Metastable Regimes for multiplexed TCP flows

François Baccelli*, M. Lelarge† and D. R. McDonald‡

October 1, 2004

Abstract

Consider the mean field limit of a model for multiple HTTP sources multiplexed through a drop-tail router. The limit may exhibit two stationary regimes. In the fluid regime the flows are independent, there are no packet losses and the average throughput is high. In the turbulent regime the flows are synchronized, there are periodic congestion epochs with packet losses and the average throughput is reduced.

In the prelimit with a finite number of sources the above regimes become metastable in the sense that we observe periodic fluctuations between the fluid and turbulent regimes. This paper outlines a general framework for describing these metastable regimes.
Key words and phrases: Mean-field, HTTP, Metastable, Large Deviations, tunnelling .

MSC 2000 subject classifications: Primary 60J25, 60K35; secondary 94C99.

1 Introduction

There have been many recent developments on the emerging science of spontaneous order, [5]. Spontaneous order or "synch" may occur among a collection of (stochastic) dynamical systems or particles due to a "coupling" between any one particle and the ensemble. This phenomenon of "synch" is quite common when observing multiple TCP/IP connections (each connection is a dynamical system) routed through a common tail-drop bottleneck router. The coupling is provided by the spurt of packet losses and the resulting rate reductions caused when the total transmission rate exceeds the link rate (i.e. the ensemble average exceeds a threshhold). Each connection suffering a loss reduces its transmission rate by half thereby synchronizing a number of connections with a relatively low transmission rate. These connections then increase their transmission rates together according to the rules of TCP (i.e. linearly) until the next spurt of losses. After a while the total transmission rate looks like the familiar saw-tooth rising up to the link rate and then falling abruptly with many connections having synchronized transmission rates.

In [1] we studied HTTP connections routed through a bottleneck router tail-drop router at some popular web site. In contrast to a long lived TCP connection transmitting a large file, HTTP connections tend to be alternate between busy and silent periods. Once a connection to a remote web site is established, the user may click on a link which causes a busy period while a page is transmitted through the bottleneck router. The silent period follows while the user reads the page. When the user is finished with the page he might click on another link to restart the busy period.

*INRIA-ENS

†INRIA-ENS and IBM T.J. Watson Research Center

‡Department of Mathematics, University of Ottawa, dmdsg@mathstat.uottawa.ca, Research supported in part by NSERC grant A4551

The transmission rate of each connection may be described as a dynamical system which increases linearly until they are reset to zero at random times. We may therefore describe N stationary dynamical systems which evolve independently as long as the total transmission rate never exceeds the link rate L. Denote the mean transmission rate of a stationary HTTP connection by α.

If the link rate is exceeded then there will be a spurt of packet losses and a resulting synchronization of the transmission rates. Nevertheless it seems intuitive that if we scale up the link rate so $L = CN$ as $N \to \infty$ and $\alpha < C$ then the average transmission rate will converge to α. Hence the link rate would not be exceeded. All the connections would then evolve independently and there will be no packet losses except for rare fluctuations. We call this the fluid regime and it corresponds to a fixed point for the limiting mean field system.

On the other hand it is equally intuitive that if $\alpha > C$ then we are back in the saw-tooth or turbulent regime because on average the link rate will be exceeded resulting in packet loss and synchronization. We have shown this corresponds to a limit cycle for the limiting mean field system.

Surprisingly if α is roughly 90% of C both regimes are possible. In other words there is both a fixed point and a limit cycle for the limiting mean field system. For a finite but large number of connections this means that there will be rare fluctuations when the system tunnels between these two regimes. This has the practical effect that the transmission rate will fluctuate between a high throughput regime where there are no losses and the sources are independent and a low throughput regime where there are packet losses and and the sources are synchronized.

In order to understand the basic mechanism behind the the HTTP example and ultimately to calculate the mean time to tunnel between different regimes we have begun to study a simpler system. We consider a system of N irreducible Markov chains each evolving on a finite state space. The only way the particles interact is through the common transition kernel $K(\mu)$ that depends on the occupation measure μ of the N particles. We can find Markov chains whose dynamics are analogous to the different regimes of TCP and HTTP described above. We show that occupation measure is a Markov chain which satisfies the conditions of [7] and we use those results to calculate the mean tunneling time between the different regimes.

2 Model

We consider a finite state space: $\mathcal{S} = \{1, 2, \ldots, S\}$. Let $\mathcal{M}$ be the set of all probability measures on $\mathcal{S}$. We denote by K the mapping that to any $\mu \in \mathcal{M}$ associates a stochastic matrix $(K_{i,j}(\mu))_{(i,j)\in\mathcal{S}^2}$. We define the map $F : \mathcal{M} \to \mathcal{M}$ such that

$$F(\mu) = \mu K(\mu) \Leftrightarrow \forall i,\ F(\mu)_i = \sum_{l=1}^{S} \mu_l K_{l,i}(\mu).$$

We assume throughout that F is continuous for the topology generated by the total variation norm $||\cdot||$. We will consider the associated dynamical system on $\mathcal{M}$:

$$(D) \quad \begin{cases} \mu_{t+1} = F(\mu_t), \\ \mu_0. \end{cases}$$

We consider now a system constituted by N interacting particles evolving on the finite state space $\mathcal{S}$. Denote the N interacting particles by $\mathbf{X}^N(t) = (X_1^N(t), \ldots, X_N^N(t))$. Denote the corresponding occupation measure by $M_t^N \in \mathcal{M}^N$ where $\mathcal{M}^N$ denote occupation

measures obtainable from N particles:

$$M_t^N(A) = \frac{1}{N}\sum_{k=1}^{N} \chi\{X_k^N(t) \in A\}.$$

If at time t, the occupation measure of the system is M_t^N then each particle which is in state i (if any), will jump at time $t+1$ independently of everything else in state j with probability $K_{i,j}(M_t^N)$. Hence the coupling between the particles only occurs thanks to M_t^N. We just defined a random map $G^N : \mathcal{M}^N \to \mathcal{M}^N$ by $G^N(M_t^N) = M_{t+1}^N$.

Given a probability measure $\mu \in \mathcal{M}$, we define $\mu^N \in \mathcal{M}^N$ as follows

$$\forall i \leq S, \quad N(\mu(1) + \cdots + \mu(i)) \leq N(\mu^N(1) + \cdots + \mu^N(i)) < N(\mu(1) + \cdots + \mu(i)) + 1. \tag{2.1}$$

With these definitions, we can introduce the random dynamical system on $\mathcal{M}$:

$$(S) \quad \left\{ \begin{array}{l} M_{t+1}^N = G^N(M_t^N), \\ M_0^N = \mu_0^N. \end{array} \right.$$

We are interested in the above defined family of Markov chains M_t^N, $N > 0$, $t \geq 0$ on the space $\mathcal{M}$ and its comportment when $N \to \infty$. We will take the following notation for its transition probabilities $P^N(\mu, .)$, $\mu \in \mathcal{M}^N$.

We can see the dynamical system (D) as a degenerated Markov chain with transition "probabilities" given by:

$$P^\infty(\mu, d\eta) = \chi\{\eta = F(\mu)\}.$$

Proposition 1 *For any continuous function f on $\mathcal{M}$, the following limit holds uniformly in $\mu \in \mathcal{M}$*

$$\mathbb{E}[f(M_1^N)|M_0^N = \mu^N] \xrightarrow{N\to\infty} \int_{\mathcal{M}} P^\infty(\mu, dy) f(y) = f(F(\mu)). \tag{2.2}$$

Proof
Since $\mathcal{M}$ is compact, we have only to prove (2.2) for a fixed $\mu \in \mathcal{M}$. Consider $\mu \in \mathcal{M}$ and define Y_k^i to be i.i.d. in k for i fixed, such that $P(Y_1^i = j) = K_{i,j}(\mu^N)$. We have,

$$G^N(\mu^N)(j) = \frac{1}{N}\sum_{i\in\mathcal{S}} \sum_{k=1}^{N\mu^N(i)} \mathbb{1}_{\{Y_k^i = j\}}.$$

Note that $\mathbb{E}\left[G^N(\mu^N)(j)\right] = F(\mu^N)(j)$ so the variance

$$\begin{aligned} \sum_{j\in\mathcal{S}} \mathbb{E}\left[(M_{t+1}^N(j) - F(\mu^N)(j))^2 | M_t^N = \mu^N)\right] &= \sum_{j\in\mathcal{S}} \mathbb{E}\left(G^N(\mu^N)(j) - F(\mu^N)(j)\right)^2 \\ &= \frac{1}{N^2}\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} N\mu^N(i) K_{i,j}(\mu^N)(1 - K_{i,j}(\mu^N)) \\ &\leq \frac{1}{N}. \end{aligned}$$

Hence thanks to Chebyshev's inequality, we have

$$\begin{aligned} \mathbb{P}\left(\|M_1^N - F(\mu^N)\| > \epsilon | M_0^N = \mu^N\right) &\leq \frac{\mathbb{E}\left[\|M_1^N - F(\mu^N)\|^2 | M_0^N = \mu^N\right]}{\epsilon^2} \\ &\leq S\frac{\sum_{j\in\mathcal{S}} \mathbb{E}\left[(M_1^N(j) - F(\mu^N)(j))^2 | M_0^N = \mu^N\right]}{\epsilon^2} \\ &\leq \frac{S}{N\epsilon^2}. \end{aligned}$$

Now thanks to the continuity of F, we have, for N sufficiently large, that $\|F(\mu^N) - F(\mu)\| < \epsilon/2$ so

$$\|M_1^N - F(\mu)\| \leq \|M_1^N - F(\mu^N)\| + \epsilon/2,$$

Hence we proved that as $N \to \infty$,

$$\mathbb{P}\left(\|M_1^N - F(\mu)\| > \epsilon | M_0^N = \mu^N\right) \to 0.$$

In other words, if we take $B(F(\mu), \epsilon) = \{\eta \in \mathcal{M},\ \|\eta - F(\mu)\| \leq \epsilon\}$, we proved that

$$P^N(\mu^N, B(F(\mu), \epsilon)^c) \to P^\infty(\mu, B(F(\mu), \epsilon)^c) = 0,$$

and the proposition follows

■

This convergence implies directly a first result:

Proposition 2 *For any compact $C \subset I\!R^+$, we have convergence in distribution (and in probability) of the process $\{M_t^N\}_{t\in C}$ to the process $\{\mu_t\}_{t\in C}$ as $N \to \infty$.*

Proof

First, since the limit is deterministic, convergence in distribution and in probability are equivalent. Hence we have to show that for any fixed t, $M_t^N \to \mu_t$ as $N \to \infty$ where the convergence is in distribution.

For $t = 0$ the result is clear since, we have convergence almost surely. Now assume that $M_t^N \xrightarrow{\mathcal{D}} \mu_t$. Thanks to the limit (2.2) which is uniform over $\mathcal{M}$, we have for any continuous function f,

$$\sup_{\mu\in\mathcal{M}} \left|\mathbb{E}[f(M_{t+1}^N)|M_t^N = \mu^N] - f(F(\mu))\right| \to 0,$$

hence thanks to $M_t^N \xrightarrow{\mathcal{D}} \mu_t$, we have as $N \to \infty$,

$$\mathbb{E}[f(M_{t+1}^N)] \to f(F(\mu_t)) = f(\mu_{t+1}),$$

and the proposition follows by induction.

■

Indeed thanks to Proposition 3.2 of [7], we have that any weak limit as $N \to \infty$ of invariant measures of Markov chain M_t^N is an invariant measure of the Markov chain with transition $P^\infty(\mu, .)$.

Let m^N be a measure on $\mathcal{M}$. It is an invariant measure for the Markov chain M_t^N if

$$\int_{\mathcal{M}} dm^N(\mu) P^N(\mu, \Gamma) = m^N(\Gamma).$$

If $m^{N_i} \xrightarrow{w} m$, then m is an invariant (probability) measure of the map F:

$$m(F^{-1}\Gamma) = m(\Gamma).$$

Theorem 3.1 of [7] gives the support of such invariant measure.

3 Examples

We will now construct three examples which mimic the different limiting regimes observed with TCP and HTTP traffic.

We consider $\mathcal{S} = \{0, 1\}$. The state 0 represents an off state and 1 represents an on state. For simplicity, for any $\mu \in \mathcal{M}$, we will write $\mu(1) = \mu$, i.e. the proportion of on's. We define the two matrices:

$$\begin{aligned} K^a &= \begin{pmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix}, \\ K^b &= \begin{pmatrix} 1 & 0 \\ 2/3 & 1/3 \end{pmatrix}. \end{aligned}$$

Denote the fixed point of K^a by $\pi^a = (2/5, 3/5)$.

We define the mapping K as follows: if $\mu < C - \epsilon/2$, then $K(\mu) = K^a$ and if $\mu > C + \epsilon/2$, then $K(\mu) = K^b$. In the interval $(C - \epsilon/2, C + \epsilon/2)$ we define K as follows in order to obtain a continuous mapping:

$$K(\mu) = \frac{(C + \epsilon/2 - \mu)}{\epsilon} K^a + \frac{(\mu - (C - \epsilon/2)}{\epsilon} K^b \text{ for } C - \epsilon/2 \leq \mu \leq C + \epsilon/2.$$

This N particle system is irreducible and has a unique steady state.

- [Fluid HTTP Case] Suppose $C > 9/13$ then $\mu_0 = (2/5, 3/5)$ is stationary for the infinite particle system; i.e. $F(\mu_0) = \mu_0$, $\mu_0(1) < C$.
- [Sawtooth Case] Suppose $3/13 < C < 3/5$ then if $\mu_0 = (10/13, 3/13)$ then $F(\mu_0) = \mu_1 = (4/13, 9/13)$ and $\mu_2 = F(\mu_1) = \mu_0$. In other words we have a limit cycle.
- [Bistable HTTP Case] If $3/5 < C < 9/13$ then the initial distribution $\mu_0 = (2/5, 3/5)$ is stationary for the infinite particle system; i.e. $F(\mu_0) = \mu_0$, $\mu_0(1) < C$. However if $\mu_0 = (10/13, 3/13)$ then $\mu_1 = (4/13, 9/13)$ and $\mu_2 = F(\mu_1) = \mu_0$. In other words we also have a limit cycle.
- [Longer cycles] Other more complicated cycles are possible. If $3/37 < C < 9/37$ and $\mu_0 = (10/37, 27/37)$ then $\mu_1 = (28/37, 9/37)$, $\mu_2 = (34/37, 3/37)$ and $\mu_3 = \mu_0$.

4 Large deviation results

It is of some interest to determine the domain of attraction of the above fixed point and limit cycle and then to determine the mean time the N particle system takes to tunnel from one domain of attraction to another. Hence we are interested in the asymptotics as $N \to \infty$ of the time spent by the Markov chain in each domain of attraction. These questions are known in the literature as problems of exit from a domain (`Freidlin-Wentzell [6], Kifer [7]`).

Suppose that $\mu \in K_{\mu^*}$, where K_{μ^*} is some compact set containing only one attractor μ^*. Define

$$\tau^{\mu,N} = \inf\{n > 0,\ M_t^N \notin K_{\mu^*},\ M_0^N = \mu^N \in K_{\mu^*}\}.$$

The transition probabilities $P^N(\cdot, d\eta)$ is well-defined on $\mathcal{M}^N$ and we extend it to $\mu \in \mathcal{M}$ by $P(\mu, d\eta) = P^N(\mu^N, d\eta)$. We can apply the results in [7] if we can check (1.1) there. We do this in the following Lemma whose proof is given in [2].

Proposition 3 *For any open set $U \subset \mathcal{M}$, we have*

$$\lim_{N\to\infty} \sup_{\mu\in\mathcal{M}} \left| \frac{1}{N} \log P^N(\mu, U) + \inf_{\eta\in U} \rho(\mu, \eta) \right| = 0,$$

where $\rho(\mu,\eta) \geq 0$ is a continuous function on $\mathcal{M} \times \mathcal{M}$ given by:

$$\rho(\mu,\eta) \quad = \quad \inf_{\overline{\pi}=\mu,\underline{\pi}=\eta} \sum_{i,j} \pi(i,j) \log\left(\frac{\pi(i,j)}{\mu(i)K_{i,j}(\mu)}\right),$$

and where we denote by $\overline{\pi}$ and $\underline{\pi}$ the two marginals of the distribution π on S^2; i.e.

$$\begin{array}{ll} \overline{\pi}(i) & = \sum_{j=1}^{S} \pi(i,j) = \mu(i), \\ \underline{\pi}(j) & = \sum_{i=1}^{S} \pi(i,j) = \eta(j). \end{array}$$

Let A_N be a function on $\mathcal{M}^N$ defined for $\Xi = (\xi_0, \ldots, \xi_{N-1}) \in \mathcal{M}^N$, $\xi_i \in \mathcal{M}$, $i = 0, \ldots, N-1$ by the formula

$$A_N(\Xi) = \sum_{i=0}^{N-2} \rho(\xi_i, \xi_{i+1}) \quad \text{for } N > 1 \text{ and } A_1 = 0.$$

For any pair of points $\mu, \eta \in \mathcal{M}$ put

$$B(\mu,\eta) = \inf\{A_n(\Xi) : \; n \geq 1,\, \Xi = (\xi_0, \ldots, \xi_{n-1}),\, \xi_0 = \mu,\, \xi_{n-1} = \eta\}.$$

Set $B = \inf_{\mu \notin K_{\mu^*}} B(\mu^*, \mu)$, then if this quantity is finite, we have for any $\delta > 0$,

$$\mathbb{P}\left(\exp(N(B-\delta)) \leq \tau^{\mu,N} \leq \exp(N(B+\delta))\right) \xrightarrow{N\to\infty} 1.$$

5 Applications

We can apply the above theorem to our two state example. We will assume the ϵ used in the definition of K is so small that we can approximate the domain of attraction of the fixed point by $\mathcal{D}^a = \{\mu | 4(C - 1/2) < \mu(1) < C\}$ (as if $\epsilon = 0$). To calculate $\rho(\alpha, \beta)$ for $\alpha, \beta \in \mathcal{D}^a$ we must find a matrix

$$\pi = \begin{pmatrix} \alpha(0) & 0 \\ 0 & \alpha(1) \end{pmatrix} \begin{pmatrix} x & 1-x \\ y & 1-y \end{pmatrix}$$

such that $\alpha(0)(1-x) + \alpha(1)(1-y) = \beta(1)$ which minimizes

$$\begin{array}{rcl} V^{\alpha,\beta}(x,y) & = & \alpha(0)x \log(\frac{x}{1/4}) + \alpha(0)(1-x)\log(\frac{(1-x)}{3/4}) \\ & & +\alpha(1)y\log(\frac{y}{1/2}) + \alpha(1)(1-y)\log(\frac{(1-y)}{1/2}). \end{array}$$

$V^{\alpha,\beta}(x,y)$ is a convex function with an unconstrained minimum at $x = 1/4$, $y = 1/2$. Using a Lagrange multiplier λ we must find the extremals of

$$V^{\alpha,\beta}(x,y) - \lambda(\alpha(0)x + \alpha(1)y - (1 - \beta(1))).$$

This yields, $\frac{3x}{1-x} = \frac{y}{1-y}$ subject to $\alpha(0)x + \alpha(1)y = (1 - \beta(1))$.

We have found numerically that if $\pi^a(1) \leq \alpha(1) < \gamma(1) < \beta(1)$ or if $\beta(1) < \gamma(1) < \alpha(1) \leq \pi^a(1)$ then

$$\rho(\alpha,\beta) \leq \rho(\alpha,\gamma) + \rho(\gamma,\beta).$$

This means the minimization used to define $B(\alpha,\beta)$ is obtained with $n = 1$. Hence, under the above conditions on α and β, $B(\alpha,\beta) = \rho(\alpha,\beta)$.

- [Fluid HTTP Case] There is only one ρ-attractor at $\pi^a = (2/5, 3/5)$. Suppose we want to calculate the asymptotics of the mean time for the M_t to exceed the threshold $C = 10/13$ at which time packets are lost. We apply Theorem 4.2 in [7] where $B = \rho(\underline{\pi}, \overline{\pi})$ and where $\underline{\pi} = (2/5, 3/5)$ and $\overline{\pi} = (3/13, 10/13)$. Calculation shows that
$$\pi = \begin{pmatrix} 2/5 & 0 \\ 0 & 3/5 \end{pmatrix} \begin{pmatrix} .125 & .875 \\ .301 & .699 \end{pmatrix}$$
and $B = .0689$. Define $\tau^{\mu_0, N}$ to be the mean time to escape $K_{\pi^a} = \{\mu : \mu(1) \leq C\}$ starting from μ_0 in the domain of attraction of π^a. Then
$$\lim_{N \to \infty} \frac{1}{N} \log E_{\mu_0}(\tau^{\mu_0, N}) = B.$$
- [Sawtooth Case] There is no fixed point and one limit cycle. The theory predicts a convergence to the limit cycle as $N \to \infty$.
- [Bistable HTTP Case] If $3/5 < C < 9/13$ it is of some interest to calculate the asymptotics of the mean time to tunnel from near the stationary state π^a to the limit cycle alternating between $(10/13, 3/13)$ and $(4/13, 9/13)$. Suppose $C = 8/13$ then the entrance states to the domain of attraction of the limit cycle are $u = (5/13, 8/13)$ and $d = (6/13, 7/13)$.
Since $B(u, \beta) = 0$ we calculate $\rho(\pi^a, u)$ so $\underline{\pi} = \pi^a$ and $\overline{\pi} = u$. Calculation shows that
$$\pi = \begin{pmatrix} 2/5 & 0 \\ 0 & 3/5 \end{pmatrix} \begin{pmatrix} .237 & .763 \\ .483 & .517 \end{pmatrix}$$
and $B(\pi^a, u) = \rho(\pi^a, u) = .0005$
Since $B(d, \beta) = 0$ we calculate $\rho(\pi^a, \beta)$ so $\underline{\pi} = \pi^a$ and $\overline{\pi} = d$. Calculation shows that
$$\pi = \begin{pmatrix} 2/5 & 0 \\ 0 & 3/5 \end{pmatrix} \begin{pmatrix} .304 & .696 \\ .567 & .433 \end{pmatrix}$$
and $B(\pi^a, d) = \rho(\pi^a, d) = .008$
If we make this argument rigorous by a series of approximations as $\epsilon \to 0$ we could conclude
$$\lim_{N \to \infty} \frac{1}{N} \log E_{\mu_0}(\tau^{\mu_0, N}) = B = 0.008.$$

References

[1] Baccelli, F., Chaintreau, A., De Vleeschauwer, D., McDonald, D (2004). HTTP Turbulence, INRIA Research Report RR-5205, 54 p.p.

[2] Baccelli, F., Lelarge, M., McDonald, D.(2005) Metastable Regimes for multiplexed TCP flows, *working paper*.

[3] Csiszar, I., Cover, T., Choi B.-S. (1987) Conditional limit Theorems under Markov Conditioning, IEEE Transactions on Information Theory, No. 6, pp. 788-801.

[4] Dembo, A., Zeitouni, O. (1998). *Large Deviations Techniques and Applications, Second Edition.* Springer Verlag, 396 p.p.

[5] Ermentrout, G.-B. (2004). Review of Sync: The Emerging Science of Spontaneous Order, AMS Notices, **51**, 312-319.

[6] Freidlin, M. I., Wentzell, A. D. *Random Perturbations of Dynamical Systems.* Springer Verlag, 1984.

[7] Kifer, Y. (1990). A discrete-time version of the Wentzell-Freidlin theory. *Annals of Probab.*, **18**, 1676-1692.