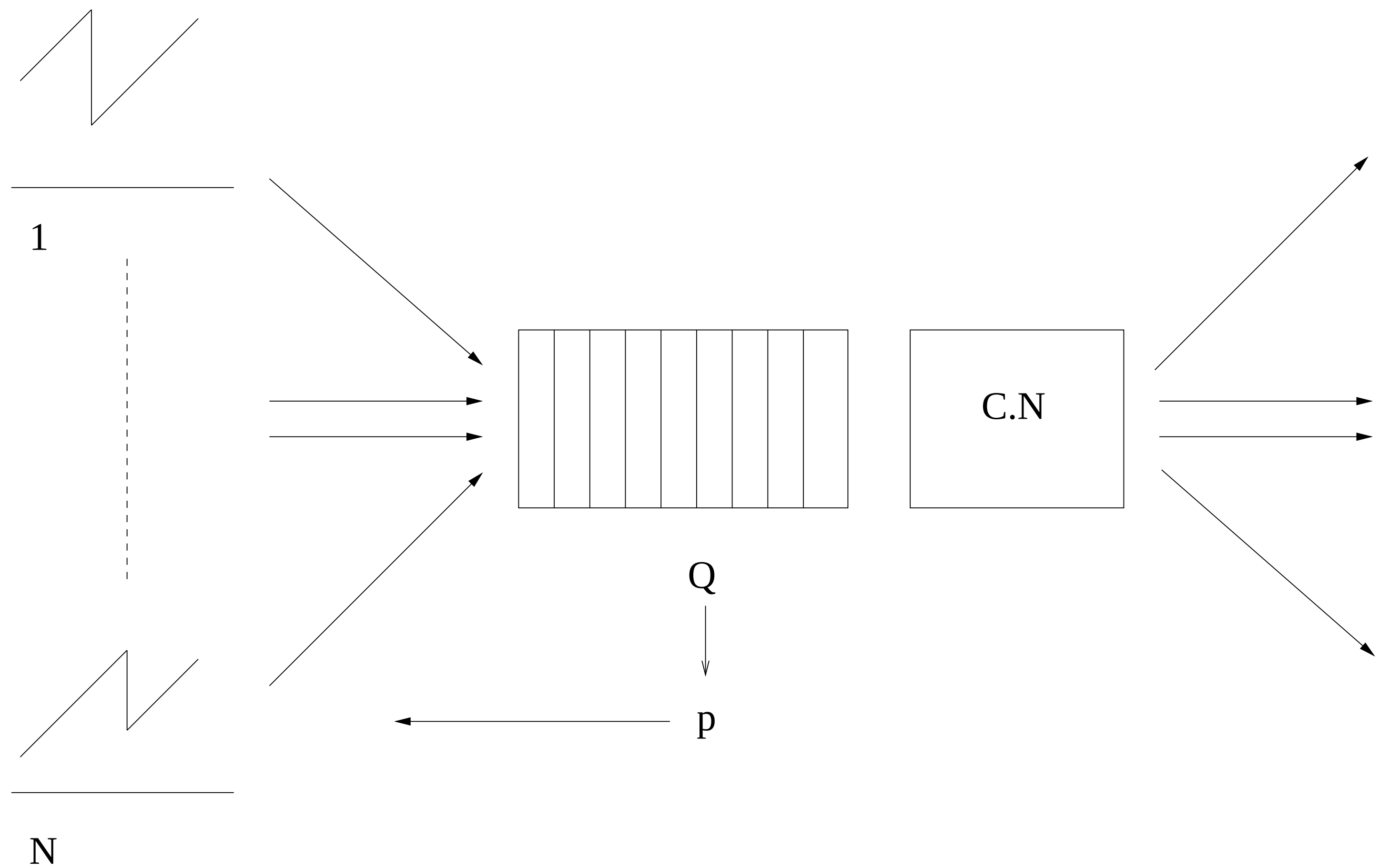
- Assume a collection of TCP flows sharing a link equipped with a buffer
- Active Queue Management AQM: packets are dropped randomly with a rate that depends on the queue size in the shared buffer
- Tail-Drop Case TD: packets are dropped when the shared buffer overflows

II. Persistent TCP flows on an AQM Link

– PDE and ODE for throughput distribution

– Positivity and uniqueness

– Square root formula for mean throughput


1
N
C.N
Q
p

PDE for the Dynamics of Throughput Distribution

- $f(z,t)$: density at time t of the throughput of a TCP connection when an AQM router drops its packets with probability $p(t)$.

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

- The rate at which its packets are dropped is β times the throughput
- mass leaves the interval $[z, z+dz]$ at rate $p(t)zf(z,t)dz$ approximately.
- mass enters this interval because of losses among throughputs in the interval $[2z, 2(z+dz)]$ at rate $p(t)2zf(2z,t)\cdot 2dz$

ODE for the Stationary Throughput Distribution

- Assume the queue size stabilizes so that $p(.)$ becomes constant p (justified via mean field in McDonald and Reynier 04).
- In statistical equilibrium, the density of throughputs satisfies

$$\frac{df}{dz}(z) = \beta(4zf(2z) - zf(z)), \quad z \geq 0,$$

with $\beta = pR^2$.

- We are looking for a solution which is a probability density on the positive half line.

Solution of the ODE

- Theorem Let $\beta = pR^2$. The unique density satisfying the ODE is

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

with

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

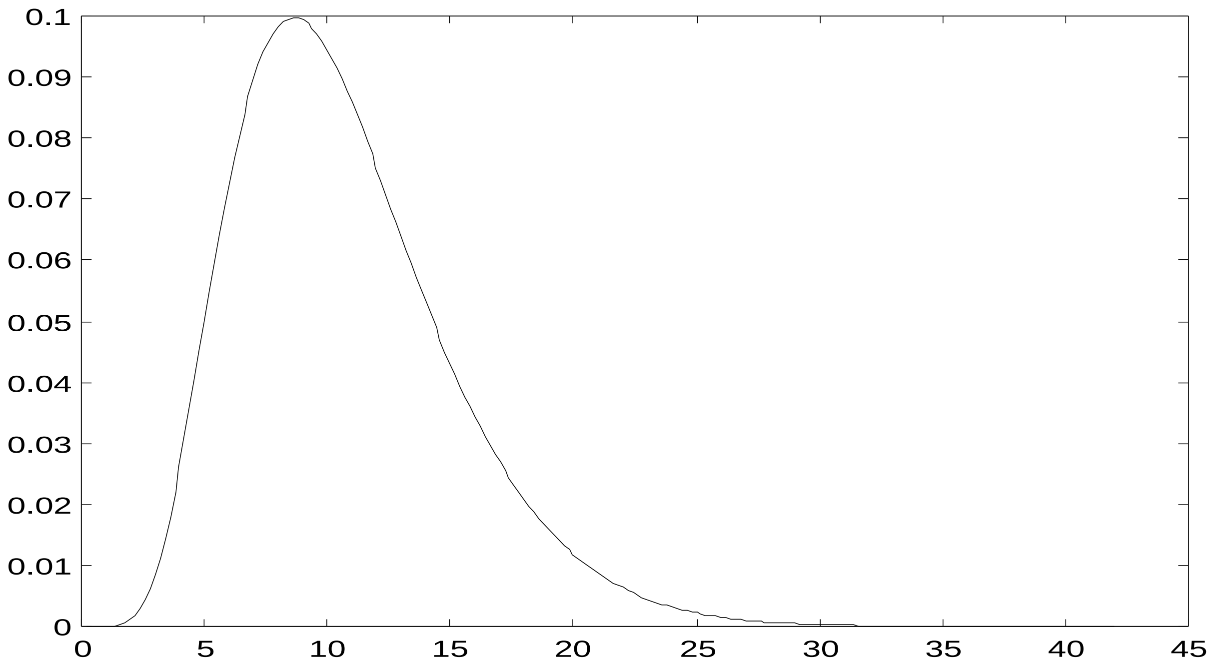
- Based on the following functional equation for the Mellin transform of f:

$$\widehat{f}(u) \;=\; \int_0^\infty f(z) z^{u-1} dz$$

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

Distribution

■ Distribution of rate, windows, or number of packets in an interval as observed in traces and simulations.



Mean Value: the Square Root Formula

- The stationary mean is

$$M = \sqrt{\frac{2}{pR^2}}\sqrt{\frac{1}{\pi}}\frac{\prod_{k\geq 1}(1-2^{-2k})}{\prod_{k\geq 1}(1-2^{-2k+1})}$$

that is approximately

$$M = \frac{1.309}{R\sqrt{p}}$$

III. Non Persistent TCP flows on an AQM Link

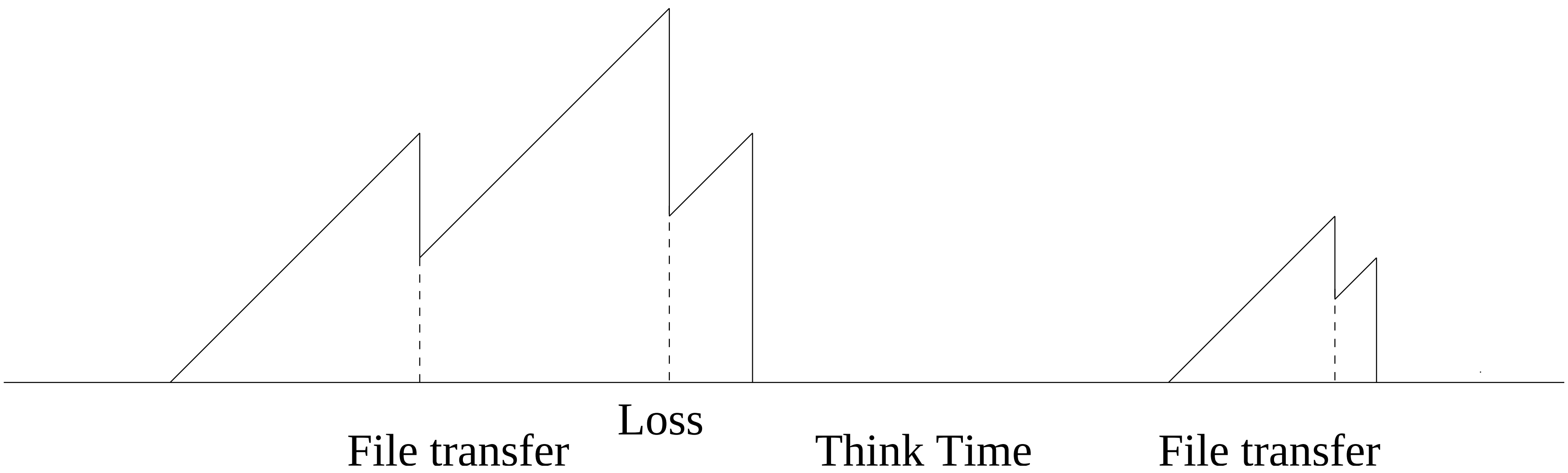
– Exponential Model

– Square Root Formula for On-Off Flows

– Heavy Tailed Model

Dynamics

- 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 $1/\beta$



Exponential Model

- File sizes are exponential with parameter μ
- Think times are exponential with parameter β

– Packet loss probability $p(t)$

– RTT R is constant.

PDE

- $(f(z,t), \nu(t))$ distribution of rates

$$\frac{\partial}{\partial t} f(z,t) + \frac{1}{R^2}\frac{\partial}{\partial z} f(z,t) \ = \ -\mu z f(z,t) + 4zp(t) f(2z,t) - zp(t) f(z,t),$$

- 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 Points

$$\frac{df(z)}{dz} = -\mu R^2 z f(z) + 4zpR^2 f(2z) - zpR^2 f(z)$$

- Boundary:

$$\int_0^\infty \mu z f(z) dz = \beta\nu.$$

- Normalization:

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

- Theorem There exists a unique probability density solution of this ODE

Solution by Mellin Transform

- Multiplying both sides of the ODE by z^u and integrating w.r.t. z, we get

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

- Let

$$\widehat{f}(u) = g(u)\Gamma(\frac{u}{2})\left(\frac{2}{(p+\mu)R^2}\right)^{\frac{u}{2}}.$$

Then

$$g(u) = g(u+2)(1-\frac{p}{p+\mu}2^{-u})$$

$$g(u) = g(\infty)\Pi_\infty(u)$$

with

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

Solution by Mellin Transform (*continued*)

$$\widehat{f}(u) \;=\; g(\infty)\frac{\Gamma(\frac{u}{2})}{\left(\frac{(p+\mu)R^2}{2}\right)^{\frac{u}{2}}}\Pi_\infty(u)$$

- Specializing to $u = 1$ and using normalization, we get $g(\infty)$ from normalization.

Mean Values

- The mean rate $M = \widehat{f}(2)$ is found to be of the form:

$$M = \frac{\frac{1}{\mu}}{\frac{1}{\beta} + T}$$

where T is the mean time T to transfer a file:

- The probability that a flow is ON is

$$\nu = 1 - M\frac{\mu}{\beta}$$

Mean Values (*continued*)

- Mean time T to transfer a file

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

- Mean rate of a stationary flow

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

Distribution

- $f(z)$ density of the rate at $z > 0$.

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

with $a_0 = 0$ and

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

- When letting the file size go to infinity, compatible with the distribution found for the persistent flow case.

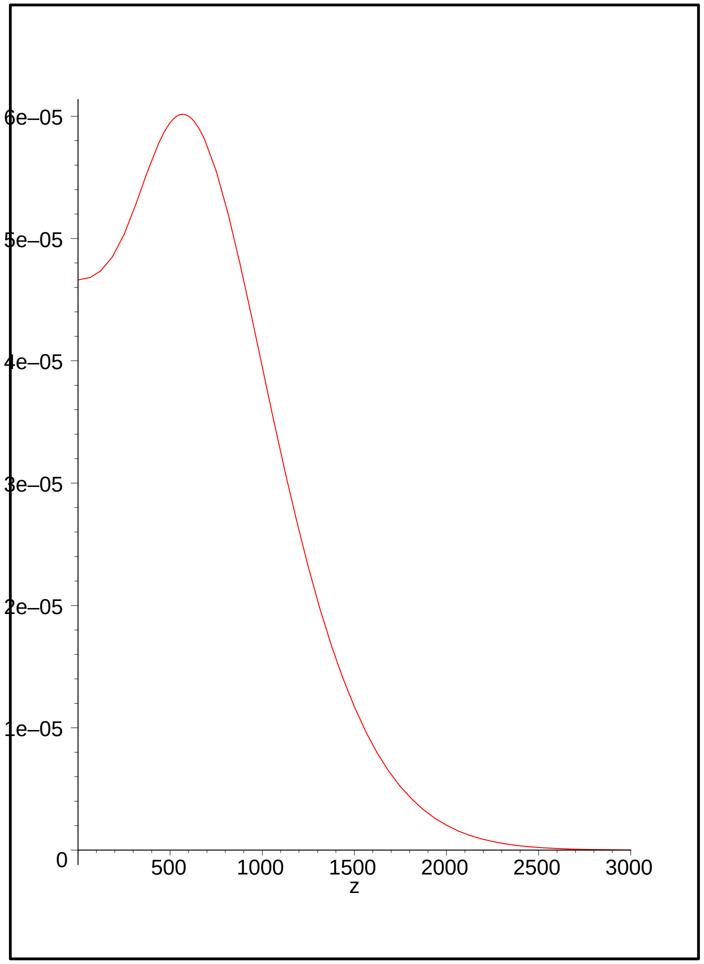


Figure 1: The density of the stationary rate (1). Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 100$ and $p = 1\%$.

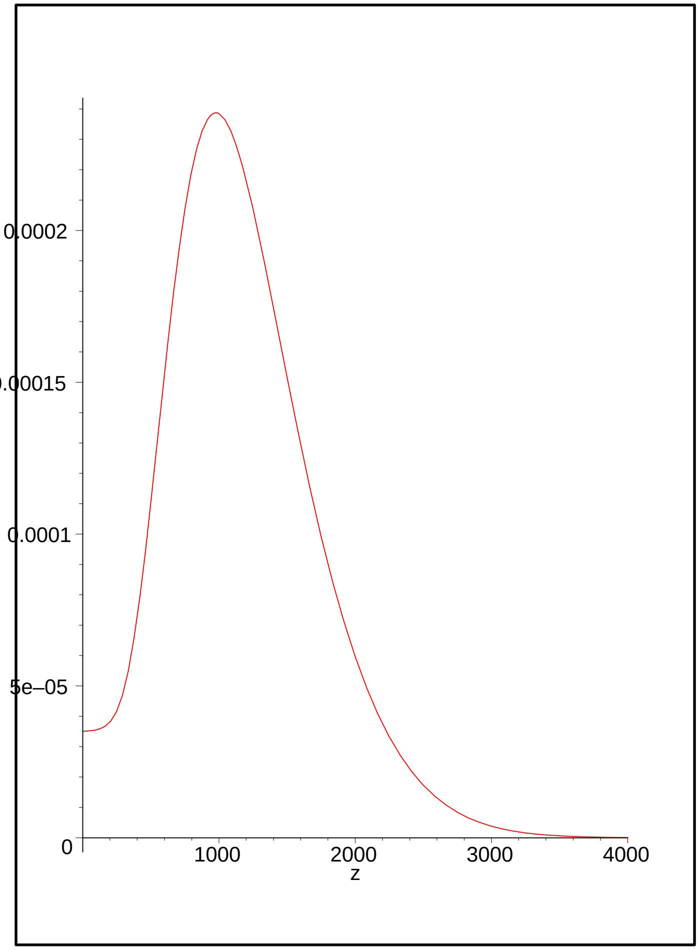


Figure 2: The density of the stationary rate (1). Here $R = 0.1$ s., $1/\beta = 2$ s., $1/\mu = 1000$ and $p = 1\%$.

IV. Non Persistent TCP flows on a TD link

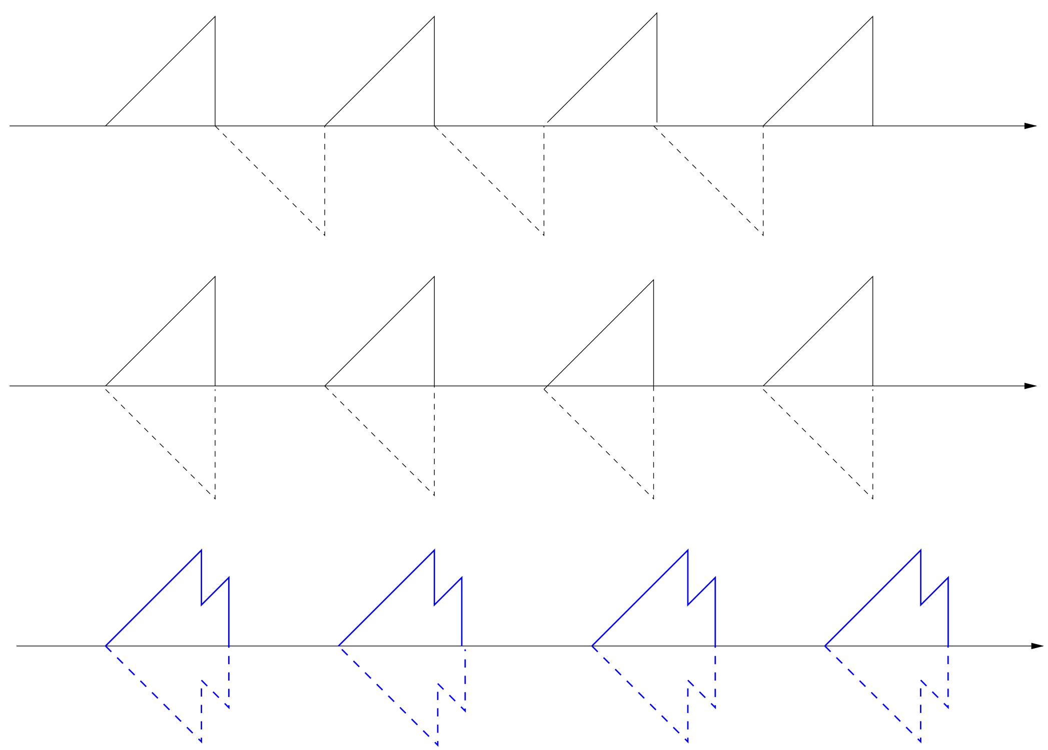
– In phase and out of phase HTTP flows

– PDE for the free regime

– Fredholm integral equation for the congestion regime

– HTTP turbulence

TD Case: In Phase and Out of Phase AIMD, On-Off Flows



Interaction of Non Persistent 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

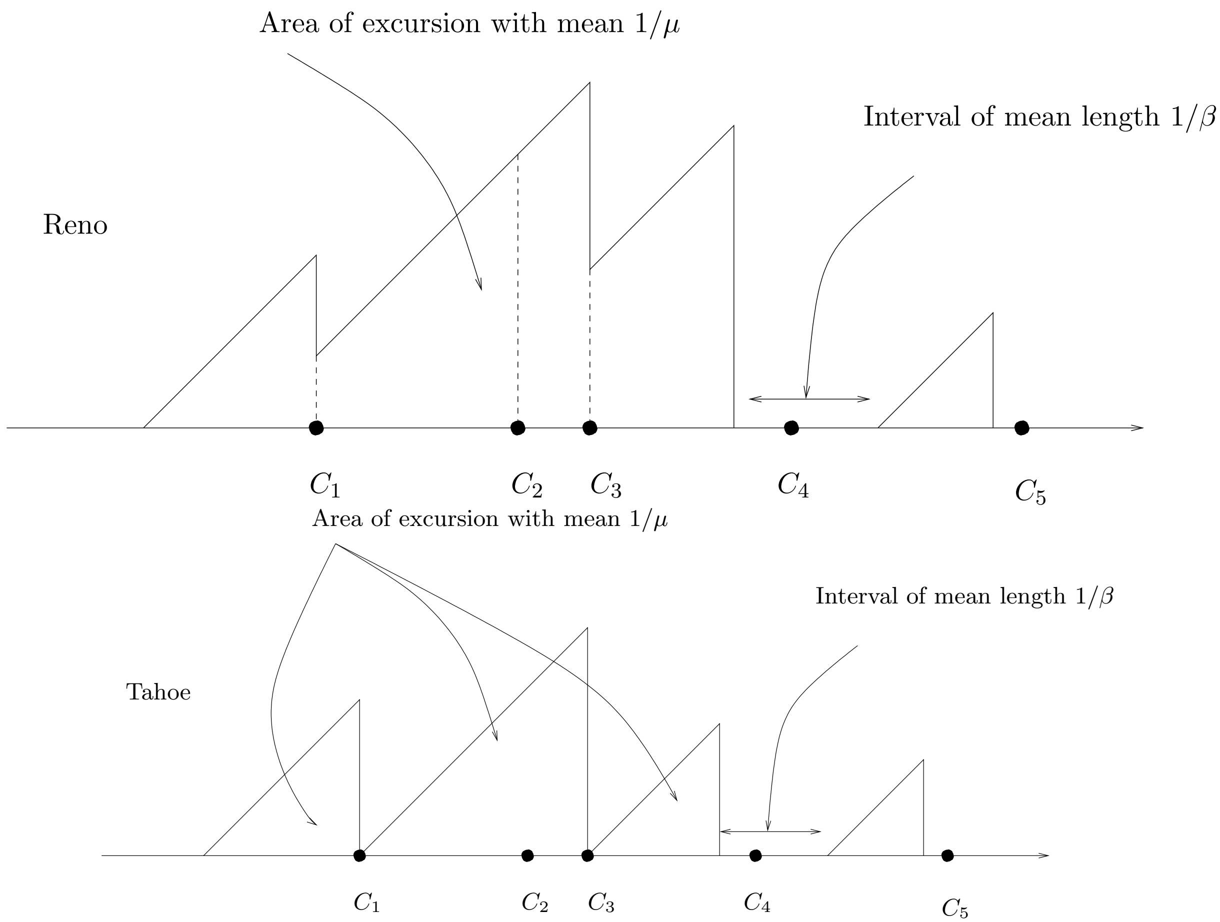
– Document sizes are i.i.d. with mean $1/\mu$

– 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 (small buffer case).

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



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

Mean Field Limit

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

– the limiting aggregated rate

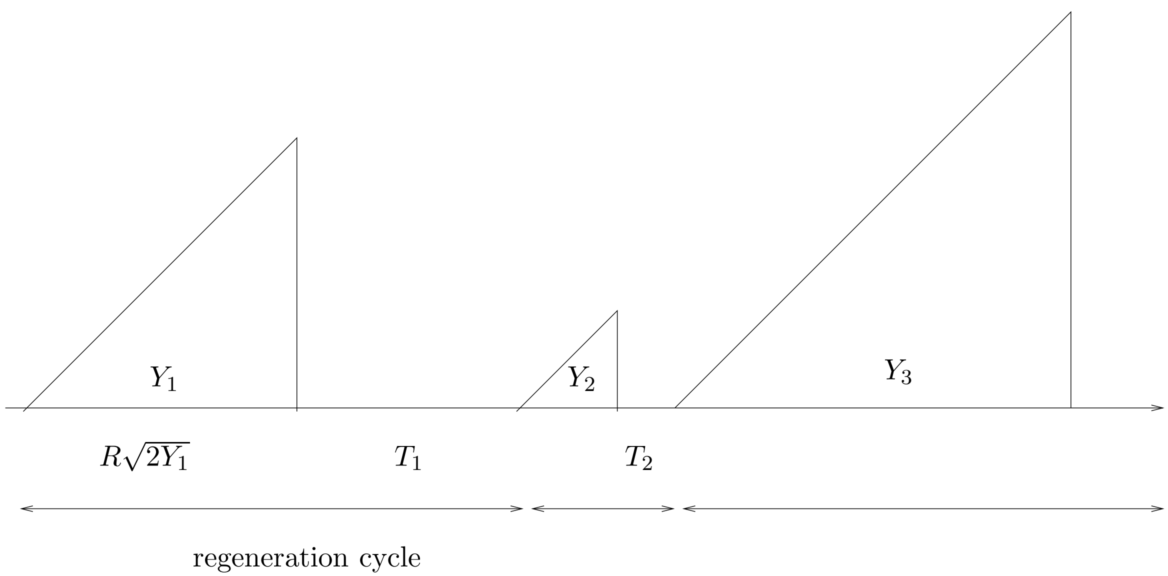
$$\alpha(t) = \lim_N \frac{1}{N} \sum X_n^N(t)$$

– the limiting distribution of rates

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

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

– $f(z,t)$ the proportion of flows active and 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 measure:

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

with

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

- Explicit solution via either Laplace transform in time or Renewal theory.

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

- Results via Laplace transform in time and Tauberian theorems

$$\nu(\infty) = \frac{\frac{1}{\beta}}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}}$$

$$f(z,\infty) = \frac{R^2 e^{-R^2 \mu z^2/2}}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}}$$

$$\alpha(\infty) = \frac{1}{\mu}\frac{1}{\frac{1}{\beta} + R\sqrt{\frac{\pi}{2\mu}}} = \rho.$$

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

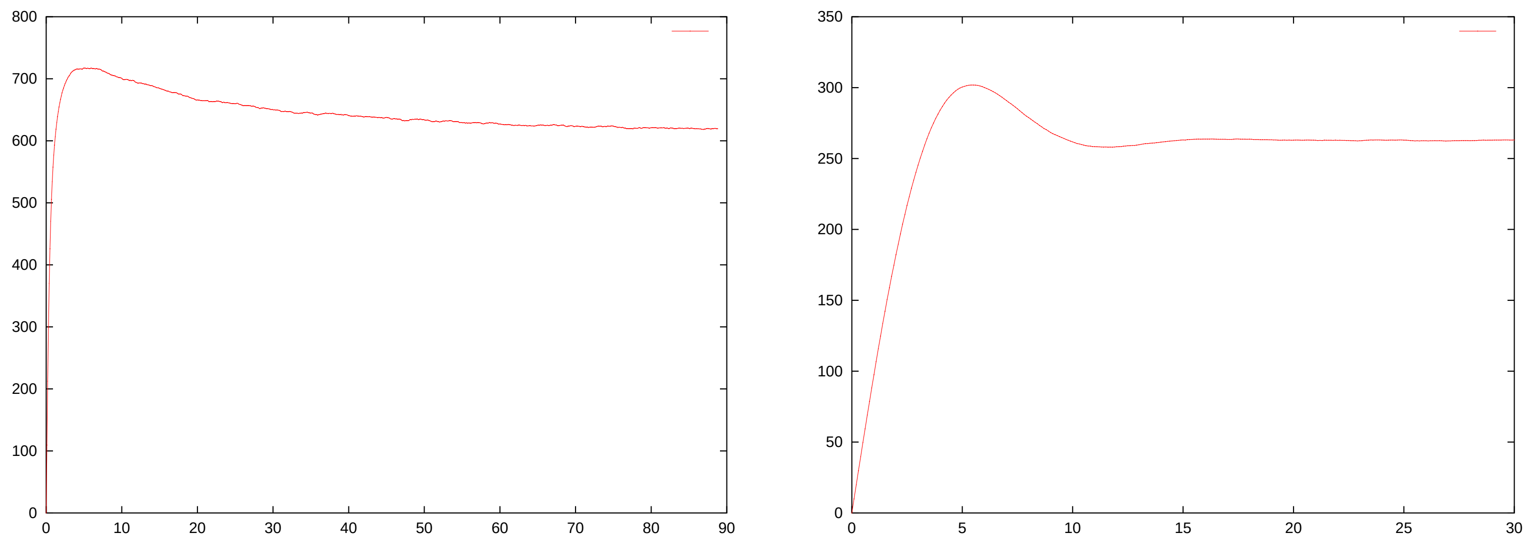
The Free Regime Mean Field Limit ($C = \infty$) (*continued*)

- Reduction to a Fredholm integral equation of the second kind

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

- Handy for numerical exploitation: reduction to linear matrix equations after discretization.

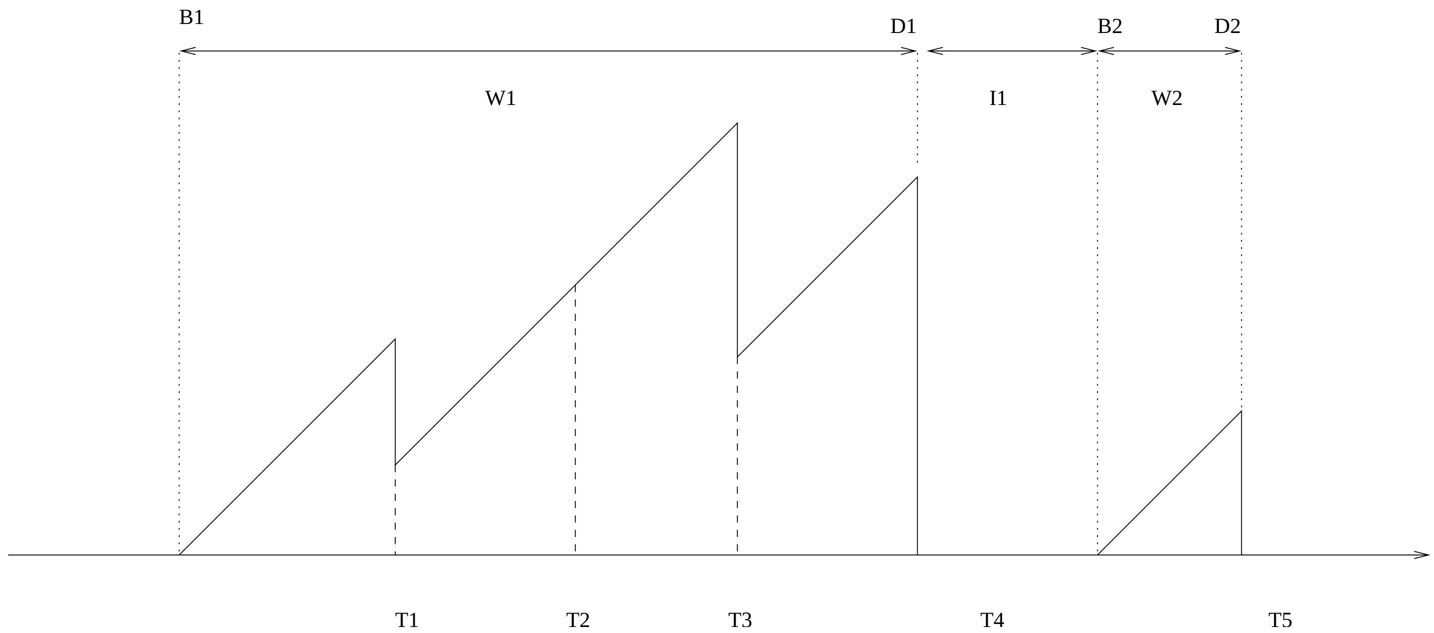
Examples of Aggregated Rates



All flows are initially active and with 0 rate;
Mean file size 2000; mean think time 2 sec.
Left: lognormal with standard deviation 4 times the mean and RTT 30 ms.
Right: exponential with RTT 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 in D. Mc Donald & J. Reynier



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

- 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:

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

- In the exponential case, all terms in this fixed point equation are computable thanks to the regenerative structure.
- Regeneration Cycle when tagged flow is inactive at a congestion epoch.

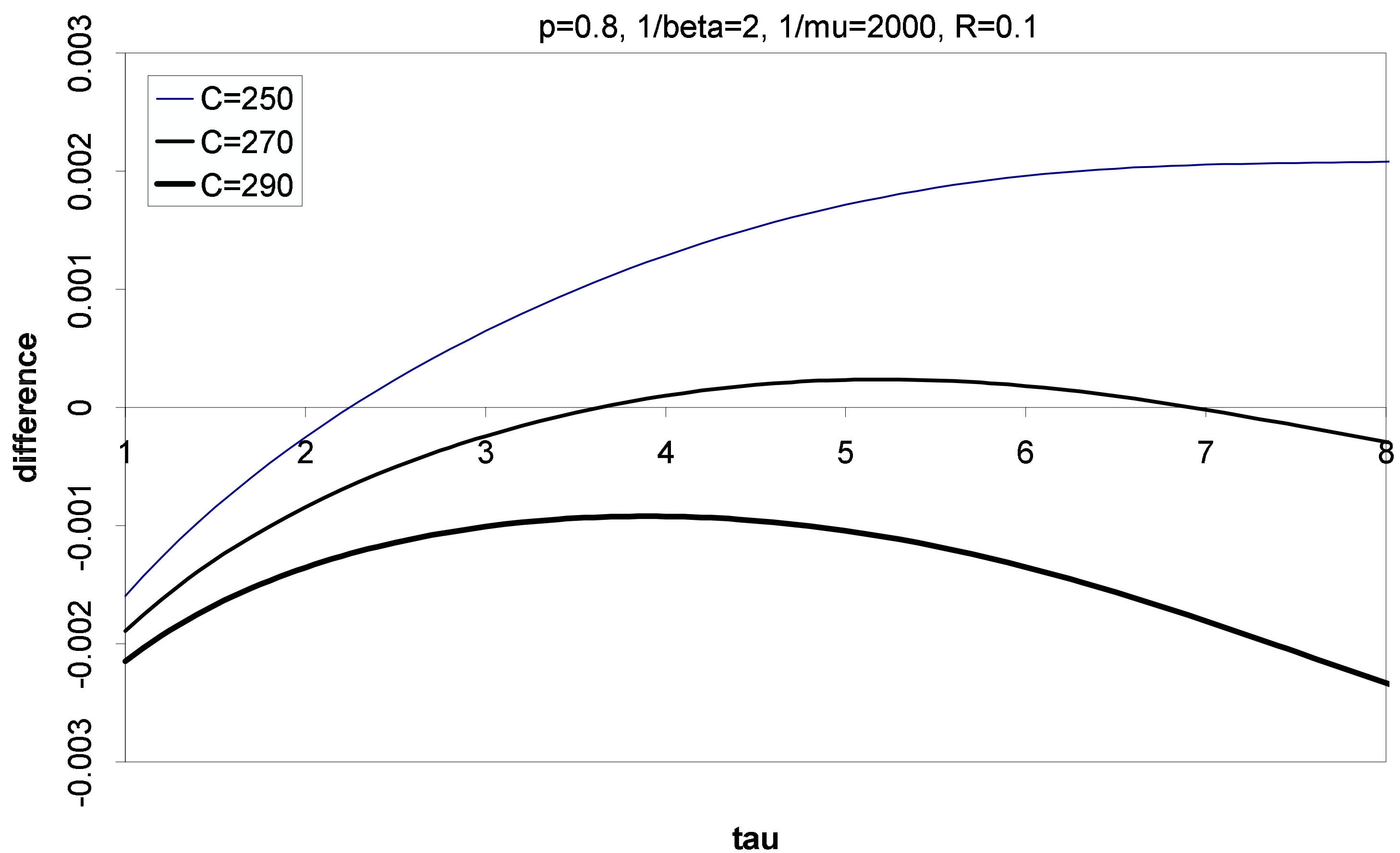


Figure 3: The 3 Cases for the Rate Conservation Principle Fixed Point Equation

Example of Computation: Mean Number of Files Transmitted in a Regeneration Cycle

- $h(u)$: expected number of files that will be transmitted by the end of the current cycle given that the tagged flow is inactive at time $0 \leq u < \tau$.

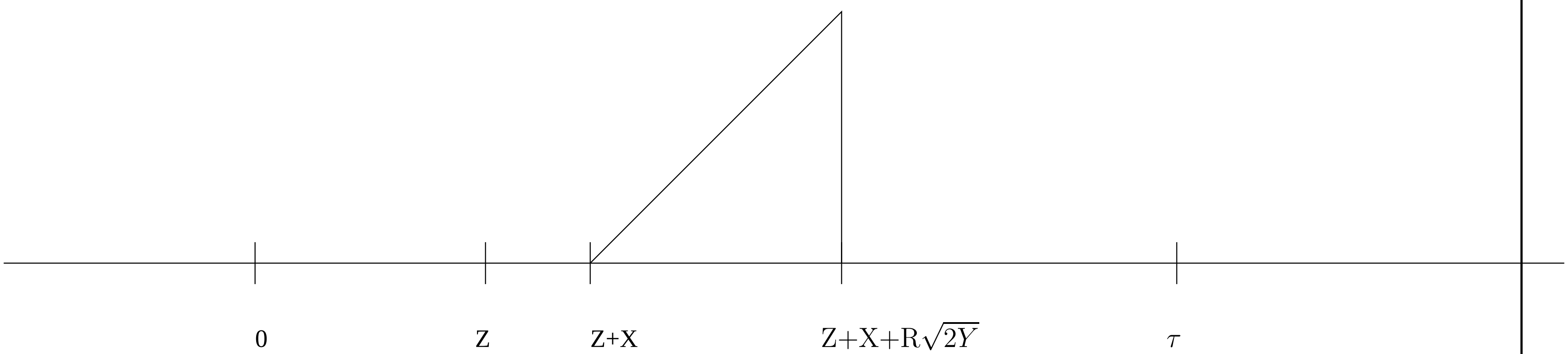
$$E(N) = h(0)$$

- $g(v)$: expected number of files that will be transmitted by the end of the current cycle given that the tagged flow is active at some congestion epoch and that the current transmission rate of the tagged flow is v.
- Fredholm type equation for (h, g).

Example of Fredholm Equations for $h(.)$

- Case 0: no file transmitted within the end of the current cycle: contributes 0;
- Case 1: at least one file transmitted and we can transmit this file before the next congestion epoch. Then the expected number of files to transfer until end of cycle is:

$$\int_0^{\tau-z} \beta e^{-\beta x} dx \int_0^{(\tau-z-x)^2/2R^2} \mu e^{-\mu y} dy (1 + h(z + x + R\sqrt{2y})).$$



Example of Fredholm Equations for $h(.)$ (*continued*)

- Case 2: same as 1 but the file cannot be transmitted by the the next congestion epoch.
- This occurs with probability $\exp(-\mu(\tau - z - x)^2/2R^2)$.
- By the congestion epoch the transmission rate of the tagged flow is $(\tau - z - x)/R^2$.
- Then the expected number of files to transfer until end of cycle is:

$$\exp(-\mu(\tau - z - x)^2/2R^2)\left(qg(\frac{\tau - z - x}{2R^2}) + (1-q)g(\frac{\tau - z - x}{R^2})\right).$$

Example of Fredholm Equations for $h(.)$ (*continued*)

- We conclude

$$h(z) = \int_0^{\tau-z} \beta e^{-\beta x} dx \int_0^{(\tau-z-x)^2/2R^2} \mu e^{-\mu y} dy (1 + h(z + x + R\sqrt{2y})) + \exp(-\mu(\tau - z - x)^2/2R^2) \left(q g(\frac{\tau - z - x}{2R^2}) + (1-q) g(\frac{\tau - z - x}{R^2}) \right).$$

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 for τ:

$$s_0(z) = (1-q)f_0(z,\tau) + qf_0(2z,\tau),$$

where $f_0(z,t)$ 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 the

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

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

- Exponential model, $1/\mu = 2000$ Pkts, $1/\beta = 2$, $q = 0.8$ 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

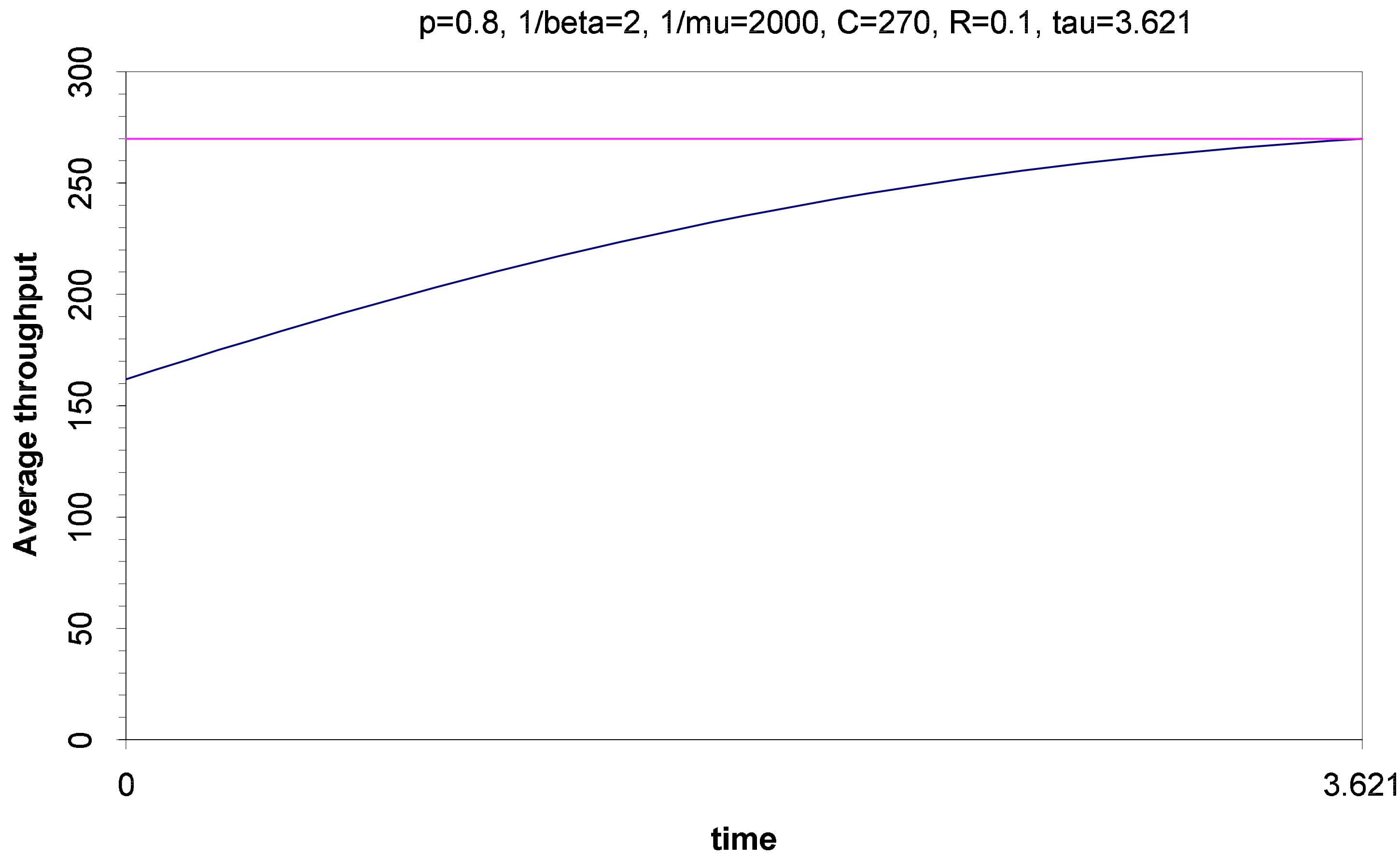


Figure 4: Evolution of the congestion-less aggregated rate with the time

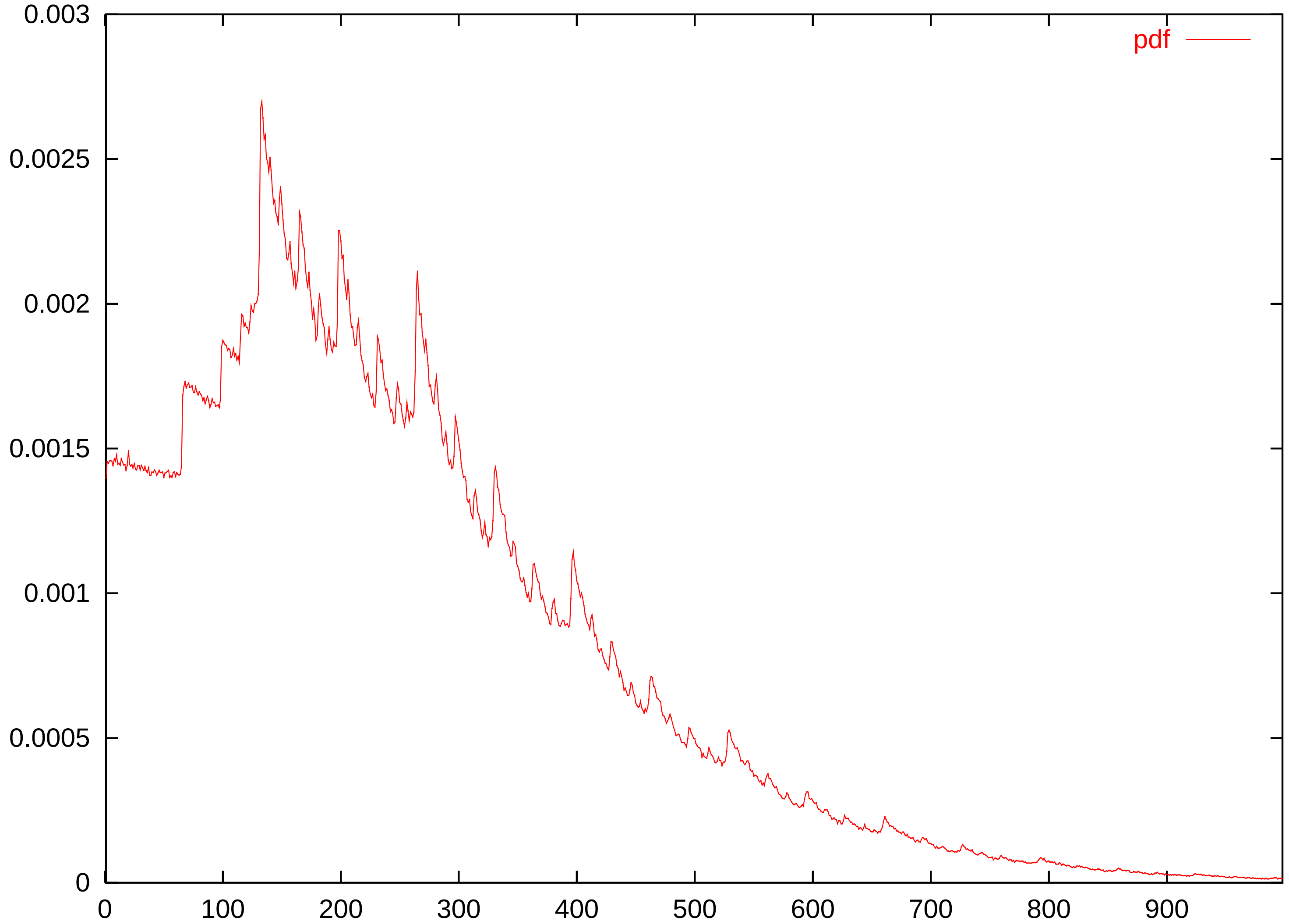


Figure 5: Invariant rate pdf

Turbulence Scenario

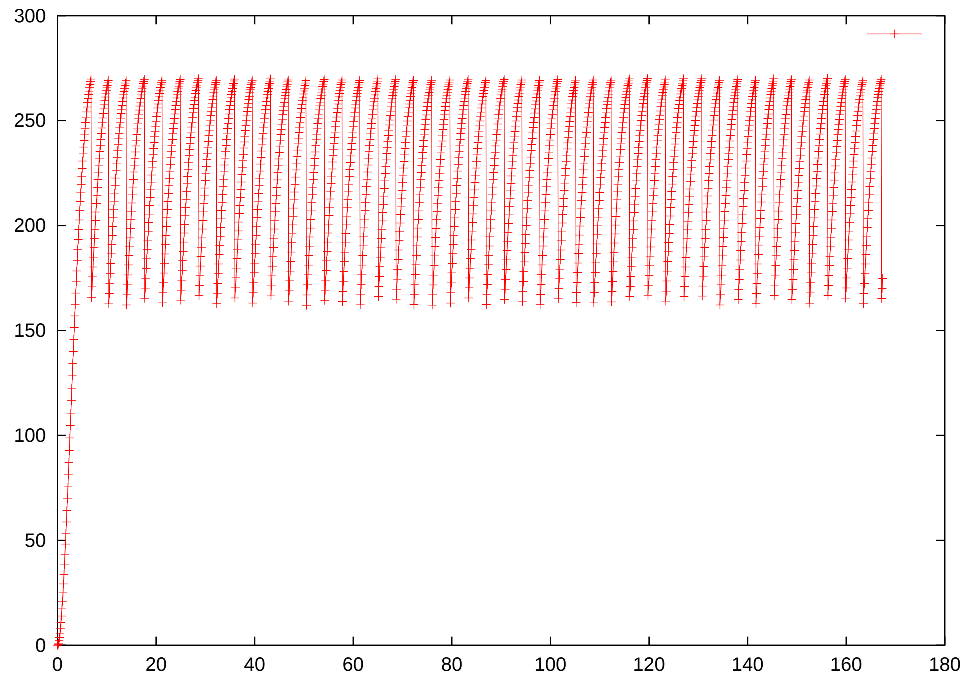
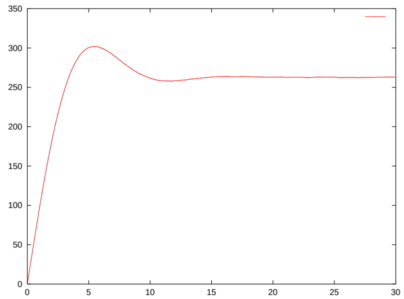


Figure 6: Evolution of aggregated rate when all flows are initially active and with null rate for $C = 270$ Pkts/sec.

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.

Turbulence: – Mathematical proof (Tahoe Case) (*continued*)

Let

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

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.

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.

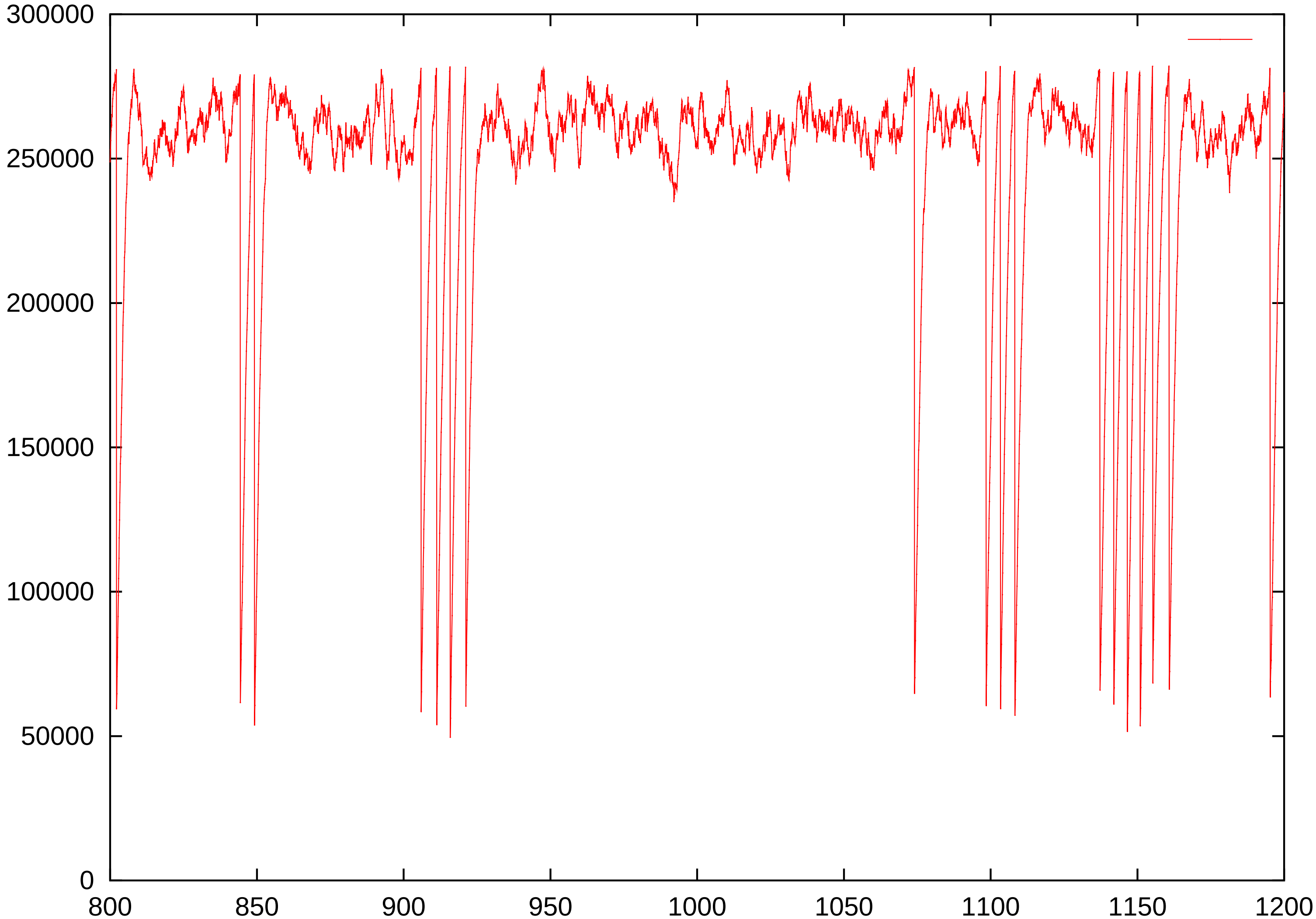


Figure 7: Bi-stability: 1000 Tahoe flows with $C = 282$.

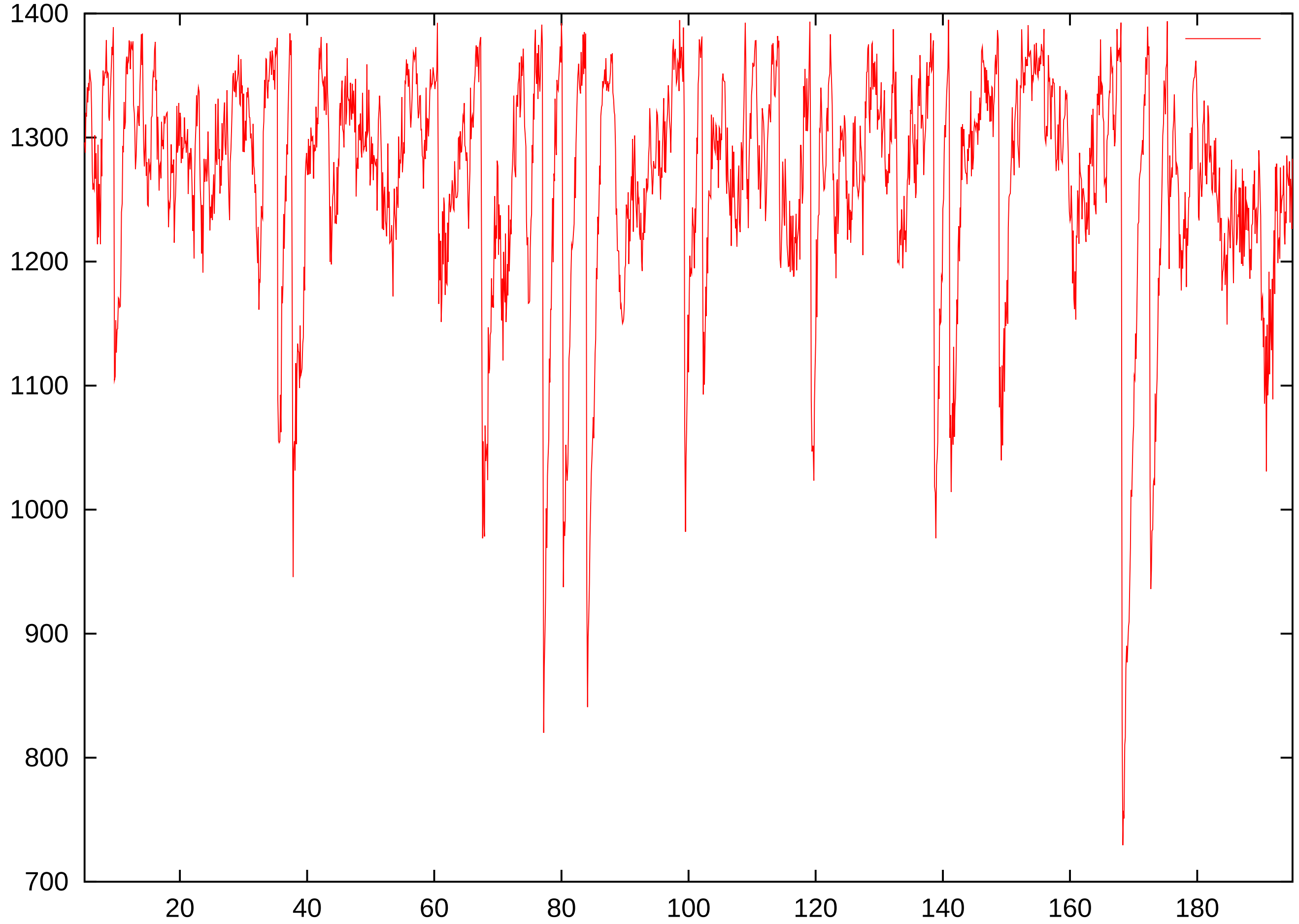


Figure 8: ns2 aggregated traffic: alternance between the two regimes.

Conclusions for TD

This mean-field model shows that

- TCP under TD may induce in phasing of http flows
- TCP under TD may induce network level fluctuations which come on top of the other short time scale fluctuations

1. Packet Level fluctuations (not studied here): slow start, packet clumps etc.
2. Flow Level (individual AIMD) fluctuations,

Outlook

- Ongoing analysis with M. Lelarge & D. McDonald of the rarity of the transitions using Kiffer's discrete version of Wentzell-Freidlin's theory.
  - PDE for heavy tailed case
  - Proof of turbulence in the Reno case
  - Multi-link case

References

- "A mean-field analysis of short lived interacting TCP flows", F.B., Chaintreau, A., De Vleeschauwer, D., McDonald, D. ACM Sigmetrics'04, New York, June 2004. Extended version in "HTTP Turbulence", INRIA Report 2004.
- "A Square Root Formula for the Rate of Non-Persistent TCP Flows", F. B. & D. McDonald, EuroNGI Conference, Roma, April 2005. Extended version in INRIA Report with the same title, 2O04.
- "A mean-field model for multiple TCP connections through a buffer implementing RED," F.B., D., McDonald, J. Reynier Performance Evaluation, vol. 49, pp. 77–97, 2002.

Mean Field Fixed Points for AQM Bandwidth Sharing

- N statistically identical ON-OFF flows with parameters (μ, β, h, R) that share a common AQM link with capacity $C \cdot N$.
- Let

$$\rho = \frac{1/\mu}{\frac{1}{\beta} + T}$$

with

$$T = \int_{z=0}^{\infty} \int_{u=0}^{\infty} h(z) \mu e^{-\mu u} \left(\sqrt{z^2 R^4 + 2uR^2} - zR^2 \right) du dz$$

mean value of the rate of one flow in the absence of packet loss

The Two Mean Field Regimes

- The stabilized congestion regime is reached by the system when $\rho > C$: a positive drop probability is required to match the load brought by the flows and the capacity of the link. The system stabilizes to a constant buffer content b, to a constant packet loss probability p and to a mean rate per flow $M[p]$, such that

$$M[p](1-p) = C$$

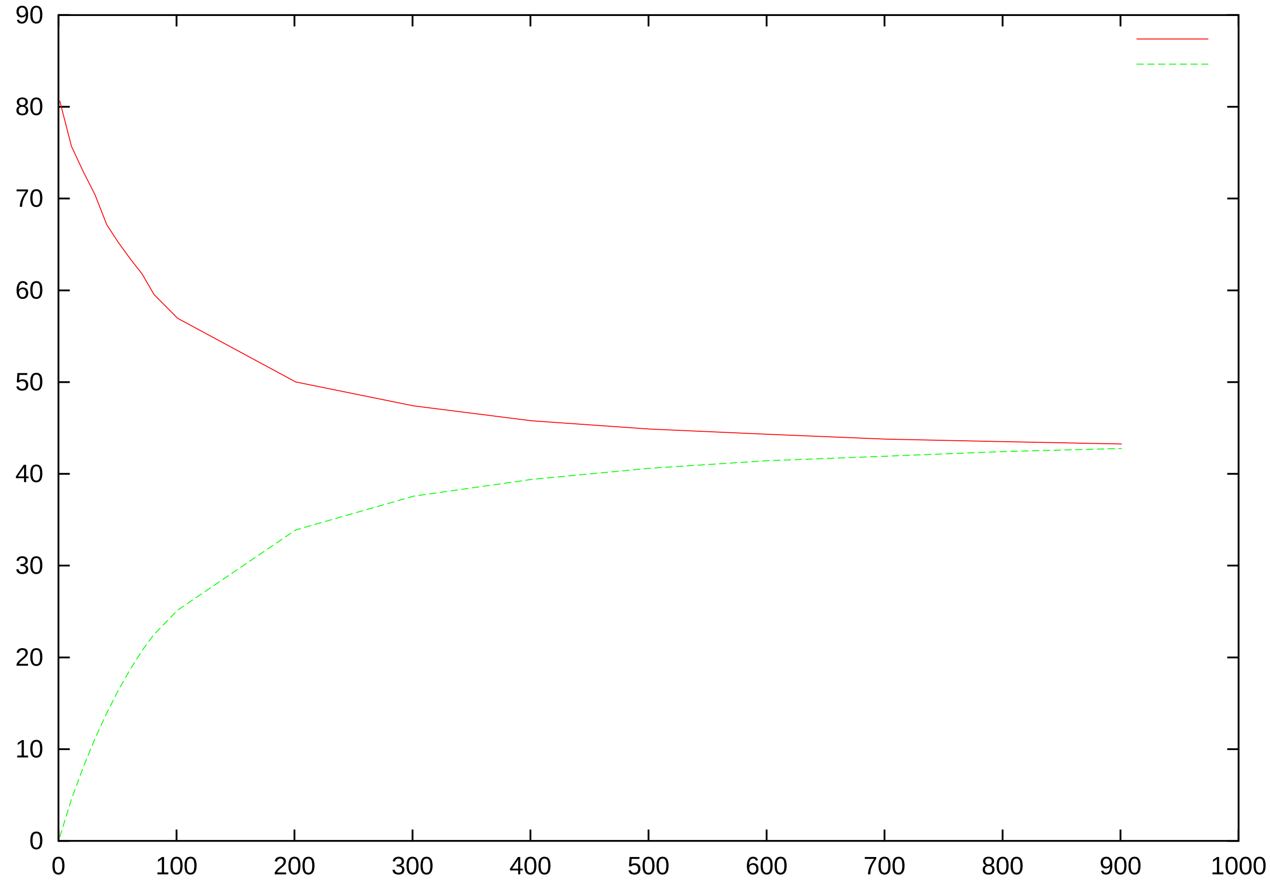
- Since the function $p \to M[p](1-p)$ is decreasing in p and tends to ρ when p tends to 0, the above equation defines a unique equilibrium point p^* whenever $\rho > C$.

The Two Mean Field Regimes (*continued*)

- The congestion-less regime is reached when $\rho < C$; the load brought by the flows is less than the link rate, and each flow gets a mean rate of ρ.
- Other and in particular oscillating regimes are possible.

Heterogeneous Traffic Example: Mice and Elephants

- The elephant class with mean file size 1000 packets and the class with file m, where m is smaller than 1000, which can be called the mouse class when m is significantly smaller than 1000.
- In the model, 50 elephant flows and 50 mice flows compete for the bandwidth of a 2000 pkts/s link. The RTT is .1 s. and the think times have a 2 s. mean. There is no slow-start.


90
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
600
700
800
900
1000

HT Case

- Arbitrary distribution for the OFF times
- file sizes with a distribution of the form

$$G(x) = \sum_{i=1}^{\infty} q_i (1 - e^{-\mu_i x}),$$

$\{q_i\}$ is a probability and $\{\mu_i\}$ a sequence of positive real numbers s.t. $\sum_i q_i \mu_i^{-1} < \infty$, which guarantees that G has a finite first moment μ^{-1}.

- An interesting instance is that with $q_i = A i^{-\alpha}$ and $\mu_i = \mu / i$ with $\alpha > 2$.

HT Case (*continued*)

- Mean rate

$$M = \frac{1}{\frac{\mu}{\beta} + \sqrt{\frac{\pi}{2}} R \sum_i q_i \left(\frac{\prod_{l=1}^{\infty}\left(1 - \frac{2p}{p+\mu_i} 4^{-l}\right)}{\prod_{l=1}^{\infty}\left(1 - \frac{p}{p+\mu_i} 4^{-l}\right)} \sqrt{p+\mu_i} \right)}.$$

- Mean almost insensitive to distribution of file size within this setting.